The Journal of Economic Perspectives

A journal of the
American Economic Association

Spring 2026

The Journal of
Economic Perspectives

A journal of the American Economic Association

Editorial offices:
Journal of Economic Perspectives
American Economic Association Publications
2403 Sidney St., #260
Pittsburgh, PA 15203
email: jep@aeapubs.org

The Journal of Economic Perspectives gratefully acknowledges the support of Macalester College.
Registered in the US Patent and Trademark Office (®).

Composed by American Economic Association Publications, Pittsburgh, Pennsylvania, USA.
Printed at Sheridan Press, Hanover, Pennsylvania, USA.

THE JOURNAL OF ECONOMIC PERSPECTIVES (ISSN 0895-3309), Spring 2026, Vol. 40, No. 2. The JEP is published quarterly (February, May, August, November) by the American Economic Association, 2014 Broadway, Suite 305, Nashville, TN 37203-2418. For details and further information on the AEA go to https://www.aeaweb.org/. Periodicals postage paid at Nashville, TN, and at additional mailing offices.

POSTMASTER: Send address changes to the *Journal of Economic Perspectives*, 2014 Broadway, Suite 305, Nashville, TN 37203. Printed in the USA.

The Journal of
Economic Perspectives

Contents

Volume 40 • Number 2 • Spring 2026

Symposium

Articles

Features

Statement of Purpose

The *Journal of Economic Perspectives* aims to bridge the gap between the general interest business and financial press and standard academic journals of economics. The journal aims to publish articles that will serve several goals: to synthesize and integrate lessons learned from active lines of economic research; to provide economic analysis of public policy issues; to encourage cross-fertilization of ideas among the fields of economics; to offer readers an accessible source for state-of-the-art economic thinking; to suggest directions for future research; to provide insights and readings for classroom use; and to address issues relating to the economics profession. Articles appearing in the journal are normally solicited by the editors and associate editors. Proposals for topics and authors should be directed to the journal office, at the address inside the front cover.

Policy on Data Availability

It is the policy of the *Journal of Economic Perspectives* to publish papers only if the data used in the analysis are clearly and precisely documented and are readily available to any researcher for purposes of replication. Details of the computations sufficient to permit replication must be provided. The Editor should be notified at the time of submission if the data used in a paper are proprietary or if, for some other reason, the above requirements cannot be met.

Policy on Disclosure

Authors of articles appearing in the *Journal of Economic Perspectives* are expected to disclose any potential conflicts of interest that may arise from their consulting activities, financial interests, or other nonacademic activities.

Journal of Economic Perspectives—Volume 40, Number 2—Spring 2026—Pages 3–16

The Emerging Role of Competition in Health Care

Paul B. Ginsburg

When I trained in health economics in the late 1960s, very few perceived that competition played a meaningful role in health care services, and perhaps even fewer saw potential that policies to employ or foster such competition might lower the cost or increase the quality of health care. Instead, the focus of the academic literature of that time is well-summarized by the titles of papers like Arrow's (1963) classic "Uncertainty and the Welfare Economics of Medical Care." Arrow, writing a few years before the enactment of Medicare and Medicaid, argued that because of uncertainty in the incidence of disease and the efficacy of treatment, the health insurance market was likely to have substantial gaps. He wrote that "society will seek to achieve optimality by nonmarket means if it cannot achieve them in the market" (947). But in this essay and many others of that time, the consistent focus was on reasons why market competition would not work well—not on reasons why including some elements of competition might be useful in the health insurance and health care industry.

In the last half-century, attitudes about and understandings of the role for competition in health care have evolved considerably. I will begin by sketching the situation soon after Medicare was implemented in 1966 and then describe major developments that followed, setting the stage for the examinations of current competition in health care in this symposium.

■ *Paul Ginsburg is Professor of the Practice of Health Policy and Management, Sol Price School of Public Policy, University of Southern California, Los Angeles, California. He is also Senior Scholar, USC Schaeffer Institute for Policy and Government Service and a Nonresident Senior Fellow in Economic Studies at the Brookings Institution. His email address is pginsbur@usc.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20251470.

The Era of Regulation

In the late-1960s and much of the 1970s, the major policy issue in health care that engaged economists was rapidly rising health care costs, especially hospital costs, a development that coincided with the implementation of expanded health insurance coverage for the elderly through Medicare and for some of the poor through Medicaid. Much later, Finkelstein (2007) would explain how the sharp increase in health insurance coverage in this period had a market-level impact on hospitals about six times larger than what actuaries had predicted based on increased utilization by individuals, because the large expansion of health insurance gave hospitals an incentive and resources to enter the market and to invest in capacity to deliver new (and more expensive) procedures. But at that time, few policymakers considered enhanced competition as a possible solution. Instead, patients were seen to be highly dependent on physician recommendations. Also, there was a widespread belief that patient responses to lower health care prices by some providers would be muted by concerns that low prices signaled low quality.

Hospital Price Regulation

Instead, policy focused on regulation. Medicare had paid hospitals on the basis of reimbursement of "reasonable" costs (meaning related to patient care) plus an additional 2 percent allowance for capital costs, following the approach used by Blue Cross insurance plans since their inception years earlier. This approach to payment was particularly passive, in that wide variation in costs across hospitals was accepted. In the 1970s, Medicare began to place limits on reimbursement per patient day, focused on routine costs (like room, board, and nursing), but not costs for specialty care. Limits were tightened over time.

Some states were particularly active in health care cost containment policy during the 1970s. Maryland, New York, New Jersey, and Massachusetts regulated rates for all insurance payers, including Medicare (through federal waivers that allowed states to experiment with different payment schemes) and state Medicaid programs, which are managed by each state with federal grants covering percentage (based on state per capita income) of program spending. Particularly notable was Maryland's establishment of its Health Services Cost Review Commission in 1974, which was guided by a philosophy of simulating the functioning of a hypothetical competitive market. Indeed, New Jersey in 1980 pioneered inpatient prospective payment using diagnosis-related groups (DRG), which likely inspired Medicare's subsequent adoption of the approach in 1983 (as discussed below) (Hsiao et al. 1986). Under DRG payment, each patient admitted is classified into a DRG on the basis of diagnosis and the presence of complications; payments per admission are adjusted for the costliness of the DRG, allowing hospitals with relatively low costs to earn surpluses.

Recent literature on the effectiveness of these programs has been generally favorable (Atkinson 2009; Murray and Gudiksen 2024). However, the combination of Medicare's adoption of inpatient prospective payment in 1983, a changing national political environment more hostile to regulation, and hospital industry

opposition led New York, New Jersey, and Massachusetts to end their programs, although Maryland has continued its program to this day with important refinements, especially in the last decade.

As Medicare limits on hospital payment per patient day became increasingly stringent, with a particularly large tightening from 1982 legislation, hospitals came to view this approach as oppressive, because it involved penalties from overly high costs but no rewards for hospitals with relatively low costs. In 1983, hospitals endorsed a shift to a prospective payment approach based on diagnosis-related groups, proposed by the Reagan administration, an approach that remains in use today. Although the approach is often described as "administered pricing," it strikes me as having important elements that might foster competition, because hospitals had incentives to reduce costs per hospital admission and hospitals with costs below the payment rate could keep the surplus. In addition, payment rates were uniform nationally across hospitals, except for adjustments for local input prices and the extent of graduate medical education activity. Subsequently, over several decades, Medicare extended this "prospective payment" approach of paying a fixed amount for a bundle of services to hospital outpatient services and services from many other types of providers.

The contrast between the adoption of Medicare inpatient prospective payment in 1983 and current efforts to further reform Medicare hospital payment under demonstration authority in the Patient Protection and Affordable Care Act of 2010 is striking. The 1983 reform involved substantial redistribution across hospitals, but nevertheless was nationally implemented over a four-year transition. Policymakers perceived New Jersey's experience to be all that was needed to roll out the approach nationally. In contrast, numerous demonstrations of approaches such as Accountable Care Organizations (in which reimbursement is linked in part to health outcomes and lower costs, not just to provision of services) and bundled payments have been proceeding for 15 years now, with most of the demonstrations voluntary and the possibility of a future mandatory policy still not clear.

Supply Restrictions

This earlier period saw supply restrictions along with price regulation. At the time, there was a widespread notion that in health care, with its extensive third-party payment and reimbursement of costs, additional supply created its own demand. An earlier study had found a positive correlation between hospital beds and patient days, which led many in the hospital industry to believe that additional beds would always be filled, calling it "Roemer's Law" (Shain and Roemer 1959). Of course, economics has become more disciplined in recent decades about not treating simple correlations as evidence of causality: in this case, it seems plausible that the correlation resulted from a major increase in market demand, driven by a large expansion of private employer-sponsored health insurance plans during the 1950s, having overwhelmed existing capacity.

But starting with New York in 1964, 26 states had enacted "Certificate of Need" laws by 1974, when the federal government required all states to do so. These laws

required hospitals to get permission from a state agency on the basis of "community needs" to construct facilities or to make major equipment purchases. Even after the federal government repealed its requirement in 1986, many states continued their Certificate of Need programs; 35 programs remain active today. Research on the effects of Certificate of Need programs has been extensive, though challenging, given the need to account for endogeneity (in this case, states experiencing a period of rapidly increasing hospital expenditures are more likely to enact Certificate of Need laws). An exhaustive review of the literature by Conover and Bailey (2020) concludes that "the evidence regarding hospital CON's effect on health expenditures is generally mixed, although one could credibly conclude that the weight of this evidence is that CON has no impact on health costs overall." Given the apparent lack of effect on health care costs, an obvious question is why so many states have continued such programs for so long. The likely explanation is that the rules can be used to protect incumbent providers from entry of competitors (Mitchell 2024). Some suspect that application of Certificate of Need laws in some states to the addition of nursing homes beds reflects state officials using it as a tool to limit Medicaid outlays for such services.

Antitrust Policy

Antitrust policy during the 1970s played little role in the health sector. Challenges to mergers between nonprofit hospitals were not successful. In 1975, a US Supreme Court decision in *Goldfarb v. Virginia State Bar* (421 U.S. 773 [1975]) established that the "learned professions" were indeed subject to antitrust law under the Sherman Act, but the main focus of the decision was on minimum fees charged for certain transactions by real estate lawyers. In 1993, the US Department of Justice and the Federal Trade Commission (1993) issued a policy statement clarifying antitrust enforcement intentions concerning collaboration among health care providers. But it would not be until 2007 that the Federal Trade Commission's ruling against Evanston Northwestern Healthcare's acquisition of Highland Park Hospital ended an 18-year streak of failed challenges to mergers between nonprofit hospitals (Federal Trade Commission 2008). I can only speculate, but I believe that some combination of awareness of rate-setting by some states, belief by some that charitable organizations would not take actions to raise prices, and the slow development of research evidence needed to win challenges to mergers might explain the lack of antitrust activity in the health care sector during the 1970s and 1980s.

Initial Steps Towards Fostering Competition

Perhaps the earliest step by the federal government to foster competition in health care was to encourage the development of health maintenance organizations, which incorporate financing and delivery of a defined set of health services. Likely inspired by the success of Kaiser Permanente on the West Coast in integrating health insurance with a provider organization and reducing health care costs, a

Minnesota physician named Paul Ellwood and other aides working in the Nixon Administration came up with the concept and name. Legislation was enacted in 1973 that provided grants to those developing health maintenance organizations, overrode state restrictions, and required those employers offering insurance coverage to provide employees choice of a health maintenance organization. The goal was to save money.

In the 1970s, Kaiser Permanente, perhaps the paradigmatic health maintenance organization, was successfully achieving lower per enrollee costs through this combination of financing and delivery. Indeed, a number of the prominent multi-specialty group practices started to offer health maintenance organizations.

But many of the health maintenance organizations formed under this program did not have the full integration of financing and delivery that Kaiser Permanente did. Despite success, some of these health maintenance organizations—for example, the Harvard Community Health Plan in the Boston area—subsequently separated the group of health care providers from the insurance part of the organization in order to allow the medical group to also serve patients with insurance from other carriers. Today, most perceive health maintenance organizations as a type of health insurance product with more restrictions on provider choice than the so-called "preferred provider organizations," rather than as organizations that deliver integrated health services.

In the early 1980s, interest in policies to foster competition in health care grew. I believe that it reflected a broad national trend, initiated during the Carter presidency and inspired further by the Reagan presidency, to use competition instead of regulation throughout the economy. Legislation enacted in 1982 initiated offering private health plans to Medicare beneficiaries to choose as an alternative to traditional Medicare (discussed further below).

Extensive discussions were held within and around Congress about the longstanding ability of employers to fund health insurance benefits without employees being taxed the way they are on wages and salaries as a factor that was undermining the possibility of a more competitive health care system. This led to discussion of limiting the amount that employers could subsidize employee health benefits with pre-tax dollars (Wilensky and Taylor 1982). During the Reagan administration, the House Ways and Means Committee had extensive discussions of this strategy and it was proposed in early versions of the Tax Reform Act of 1986.

Eventually, several decades later, the "Cadillac Tax" (so-named because it focused on taxing only the highest-premium health insurance plans) was enacted as part of the Patient Protection and Affordable Care Act of 2010 (Glied and Striar 2016). Although limiting the extent to which employer and employee contributions were excluded from employee taxable income was a more direct way to achieve the goal of making the health care system more competitive, the Cadillac tax instead taxed the insurers (or employers for self-insured plans) with high premiums, presumably for more political appeal. The stated goal of the tax was to increase federal revenue to pay for provisions expanding health insurance coverage, but a later Congress repealed the provision before it was implemented. Support for limits on funding of

employer-provided health insurance with pre-tax earnings seems almost universal among economists regardless of political persuasion, but it has never had support from politicians.

However, it was Alain Enthoven's (1978a, b) influential work that brought the terminology of "managed competition" into health care policy circles. He envisioned the health maintenance organization as a vehicle for competition in health care, with employers offering a choice of plans to employees and a fixed contribution so that employees choosing more expensive plans paid the difference in premiums. Some leading large employers, such as General Electric, adopted this approach, offering employees a choice of carefully-selected local health maintenance organizations in areas with large concentrations of employees, with financial incentives to choose those with lower premiums.

Approaches to managed competition were further developed in the late 1980s and early 1990s through discussions of the Jackson Hole Group, co-led by Paul Ellwood and Alain Enthoven, which convened leading health care professionals, government officials, business leaders, academics, and other experts in Elwood's living room in Teton Village, Wyoming. Many of its ideas on managed competition were incorporated into President Bill Clinton's health care reform proposal, although the group ultimately opposed the legislation because of the degree of government regulation and price controls.

However, during the 1990s, many large employers opted for a single health insurance carrier for their employees throughout the nation. Although employees in large firms tend to have a choice of health plans, often the choice is only between a health maintenance organization and a preferred provider organization from the same carrier.

In contrast to the experience with US employers, Enthoven's ideas about managed competition got more traction in those European countries where national health insurance is provided through competing health insurers like the Netherlands, Germany, and Switzerland (as discussed in the article in this symposium by Kauer, McGuire, Schillo, and van Kleef).

Although US employers did not pursue Enthoven's managed competition model, they did change their health insurance offerings from products that passively paid health care bills—albeit with a mix of deductibles, coinsurance, copayments and, more recently, limits on patient out-of-pocket responsibility—to health insurance offerings designed to influence the delivery of care. This new approach is usually what is now meant by "managed" care, to reflect various restrictions and incentives applied to the use of health care services and choice of providers.

These changes were based on a recognition that health insurance is quite distinct from other types of casualty insurance. Most casualty insurance, such as fire insurance, pays cash to offset a portion of losses, such as damage to structures, regardless of the purpose for which the insured uses the payment. Health insurance, in contrast, does not pay for loss of health, but pays a portion of the cost of medical services used to diagnose or treat a medical problem. While the insurance term "moral hazard" usually refers to insurance leading individuals to be less careful

about preventing fires in their home, its application in health insurance, as pointed out by Pauly (1968), refers to seeking more medical care because insurance will underwrite much of the cost.

Another distinction is that many large employers are now self-insured, meaning that the company takes the insurance risk for overall health care spending while hiring a third-party administrator, many of which are divisions of insurance companies, to process the claims and access their provider network agreements. Even if employees are shielded to some extent from health care costs, a self-insured employer knows that rising health care costs of its employees will increase compensation costs unless offset by reductions in other components of compensation, such as wages or other fringe benefits. The large proportion of employees covered under self-insured plans has created barriers to state policies affecting health insurance because the federal Employee Retirement Income Security Act of 1974 (ERISA) bars regulation of such plans by states.

The version of "managed care" that has prevailed in recent decades has typically involved employee choice between a health maintenance organization or a preferred provider organization. The main difference is that health maintenance organizations are more restrictive when it comes to provider choice. For example, in some cases, they may cover only health care providers that have a network agreement with the insurer, with no allowance for coverage of out-of-network care. Some health maintenance organizations also require that patients first see a primary care provider and get a referral to a specialist if warranted. Typically, preferred provider organizations have a broader network of providers to choose from. The fewer restrictions in preferred provider plans in choice of provider come at a higher cost to enrollees.

These forms of managed care grew rapidly in the late 1980s and have been the norm for employer-provided health insurance for some time. These plans reduced costs for employers, but the restrictions also angered many employees, leading to a backlash. When labor markets became tight, employers made managed care less restrictive. When labor markets softened, some additional restrictions were added. Over time, the techniques of managed care have evolved. Today, incentives to use in-network providers are much stronger and tools other than financial incentives, such as requiring prior authorization for especially expensive procedures or prescription drugs, are used more extensively.

The concepts of managed competition have also been reflected in public health insurance programs. Starting in 1985, Medicare allowed beneficiaries to opt out of traditional Medicare into a privately-administered plan that would provide benefits at least as good as those in traditional Medicare, but could offer additional benefits. Private health plans in Medicare have gone through a number of important iterations (and name changes) since then. The latest version, called Medicare Advantage, has been so successful that in 2026 it accounted for more than half of those beneficiaries eligible to enroll (they must be enrolled in both the hospital insurance Part A and the doctor and outpatient insurance Part B), with large numbers of plans competing in most areas of the country (as discussed in this

symposium by Layton, Maini, and McWilliams). However, an important part of this success has been due to the fact that payment to Medicare Advantage plans for a given patient is on average substantially higher than the cost of covering that same patient in traditional Medicare, with favorable selection the most important factor (MedPAC 2026).

Most Medicaid programs have also opted for a managed competition approach, requiring enrollees to choose a managed care plan. Because Medicaid is focused on those with low incomes, choice of plan is not based on premiums. Instead, states employ competition by having plan sponsors compete to be chosen as one of a small number of entities permitted to offer plans to enrollees (as discussed by Shepard and Wallace in this symposium). Insurers active in this market tend to specialize, whether Medicaid is the principal focus of the organization or a separate division of very large firms with a broad focus (Hinton and Raphael 2025).

The state-run Marketplace exchanges for the purchase of individual health insurance created by the Patient Protection and Affordable Care Act of 2010 probably come closest to Enthoven's vision for managed competition (for an overview, a readable starting point is DeSilver 2026). Through these exchanges, plans that meet required qualifications compete for individual and family enrollments with benefit designs conforming to one of four actuarial values, as defined by the proportion of spending for covered services in a population that would have been paid by the plan. The schedule of subsidies for those with low incomes is based on the plan with an actuarial value of 70 percent that has the second-lowest premium.

Recent Developments: Both More Competition and More Regulation

In recent years, governments have expanded their efforts to foster or make use of competition in health care, but have also have begun to use regulation more actively to address issues for which competition is unlikely to be effective.

Medicare Physician Payment Reform

An important policy development that is difficult to categorize between competition and regulation is a complete restructuring of how Medicare sets payment rates for physician services, enacted in 1989 and implemented starting in 1992. It was motivated by broad concerns about the structure of physician payment, with procedures paid too much in relation to visits and services in rural areas paid to little relative to urban areas. As physician services had begun to be covered by health insurance, both public and private insurers used the highly passive approach of screening physician charges on the basis of whether they were "usual, customary, and reasonable." "Reasonable" meant that the charge for the service was below a certain percentile in the distribution of physician charges in an area.

The reform involved replacing the system with a fee schedule with the relative values based on a large research study funded by the Medicare program. With

support from the American Medical Association, a research group at Harvard University used surveys of physicians to estimate the time and "work" involved in providing different medical services (Hsiao et al. 1988). The structure of payment emerging from study and work on other components of the fee schedule developed by the Physician Payment Review Commission, an entity created by Congress to develop reforms in Medicare physician payment.[1] The Commission's recommendations were consistent with Congressional aspirations, leading to passage of legislation. Implementation of the legislation led to the desired changes in the structure of payment, although a flawed process used over many years to update relative values led to a portion of the increased relative payment for visits being eroded over time. Private insurers quickly embraced Medicare's fee schedule, perceiving that the credibility of Medicare would make changes that they believed desirable to be more accepted by physicians. The fee schedule also facilitated negotiations between private insurers and physicians, allowing them to focus on a single number for thousands of procedures—percentage of the Medicare rate.

Price and Quality Transparency

Reflecting an increasingly broad societal interest in transparency, some governments and insurers have sought to improve transparency of health care prices and quality, hoping to encourage patients to select lower-priced and higher-quality health care. Some state governments have published hospital prices negotiated with insurers. The data have shown large variation in hospital prices—apparently not highly correlated with quality. The federal government now requires both hospitals and insurers to report negotiated prices, but at this point, the data are difficult for interested consumers to access.

Insurers have made more data on what patients would have to pay at different providers readily available to enrollees for services that are "shoppable," such as magnetic resonance imaging (MRI) scans, but take-up seems limited. Studies on employer plans with large deductibles and good data on what patients have to pay have found very limited responses by patients (Chernew et al. 2021).

While a lot of effort has gone into measuring quality at either the plan level or the provider level, it does not appear that the current state of quality transparency is likely to have much effect on actions by consumers in choosing plans or on patients in choosing providers. The problem with both price and quality transparency tools might be that using them effectively is a heavy lift for many consumers, especially at times when they are not feeling well and worried about their health. In addition, the provision of useful information on what is most important—health outcomes—is very challenging, due to needed risk adjustment and lack of uniform data. For example, a high-quality hospital may end up treating a higher proportion of patients

[1] In 1997, the Commission was merged with a similar commission focused on Medicare hospital payment to become the Medicare Payment Advisory Commission, which continues to be an influential voice on Medicare payment issues.

with the most severe conditions, and for this reason end up with higher-than-average mortality rates and costs.

However, greater transparency may offer tools that insurers can use to nudge patients toward choices that are more sensitive to price and quality. For example, insurers can offer "tiered" networks, in which a subset of providers in a plan's network are identified as "preferred" and require lower cost-sharing amounts from patients (as discussed in this symposium by Sinaiko). Another option is "reference pricing," where limits are placed on how much an insurer will pay for a service: a reference pricing limit might be set based on some percentile, perhaps the median, of rates charged by network providers. Given the difficulties employers have experienced with offering health plans with narrow provider networks—due to lack of choice of health plans—including a tiered network within the choice of a health maintenance or preferred provider organization offered might be a particularly effective tool to foster price competition. But powerful health care provider systems have often demanded the insurers place them in the preferred tier, undermining the effectiveness of this approach. Massachusetts banned such "anti-tiering" clauses in 2012, leading to greater use of this strategy.

One issue with price transparency has not been confronted by policy makers: In a market with concentrated firms, price transparency might increase prices. In health care, many anecdotes have appeared about hospitals with relatively low prices raising them in response to learning about higher prices at hospitals they consider to be their peers or inferiors (Ginsburg 2007).

Fostering More Rational Consumer Choices

Efforts to improve consumers' choice of a health insurance plan have also drawn on findings from behavioral economics, like consumers making more informed choices when given only a limited number of options (the so-called "paradox of choice" as explained in Decision Lab n.d.). For example, the Covered California Marketplace exchange for the purchase of individual health insurance limits the number of insurers allowed to offer plans in each region and has insurers compete for that privilege. California has also standardized the benefit designs at each of the four actuarial value levels. Lambrew and Young (2025) discuss how simplifying choices in the state-run Marketplaces can improve health insurance choices.

Other findings about common consumer biases have influenced the design of policies relevant to health insurance as well, like consumers' tendency to undervalue protection against catastrophic expenses and overvalue coverage for predictable expenses. These findings likely were behind the requirement in the Affordable Care Act that benefit structures include an out-of-pocket maximum. In contrast, the original design of the Medicare Part D drug benefit structure in 2006 had a deductible of $250 per year and coinsurance of 25 percent, but a coverage limit of $2,250 per year until the threshold for catastrophic coverage of $5,100 was reached. Spending between $2,250 and $5,100—called the "donut hole"—was not covered. Plans were permitted to revise the benefit structure (as long as the actuarial value was not reduced), but as one insurer painfully found, offering a premium plan that

did not have a donut hole was not viable due to the degree of adverse selection that resulted. Through a number of later pieces of legislation, over a number of years, the donut hole in Part D health insurance was reduced and finally eliminated.

Recent Antitrust Challenges

The traditional focus of antitrust in a health care context has been hospital mergers, but recent experience with consolidation across the US health care industry has shown the need for antitrust policy to broaden its focus. For example, there has been widespread vertical integration of health care providers, like combinations of hospitals, outpatient facilities, and physician practices, as well as mergers of health care providers across geographic markets (as discussed in this symposium by Sinaiko). The health insurance industry has become more concentrated as well (as discussed in this symposium by Gaynor and Starc). There have long been competition issues raised by whether the restrictions imposed on the care that nurse practitioners and physician assistants are allowed to provide are focused on providing cost-effective and quality care for patients, or whether they are instead protecting the incomes of physicians (as Gottlieb and Nicholson discuss in this symposium). Developing shortages of primary care physicians have led to expansions in allied health professions' scope of practice.

The subject of site-neutral payment (discussed by Richards and Whaley in this symposium) offers a vivid example of challenges to competition. The Medicare system for determining payment rates for a number of procedures is divided into a professional component and a facilities component. As a result, the facility component for physician services provided in a hospital outpatient department is determined in a different way—and is often substantially higher—than for a service performed in a physician office. One result is that hospitals can use their higher rates to recruit physicians to shift from independent practice to hospital employment. Hospital employers can encourage physicians to steer patients to hospital-employed specialists. Moreover, integrated hospital systems have greater clout in negotiating rates for both hospital and physician services.

Public Option

Some states are attempting to spur competition by creating a "public option" for the provision of individual health care insurance (as discussed in Chen 2024). Washington and Colorado have just started such plans, while Minnesota and Nevada have plans to do so. However, the state-level approach to providing a "public option" has typically involved the state hiring a private insurer not active in the state to administer the public option—with a mandate to set provider rates lower than the norm for private insurers. The strategy is seen as a combination of (1) adding another insurer to the market to increase insurer competition and (2) lowering provider payment rates through that insurer setting rates as a percentage of Medicare rates.

Another group of states has been taking a different approach to provider rates, setting targets for growth in either statewide or individual spending per capita that are based on projected or actual growth in state gross domestic product (as reported

by Flaherty and Angeles 2025). The approach has been moving from "naming and shaming" providers who exceed the targets toward potential financial penalties for overshooting the targets. A key question is whether there is enough political clout in these states pursuing this approach to significantly slow growth in spending.

Conclusion

The US health care industry surely lies far from any textbook model of perfect competition, but nonetheless, competition does play a major role. Insurance firms compete with each other to provide employer-sponsored health insurance to households. Hospitals and physician practices compete to be part of provider networks. Even in public health insurance programs like Medicare and Medicaid, a majority of enrollees are enrolled with a private company that bids for their business. The state-run Marketplaces for individual health insurance connect individual buyers with private insurance companies.

Economists are deeply familiar with the idea that competition is a two-edged set of incentives, especially in markets characterized by imperfect information and imperfect competition. On one side, the incentives of competition can encourage providers to provide lower costs and higher quality; on the other side, they can encourage providers to game the competitive mechanisms design to create such a market or cut costs by limiting access to care. The increased reliance on competition may have motivated additional merger activity to blunt competition's effectiveness.

The ongoing research challenge in US health care markets is to understand the consequences of the evolving blend of competition and regulation. The ongoing policy challenge is to build on the research results for more effective forms of managed competition in the health care industry.

References

Arrow, Kenneth J. 1963. "Uncertainty and the Welfare Economics of Medical Care." *American Economic Review* 53 (5): 941–73.

Atkinson, Graham. 2009. "State Hospital Rate-Setting Revisited." Commonweath Fund Issue Brief, October 27. https://www.commonwealthfund.org/publications/issue-briefs/2009/oct/state-hospital-rate-setting-revisited.

Chen, Lanhee J. 2024. "Early Lessons from State-Based Public Option Plans." *JAMA Health Forum* 5 (3): e240973.

Chernew, Michael, Zack Cooper, Eugene Larson Hallock, and Fiona Scott Morton. 2021. "Physician Agency, Consumerism, and the Consumption of Lower-Limb MRI Scans." *Journal of Health Economics* 76: 102427.

Conover, Christopher J., and James Bailey. 2020. "Certificate of Need Laws: A Systematic Review and Cost-Effectiveness Analysis." *BMC Health Services Research* 20: 748.

Decision Lab. n.d. "What Is the Paradox of Choice." thedecisionlab.com/reference-guide/economics/the-paradox-of-choice (accessed March 9, 2026).

DeSilver, Drew. 2026. "What the Data Says about Affordable Care Act Health Insurance Exchanges." Pew Reseach Center, January 22. https://www.pewresearch.org/short-reads/2026/0½2/what-the-data-says-about-affordable-care-act-health-insurance-exchanges/.

Enthoven, Alain C. 1978a. "Consumer-Choice Health Plan—Inflation and Inequity in Health Care Today: Alternatives for Cost Control and an Analysis of Proposals for National Health Insurance (First of Two Parts)." *New England Journal of Medicine* 298 (12): 650–58.

Enthoven, Alain C. 1978b. "Consumer-Choice Health Plan—A National-Health-Insurance Proposal Based on Regulated Competition in the Private Sector (Second of Two Parts)." *New England Journal of Medicine* 298 (13): 709–20.

Federal Trade Commission. 2008. "Evanston Northwestern Healthcare Corporation and ENH Medical Group, Inc." FTC Cases and Proceedings, April 29. https://www.ftc.gov/legal-library/browse/cases-proceedings/0110234-evanston-northwestern-healthcare-corporation-enh-medical-group-inc.

Finkelstein, Amy. 2007. "The Aggregate Effects of Health Insurance: Evidence from the Introduction of Medicare." *Quarterly Journal of Economics* 122 (1): 1–37.

Flaherty, Grace, and January Angeles. 2025. "Beyond Public Reporting: Strengthening Accountability to States' Cost Growth Targets and Leveraging Targets in Health Care Oversight." Milbank Memorial Fund Issue Brief, June 24. https://www.milbank.org/publications/beyond-public-reporting-strengthening-accountability-to-states-cost-growth-targets-and-leveraging-targets-in-health-care-oversight/.

Ginsburg, Paul B. 2007. "Shopping for Price in Medical Care." *Health Affairs* 26 (S2): W208–16.

Glied, Sherry A., and Adam Striar. 2016. "Looking under the Hood of the Cadillac Tax." Commonwealth Fund Issue Brief, June 8. https://www.commonwealthfund.org/publications/issue-briefs/2016/jun/looking-under-hood-cadillac-tax.

Hinton, Elizabeth, and Jada Raphael. 2025. "10 Things to Know about Medicaid Managed Care." KFF, February 27. https://www.kff.org/medicaid/10-things-to-know-about-medicaid-managed-care/.

Hsiao, William C., Peter Braun, Douwe Yntema, and Edmund R. Becker. 1988. "Estimating Physicians' Work for a Resource-based Relative-Value Scale." *New England Journal of Medicine* 319 (13): 835–41.

Hsiao, William C., Harvey M. Sapolsky, Daniel L. Dunn, and Sanford L. Weiner. 1986. "Lessons of the New Jersey DRG Payment System." *Health Affairs* 5 (2): 32–43.

Lambrew, Jeanne M., and Christen Linke Young. 2025. "Lessons from the ACA: Simplifying Choices to Optimize Health Coverage." Commonwealth Fund Issue Brief, December 2. https://www.commonwealthfund.org/publications/issue-briefs/2025/dec/lessons-aca-simplifying-choices-optimize-health-coverage.

MedPAC. 2026. "The Medicare Advantage Program: Status Report." In *Report to the Congress: Medicare Payment Policy*, 343–472. MedPAC.

Mitchell, Matthew D. 2024. "Certificate of Need Laws in Health Care: Past, Present, and Future." *INQUIRY: The Journal of Health Care Organization, Provision, and Financing* 61: 00469580241251937.

Murray, Robert M., and Katherine L. Gudiksen. 2024. *U.S. State-Based Hospital Rate Setting: What Worked, What Didn't, and What We Need to Do Now.* The Source on Healthcare Price and Competition.

Pauly, Mark V. 1968. "The Economics of Moral Hazard: Comment." *American Economic Review* 58 (3): 531–37. https://www.jstor.org/stable/i331465

Shain, Max, and Milton I. Roemer. 1959. "Hospital Costs Relate to the Supply of Beds." *Modern Hospital* 92 (4): 71–3.

US Department of Justice and the Federal Trade Commission. 1993. "Antitrust Enforcement Policy Statements Issued for Health Care Industry." September 15. https://www.justice.gov/archive/atr/public/press_releases/1993/211661.htm.

Wilensky, Gail R., and Amy K. Taylor. 1982. "Tax Expenditures and Health Insurance: Limiting Employer-Paid Premiums." *Public Health Reports* 97 (5): 438–44.

Journal of Economic Perspectives—Volume 40, Number 2—Spring 2026—Pages 17–42

Competition in Health Insurance Markets

Martin Gaynor and Amanda Starc

The United States relies on markets for the majority of its health insurance system. Approximately 238 million Americans—78 percent of those with health insurance and 70 percent of the total population—obtain their health insurance from a private health insurance company. Most of these are people receiving private health insurance through their employer, but an increasing share of consumers receive private health insurance that is publicly-subsidized—for example, more than half of all Medicare enrollees participate in the Medicare Advantage program, in which the federal government pays premiums to a private insurance company to provide Medicare coverage. However, the data clearly suggest that each of these US health insurance markets is highly concentrated, and that these markets appear to be becoming more concentrated over time.

Why is this the case? One reason is that health insurance, like many US industries, has experienced a rise in concentration in the past several decades (for example, Autor et al. 2020; Covarrubias, Gutiérrez, and Philippon 2019; Ganapati 2021).[1] Beyond general economy-wide trends, health insurance markets may exhibit natural tendencies toward concentration (Dafny 2026), due to the inverse relationship between size and risk, economies of scale in claims processing, the impact of

■ *Martin Gaynor is University Professor of Economics and Public Policy Emeritus, Heinz College, Carnegie Mellon University, Pittsburgh, Pennsylvania. Amanda Starc is Professor of Strategy, Kellogg School of Management, Northwestern University, Evanston, Illinois. Both authors are Research Associates, National Bureau of Economic Research, Cambridge, Massachusetts. Their email addresses are mgaynor@cmu.edu and amanda.starc@kellogg.northwestern.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20251459.

[1] For an alternative view and measures, see Benkard et al. (2023).

rising fixed and sunk costs, size advantages in contracting, and asymmetric information. Insurers with larger numbers of enrollees bear less risk and can spread risk over a larger number of policy holders, because sample variance declines with sample size. Larger insurers can spread fixed administrative costs across a broader enrollment base. Fixed and sunk costs may also be increasing due to the cost of investments in information technology, which are particularly important in health insurance (for example, Berry, Gaynor, and Scott Morton 2019; Sutton 1991). If so, then insurers with a large customer base will be better able to make these investments and defray the costs. Larger insurers may also achieve better negotiating positions with hospitals and physicians, either via better management/negotiation skills or simply by their sheer size, creating cost advantages that smaller competitors struggle to match. Increases in concentration may also occur due to mergers fueled by the desire to reduce competition and increase market power.

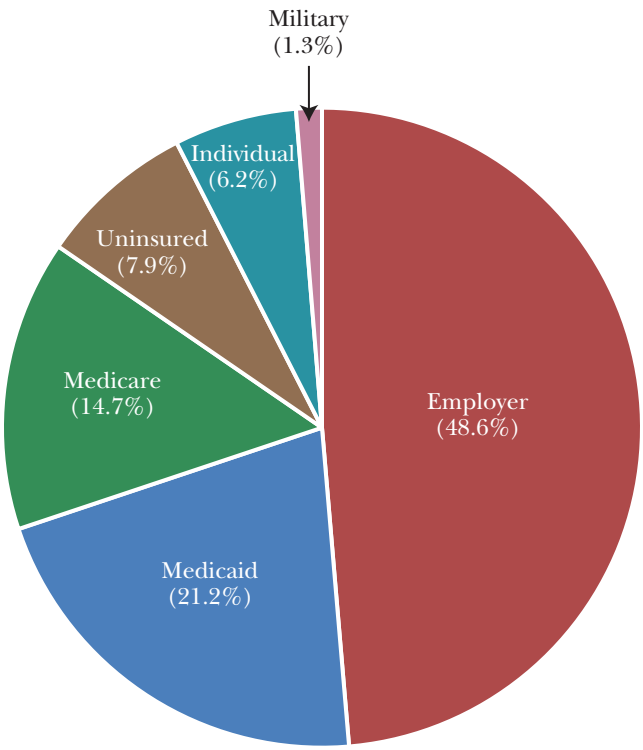
What does this mean for consumers? The benefits of markets flow from the incentives that arise when competitive sellers face pressures from buyers for lower prices and improved quality. In health insurance markets, these incentives are muted, which could mean higher premiums and less generous coverage. Indeed, insurers seem to be able to exert market power (Dafny 2010; Starc 2014) and prices rise following mergers (Dafny, Duggan, and Ramanarayanan 2012).

In this paper, we argue that the US health insurance system will work only as well as the health insurance markets that underpin it. But the challenges of policy support for a competitive health care market go beyond straightforward questions of a concentrated market. Health insurance, along with the health care industry as a whole, is beset by issues of imperfect information, particularly adverse selection. For insurance companies, these issues create a number of incentives to behave in ways that undercut the performance of markets: for example, structuring insurance policies that tend to attract healthier-than-average enrollees; finding ways to stint on care that may be difficult to observe; cutting prices in the hope of attracting healthier-than-average enrollees; and more. As concentration rises in commercial and government-sponsored markets, the resulting dynamics of competition in the context of imperfect information are critical for premiums, coverage, plan choices, and ultimately consumer welfare. Much remains to be done in formulating "rules of the road" to usefully guide competition in health insurance markets.

Market Structure

We begin by discussing concentration in US health insurance markets. Figure 1 summarizes the sources of insurance for Americans in 2023 (the latest year for which comprehensive data are available). Nearly half of the population had private employer-sponsored (commercial) insurance. Most of the remainder were covered by the two large public insurance programs—Medicare (14.7 percent) and Medicaid (21.2 percent). Individual and military coverage were less common, and approximately 7 percent of Americans remained uninsured. We begin with the

Figure 1
Health Insurance Share: Total Enrollment by Type



Source: KFF (2025a).
Note: This shows the total enrollment US share among different types of health insurance providers. Dual-eligible individuals are included in the "Medicaid" share.

commercial market, categorizing it into three main groups: large (employer) group, small (employer) group, and individuals. We then turn to government-sponsored insurance programs. We also provide some evidence on vertical integration.

Commercial Health Insurance Markets

Americans with private health insurance mainly obtain it through their employers. As of 2024, about 54 percent of American companies offered employer-sponsored insurance. These plans covered approximately 178 million people, accounting for around 60 percent of the nonelderly population or 63 percent of working-age adults (KFF 2025c; Keisler-Starkey and Bunch 2024). Large-group employer-sponsored health insurance plans are generally "self-insured," meaning the employer pays medical claims directly, while the insurer provides claims processing and connections to a provider network for a fee. In contrast, small-group employer-sponsored insurance plans—often provided by businesses with fewer than

50 employees—are typically "underwritten," meaning the insurer bears the risk and pays claims. The small-group plans tend to incur higher administrative costs per enrollee and exhibit greater variability in plan choices, which can result in higher premiums and reduced coverage options for workers in smaller firms (Cubanski and Damico 2024).

For individuals without employer-sponsored coverage, such as the self-employed or those who are unemployed, individual health plans from the state-level Marketplaces established by the Patient Protection and Affordable Care Act of 2010 are the primary private insurance option. Marketplace plans are regulated and subsidized based on income, making private insurance more affordable. The Centers for Medicare & Medicaid Services (CMS 2025) reported in January that nearly 24 million people had enrolled in Marketplace plans for 2025.

To characterize concentration in the commercial market, we rely on data reported by the Kaiser Family Foundation. Because health insurance is regulated at the state level (and thus data on market shares are most readily available by state), we report statistics on market structure at the state level.[2] In the large group market, the average market share of the largest insurer in 2023 was 64 percent. The analogs for the small group and individual markets are 65 percent and 53 percent, respectively. When expanding to the top three insurers, the analogous percentage for the large group market is 91 percent, the small group market is 94 percent, and the individual market is 79 percent (89 percent). By any metric, commercial markets are highly concentrated (KFF 2025a).

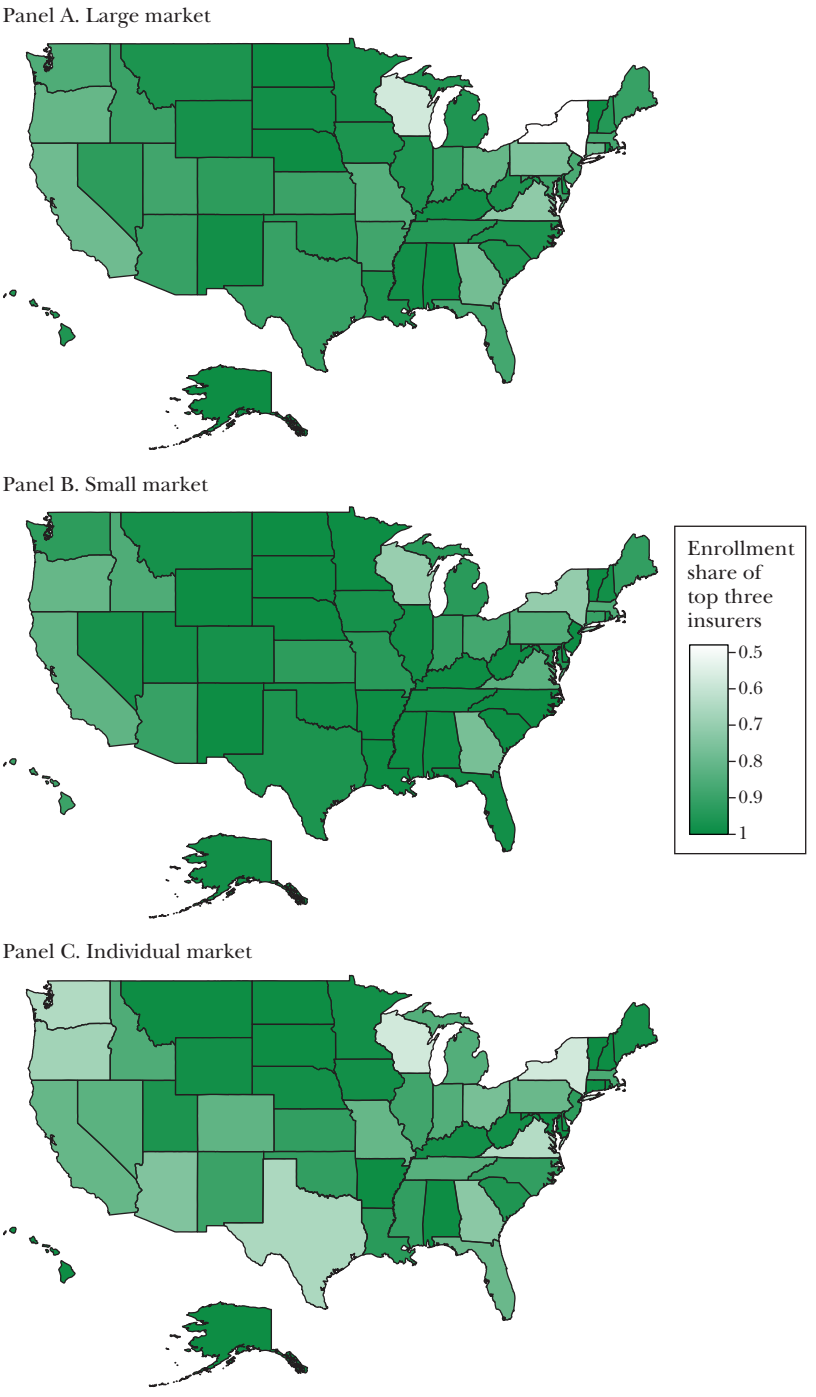
In addition, there is little entry of new health insurance firms—instead, the market is dominated by a small number of firms that have been in the market for years. New entrants attract enrollees and grow by offering low premiums. However, new entrants are unable to negotiate favorable rates with health care providers because they have relatively few enrollees to direct to providers (as a *quid pro quo* for offering lower rates). As a consequence, new entrants have higher expenses than established firms, and are unable to offer low premiums without losing money. This makes de novo entry into health insurance exceedingly difficult.

There is also heterogeneity across states, as shown in Figure 2, which displays the combined enrollment share of the three largest insurers in each state for individual, small group, and large group markets, respectively. Each state is shaded according to the percentage of the market these top insurers control. Darker shades represent states where the top three insurers have a higher market share, sometimes exceeding 80 percent or even 90 percent of the market. Lighter shades show areas with more evenly distributed market shares among insurers. The maps indicate limited potential competition, especially in the small group market. These maps provide a quick overview of market concentration differences across states and segments. The market share of the largest insurer in the large group market

[2]For purposes of analyzing the extent of competition by health insurers in states, we recognize that markets are not necessarily states, and are likely smaller. Among other reasons, health care plan provider networks are mostly local, making plans in different local areas poor substitutes.

Figure 2

Enrollment Share of Three Largest Large-Market Insurers by State, 2023



Source: KFF (2025b).

Note: These maps show the market share of the largest three insurers by cumulative enrollment for each state and market type.

ranges from 17 percent in New York to 94 percent in Alabama, where Blue Cross/Blue Shield has a *de facto* monopoly. While the cross-state variation is nontrivial, the vast majority of states have market shares concentrated among a small number of insurers.

Government-Subsidized Provision of Private Health Insurance

Publicly-subsidized health insurance operates through both federal or state programs. Medicare is a federal program for individuals aged 65 and older, as well as some individuals with disabilities or certain diseases who are younger than 65. By 2023, the Medicare program enrolled approximately 62.6 million Americans (Keisler-Starkey and Bunch 2024). Traditional Medicare consists of hospital insurance in Part A and coverage for physicians, outpatient care, and other services in Part B. In 1997, Medicare Part C or Medicare Advantage was created, which allowed beneficiaries the option to enroll in a privately administered health insurance plan—with the federal government paying the premiums. In 2006, Medicare Part D was added, providing insurance coverage for drugs.

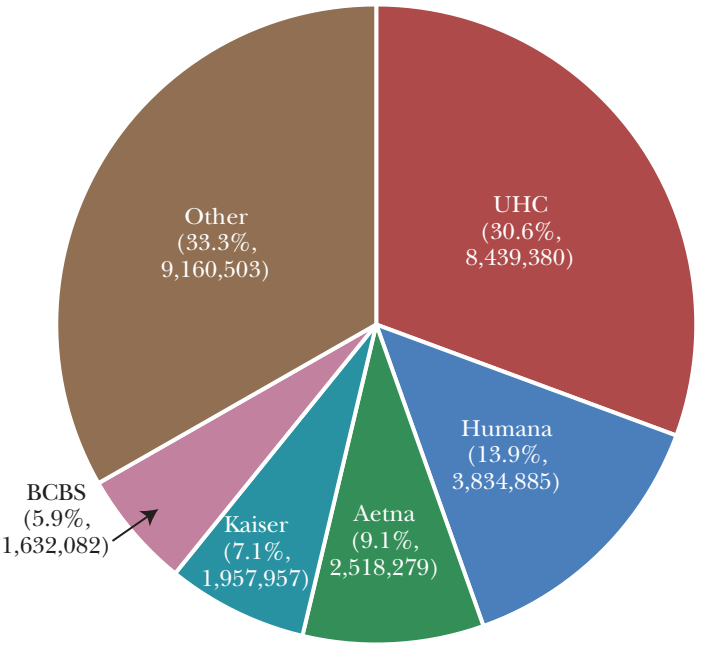
Medicare Advantage has become steadily more popular over time. Nearly 51 percent of all Medicare beneficiaries—about 35.7 million people—are projected to be enrolled in Medicare Advantage in 2025 (CMS 2024). In addition, most Medicare Advantage plans include integrated prescription drug benefits, making them the main option for drug coverage under Medicare. Over half (57 percent) of all Part D enrollees are in Medicare Advantage drug plans, while stand-alone prescription drug plans account for about 43 percent (Cubanski and Damico 2024).

For Medicare Advantage, the relevant geographic market is the county, as defined by the Centers for Medicare and Medicaid Services, the federal government agency that runs Medicare. For Part D prescription drug coverage, only private insurers sell plans; the relevant geographic market is a region.[3]

Geographic differences heavily affect plan availability and enrollment in Medicare Advantage. Beneficiaries in cities have many more options (45 plans on average) than those in rural areas (27 plans). For instance, some counties in Pennsylvania or Ohio offer over 50 Medicare Advantage plans, while some rural counties and Alaska may have very few or none. Enrollment rates mirror this pattern, with higher enrollment in urban areas. Major insurers, such as UnitedHealthcare, Humana, and CVS Health, are present in almost all counties (Cubanski and Damico 2025; Hammond, Johnson, and Serna 2025). Yet large national players play a disproportionately large role. We construct national market shares by insurer for health maintenance organizations, preferred provider organizations, and private fee-for-service plans within the Medicare Advantage program. Again, while this program is publicly financed, the program is administered by private

[3]There are 34 regions for Medicare prescription drug plans comprising one or multiple states (see https://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovGenIn/downloads/PDPRegions.pdf).

Figure 3
Medicare Advantage National Enrollment Totals, July 2025



Source: CMS (2026).
Note: This chart shows the national enrollment totals of the largest Medicare Advantage insurers as of July 2025. The enrollment of subsidiary organization has been added to that of its respective conglomerate (for example, Sierra Health and UnitedHealthcare). Chart includes only preferred provider organizations, health maintenance organizations, and private fee-for-service plans.

insurers. As is shown in Figure 3, UnitedHealthcare captures nearly one-third of the market, and large players as a group hold over two-thirds of the market.

Both the Medicare Advantage program and the Part D drug insurance program have seen significant consolidation. The five largest plan sponsors cover about 73 percent of enrollees. The number of plans and sponsors has decreased, especially among standalone prescription drug plans. This concentration allows larger firms to charge prices higher due to less competition. Medicare Advantage plans can use rebates in the form of additional covered services or lower premiums,[4] so most of those who enroll in prescription drug plans via Medicare Advantage pay little or nothing for their Part D premiums. Part D prescription drug plans' premiums

[4]Medicare sets a "benchmark" rate for Medicare Advantage premiums in each county. Plans that set their premiums (called "bids") below the benchmark rate receive a portion of the difference from Medicare. The plans are then required to use that to provide additional benefits to enrollees (for example, coverage for prescription drugs, dental care, vision, hearing, and ancillary benefits) or to reduce the premium enrollees pay.

average $39, while premiums for those who enroll in Part D plans via Medicare Advantage are around $7 in 2025. This premium difference has contributed to the high market share of Medicare Advantage plans (Cubanski and Damico 2025).

Consolidation and the resulting market concentration in Medicare Advantage has drawn the attention of antitrust enforcers. In 2016, the Antitrust Division of the Department of Justice challenged a proposed merger between the insurance companies Aetna and Humana (US Department of Justice 2018). A central issue in the case was the extent to which Medicare beneficiaries treat Medicare Advantage and traditional Medicare as substitutes. The government alleged that they are not close substitutes (Medicare beneficiaries tend to choose one and stick with it) they are highly unlikely to switch from one to the other (which means that Medicare Advantage and traditional Medicare are separate markets). The court agreed, found that Aetna and Humana were close competitors in Medicare Advantage without other firms serving as significant alternatives, and blocked the merger (*United States v. Aetna Inc. and Humana Inc.*, Civil Action No. 16-1494 (JDB), Document 306 (2017)).[5]

The Medicaid program provides health coverage to low-income individuals and families. Taken together, Medicaid and the Children's Health Insurance Program cover over 78 million people nationwide as of May 2025 (Medicaid.gov 2025). They collectively accounted for nearly one-fifth of national health care spending in 2022 (MACPAC 2026). In most states, Medicaid beneficiaries are enrolled in Medicaid managed care plans provided by private firms. Private plans compete for contracts with the state Medicaid program via competitive bidding, then beneficiaries choose among the plans that have been awarded contracts. As of 2022, approximately 85 percent of Medicaid beneficiaries were enrolled in some type of managed care plan in the United States (CMS 2022).

As also shown in Figure 1, there are two plans covering military service members. TRICARE is a health insurance program for active-duty service members, their families, and retirees, which often pays for care provided by civilian doctors. Veterans Affairs health benefits are provided in VA facilities and serve eligible veterans—and in some cases, their dependents (US Department of Veterans Affairs 2025).

Vertical Consolidation

Along with the horizontal consolidation of health insurance markets, vertical consolidation has become increasingly common, as insurers acquire physician groups, data firms, pharmacy benefit managers, and specialty providers. The largest insurance companies have assembled health care conglomerates with extensive ties across the care continuum (Brot-Goldberg, Che, and Handel 2022).

For example, UnitedHealth Group, through its Optum subsidiary, has significantly expanded its footprint in the US health care market over the past decade

[5] The demonstratives used by the governments' expert witnesses in their testimony for this case, particularly those of the economics experts Aviv Nevo and Richard Frank, are available at https://www.justice.gov/atr/us-and-plaintiff-states-v-aetna-and-humana.

by acquiring numerous physician practices. A pivotal acquisition was Optum's $4.3 billion purchase of DaVita Medical Group in 2019, an acquisition that was subject to review by the Federal Trade Commission and led to DaVita's integration into OptumCare, dramatically expanding Optum's reach in direct patient care. Subsequent deals, such as the $236 million acquisition of Atrius Health in Massachusetts and the absorption of CareMount Medical, ProHEALTH, and Riverside Medical Group into the Optum Tri-State network, have further solidified Optum's presence in multiple regions (UnitedHealth Group 2019; Federal Trade Commission 2019; Japsen 2019; Morse 2022).

These physician practice acquisitions have not only fueled UnitedHealth Group's revenue growth—from $130.47 billion in 2014 to over $400 billion in 2024—but also shifted the employment landscape for US physicians. By 2022, Optum was employing tens of thousands of doctors, contributing to an ongoing trend where independent physician practices have become increasingly rare. As recently as 2025, Optum was continuing its run of physician practice acquisitions.[6]

In addition to UnitedHealth Group, the largest health insurers in the country, Humana, Elevance Health (formerly Anthem), and CVS Health (with Aetna), have also increased their degree of vertical integration. Each organization has made significant acquisitions to combine insurance, pharmacy, care delivery, and home health services, as well as being major holders of health care data and providers of data analytics. Humana targeted the senior market and Medicare Advantage; Elevance Health expanded its presence in Medicare Advantage as well. CVS Health's $69–$77 billion acquisition of Aetna in 2018 was a major vertical merger. It combined CVS's retail pharmacy, MinuteClinics, specialty distribution, and Caremark pharmacy benefit managers with a major insurer. CVS also expanded care delivery by acquiring Signify Health (in-home clinical care for $8 billion) and Oak Street Health (value-based primary care for $10.6 billion) (CVS Health 2018, 2022, 2023).

On one side, these integrated models may offer the potential for enhanced efficiency by having everything "under one roof," as well as by giving insurance companies greater leverage to negotiate for lower prices. But they also raise concerns. These huge conglomerates may lead to greater market power for them as insurers. They might also leverage control over related services (possibly the entire health care ecosystem), potentially foreclosing competition and extracting additional profits at the expense of consumers (Kanter and Gaynor 2025).

There is another aspect to this integration that stems from the medical loss regulation (MLR) on health insurance introduced by the Affordable Care Act. The MLR requires insurers to have medical expenses that are at least a minimum proportion of premiums—85 percent for large group plans. The fact that the MLR requires insurers to have a minimum proportion of medical expenses relative to premiums gives insurers an incentive to acquire medical providers and then inflate their payments to those providers. This enables an insurer to be in compliance with

[6] In August 2025, Optum acquired Tennessee-based Holston Medical Group (Cass 2025).

the MLR requirement, but without actually spending more money, as the additional amounts paid to providers owned by insurers are actually profits kept "in-house" at the firm.

Insurers then have an opportunity to exploit market power that they could not do if constrained by the medical loss ratio rules. To illustrate, consider an insurer with market power that is constrained by the medical loss regulation. This insurer cannot raise their price to exploit the market power they possess without being in violation of the MLR.[7] However, if the insurer acquires providers and then inflates their payments, the MLR constraint is no longer binding (medical expenses have increased while premiums have not). The insurer can thus exploit their market power and increase premiums while being in compliance with the MLR. Insurer acquisitions of medical providers in order to exercise market power can thus constitute a form of regulatory evasion (Baker et al. 2019).

Antitrust authorities have scrutinized several of these deals, occasionally imposing conditions aimed at preserving market competition. For example, the Department of Justice sued to block UnitedHealth Group's acquisition of home health and hospice care provider Amedisys (US Department of Justice 2024). The precise nature of the ownership structures of these large entities is often opaque and hard to ascertain, making analysis of, and policy toward, the impacts of ownership changes challenging.

Asymmetric Information and Competition

How Does Asymmetric Information Contribute to Health Insurance Consolidation?

Health insurance markets are fundamentally characterized by asymmetric information, which shapes how health insurance is structured and how markets for health insurance perform. The two types of information asymmetries are moral hazard and adverse selection.

The tradeoff between risk protection and moral hazard is key in designing insurance contracts (Zeckhauser 1970), especially in health insurance. On the one hand, insurance provides risk protection—it helps individuals avoid significant, unexpected financial losses. Good insurance allows people to manage expenses, avoid uncertainty, and lessen the hardship from bad events. On the other hand, moral hazard occurs as a matter of course, because insurance reduces the marginal cost paid for health care. Because individuals rationally consume to the point where marginal benefits equal the marginal costs they bear, lower effective prices lead consumers to use more services than is socially optimal. A (second-best) optimal contract must strike a balance between the additional risk protection from more extensive insurance coverage and the increased moral hazard that accompanies it.

[7] If insurers exploit their market power, premiums will increase and the ratio of medical expenses to premiums will fall below the minimum required by the regulation.

It is unclear how competition affects insurance contracts in the presence of moral hazard—in particular, does competition promote optimal contracts, or does it perhaps result in insurers offering excessive coverage in an attempt to attract consumers?[8] The latter results in greater moral hazard than is optimal, driving up spending and excess utilization. If this is the case, then competitive health insurance markets will not result in optimal coverage. We do not know of research on this topic; both theory and empirics are needed to cast light on this important issue.

Adverse selection refers to a situation where one party knows something about themselves that affects relevant outcomes, but the other party does not. In health insurance, consumers are likely to have better knowledge than an insurance company how whether they have an idiosyncratic risk that is likely to lead to future health care expenses. Insurers know the expected health care expenses on average in the population, but not for any individual.

In this situation, markets can unravel or even fail to exist, in what is sometimes called an insurance "death spiral." The scenario unfolds this way. Insurers need to set premiums for the health insurance policies they sell that cover the expected health care expenses of those policies (at a minimum). As a simple example, say that the insurer sets a single premium, which in a competitive market is equal to the average population's expected expenses.[9] However, individuals with (private) information that their health expenses will be low will not purchase insurance at this average price. As a result, insurers' expected expenses end up being higher than the premiums they charge, because only people with higher expected expenses buy the insurance. Insurers then need to increase the premiums to cover the expected expenses of the portion of the population that purchases health insurance. But now, only a portion of that remaining population will find it beneficial to purchase health insurance at the new, higher, premium. This cycle can continue, with the market shrinking smaller and smaller, until it ceases to exist (the so-called "death spiral").

In this dynamic, adverse selection in the presence of competition can lead to missing markets, in which sellers become unwilling to participate in the market (Rothschild and Stiglitz 1976).[10] Even when markets exist, plans are likely to be mispriced (Einav, Finkelstein, and Cullen 2010). Consider an insurer with some ability to set prices strategically, who is contemplating raising the price of an insurance policy by $1. The insurer knows that it will lose some consumers, as is the case

[8]We speculate that if firms take account of consumer responses to the coverage they offer, but do not (fully) account for other firms' actions, competitive markets will provide excessive coverage in equilibrium (for example, Mankiw and Whinston 1986; Gaynor 2007). It may also be the case that consumers with employer-sponsored insurance put much more weight on coverage than premiums, because they only pay for a fraction of premiums directly. This could also lead to a competitive equilibrium with supra-optimal coverage.

[9]For simplicity in exposition, we ignore insurers' cost of doing business or market power and assume that premiums are set equal to expected expenses—what is called the "actuarially fair premium."

[10]A long literature highlights that selection can lead to "missing markets" (for an empirical example, see Cutler and Reber 1998).

in any market with downward-sloping demand. However, the consumers the firm loses are not a random subset of the firm's consumers. They are those consumers with the lowest willingness-to-pay (or the most sensitive to price). Under adverse selection, these consumers are likely also the least costly to insure. As a result, adverse selection puts downward pressure on markups.

Adverse selection can also shape market structure and competitive dynamics. In imperfectly competitive selection markets, insurers who lower prices to attract more customers will also disproportionately draw in healthier, lower-cost enrollees. A $1 price cut can reduce an insurer's average costs by more than $1, intensifying the incentive to undercut rivals. Indeed, this aggressive price competition can push prices below the level needed to cover the costs of insuring the entire risk pool. As a result, even with multiple potential entrants, the market may only be able to sustain a single firm. Kong, Layton, and Shepard (2024) discuss this dynamic and refer to it as an "unnatural monopoly."[11] Essentially, adverse selection amplifies the impact of price sensitivity—the most price-sensitive consumers are also the healthiest, making price-cutting even more attractive and further reducing the equilibrium number of firms.

How Can Insurance Firms Respond to Information Asymmetries?

In the simple example above, health insurers set a single price for the market. But of course, insurance companies can set a variety of insurance prices (if allowed to do so by regulation) with variations based on factors like age, smoking status, and previous health history. This step will ameliorate the adverse selection dynamics, but may also end up leading to large differences in premiums. Indeed, the reason for the existence of the Medicare program is that, without it, health insurance would be priced in a way that would make it unavailable for many of the elderly.

Of course, pricing for a given policy is only one of the strategic tools insurers have at their disposal. Insurers can also design a menu of plans, seeking to screen consumers and exclude some consumers from the market. This outcome similarly leads to a deadweight loss, but one that is different in nature than an adverse selection "death spiral" in a perfectly competitive market.

When insurers control prices and contract menus, deadweight loss arises from two main factors: exclusion (where some consumers lack coverage) and screening (where insurers use more contracts to segment the market than is socially ideal). Using a graphical approach, Chade et al. (2024) show how to analyze multiple "levels" (or "margins") of vertically differentiated insurance. Both the intensive (generosity) and extensive (coverage) margin have an impact on optimal regulation (Geruso et al. 2023). Analyzing plan menus is therefore critical, despite the theoretical challenges. If technical conditions are met, the Chade et al. (2024)

[11] They formalize this intuition using a Salop (1979) circle model with horizontal differentiation. They use data from Massachusetts's health insurance exchange to test the model's predictions. They show that lower premiums attract healthier enrollees and reduce average costs nearly one-for-one. Simulations show that, without corrective policies, the market collapses to monopoly.

approach suggests that a monopolist will exclude more consumers and offer a more detailed menu than a social planner.

Gottlieb and Moreira (2023) also examine how market power affects insurance coverage. They similarly demonstrate that market power leads to systematic exclusion. A monopolist typically leaves a portion of consumers without coverage, while competitive markets usually provide some (perhaps minimal) coverage for all. A monopolist can profit more by excluding low-willingness-to-pay customers and raising prices for those who value insurance more. By contrast, competitive firms, who are forced to fear the loss of customers to rivals, reduce coverage for high-risk consumers, but with the ability to offer certain menus and prices, do not entirely exclude them.

Finally, large consolidated insurers have some ability to wield monopsony power. Health care provider markets, including hospitals and physician practices, have also consolidated. Hospital mergers have raised prices, with little evidence of improvements in quality (Gaynor and Town 2012; Gaynor, Ho, and Town 2015). In recent years, consolidation among hospitals and physician practices has grown (Brot-Goldberg et al. 2024; Venkatesh 2024). In this context, negotiations between insurers and providers are key in setting prices. As consolidated insurers gain more bargaining power, they can negotiate lower payments to providers. If these savings are passed on, consumers could see lower out-of-pocket costs or reduced insurance premiums. However, whether consumers benefit depends on the structure of provider and insurer markets and various regulatory and contractual factors.

The same lesson, that while insurer consolidation can sometimes result in lower provider prices, any benefits to consumers will depend on the economic context, emerges from the Ho and Lee (2017) study of bargaining between insurers, employers, and hospitals in California's large-group employer-sponsored health insurance market. Using data from the California Public Employees' Retirement System (CalPERS), they model and simulate the effects of reducing the number of insurers in the market on consumer and overall welfare. Removing a major insurer, like Kaiser Permanente, often raises premiums for the remaining insurers, because as competitive pressure decreases, employers lose bargaining power. However, the impact on hospital prices varies—in some markets, prices increase, while in others, they decrease as remaining insurers gain more bargaining leverage over providers. The study also points out that employer bargaining constraints, like those from large buyers such as CalPERS, help limit premium hikes.

Empirical Evidence

As a starting point, it is worth asking if insurers can price strategically and charge higher premiums to buyers willing to pay more. Using detailed data on large, multisite employers, and controlling for plan design, employee demographics, and local market conditions, Dafny (2010) documents higher growth in health insurance premiums for firms purchasing insurance for their employees that experience

positive profit shocks. The effect is strongest in markets with fewer insurers. Premium increases are not due to changes in plan generosity or employee risk. Instead, she suggests they stem from insurers' ability to extract surplus from buyers with a higher willingness to pay. These findings challenge the idea that health insurance markets are highly competitive. It also suggests further work exploring the relationship between insurer competition and premiums.

Effects of Horizontal Consolidation

If insurers are not simply price takers, changes in market structure are likely to be associated with changes in premiums. For example, the 1999 merger between Aetna and Prudential was a national event that had different effects across local markets, based on each firm's presence in the local market before the merger. To address endogeneity concerns, like whether rising premiums attract mergers, Dafny, Duggan, and Ramanarayanan (2012) focus on the local market concentration changes caused by the merger. On average, large-group premiums in the markets most affected by the Aetna-Prudential merger rose by about 7 percent by 2007, compared to what would have happened without the merger. This increase was not just for the merged firm—rival insurers also raised prices. Insurers also benefited from lower provider reimbursement rates, due to greater bargaining power. In markets with larger concentration increases, physician earnings growth slowed. Thus, insurers in concentrated markets may be able to exercise both monopoly and monopsony power.

A similar methodology can be applied to the decision by United HealthCare in 2014 to exit from the state-level individual insurance Marketplaces. Again, this business decision reduced competition across markets, but the effect was heterogeneous, depending on UnitedHealthcare's earlier market share. Dafny, Gruber, and Ody (2015) find that in markets where UnitedHealthcare would have been a major competitor, premiums for benchmark "silver plans" rose by 5 percent to 11 percent. These results align with those in the large-group market and highlight how crucial insurer competition is for controlling premiums.

Combined Effects of Selection and Competition

High medical costs, selection, and imperfect competition can all drive high premiums. How might these factors interact? The link between imperfect competition and selection matters both theoretically (as described above) and empirically. Consider a situation in which firms have different levels of market power, such that policies aimed at addressing one market failure can inadvertently exacerbate the other. Working with this model, Mahoney and Weyl (2017) find that risk adjustment can hurt consumer welfare and social surplus when firms have market power.

Risk adjustment is a process where a government entity adjusts the premiums insurers receive up or down, depending on the expected expenses of their enrollees. If an insurer's enrollee population is more expensive than average, their premium is adjusted upward to address the higher expense, while an insurer's premium is adjusted downward if their enrollees are less expensive. In this way, risk adjustment

removes incentives for firms to lower prices to attract low-cost consumers, thereby weakening competition. The authors apply their model to both health insurance and credit markets. In health insurance, typical risk adjustment policies often hurt employees and employers, while benefiting insurers. By contrast, in auto lending, competition can lead to excessive lending. In many settings, it is essential to incorporate the presence of selection into models that analyze insurer competition.

In the Medicare program, it is common for enrollees to purchase additional private insurance to cover health care costs or services that Medicare does not—for example, out-of-pocket costs that are part of Medicare plans, or certain hospital, hospice, or foreign medical care services. In a study of these supplemental or "Medigap" policies, Starc (2014) finds that while the policies are often contractually identical (after all, the policies cover the same gaps in Medicare), premiums can vary greatly across insurers. Furthermore, during the period of the study, two firms controlled about 70 percent of the market. There is also evidence of selection—sicker consumers tend to buy more expensive Medigap plans, especially from trusted brands. As a result, there is a strong link between premiums and claims. In addition, demand is not very responsive to price. The low elasticity, driven by brand loyalty and high search costs, allows insurers to keep prices high. These results suggest that imperfect competition, not just selection, drives high Medigap premiums.

Insurance companies that participate in the Medicare Advantage system have an incentive to find ways to make themselves attractive to enrollees who are more likely to have low health care costs. To push back against this selection problem, insurers that wish to participate are required to submit an estimate of the cost of covering a standard beneficiary for each individual plan (a given insurer can offer multiple Medicare Advantage plans). If this estimate is lower than the county-level benchmark cost, then the proposed plan receives a rebate; if higher, the plan must charge the extra amount to beneficiaries. The goal is to provide a financial incentive for enrollees to choose low-cost, high-quality plans that also accord with their preferences. Curto et al. (2021) estimate that this system generates significant economic surplus. Critically, their model incorporates both market power and risk selection. They emphasize that residual unobserved risk selection may still distort plan incentives and enrollment patterns. But while private Medicare Advantage plans can deliver cost savings over traditional Medicare, much of the benefit (about two-thirds) accrues to the insurers themselves. The authors then demonstrate that modifying market design—such as increasing the rebate pass-through rate and reducing benchmark payments—can enhance consumer surplus and competition, without incurring additional taxpayer costs.

Other work on the interaction of adverse selection and competition involves the state-run individual health insurance Marketplaces established by the Affordable Care Act. In Massachusetts, the Marketplace used a modified community rating system, which limits the potential for selection by also limiting how much more older consumers could be charged compared to younger ones. However, younger consumers respond to price changes more than older ones. Even with perfect risk adjustment, differences in preferences lead to different markups, giving insurers

a reason to price discriminate. Ericson and Starc (2015) present simulations show that current age rating bands raise premiums for younger consumers by a significant amount. The effect is larger when insurers price strategically. Without a strong mandated requirement that young people must have health insurance, healthier, price-sensitive consumers might leave the market.

In the California Marketplace, market power and adverse selection interact as well. Inertia, defined in this case as consumers sticking with their current plan, bestows insurers with significant market power (Saltzman, Swanson, and Polsky 2021). Switching costs amount to nearly half the annual premium, and four firms dominate the market. Removing inertia could lower average premiums and raise annual per-capita welfare substantially. However, this effect is negligible in markets without strategic pricing or risk adjustment. Ultimately, improving consumer welfare requires attention to both adverse selection and insurer pricing strategies, as well as the barriers that limit effective competition.

Competition can also influence the efforts of insurers to engage in "cream-skimming" by minimizing their enrollment of those likely to have high health care costs. In the context of the Massachusetts Health Insurance Exchange, Shepard (2022) shows that competition in hospital networks is affected by adverse selection. Plans that exclude costly "star" hospitals attract healthier patients. These "star" hospitals often draw sicker, higher-spending patients, even with risk adjustment—and so for the insurance companies, this form of cream-skimming is profitable.

In a study of health insurance in Colombia, Serna (forthcoming) uses a structural model to show that insurers' choices about network size balance risk selection and administrative costs. Insurers may exclude unprofitable patient groups by limiting access to costly services. In tightly regulated markets, some insurers offer broad networks due to cost efficiencies. This partially counters the exclusion effects of risk selection. Again, the intersection of competition among insurers and pressures for selection involves complex trade-offs.

Effects of Vertical Consolidation

Despite a well-developed literature on the effects of horizontal consolidation in the health insurance industry, there is less understanding of vertical relationships between insurers and providers. In seminal theoretical work on this topic, Ma (1997) models the impacts of vertical mergers between upstream and downstream firms when the upstream firms' products are differentiated. This applies to health care providers (upstream) and health insurers (downstream). He shows that a merger between a health insurer and a health care provider can lead to foreclosure—that is, limiting access of other health care insurers to a full range of providers—harming competition in the downstream health insurance market and leaving consumers worse off.

However, in the US health insurance industry, these vertical relationships are, for the most part, relatively recent and data on the contracts are limited. Models that address the details of these for complex markets are still being developed, and

empirical work is only now beginning to accumulate across a variety of health insurance contexts.

Most empirical studies about vertical integration between health care insurers and providers have examined how insurance design affects care delivery, with few investigating competitive dynamics. Among these, research on competition and vertical integration has mainly focused on pharmacy benefit managers—that is, companies that contract with insurance companies to manage pharmaceutical benefits, a task that includes both deciding what drugs will be available to consumers and negotiating prices with drug manufacturers. In a thorough overview of this research, Brot-Goldberg, Che, and Handel (2022) highlight the limited understanding of the vertical ties between pharmacy benefit managers, insurers, and drug manufacturers. However, the top three pharmacy benefit managers now cover nearly 80 percent of the market and are vertically integrated with major insurers. They argue that standard policy analysis tools fail in these multi-sided markets. In their preferred framework, the welfare effects of integration between pharmacy benefit managers and insurers are unclear and depend on factors like bargaining power, pass-through rates, and regulations. They also note that the current literature struggles due to data opacity, especially concerning rebates and contracts.

Gray, Alpert, and Sood (2026) provide additional, reduced-form evidence on vertical integration between insurers and pharmacy benefit managers. They track market outcomes before and after the major 2015 merger between UnitedHealth and Catamaran. By linking a unique dataset on contracts between insurers and pharmacy benefit managers to plan-level enrollment and pricing, they show that as vertical integration increased, nonintegrated insurers using a rival's pharmacy benefit managers faced much higher premiums. Notably, they find no premium reductions for consumers of the merged firm. Their evidence suggests that as pharmacy benefit managers consolidate and integrate vertically, competitive harms to nonintegrated rivals increase, while consumer benefits, in the form of lower prices, remain limited.

In Medicare Part D, enrollees are offered an optional drug benefit, in which people may purchase drug insurance in exchange for a monthly premium (which varies by income and location), along with deductibles and copays. Insurance companies that offer such plans may work with or own a pharmacy benefit manager. Yde (2025) develops an empirical model of the Medicare Part D market. Using new data on relationships between insurers, pharmacy benefit managers, and pharmacies themselves, he finds evidence of "profit tunnelling." The insurance companies that offer Part D plans face government rules that tend to limit their profits, including cost-sharing rules and medical loss ratio rules, which require that a certain share of premiums paid be spent on benefits for enrollees. However, pharmacies owned by the insurance companies do not face such rules, and so the parent insurer has an incentive to shift profits from regulated insurance to less-regulated pharmacy operations. His simulations suggest that separating vertically integrated pharmacies could lower drug prices, slightly raise average premiums, and increase consumer surplus. In related work using a vertical model of drug and insurance

demand, Che (2025) shows that bargaining over drug rebates between pharmacy benefit managers and drug manufacturers can influence both level and dispersion of drug prices.

Chile has also experienced a wave of vertical integration between insurers and hospitals. Cuesta, Noton, and Vatter (2025) examine this and show that integrated insurers preferentially include their own hospitals in their networks and limit access to rival hospitals. They then offer more attractive insurance products, lowering cost-sharing and premiums, to attract patients. Nonintegrated insurers respond by excluding integrated hospitals from their preferred networks, fragmenting the market and lessening price competition. While vertical integration does lead to some efficiencies (it reduces double marginalization and in general gains in efficiency are expected from mergers of complements), the gains are offset by distorted plan design and decreased hospital competition, resulting in lower consumer welfare.

As a final example, a health care "platform" is the name given to entities that incorporate health insurers, health care providers of multiple types, health care data holding and analysis entities, and other related firms. As mentioned previously, Kanter and Gaynor (2025) raise the issue that such platforms hold the potential for serious harm to competition by substantially raising the costs of entry, creating (or expanding) strong incentives for self-preferencing, and expanding opportunities for exploiting market power via regulatory gamesmanship. Their discussion is related to concerns about tech platforms (for example, Stigler Center 2019)—a rigorous exploration of these potential impacts is an important avenue for theoretical research.

Public Policies for Health Insurance: Rules of the Road?

Health insurance is fundamentally designed to mitigate uncertain and potentially catastrophic medical expenses through risk pooling and financial protection. In addition, health insurance is more than a financial instrument—it shapes access to care and can produce meaningful effects on health outcomes, both at the extensive margin (who gets insured) and the intensive margin (how much care is used) (Abaluck et al. 2021; Acquatella and Marone 2025). In the United States, health insurance is provided mainly through markets, and how competition functions in health insurance markets is a key factor in determining how well the US system of health care works.

However, these markets do not work as well as they could, or should. The substantial market imperfections in this sector, both due to asymmetric information and due to market power, require significant monitoring and oversight to improve market performance.

Antitrust enforcement is one important policy tool to improve market performance. Recent years have seen extensive consolidation among insurers, both from horizontal mergers and vertical integration with provider groups, resulting in heavily concentrated health insurance markets. Antitrust enforcers monitor

proposed mergers and acquisitions and seek to prevent those assessed as harmful to competition. They also monitor and seek to prevent anticompetitive behavior: collusion, refusals to deal, exclusive dealing, tying, and so on. The purpose of antitrust enforcement is intended to deter other firms from seeking to engage in actions harmful to competition, so the total impact is larger than just the specific cases where the antitrust authorities intervene.

In antitrust cases involving the health insurance industry, modeling the behavior of imperfectly competitive insurers is critical, as the nature of the welfare loss changes when firms price strategically, rather than taking the market price as given. A market with selection pressures will have different incentives for post-merger price changes than a market without. Thus, total welfare can be higher or lower under monopoly compared to competition. When insurers have market power, analysis and welfare implications are complex.

However, antitrust is only one element of public policy towards markets, especially health insurance markets. Regulatory policies that establish the "rules of the road" for insurance markets can attempt to reduce the potential for problems with competition and thereby minimize the need for antitrust enforcement. When determining these rules of the road, regulators must consider the strategic behavior of insurers—optimal market intervention might change if insurers are not perfectly competitive.

For example, earlier discussion pointed out that insurance firms with market power become more likely to use details of the menu design of their benefit package as a way of both holding down costs and also excluding customers who are more likely to incur high health care costs. Health care consumers and providers also face time-consuming and sometimes confusing administrative hurdles imposed by insurance providers, including billing, prior authorizations, and navigating networks. Some administrative burdens may help to contain costs or reduce unnecessary utilization. Yet it is often difficult to determine which are efficient and which are simply wasteful or harmful to patient welfare. It would be useful to have rules of the road that distinguish between menu choices that encourage cost-efficiency and allow a range of insurance plans for heterogenous consumers, but that set limits on menu choices focused on selecting consumers who are likelier to be healthy. A possible approach to this is requiring standardized choices—a fixed menu of alternatives for all consumers. Standardizing and fixing choices has the virtue of making it easier for consumers to understand their alternatives, and of limiting attempts by insurers to obfuscate choice. Fixed menus are used in the market for private Medicare supplemental ("Medigap") plans, the state-level Marketplaces, and in countries like the Netherlands and Switzerland. Of course, fixing the alternatives limits the ability of insurers to accommodate diverse consumer preferences—so this tradeoff has to be considered.

Risk adjustment is often proposed as tool for addressing selection issues. But as stated previously, an overly simple form of risk adjustment will eliminate incentives for an insurer to seek out cost savings from greater efficiencies and healthier consumers.

Informational interventions may also be important. Insurance is complex, and various choice frictions can limit competition's impact. These frictions may be behavioral—such as inertia or limited attention—or may stem from a lack of information about benefits (Handel and Kolstad 2015). Such frictions may be exacerbated in imperfectly competitive markets. Well-functioning competition is disciplined by feedback from customers, but consumers may have limited information about the underlying quality of plans themselves. For example, Abaluck et al. (2021) argue that consumer willingness-to-pay for plan quality is strikingly low. In addition, consumers usually lack information about insurers' administrative performance when choosing health plans. Enhancing transparency and aligning incentives for genuine quality improvement could enable healthcare markets to deliver better outcomes for beneficiaries.

Another potentially promising arena for policy is the enabling and promotion of long-term health insurance contracts. Health insurance contracts in the United States are short term (usually one year) and as a consequence do not protect individuals against "reclassification risk"—that is, the risk that their health will decline and their health insurance premiums will therefore rise in subsequent years. Long term health insurance contracts cover an individual over a long period of time (potentially a lifetime) and therefore insure against this risk. Such contracts exist in Germany and Chile and are quite efficient (Atal et al. 2025). Understanding how those markets work and formulating policy to enable long term health insurance contracts for the United States seems like a promising and important area for policy.

Given the multiple issues with health insurance markets and the possible ways they can interact, having a single entity responsible for monitoring, oversight, and policy implementation in health insurance markets should be considered. Because insurance is regulated by the states, this likely means separate entities for every state (although the federal government is the relevant entity for health insurance for federal employees, such as the Federal Employees Health Benefits Program or TRICARE for military service members and their families).

Others have proposed more creative solutions. For example, potential entrants in insurance markets face significant barriers to entry; lowering those barriers could improve market outcomes. This phenomenon has been noted in the context of the geographic expansion of integrated delivery systems like Kaiser (Ho 2009). More generally, low premiums attract consumers and help new companies grow. However, new entrants often start with low enrollments, making it hard for them to negotiate discounts with providers and offer low premiums without risking losses. One suggestion (from Leemore Dafny) is to have health care providers agree to accept payments from new entrants at the nth-lowest market rate, with n set by a regulator. Other options include regulatory structures discussed by Einav and Finkelstein (2023). Under their proposal, all Americans would receive automatic, basic, and free universal coverage. Then, similar in spirit to the Australian and some European systems, there would be an option to buy additional, supplemental coverage. Such "out-of-the-box" reform proposals may provide a way to harness the power of markets to deliver coverage, provided they account for insurer incentives.

Because it is unlikely that competition alone will lead to socially desirable outcomes in health insurance market, effective regulation remains essential. Researchers are still developing a fuller understanding of the interactions between two sets of market failures in health insurance markets: firms large enough to wield market power in a setting of asymmetric information. But while a significant body of research examines competition and consolidation, major gaps remain, especially regarding insurer acquisitions of providers and the theoretical and empirical effects of these new industry structures.

■ *This paper is dedicated to the memory of Kate Ho, a brilliant scholar and a wonderful colleague. She is gone far too soon. We are grateful to the editors, Jeff Kling, Timothy Taylor, and Heidi Williams, and to David Dranove, Joe Starc, and Bob Town for comments and suggestions that substantially improved the quality of this paper. Josh Byrne provided excellent research assistance.*

References

Abaluck, Jason, Mauricio Caceres Bravo, Peter Hull, and Amanda Starc. 2021. "Mortality Effects and Choice across Private Health Insurance Plans." *Quarterly Journal of Economics* 136 (3): 1557–610.

Acquatella, Angelique, and Victoria Marone. 2025. "The Risk Protection Value of Moral Hazard." NBER Working Paper 34156.

Atal, Juan Pablo, Hanming Fang, Martin Karlsson, and Nicolas R. Ziebarth. 2025. "German Long-Term Health Insurance: Theory Meets Evidence." *Journal of Political Economy* 133 (6): 1840–85.

Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen. 2020. "The Fall of the Labor Share and the Rise of Superstar Firms." *Quarterly Journal of Economics* 135 (2): 645–709.

Baker, Jonathan B., Nancy L. Rose, Steven C. Salop, and Fiona Scott Morton. 2019. "Five Principles for Vertical Merger Enforcement Policy." *Antitrust* 33 (3): 12–19.

Benkard, C. Lanier, Ali Yurukoglu, and Anthony Lee Zhang. 2023. "Concentration in Product Markets." NBER Working Paper 28745.

Berry, Steven, Martin Gaynor, and Fiona Scott Morton. 2019. "Do Increasing Markups Matter? Lessons from Empirical Industrial Organization." *Journal of Economic Perspectives* 33 (3): 44–68.

Brot-Goldberg, Zarek C., Catherine Che, and Benjamin R. Handel. 2022. "Pharmacy Benefit Managers and Vertical Relationships in Drug Supply: State of Current Research." NBER Working Paper 29959.

Brot-Goldberg, Zarek, Zack Cooper, Stuart V. Craig, Lev R. Klarnet, Ithai Lurie, and Corbin L. Miller. 2024. "Who Pays for Rising Health Care Prices? Evidence from Hospital Mergers." NBER Working Paper 32613.

Cass, Andrew. 2025. "Optum Acquires 200-Provider Medical Group." Becker's Payer Issues, August 13. https://www.beckerspayer.com/m-and-a/optum-acquires-200-provider-medical-group/.

Chade, Hector, Victoria R. Marone, Amanda Starc, and Jeroen Swinkels. 2024. "Multidimensional Screening and Menu Design in Health Insurance Markets." NBER Working Paper 30542.

Che, Catherine. 2025. "Do Middlemen Raise Drug Costs? Countervailing Market Power Meets Agency Frictions." Unpublished.

CMS (Centers for Medicare and Medicaid Services). 2022. *Medicaid Managed Care Enrollment and Program Characteristics: 2022*. US Department of Health and Human Services.

CMS. 2024. "Medicare Advantage and Medicare Prescription Drug Programs to Remain Stable as CMS Implements Improvements to the Programs in 2025." CMS Newsroom, September 27. https://www.cms.gov/newsroom/fact-sheets/medicare-advantage-and-medicare-prescription-drug-programs-remain-stable-cms-implements-improvements.

CMS. 2025. "Nearly 24 Million Consumers Have Selected Affordable Health Coverage in ACA Marketplace, with Time Left to Enroll." CMS Newsroom, January 8. https://www.cms.gov/newsroom/press-releases/nearly-24-million-consumers-have-selected-affordable-health-coverage-aca-marketplace-time-left#_ftn1.

CMS. 2026. "Medicare Advantage/Part D Contract and Enrollment Data." https://www.cms.gov/data-research/statistics-trends-and-reports/medicare-advantagepart-d-contract-and-enrollment-data.

Covarrubias, Matias, Germán Gutiérrez, and Thomas Philippon. 2019. "From Good to Bad Concentration? US Industries over the Past 30 Years." In *NBER Macroeconomics Annual*, Vol. 34, edited by Martin Eichenbaum, Erik Hurst, and Jonathan A. Parker, 1–46. University of Chicago Press.

Crew, Michael. 1969. "Coinsurance and the Welfare Economics of Medical Care." *American Economic Review* 59 (5): 906–08.

Cubanski, Juliette, and Anthony Damico. 2024. "Medicare Part D in 2025: A First Look at Prescription Drug Plan Availability, Premiums, and Cost Sharing." KFF, November 22. https://www.kff.org/medicare/medicare-part-d-in-2025-a-first-look-at-prescription-drug-plan-availability-premiums-and-cost-sharing/.

Cubanski, Juliette, and Anthony Damico. 2025. "Key Facts about Medicare Part D Enrollment, Premiums, and Cost Sharing in 2025." KFF, July 16. https://www.kff.org/medicare/key-facts-about-medicare-part-d-enrollment-premiums-and-cost-sharing-in-2025/.

Cuesta, José Ignacio, Carlos E. Noton, and Benjamin Vatter. 2025. "Vertical Integration and Plan Design in Healthcare Markets." NBER Working Paper 32833.

Curto, Vilsa, Liran Einav, Jonathan Levin, and Jay Bhattacharya. 2021. "Can Health Insurance Competition Work? Evidence from Medicare Advantage." *Journal of Political Economy* 129 (2): 570–606.

Cutler, David M., and Sarah J. Reber. 1998. "Paying for Health Insurance: The Trade-Off between Competition and Adverse Selection." *Quarterly Journal of Economics* 113 (2): 433–66.

CVS Health. 2018. "CVS Health Completes Acquisition of Aetna, Marking Start of Transforming Consumer Health Experience." CVS Health Newsroom, November 28. https://investors.cvshealth.com/news/news-details/2018/CVS-Health-Completes-Acquisition-of-Aetna-Marking-the-Start-of-Transforming-the-Consumer-Health-Experience/default.aspx.

CVS Health. 2022. "CVS Health to Acquire Signify Health." CVS Health Newsroom, September 5. https://investors.cvshealth.com/news/news-details/2022/CVS-Health-to-Acquire-Signify-Health/default.aspx.

CVS Health. 2023. "CVS Health Completes Acquisition of Oak Street Health." CVS Health Newsroom, May 2. https://www.cvshealth.com/news/company-news/cvs-health-completes-acquisition-of-oak-street-health.html.

Dafny, Leemore S. 2010. "Are Health Insurance Markets Competitive?" *American Economic Review* 100 (4): 1399–431.

Dafny, Leemore. 2026. "Health Insurance after Corporatization—What Next?" *New England Journal of Medicine* 394 (6): 521–23.

Dafny, Leemore, Mark Duggan, and Subramaniam Ramanarayanan. 2012. "Paying a Premium on Your Premium? Consolidation in the US Health Insurance Industry." *American Economic Review* 102 (2): 1161–85.

Dafny, Leemore, Jonathan Gruber, and Christopher Ody. 2015. "More Insurers Lower Premiums: Evidence from Initial Pricing in the Health Insurance Marketplaces." *American Journal of Health Economics* 1 (1): 53–81.

Einav, Liran, Amy Finkelstein, and Mark R. Cullen. 2010. "Estimating welfare in insurance markets using variation in prices." *Quarterly Journal of Economics* 125(3): 877–921.

Einav, Liran, and Amy Finkelstein. 2023. *We've Got You Covered: Rebooting American Health Care*. Portfolio/Penguin.

Elevance Health. 2017a. "Anthem to Acquire HealthSun." Elevance Health Newsroom, September 20. https://www.elevancehealth.com/newsroom/anthem-to-acquire-healthsun.

Elevance Health. 2017b. "Anthem Launches IngenioRx, New Pharmacy Benefits Manager." Elevance

Health Newsroom, October 18. https://www.elevancehealth.com/newsroom/anthem-launches-ingeniorx-new-pharmacy-benefits-manager.

Elevance Health. 2018. "Anthem Completes Acquisition of America's 1st Choice." Elevance Health Newsroom, February 15. https://www.elevancehealth.com/newsroom/anthem-completes-acquisition-of-america-s-1st-choice.

Ericson, Keith M. Marzilli, and Amanda Starc. 2015. "Pricing Regulation and Imperfect Competition on the Massachusetts Health Insurance Exchange." *Review of Economics and Statistics* 97 (3): 667–82.

Federal Trade Commission. 2019. "FTC Imposes Conditions on UnitedHealth Group's Proposed Acquisition of DaVita Medical Group." Press release, June 19. https://www.ftc.gov/news-events/news/press-releases/2019/06/ftc-imposes-conditions-unitedhealth-groups-proposed-acquisition-davita-medical-group.

Ganapati, Sharat. 2021. "Growing Oligopolies, Prices, Output, and Productivity." *American Economic Journal: Microeconomics* 13 (3): 309–27.

Gaynor, Martin. 2007. "Competition and Quality in Health Care Markets." *Foundations and Trends in Microeconomics* 2 (6): 441–508.

Gaynor, Martin, Deborah Haas-Wilson, and William B. Vogt. 2000. "Are Invisible Hands Good Hands? Moral Hazard, Competition, and the Second-Best in Health Care Markets." *Journal of Political Economy* 108 (5): 992–1005.

Gaynor, Martin, Kate Ho, and Robert J. Town. 2015. "The Industrial Organization of Health-Care Markets." *Journal of Economic Literature* 53 (2): 235–84.

Gaynor, Martin, and Robert Town. 2012. "The Industrial Organization of Health Care Markets." In *Handbook of Health Economics* Vol. 2, edited by Mark V. Pauly, Thomas G. McGuire, and Pedro P. Barros, 1407–61. North-Holland.

Geddes, Eilidh. 2025. "The Effects of Price Regulation in Markets with Strategic Entry: Evidence from Health Insurance Markets." Unpublished.

Geruso, Michael, Timothy J. Layton, Grace McCormack, and Mark Shepard. 2023. "The Two-Margin Problem in Insurance Markets." *Review of Economics and Statistics* 105 (2): 237–57.

Gottlieb, Daniel, and Humberto Moreira. 2023. "Market Power and Insurance Coverage." Financial Markets Group London School of Economics Discussion Paper 890.

Gray, Charles, Abby Alpert, and Neeraj Sood. 2026. "Disadvantaging Rivals: Vertical Integration in the Pharmaceutical Market." *American Economic Journal: Applied Economics* 18 (1): 286–304.

Hammond, Stuart, Andy Johnson, and Luis Serna. 2025. *The Medicare Advantage program: Status report.* MedPAC. https://www.medpac.gov/wp-content/uploads/2025/01/Tab-M-MA-status-report-January-2025-SEC.pdf.

Handel, Benjamin R., and Jonathan T. Kolstad. 2015. "Health Insurance for 'Humans': Information Frictions, Plan Choice, and Consumer Welfare." *American Economic Review* 105 (8): 2449–500.

Herman, Bob. 2017. "Humana, Private Equity Firms Buy Out Kindred Healthcare." *Axios*, December 19. https://www.axios.com/2018/01/05/humana-private-equity-buys-out-kindred-1515110688.

Ho, Katherine. 2009. "Barriers to Entry of a Vertically Integrated Health Insurer: An Analysis of Welfare and Entry Costs." *Journal of Economics and Management Strategy* 18 (2): 487–545.

Ho, Kate, and Robin S. Lee. 2017. "Insurer Competition in Health Care Markets." *Econometrica* 85 (2): 379–417.

Humana, Inc. 2018. "Humana Acquires Family Physicians Group in Central Florida." Humana Financial Press Releases, April 10. https://humana.gcs-web.com/news-releases/news-release-details/humana-acquires-family-physicians-group-central-florida.

Humana, Inc. 2019. "Humana to Acquire Enclara Healthcare." Humana Newsroom, December 16. https://humana.gcs-web.com/news-releases/news-release-details/humana-acquire-enclara-healthcare.

Humana, Inc. 2021a. "Humana Announces Agreement to Acquire Remaining 60 Percent Interest in Kindred at Home, Accelerating Integration of the Nation's Largest Home Health Provider into Humana's Payer-Agnostic Healthcare Services Platform." *Business Wire*, April 27. https://www.businesswire.com/news/home/20210427006057/en/Humana-Announces-Agreement-to-Acquire-remaining-60-Percent-Interest-in-Kindred-at-Home-Accelerating-Integration-of-the-Nations-Largest-Home-Health-Provider-into-Humanas-Payer-Agnostic-Healthcare-Services-Platform.

Humana, Inc. 2021b. "Humana Announces Agreement to Acquire onehome." *Business Wire*, June 14. https://www.businesswire.com/news/home/20210614005231/en/Humana-Announces-Agreement-to-Acquire-onehome.

Humana, Inc. 2022. "Humana Launches CenterWell Pharmacy and CenterWell Specialty Pharmacy."

Humana Newsroom, June 13. https://news.humana.com/press-room/press-releases/2022/humana-launches-centerwell-pharmacy-and-centerwell-specialty.

Japsen, Bruce. 2019. "UnitedHealth Group Wins FTC Approval of DaVita Deal." *Forbes*, June 19. https://www.forbes.com/sites/brucejapsen/2019/06/19/unitedhealth-group-wins-ftc-approval-of-davita-deal/.

Kanter, Jonathan, and Martin Gaynor. 2025. "The Rise of Health Care Platforms." *Journal of the American Medical Association* 333 (20): 1773–74.

Keisler-Starkey, Katheine and Lisa N. Bunch. 2024. *Health Insurance Coverage in the United States: 2023*. US Census Bureau.

KFF. 2025a. *State Health Facts: Total Population by State*. KFF State Health Facts. https://www.kff.org/state-health-policy-data/state-indicator/total-population/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D (accessed March 10, 2026).

KFF. 2025b. "Health Insurance and Managed Care Indicators: Insurance Market Competitiveness." KFF State Health Facts. https://www.kff.org/state-category/health-insurance-managed-care/insurance-market-competitiveness/.

KFF. 2025c. "Health Policy 101: Employer-Sponsored Health Insurance." Health Policy 101, October 8. https://www.kff.org/health-costs/health-policy-101-employer-sponsored-health-insurance/?entry=table-of-contents-what-is-employer-sponsored-health-insurance.

Kong, Edward, Timothy Layton, and Mark Shepard. 2024. "Adverse Selection and (Un)Natural Monopoly in Insurance Markets." NBER Working Paper 33187.

Ma, Ching-To Albert. 1997. "Option Contracts and Vertical Foreclosure." *Journal of Economics and Management Strategy* 6 (4): 725–53.

MACPAC. 2026. "Spending." https://www.macpac.gov/medicaid-101/spending.

Mahoney, Neale, and E. Glen Weyl. 2017. "Imperfect Competition in Selection Markets." *Review of Economics and Statistics* 99 (4): 637–51.

Mankiw, N. Gregory, and Michael D. Whinston. 1986. "Free Entry and Social Inefficiency." *RAND Journal of Economics* 17 (1): 48–58.

Medicaid.gov. 2025. "May 2025 Medicaid and CHIP Enrollment Data Highlights." https://www.medicaid.gov/medicaid/program-information/medicaid-and-chip-enrollment-data/report-highlights.

Hammond, Stuart, Andy Johnson, and Luis Serna. 2025. *The Medicare Advantage Program: Status Report*. MedPAC. https://www.medpac.gov/wp-content/uploads/2025/01/Tab-M-MA-status-report-January-2025-SEC.pdf.

Morse, Susan. 2022. "Massachusetts AG Agrees to $236 Million Optum and Atrius Health Merger." *Healthcare Finance News*, April 25. https://www.healthcarefinancenews.com/news/massachusetts-ag-agrees-236-million-optum-and-atrius-health-merger.

Pauly, Mark V. 1968. "The Economics of Moral Hazard: Comment." *American Economic Review* 58 (3): 531–37.

Rothschild, Michael, and Joseph Stiglitz. 1976. "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information." *Quarterly Journal of Economics* 90 (4): 629–49.

Salop, Steven C. 1979. "Monopolistic Competition with Outside Goods." *Bell Journal of Economics* 10 (1): 141–56.

Saltzman, Evan, Ashley Swanson, and Daniel Polsky. 2021. "Inertia, Market Power, and Adverse Selection in Health Insurance: Evidence from the ACA Exchanges." NBER Working Paper 29097.

Serna, Natalia. Forthcoming. "Determinants of Healthcare Provider Networks: Risk Selection vs. Administrative Costs." *American Economic Journal: Microeconomics*.

Shepard, Mark. 2022. "Hospital Network Competition and Adverse Selection: Evidence from the Massachusetts Health Insurance Exchange." *American Economic Review* 112 (2): 578–615.

Starc, Amanda. 2014. "Insurer Pricing and Consumer Welfare: Evidence from Medigap." *RAND Journal of Economics* 45 (1): 198–220.

Stigler Center. 2019. *Stigler Committee on Digital Platforms: Final Report*. Stigler Center for the Study of the Economy and the State, Chicago Booth School of Business.

Sutton, John. 1991. *Sunk Costs and Market Structure: Price Competition, Advertising, and the Evolution of Concentration*. MIT Press.

UnitedHealth Group. 2019. "Optum DaVita Medical Acquisition." UnitedHealth Group Newsroom, June 19. https://www.unitedhealthgroup.com/newsroom/2019/2019-06-19-optum-davita-medical-acquisition.html.

US Department of Justice. 2018. "*United States v. CVS and Aetna*: Questions and Answers for the General Public." https://www.justice.gov/d9/press-releases/attachments/2018/10/10/cvs-aetna_-_public_qa_final_0.pdf.

US Department of Justice. 2024. "Justice Department Sues to Block UnitedHealth Group's Acquisition of Home Health and Hospice Provider Amedisys." Press Release, November 12. https://www.justice.gov/archives/opa/pr/justice-department-sues-block-unitedhealth-groups-acquisition-home-health-and-hospice.

US Department of Veterans Affairs. 2025. "About VA Health Benefits." https://www.va.gov/health-care/about-va-health-benefits/.

Venkatesh, Shruthi. 2024. "The Effects of Physician Practice Acquisitions on Hospital Competition." Unpublished.

Yde, Eric. 2025. "Vertical Integration and Regulated Profits in Pharmacy Benefits Markets." Unpublished.

Zeckhauser, Richard. 1970. "Medical Insurance: A Case Study of the Tradeoff between Risk Spreading and Appropriate Incentives." *Journal of Economic Theory* 2 (1): 10–26.

Journal of Economic Perspectives—Volume 40, Number 2—Spring 2026—Pages 43–68

Regulated Competition in Health Insurance Markets on Two Sides of the Atlantic

Lukas Kauer, Thomas G. McGuire, Sonja Schillo, and Richard C. van Kleef

Alain Enthoven (2018) once wrote: "Universal health insurance is not synonymous with 'single payer'." Instead, many high-income countries implement their policy of universal health insurance by individual health insurance in combination with regulated or "managed" competition among insurers to finance health care and protect enrollees against the financial risk of illness. In varying degrees and depending on the country, insurers are also expected to limit health care costs, encourage access to cost-effective care for both the sick and the healthy, charge community-rated premiums, maintain open enrollment, offer choice of the form and extent of insurance coverage, foster innovation in health care delivery, undertake population-based preventive programs, and promote coordination of care. Some of these objectives run counter to the short-term business interests of competing health insurers and must be enforced, either by regulation or by pressure to adhere to social norms.

An individual health insurance market may seem an odd choice to implement social health policy. After all, insurance markets are notoriously subject to market

■ *Lukas Kauer is Senior Lecturer and Researcher at the Faculty of Health Sciences and Medicine, University of Lucerne, Switzerland. Thomas G. McGuire is Professor Emeritus in the Department of Health Care Policy, Harvard Medical School, Boston, Massachusetts. He is also a Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts. Sonja Schillo is Senior Researcher at the Federal Office for Social Security, Bonn, Germany. Richard C. van Kleef is Associate Professor at the Erasmus School of Health Policy & Management, Erasmus University, Rotterdam, The Netherlands. McGuire is the corresponding author at mcguire@hcp.med.harvard.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20251474.

failures (Einav, Finkelstein, and Fisman 2023), as are the underlying markets for the goods and services for which health insurance pays (Arrow 1963). Nonetheless, regulated competition is the basis for financing health care in Germany, the Netherlands, and Switzerland, among others, and has been emulated in health financing reform proposals in Australia (Henriquez et al. 2026), Chile (Henriquez et al. 2023), and elsewhere.

In the United States, employer-sponsored group health insurance covers about 163 million people, roughly 60 percent of the US population under age 65 (Kaiser Family Foundation 2025). Employer-sponsored health insurance is also regulated and features competition among private insurance companies for contracts with employers, although this market organization is not usually referred to as "regulated competition." Overall, US health insurance contains a mix of approaches to financing health care: sectors in which the government owns and runs the health insurance system (as in traditional Medicare); sectors in which the government hires and owns the health care providers (like the Veterans' Administration health care system); parts of Medicare and Medicaid in which the government purchases health insurance from a private company on behalf of the consumer (like Medicare Part C, or "Medicare Advantage"); and an individual-level health insurance market—now at around 25 million—that covered only about 10 million people before the passage of the Patient Protection and Affordable Care Act of 2010 overhauled this market with the creation of state-level "Marketplaces" for individual health insurance.

Those individual health insurance markets on both sides of the Atlantic that are organized according to principles of regulated competition share many significant features. Consumers choose from among numerous competing sellers. Premiums are independent of health status (though they are age-related in the US Marketplaces) and subsidies or premium policies reduce the effective premium for lower-income groups. Regulated benefit packages are comprehensive, typically covering primary care, hospital care, pharmaceutical care, and more. Insurers participate in a sophisticated risk-adjustment system that compensates them for enrolling individuals with high expected health care costs. Together, such regulations help enforce the cross subsidies (from the healthy to the sick, and from high- to low-income people) that are necessary for making health insurance coverage affordable and accessible for all members of the population of interest.

Although this essay refers to similarities, it primarily focuses on the role of three major *differences* among the institutions of regulated competition on opposite sides of the Atlantic. In particular, we compare the national systems in Germany, the Netherlands, and Switzerland to the US Marketplaces. Our essay begins with a review of the intellectual foundations of regulated competition and a brief narrative on the evolution of regulated competition in our four countries. We then turn to a discussion of three main divides separating the approaches to market organization in Europe and in the US Marketplaces: (1) mandatory and universal versus voluntary and partial (applying to only one sector of health insurance); (2) greater or lesser profit orientation of insurers; and (3) reliance on

markets or regulation to contain costs. We conclude with comments on the merits of these differing institutional choices.

Intellectual Foundations of Regulated Competition: Market Failure, the Enthoven Ideal, and Solidarity with Europe

Regulated competition in individual health insurance markets can in principle contend with fundamental market failures in health care; and further, supported by public intervention, it can smooth out some inequities in the financial consequences of ill health and in the ability to pay for health insurance.

Market Failure in Health Care and Health Insurance

Early papers in health economics, notably those by Arrow (1963), Pauly (1968), and Zeckhauser (1970), dealt with market failure and the theory of "optimal health insurance." Arrow (1963) applied axiomatic general equilibrium theory to health care, arguing that "missing markets" for health care, defined contingent on an individual's true health state and need, were a fundamental cause of market failure. In general equilibrium theory of the 1950s, a "contingent commodity" is supplied to a consumer contingent upon the state of the world. Arrow was pointing out the infeasibility of an insurance contract that specifies and pays for exactly what a patient needs contingent upon their true health state in the future—nothing more and nothing less. Arrow (p. 945) wrote: "[I]t is impossible to draw up insurance policies which will sufficiently distinguish among risks, particularly since observation of the results will be incapable of distinguishing between avoidable and unavoidable risks, so that incentives to avoid losses are diluted." Following usage in the insurance industry literature, Arrow introduced the term "moral hazard" to health economics, referring to the insurance-caused dilution of incentives to avoid losses or to control costs in the event of a loss. Arrow (pp. 961–962) implicated both the physician and the patient in the exercise of moral hazard.[1]

Markets for health insurance are themselves subject to market failure as a result of sorting behavior by consumers and insurers. When consumers know more about their health care cost risk than is reflected in premiums (either because of inherent asymmetries in information or from restrictions on premiums stipulated

[1] Pauly (1968), in a comment on Arrow, portrayed moral hazard as exclusively due to consumer demand responding to an out-of-pocket price lowered by insurance. There is no physician, no supply side in Pauly. Pauly's demand-only model of health care utilization set the interpretation of the connection between health insurance and health care utilization for a generation. Zeckhauser (1970), in the same spirit, showed that (in insurance markets where everyone is charged their actuarially fair premium) competitive insurance supply leads to the second-best health insurance coverage balancing the trade-off between risk protection and inefficient consumption. Moral hazard is again solely due to patient behavior. There is no physician, no supply side in Zeckhauser. One approach to demand-side moral hazard is to require modest copayments for the receipt of services, which was the basis of the RAND Health Insurance Experiment (Newhouse 1993; Aron-Dine, Einav, and Finkelstein 2013).

by regulation), health insurance markets can fail because of actions of consumers on the demand side or insurers on the supply side. From the demand side, the influential work by Einav and Finkelstein and colleagues, building on earlier work by Cutler and Reber (1998), explains the inefficiency which results from sicker individuals tending to join more generous plans (in this journal, Einav and Finkelstein 2011; Einav, Finkelstein, and Cullen 2010). This adverse selection forces the more generous plans to increase premiums to cover costs of the sicker enrollees, not just to cover costs due to more generous benefits. Consequently, the premium for the generous plan is actuarially "too high" so that too few of the healthier consumers will choose to pay it.

From the supply side, even in the presence of effective open enrollment regulations requiring plans to accept all applicants, insurers can seek to attract/deter individuals who are financial winners/losers, an activity referred to as "cream-skimming," "service-level selection," or "indirect selection" (Frank, Glazer, and McGuire 2000). In addition, insurance plans can favor or disfavor certain population groups or health care services through network structure, provider payment, managed care algorithms, marketing, and other measures. Examples include not contracting with providers that are particularly attractive to people with specific conditions (Shepard 2022), target marketing efforts at financially favorable groups (Stolper et al. 2022), being less responsive to inquiries from potential enrollees living in higher-cost regions (Bauhoff 2012), and setting high copayments for drugs typically used by people likely to be unprofitable to insure and lower copayments for drugs typically used by profitable people (Carey 2017; Geruso, Layton, and Prinz 2019).

In the context of individual health insurance markets, an obvious solution to adverse selection is complete risk rating of the premiums faced by enrollees, charging everyone their full expected costs under a plan. Person-specific premiums set at expected cost would lead to fully informed consumers facing the marginal cost of their choices and sorting efficiently among plans (and between the yes/no insurance decision), eliminating incentives for health plans to discourage enrollment by those with high expected costs. However, risk rating of premiums grossly violates social norms related to affordability of health plans for the sick (Buchner and Schut 2026), equity in health plan pricing, and the efficiency goal of protecting consumers against the "reclassification risk" of deteriorating health status (Handel, Hendel, and Whinston 2015) to which we all are eventually subject.

Regulated competition can address insurance market failures, both moral hazard and adverse selection, and allows for accommodation of fairness goals, as we explain next.

The Enthoven Ideal and Economic Efficiency

The intellectual roots of regulated competition in health insurance trace to Alain Enthoven (1980), who sought to "change financial incentives by creating a system of competing health plans in which physicians and consumers can benefit from using resources wisely." Enthoven elaborated on his ideas (for example, Enthoven

and Kronick 1989), and the concepts were further refined over time by Enthoven and others (for example, Enthoven and van de Ven 2007; van de Ven et al. 2013). Enthoven (1980, p. xxii, emphases in original) described his approach as follows:

> The most important principles . . . . are *multiple choice* and *fixed-dollar subsidies*. . . . The amount of financial help each family gets toward the purchase of its health plan membership—from Medicare, Medicaid, employer, or tax laws—would be the same whichever plan it chooses. The subsidy might be more for poor than for nonpoor, for old than young, for family than individuals, *but not more for people who choose more costly health plans*. The family that chooses a more costly plan would pay the extra cost itself. Thus, it would have an incentive to consider the cost. In addition, physicians would be organized in competing economic units (most would participate in one or another alternative delivery system), so that the premium each group charged would reflect its ability to control costs.

Enthoven's original vision differs from contemporary models of regulated competition in two respects. First, the health plans in the original Enthoven model were paid directly by enrollees. A central authority might subsidize purchases, but there was no sponsor or regulator collecting funds, risk-adjusting the funds, and then disbursing them to plans. In the late 1970s, when Enthoven first developed his ideas, the risk-adjustment formulas now used as a basis for health plan payment were not available. Instead, the Enthoven model relied on a simple system of demand-side risk rating; that is, allowing insurers to charge more "to people in [age] categories with higher average medical costs."

Second, Enthoven had been influenced by the early health maintenance organization (HMO) movement, in which a health care provider received a fixed payment for each insured person, providing an incentive for the provider to control costs. In that spirit, Enthoven envisioned closed network provider groups forming the backbone of the competing health plans. Today, health care systems in Germany, the Netherlands, and Switzerland consist of two distinct markets, one for health insurance (consumer-insurer contracts) and one for health care purchasing (insurer-provider contracts). In these countries, and in other settings including the US Marketplaces, regulated competition is primarily oriented to the health insurers, not the health care delivery system. While Kaiser Permanente (whose membership included Alain Enthoven when he wrote *Health Plan*) and other traditional health maintenance organization plans exist in the United States, they compete with plans with large provider networks and account for a relatively small share of the national market.

The Enthoven model has evolved as its ideas have been applied in particular institutional contexts. For example, the development of sophisticated risk-adjustment models meant that rather than a system of risk-rated premiums by age set by insurers, regulators nowadays rely primarily on risk adjustment to pay insurers more for higher-risk enrollees. In Germany and the Netherlands, contributions by the

enrollees are (re)distributed across insurers on the basis of health risks. In Switzerland and the US Marketplaces, transfers among plans at the end of the annual contract period accomplish the same end. By transferring funds to a more generous plan when sicker individuals enroll, risk adjustment dampens the component of plan premium differences due to selection, thus mitigating the Einav-Finkelstein sorting problem.[2] Risk adjustment can also mitigate an insurer's incentives to orient services to attract the good risks and deter the bad risks by reducing or eliminating the profits and losses associated with the good and bad risks. A large theoretical and empirical literature addresses the design of these risk-adjustment algorithms: for reviews and references, useful starting points are van de Ven and Ellis (2000) and Ellis, Martins, and Rose (2018).

The original Enthoven model was designed to deal with the moral hazard problem introduced by full or nearly full health insurance coverage. Paying provider groups (like Kaiser Permanente) by a prospectively set per-head or capitation payment shifts health care cost risk to the provider/insurer and conveys incentives for controlling costs. Moral hazard in health economics is often portrayed as stemming from demand decisions, but in reality, provider groups or insurers employing managed care practices in regulated competition have a number of tools (discussed below) to limit costs for fully or nearly fully covered consumers.

Solidarity and Fairness in Europe

By disconnecting what the enrollee pays in a health insurance premium from what the insurer receives for that enrollee, regulated competition permits fairness/equity objectives to play a major role in the regulation of health insurance. In continental Europe, Enthoven's ideas on market organization overlaid longstanding traditions in social health insurance emphasizing "solidarity," a term central to debates in Europe about health policy. The then–Minister of Health in Germany, Ulla Schmidt, explained solidarity in this way: "My overarching personal goal as Minister of Health has been to preserve for Germany's health system the principle of social solidarity, by which we mean that everyone in Germany should have guaranteed access to state-of-the-art medical care and contribute to the financing of this guarantee on the basis of the household's ability to pay" (as quoted in Cheng and Reinhardt 2008, p. w205). In Germany, the Netherlands, and Switzerland, regulators have applied a blend of regulatory tools to achieve solidarity, including insurance mandates, standardization of benefits, premium-rate restrictions, open enrollment, risk adjustment, and income-based subsidies/contributions.

Although the term "solidarity" is largely absent from American debates about social policy, equity considerations familiar to American audiences play a similar

[2]Using risk-adjusted payments to insurers paired with community rating for enrollees leaves an inefficiency unaddressed by regulated competition. With fully risk-rated premiums charged to enrollees, individuals face the social marginal cost of their choice of plan. Community rating of premiums cannot set up incentives for fully efficient sorting among plans. With community rating, healthy, low-cost enrollees face too great a price increment for a more generous plan and unhealthy, high-cost enrollees face too low a price (Glazer and McGuire 2011).

role. For example, James Tobin (1970) advocated for "specific egalitarianism," such that necessities, including health care, should be equally accessible to all regardless of their economic status. In a solidarity-friendly approach, Enthoven's original proposal included subsidies for low-income families to help the poor and subsidies for the elderly to help with the higher premium cost for those with higher expected health care costs. All major US health financing systems involve redistributive government subsidies. Some government subsidies are explicit, like the premium- and cost-sharing subsidies in the Marketplace plans, whereas others, such as the tax subsidy for employer-based health insurance, are less explicit.

Individual Health Insurance Markets in Germany, the Netherlands, Switzerland, and the United States

In this section we sketch the history and main features of individual health insurance markets in Germany, the Netherlands, Switzerland, and the US Marketplaces. We focus on similarities and differences among these systems in terms of regulation, consumer choice and competition.

Germany

Regulation of health insurance in Germany dates to 1883 when Otto von Bismarck introduced social health insurance for low-income people, a policy later followed by other countries including Belgium, France, and the Netherlands (Schut, Henschke, and Or 2023). In the Bismarck system, employers and employees made a mandatory contribution to so-called "sickness funds" that in return provided enrollees with access to primary care and some forms of hospital care. Over time, both the eligible population and the benefits package have expanded. Today, health insurance in Germany is mandatory and about 90 percent of the population are part of this statutory system, which covers a comprehensive set of services including primary care, hospital care, and pharmaceutical care. The remaining 10 percent, which consists of self-employed persons, civil servants, and people with an income above a threshold, can opt-out of the social health insurance system and buy private health insurance. In what follows, we focus on the large statutory system.

Until the late twentieth century, German regulation of health insurance was primarily focused on achieving universal access to medical care and controlling price and supply. Since the 1990s, reforms have also aimed to increase efficiency and patient choice by introducing elements of competition among sickness funds and health care providers. For example, premiums are income-based and composed of two parts: a uniform contribution rate of 14.6 percent and a sickness-fund-specific contribution rate that varies between 2.2 percent and 4.4 percent (as of February 2026). Both apply up to an income threshold. Insurers are risk-bearing and primarily compete on price, which in this case refers to the level of the sickness-fund-specific contribution rate, as well as on customer service.

The Netherlands

Health care reform in the Netherlands can be characterized by three waves: regulation to achieve universal access to care (1940–1970), control of volumes and prices of care by the government (1970–1990), and implementation of regulated competition (1990–onwards) (van de Ven and Schut 2008; see also, Cutler 2002). Regulated competition was first proposed by the Dekker Committee in 1987, inspired by the work of Alain Enthoven, and motivated by problems with stringent supply-side regulation of health care in the 1970s and '80s. Direct price and volume regulation by the government had become too complex and lacked incentives for innovation and efficiency, for both insurers and health care providers.

In the early 1990s, the Netherlands started down the road to competition among insurers by gradually introducing consumer choice and financial responsibility for sickness funds. These steps culminated in the Health Insurance Act of 2006, unifying the former sickness fund scheme (for low- and middle-income people) with the parallel private insurance scheme (for high-income people). The Health Insurance Act introduced specific regulations to enhance affordability and accessibility of insurance coverage and health care while leaving room for competition between health insurers with the goal to enhance efficiency. Parallel to the reforms in the insurance system, the Dutch government gradually deregulated price and supply in health care delivery. The idea behind these parallel reforms was to provide health insurers with both incentives (competition and financial responsibility) and some discretion for managing care. Insurers compete on price (the out-of-pocket premium), plan design (for example, the network of providers and coverage of out-of-network spending), and customer service (for example, claims handling and waiting-list mediation).

Switzerland

Core Swiss values of local-first governance and self-responsibility shaped its first federal health insurance law enacted in 1914. The Swiss had followed the German approach where hundreds of small, nonprofit sickness funds provided some risk pooling. However, the law did not include an insurance mandate and insurance was not tied to employment (FSIO 2013). By the early 1990s, the Swiss situation was similar to individual health insurance markets in the United States before the Marketplaces created by the Affordable Care Act of 2010: no mandate for individuals to purchase insurance, risk-rated premiums (by gender and age at entry), and denial of coverage for pre-existing conditions—which made it virtually impossible for high risks to switch insurance.

In 1996, a new federal law established regulated competition with a mandate to purchase insurance, a standard and broad benefits package, open enrollment, income-related subsidies, more choice, community rating (per plan and region, with children and young adults receiving rebates), and a risk-adjustment system. Spending is financed by out-of-pocket premiums and cantonal payments for about half of inpatient hospital costs, a form of risk sharing. Apart from the cantonal payments and copayments for consumers, insurers are financially responsible for

spending under the benefits package. Insurers compete on price (the out-of-pocket premium), customer service (for example, claims handling), and—in case of managed care plans—on plan design (like the provider network). Unlike Germany and the Netherlands, the Swiss system is largely operated at the level of its 26 cantons, rather than the national level. While the federal law defines the framework and regulations, cantons license providers, execute a within-canton risk-adjustment system, and can decide how many hospitals need to be covered by the insurers. There are no cross subsidies between cantons.

US Individual Health Insurance Marketplaces

Employer-based health insurance became widespread in the US economy during the 1940s as wartime wage controls diverted competition for labor into fringe benefits, the most important of which, then and now, was health insurance. Employer contribution to worker health insurance premiums is untaxed to workers, providing a continuing impetus linking health insurance to employment. Publicly supported health insurance programs for the elderly (Medicare) and the poor (Medicaid) were part of the Great Society legislation in the 1960s. Those not covered by their employer or eligible for the public programs faced an individual health insurance market with high premiums and poor access for those most in need of health insurance. Insurers were not required to accept applicants and could terminate coverage, leaving a market with only about ten million enrollees and 16 percent of the US population without health insurance in 2010.

The Affordable Care Act of 2010 expanded Medicaid and introduced subsidies and reforms into individual health insurance markets, to be run publicly at the state level. The US Marketplaces provide an option of health insurance coverage for those who do not receive insurance through their employer or through public programs. Although the original 2010 law imposed penalties for absence of insurance, those financial penalties have since been removed, making enrollment in a Marketplace plan a voluntary decision. (Some US states—California, Massachusetts, New Jersey, Rhode Island, and Vermont—do continue to mandate health insurance coverage and penalize noncompliance.) Between 2014 and today, the number of consumers selecting Marketplace plans has grown from 8 million to 23 million (down slightly from 2025). A small (about 1.6 million) "off-Marketplace" individual health insurance remains, made up largely of those who do not qualify for subsidies. Many but not all off-Marketplace plans are subject to the same consumer protection regulations as the Marketplace plans, such as guaranteed issue and limited risk rating. Enrollment growth has been especially strong among Black and Hispanic enrollees in recent years (ASPE 2024). Marketplace enrollment coupled with the expansion of Medicaid has reduced current rates of uninsurance to half the value before 2010. Unlike Germany and the Netherlands, but like Switzerland, the Marketplaces are not operated at the national level but at the state level. Within states, insurers compete on price (the out-of-pocket premium), plan design and generosity (for example, provider network and drug formulary), and customer service.

Similarities and Differences in Market Organization

Similarities across the Four Systems

Table 1 summarizes some of the main similarities across the four systems as they look today: for further information on these countries as well as many others with some form of regulated competition, a useful starting point is McGuire and van Kleef (2018). All four systems mandate coverage of a comprehensive set of benefits including hospital care, pharmaceutical care, and primary as well as specialist care. However, the minimum benefit packages do not cover the entire spectrum of health care services. For example, dental care is outside the minimum benefit package in the Marketplaces and only partially covered in the other countries. Perhaps the most notable exclusion from individual health insurance is long-term care. In Germany and the Netherlands, a universal public scheme covers long-term care. In Switzerland, long-term care is paid mainly privately, a minor fraction paid by health insurance, and substantial subsidies from the state. In the United States, long-term care is generally paid privately or by the public Medicaid program.

Across the four systems, similar regulations and measures seek to enhance affordability and accessibility of coverage such as open enrollment, standardized benefits, premium-rate restrictions, and income-related contributions/subsidies. The funding of all four systems involves income-based and/or government-subsidized premiums. The US Marketplace system is the only one in which the health insurance premiums paid by individuals are risk-rated by multiple factors like age, geography, or smoking status, but all four systems rely on risk adjustment to compensate insurers for predictable profits and losses to mitigate incentives for risk selection. Empirical simulations have shown that the existing risk-adjustment formulas are generally effective, but not perfect (McGuire and van Kleef 2018). Examples of ongoing issues include remaining incentives for risk selection, and the existence of incentives for providers to engage in manipulation like "upcoding," in which certain services are recategorized in ways that can lead to higher reimbursement.

Another similarity is the presence of risk-sharing arrangements between insurers and the regulator. In a health insurance context, risk sharing refers to features of the payment system to insurers that shift some of the health cost risk to a regulator or an insurance pool. For example, a government might supplement a capitation payment to insurers with a "reinsurance" provision by which the government compensates an insurer for a share (say, 80 percent) of the costs of an individual above a high threshold (say, $200,000). Risk sharing mitigates selection incentives and protects insurers from risks of high-cost cases. In the Netherlands and the US Marketplaces, risk-sharing arrangements have also protected insurers from "system uncertainty" in times of impactful reforms. The US Affordable Care Act of 2010 temporarily protected insurers with reinsurance (paying for a share of an enrollee's costs after a threshold) and "risk corridors" (limiting insurer-level gains and losses). After the implementation of the Health Insurance Act in the Netherlands, the regulator relied on risk corridors, proportional risk sharing and high-cost pools to protect insurers from unpredictable reform impacts and to mitigate

Table 1
Individual Health Insurance Systems: Selected Characteristics

	Germany	Netherlands	Switzerland	United States (Affordable Care Act Marketplaces)
Current system in place since	1996	2006	1996	2014
Prior system	No free choice of sickness fund for the insured; mandatory sickness fund insurance for people below income threshold and voluntary private insurance for the rest	Mandatory sickness fund insurance for people below income threshold (2/3 of population) and voluntary private insurance for the rest	Individual/small group market with risk rating, denials, low rates of uninsurance due to local mandates and federal subsidies	Individual/small group market with risk rating, denials, and high rates of uninsurance
Mandatory insurance purchase	Yes	Yes	Yes	No
Eligible population	Citizens and legal residents with some exceptions (like asylum seekers)	People who live and/or work in the Netherlands	Citizens and legal residents	Citizens and legal residents not incarcerated
Covered lives	75m	18m (2025)	9m (2024)	23m (2026)
Open enrollment	Yes	Yes	Yes	Yes
Sources of funding	Income-related contribution, additional income-related premiums, taxes	Premiums, income-related contribution and taxes	Premiums, cantonal payments for inpatient care (risk sharing), cantonal and federal subsidies	Premiums, state and federal subsidies
Premium regulation	Community-rated per plan	Community-rated per plan	Community-rated per plan and region; regulated rebates for limited coverage	Premiums can vary by smoking and age within limited bands
Risk adjustment and risk sharing	Sophisticated risk adjustment; risk sharing through high-cost pool	Sophisticated risk adjustment; some risk sharing	Simple risk adjustment; extensive risk sharing of hospital inpatient costs	Sophisticated risk adjustment; risk sharing for very high-cost cases via transfer formula
Consumer choice of plan	Uniform product	Higher deductible; limited networks	Higher deductible; limited networks	Tiers ranging in actuarial value
Insurer spending per person	€4.395 (2024)	About €3,100 (2024)	CHF 4,060 (2024) (€4,345)	$7,236 (2024) (€6,132)

Source: McGuire and van Kleef (2018); ASPE (2024); BAG (2025); Zorginstituut (2026).
Note: US dollar spending is average plan premiums. Swiss franc and US dollar conversions to euros are made at the rate prevailing on September 23, 2025. According to Zorginstituut (2026), mean per person spending on services covered by the Dutch basic health insurance in 2024 was €3,316.96 euros. This includes out-of-pocket spending by consumers under the deductible. The amount in the table (€3,100) has been roughly corrected for out-of-pocket spending using information from van Kleef and van Vliet (2025).

predictable profits and losses under the imperfect risk-adjustment system that was in place. Germany makes use of a "high-cost pool" that serves the same function as reinsurance. As mentioned earlier, Swiss cantons presently share risk of hospital costs with insurers, a setup scheduled to change in 2028 to canton cost sharing for

all costs. Although risk sharing can be helpful, it also has a disadvantage: it reduces the financial responsibility for insurers and thereby mitigates incentives for cost control. For the regulator in the Netherlands, this was an important reason for gradually reducing risk-sharing measures as system uncertainties diminished and the risk-adjustment formula became more sophisticated.

Insurers also share risks with consumers and health care providers. In all four countries, there is consumer cost sharing in the form of deductibles, coinsurance, and/or copayments. In the United States, levels of consumer cost sharing are typically higher than in the European countries. And insurers in all four countries can and do (in different magnitudes) apply bundled-payment methods for paying providers, which may involve flat capitated payments to providers (who must then manage costs) or flat payments for a certain type of procedure that combines a number of different health care services. These payment methods inherently transfer (some) financial risk from insurers to providers of care.

Next to these similarities, there are also important differences in the details of regulation and competition. One important difference was already noted above: while the system in the Netherlands is operated at the national level, the German system is a mix of national and regional, and the systems in Switzerland and the US Marketplaces are largely operated at the cantonal/state level. Below we describe and discuss three other main divides between the four systems.

Divide #1: Mandatory and Universal Versus Voluntary and Partial

The most obvious difference between individual health insurance markets on the two sides of the Atlantic is that the social health insurance systems in Germany, the Netherlands, and Switzerland are *mandatory* and *universal* (subject to the definition of who is included within "universal" in each country), which is a broadly appreciated element of the social contract in these countries. Conversely, the state-based Marketplaces in the United States are subject to eligibility qualifications, are *voluntary*, and apply to only a small share of the US population. This divide has several important implications for market organization and competition.

First, selection in and out of the health insurance market (sometimes called the "extensive margin") is handled by a mandate in Europe, whereas the US Marketplaces must rely on subsidies to encourage participation.[3] Incentives are expensive relative to effective mandates. Subsidies must be high enough to lower effective prices to the young and healthy to keep them in the risk pool. The US federal government spent $91 billion on Marketplace subsidies in 2023, going to 92 percent of enrollees (Peterson Foundation 2024), and some states contribute additional subsidies. With a mandate, higher-income groups can simply be required to pay

[3]Indeed, the European literature on selection in health insurance primarily focuses on the intensive margin (that is, selection within the health insurance market) while the US literature on selection in health insurance also looks at selection on the extensive margin of having or not having health insurance (van Kleef et al. 2024).

higher premiums (as in Germany with premiums set as a percentage of income and in the Netherlands where about 50 percent of the total funds come from income-related contributions). The Germans, Dutch, and Swiss all subsidize lower-income enrollees, but in the context of an effective mandate, these can be designed to serve equity (solidarity) goals without concern for their effect on the composition and magnitude of enrollment in the risk pool.

Second, enrollment in the European insurance markets is relatively stable, both at the market level because of universality and at the insurer level as well. In recent years, annual turnover rates vary between 5 and 9 percent in the Netherlands and Switzerland. Germany is even lower. However, turnover in the US Marketplaces is on the order of over 30 percent annually (Kong, Shepard, and McIntyre 2022). The mandatory and universal markets in Europe make planning and projection easier. Individuals cannot leave the market, which eliminates changes in market-level risk due to systematic movement of high/low risks in/out of the market.[4] More reliable actuarial predictions allow insurers to avoid building a high safety margin premium into their pricing. Predictability of enrollment and the risk characteristics of enrollees mean providers are also more willing to accept risk delegated to them from the insurer. In the Netherlands, for instance, most insurers negotiate a "global payment" or "cost ceilings" with hospitals, sometimes with risk mitigation measures (Gajadien et al. 2023). Such contracts move the financial risk for hospital costs to the parties in the best position to manage those costs. Stability has other advantages. Insurers with a stable population have more incentive to invest in prevention. Less movement among insurers and in and out of the market reduces costs associated with turnover and recruitment.

Third, because insurers in Europe have responsibility for the entire population (or the vast majority as in Germany), they have less scope for "cost shifting," a popular sport among US health insurers. In the US context, for example, Medicare can keep its payments to hospitals low knowing that private insurers will end up paying higher margins, maintaining hospital viability and incentives for capital investment (Glazer and McGuire 2002). All payers playing by the same rules eliminates opportunities to shift cost, along with the socially wasteful effort put into cost shifting. Universal plans can solve the "public good problem" financing health care joint costs by collective action guided by regulation applying to all insurers.

Divide #2: Greater or Lesser Profit Orientation of Insurers

The business functions of health insurers are similar across our four countries: set premiums, contract with providers, structure products, and adjudicate and pay claims. Nonetheless, health insurers on one side of the Atlantic are different

[4]In the US Marketplaces, premiums have been forecast to rise by an average of 18 percent in 2026 partly attributable to reduced premium subsidies, which in turn is projected to lead to disproportionate disenrollment of healthier individuals (Ortaliza et al. 2025). As the lower-risk enrollees exit, premiums would need to be set higher next year. Recent Congressional (in)action in the United States caused federal subsidies to revert to lower 2021 levels.

Table 2
Insurers and Their Role

	Germany	Netherlands	Switzerland	United States (Affordable Care Act Marketplaces)
Number of insurers in 2025	94	20	37 (varies by canton)	Varies by state from 1–10
Conditions of entry	Strong requirements set by federal law	Strong requirements related to licensing, solvency, and governance among others	Permission requirements set by federal law	Active Marketplaces limit entry; passive Marketplaces allow any qualified insurer
Can insurers make profits?	No	Yes	No	Yes
Can insurers pay out profits?	No	Yes, not forbidden by law	No	Yes
Can insurers restrict provider networks?	Yes, but participation in networks is optional for the insured and providers	Yes	Yes, but outpatient sector only	Yes
Can insurers limit coverage for out-of-network spending?	No, but the insured loses benefits from the network	Yes	Yes	Yes
Can insurers apply alternative payment models for care?	Yes	Yes	Yes	Yes
Can insurers vertically integrate?	No	Yes, but very rare	Yes, but very rare	Yes
Range of enrollment for individual insurer	2.200–12 m (2025)	90 k–3.5 m (2025)	2.568–1.5 m (2024)	Very small–5.7 m (2025)

Source: McGuire and van Kleef (2018); ASPE (2024); BAG (2025); Zorginstituut (2026).

creatures with different behavior than those on the other. Health insurers in Germany, the Netherlands, and Switzerland are *free-standing and mostly nonprofit organizations*, ranging in size from a few thousands up to around twelve million enrollees. Private US health insurers can be nonprofit and small, but also can be *for-profit divisions of very large vertically integrated health care organizations*. In terms of total enrollment (in and outside the Marketplaces), the largest private insurers in the United States are an order of magnitude larger than those in Europe. Opportunities to make and keep profits differ across countries, being conditioned on minimum payout ratios, profit limitations, reserve and solvency requirements, payment system features, and societal expectations. Table 2 spells out some of the differences.

US insurers are a mix of profit and nonprofit organizations. Some are massive, diversified health care companies, such as UnitedHealthcare, the insurance arm of UnitedHealth Group, a for-profit conglomerate with $400 billion in revenue in 2024. It insures only two million Marketplace enrollees, but many more in private health insurance and the Medicare Advantage plans (in which the federal

government pays a private insurer a flat amount to provide Medicare coverage). UnitedHealth Group also owns Optum, which itself employs 90,000 physicians (roughly one-tenth of all US physicians), as well as surgery centers and a large pharmacy-benefit manager. This diversification creates economic complementarity among UnitedHealth products. Another enrollee in a UnitedHealth Marketplace plan is another enrollee in Optum. The Affordable Care Act of 2010 set out to encourage nonprofit mission-driven alternatives to commercial insurers, called Consumer-Operated and Oriented Plans (COOPS). However, only a handful of small COOPS have survived competition against the information systems of large insurers, which are better positioned to contract with providers and collect and report the data needed to secure risk-adjusted payments.

For basic individual health insurance, most insurers in the Netherlands and Switzerland are foundations or mutuals, which means that they are usually formally "owned" by policyholders. In Germany, sickness funds are public law corporations with self-governing structures. This status means they are legally independent entities that fulfill public tasks under state supervision. In Germany and Switzerland, insurers must be nonprofit organizations. Any profits go into financial reserves, to healthcare innovation, and/or to lower premiums. In the Netherlands, the law allows for-profit insurers, but because of low profit margins, the Dutch market is not attractive for commercial insurers and most insurers are nonprofit. In fact, since the introduction of the Dutch Health Insurance Act in 2006, there have been only two new entrants (via takeover), one of which exited the market after two years (Jeurissen and Maarse 2021).

It is common for insurers in the Netherlands and Switzerland to sell supplemental insurance (covering services not included in the basic insurance such as dental care for adults) on a for-profit basis. Although basic and supplementary health insurance are separated by law, most insurers offer both products and almost all consumers with supplementary health insurance (about 80 percent of the Dutch population, for example) buy their basic health insurance from the same insurer. Supplemental insurance serves as a complementary product to basic insurance, playing a similar role to the Optum pharmaceutical benefit manager in the case of UnitedHealth. Compared to the basic benefits package, the size (in euros or Swiss francs) of supplementary health insurance is relatively small in all three European countries.

A difference in organizational form may or may not imply a difference in behavior. With the rise of private-equity ownership in the United States, evidence has recently emerged that ownership matters to business objectives and behavior: for example, Kannan and Song (2025) find that hospitals acquired by private-equity firms reduce salary expenditures, even as similar hospitals not acquired by such firms raise salaries, and Singh et al. (2025) find that when private-equity firms acquire gastroenterology practices, both professional fees and prices increase. Less research compares organizational forms for the health insurance sector, though Dafny (2019) asked "Does it Matter if Your Health Insurer is For-Profit" and found that for-profit insurers with market power charged higher prices than nonprofits

with seemingly equivalent market power. Douven and Schut (2011) found evidence that Dutch insurers pursued financial stability over profit. Nonprofit insurers may be oriented to size, not profitability, recalling Baumol's (1959) observation that managerial compensation and prestige is associated with being big, a force that may apply to insurers in all our countries. Size also increases political influence to shape the market regulations. In the Netherlands and Switzerland, growth means more opportunity to sell profitable supplemental insurance to more people.

As far as we know, no research compares the behavior of European insurers with different organizational forms, perhaps because the variation in organizational forms across European insurers is less than in the United States. In our experience, insurers in Europe take their "social mission" seriously—to act as a prudent purchaser on behalf of their enrollees. For example, health insurers in the Netherlands are actively involved in working groups to improve the risk-adjustment model, to increase transparency in quality reporting, and on other matters of collective interest. Most insurers in Switzerland are owned by foundations with explicit social mission statements. The "social mission" of German health insurers in the statutory health insurance is stated in the law (§ 1 SGB V): It is their role to support the insured in maintaining, restoring, or improving their health through education, counseling, and benefits.

In addition to organizational form and innate business objectives, regulation mitigates profit incentives to different degrees in our four countries. As a thought experiment, imagine that a for-profit insurer in the US Marketplaces has discovered an innovation that reduces costs for its covered population by $100,000 per year. How much of the ongoing $100,000 efficiency saving will feed into higher corporate profits? Premium-setting regulations are intended to require the insurer to pass through some or all of the ongoing savings to lower premiums, and payout requirements may take back some of the "surprise" first-year savings. In the state-level Marketplaces, each insurer develops an "index rate" that can incorporate average claims for mandated benefits for the insurer's anticipated risk pool as well as other cost-related elements (Layton, Montz, and Shepard 2018). Once a benefit year has played out, "medical loss ratio" provisions require an insurer to spend at least 80 percent of premium income on health care claims and quality improvement (calculated over a three-year average), leaving the remainder for administration, marketing, and profit (Ortaliza and Cox 2024), so that an insurer spending less than the mandated ratio on benefits must pass 100 percent of any savings on to rebates to consumers. Lower premiums can attract greater enrollment, making a positive contribution to total profits. Even if these rules allow some portion of cost savings to end as higher profits, they surely dilute the incentives of insurers to seek such savings.

Take the same thought experiment to Europe: In Germany, the Netherlands, and Switzerland, cost reductions will largely be used for increasing reserves and/or reducing the premium or investing in innovation. Consequently, profit margins in the European countries are relatively small or not allowed by the regulator, which might help explain why there has been hardly any new market entry in any of the

three European health insurance markets since the current regulatory frameworks have been put in place. Moreover, the requirements for entry are formidable. In the Netherlands, for instance, insurers must meet heavy requirements related to licensing, solvency, risk management systems, governance and internal controls, and more. Several years ago, the Netherlands Authority for Consumers and Markets (2017) concluded that these requirements lead to high entry barriers. In addition, each plan offered by an insurer needs to be approved by the Dutch Healthcare Authority, which requires that a plan needs to be in accordance with the Health Insurance Act and needs to meet criteria related to transparency. Similar regulations apply in Germany and Switzerland.

Interestingly, there has been considerable consolidation and exit in the European markets. The number of insurers in Germany has fallen from 1,147 in 1990 to 93 in 2026 (GKV-Spitzenverband 2025). In the Netherlands, the number of insurers decreased from 33 in 2006 to 20 in 2025 (Jeurissen and Maarse 2021; Vektis 2025). Switzerland has seen a decrease in insurers from 145 in 1996 to 37 in 2025 (BAG 2025). The common pattern of no entry, consolidation, and exit is consistent with the idea that the advent of regulated competition and risk-bearing introduced competition and economies of scale into the European insurance markets. Regulation-enforced low (or no) profit margins and high barriers to entry might have little downside when a sector has more than enough suppliers. The welfare effects could well differ if entry and innovation were called for.

In contrast, entry and exit in US Marketplaces both occur at high rates. The Kaiser Family Foundation tracks insurer participation in Marketplaces and reported in 2021, for example, that 30 insurers entered markets across 20 states and an additional 61 expanded their service area within states in which they already operated (McDermott and Cox 2020). A state regulatory body can take an active role in managing entry, reviewing applications to enter and sell, or delegate operation of the Marketplace in their state to the federal government (ASPE 2024). Two large, successful Marketplaces, Covered California and the Massachusetts Connector, both limit entry and also negotiate pricing, network coverage, and other terms with potential suppliers.

Divide #3: Markets Versus Regulation to Contain Costs

Comparison of the health care systems in the United States and other countries often begins with an observation about higher US health care costs. Health care costs are higher in the United States than Europe because of higher prices, not higher rates of utilization. Hospital days and number of physician office visits per person are lower in the United States (OECD 2025). This is not a new phenomenon. According to a recent Peterson-KFF publication, "[T]he evidence continues to support the finding that higher prices—as opposed to higher utilization—explain the United States' high health spending relative to other high-income countries" (Tevis et al. 2025, p.1). For example, an inpatient coronary angioplasty (a procedure to widen heart arteries, performed much less frequently in the United States) cost on average $17,000 and $34,000 to public and private payers respectively in

2022, compared to $9,000 in Switzerland and $4,000 in Germany[5] (Tevis et al. 2025; Dutch cost data were not available for comparison).

The drug industry furnishes other examples. US insurers pay much more for the same brand-name drug than their counterparts in Europe. A Congressional report (Ways and Means Commitee Staff 2019) found that average US insulin prices were more than three times the international average, about $35 per dose compared to the international average of less than $11 per dose. Generally, brand-name drugs cost three times as much in the United States as in Europe (Mulcahy, Schwam, and Lovejoy 2024).

The process for setting drug prices illustrates the divide between the Marketplaces and Europe. The United States relies on competition among insurers to negotiate brand-name drug prices. No national drug price controls shelter Marketplace insurers from the market power of drug manufacturers. Instead, the Marketplace insurers have (and need) scope to select drugs for inclusion on their formularies. Insurers are free to set the drug "formulary" (the list of drugs covered), as long as a plan covers at least one drug in every therapeutic category that is included in the minimum benefits package. Insurers can also apply various forms of utilization management including prior authorization and "step therapy" (a requirement to try lower-cost medications before higher-cost ones can be authorized) to manage drug utilization and cost. Drug manufacturers will sometimes grant an insurer a lower price if they agree *not* to apply these quantity restrictions.

Regulation in the three European countries leaves less room for insurers to negotiate when it comes to drug formularies and prices, substituting collective action. The list of covered drugs in Europe is largely determined by the national government or some regulatory body. In the Netherlands, similar drugs are grouped into a "cluster," and insurers can choose to cover only one drug per cluster. In the years following the introduction of the Dutch Health Insurance Act (2006), more and more insurers indeed decided to cover only the generic option (if available) which led to substantial savings (Boonen et al. 2010). In addition, the Dutch government sets maximum prices for registered drugs based on the weighted average prices in four reference countries: Belgium, France, Norway, and the United Kingdom. In Switzerland, prices are set with a similar mechanism and if a drug is on the national list of covered drugs, an insurer must cover it. Insurers have limited discretion in managed care plans, over the copayment (for example, higher for branded drugs if significantly cheaper generics are available), and over coverage for drugs that are experimental, not yet on the list, or off-label. In Germany, health insurers must cover all prescribed drugs approved by the regulatory authority and listed in the reimbursement catalog. Insurers can negotiate rebates with pharmaceutical companies

[5] In an editorial in the *Journal of the American Medical Association*, Baicker and Chandra (2018) point out that decomposing differences into price and quantity is not as straightforward as it might seem. A hospital day might have a higher price tag because of higher intensity of use of hospital inputs. The patient might "get more" from a higher intensity of inputs, making some of what might appear as a price difference to be a difference in quantity.

and may steer patients towards specific drugs within a reference or cluster group, with physicians able to override this choice.

The reliance on competition and managed care in the United States versus collective action applies to other healthcare services as well. In terms of product design, as we have seen, covered benefits are highly regulated everywhere, channeling quality competition into the *administration* of the mandated benefits. In the United States, health insurers actively manage care, restricting choices of providers and therapies, and even sometimes overriding decisions of doctors and patients. In principle, a well-designed and administered managed care plan with little or no cost sharing can protect the consumer against financial risk and address the moral hazard problem by limiting access to only cost-effective care (McGuire 2012). Consumers value health plans that contend with moral hazard, preventing enrollees from using care not worth the cost and keeping premiums down. Consumers rationally choose such plans even while recognizing that, when sick, they may be frustrated and not get everything they demand at the insured price.

In the Marketplaces, within wide bounds, insurers can restrict provider networks (subject to hard-to-enforce network adequacy regulations) and fully exclude out-of-network providers from coverage (except for emergency care). Restricting provider networks (from the standpoint of the insurer and its enrollees) yields leverage to insurers in contract negotiation. Participation in an insurer's restricted network guarantees a flow of business in exchange for which a provider may offer better pricing or other favorable contract terms, like keeping a practice open for patients of the insurer, accepting terms of utilization management, quality reporting, and so on. When contracting with providers, insurers can negotiate the terms including the payment formula: for example, fee-for-service, bundled payment, capitation, pay-for-performance, shared savings, or something else. In the United States, insurers in the Marketplaces rely on network restriction to bargain effectively with providers over price.[6]

In the three European countries, insurers generally have less scope to manage care—with some within-Europe variation. In all three countries, in contrast to the Marketplaces, services within the benefit package are guaranteed. There may often be waiting times for specific services, and it may not be possible to get treated by the preferred doctor, but the insured can depend on medically approved services being covered by their insurance. Also, insurers normally cannot override decisions of doctors and patients, although some specific services do require prior authorization (for example, dental braces in Germany).

In general, insurers in Germany are restricted to collective agreements between insurers and providers for contracting care. All authorized and accredited healthcare providers and services must be accepted and covered by the sickness funds. In Germany, insurers can offer restricted provider networks that the insured

[6]Along with average prices, variation in prices affects the return to selective contracting. Price variation is known to be high in the United States. National price schedules in Europe leave less provider-to-provider variation in the price of the same service.

can subscribe to in exchange for some advantages like reduced waiting times, but enrollees can leave these contracts at any time. German insurers' means to restrict provider networks, to negotiate prices, and to manage health care, are the least among the four countries. So, whereas insurers in Germany compete as individual sellers on the health insurance market, they largely act as a collective of buyers on the health care delivery market (van de Ven et al. 2024).

In the Netherlands, insurers can restrict provider networks but are obliged to cover out-of-network care sufficient to keep it "affordable" for enrollees. So far, however, Dutch insurers have been reluctant to engage in selective contracting, because the political and societal trust in insurers as purchasers of care turned out to be fragile (Schut, Henschke, and Or 2023). Antitrust regulations largely forbid collective price negotiations by Dutch insurers. Instead, insurers negotiate individually with provider organizations. The government, however, has some legal instruments to control spending: for example, it sets maximum prices for various types of care including hospital care and registered drugs.[7]

In Switzerland, insurers must offer a standard plan that provides free access and includes every approved provider. Next to the standard plan, however, insurers can offer managed care plans with network restrictions, but most are only gatekeeping plans restricting direct access to specialists. Fee schedules and prices are negotiated between insurer and provider associations in Switzerland and need to be approved by the government. If the two parties fail to agree, the government sets the schedules and prices. These schedules and prices are binding except in managed care plans, where the insurer can freely negotiate reimbursement of providers.

The relatively high reliance on markets and managed care in the United States in comparison to collective action in Europe helps to explain the relative prices and quantities in the two systems. Collective action may be better at constraining prices, but negotiation and managed care may be more effective at reducing utilization by counteracting moral hazard on the part of patients and providers.

Reliance on managed care for cost control comes at a price. Discretion about benefit administration equips an insurer with new means to discriminate in favor or against certain services and enrollees, a power which might be used not for purposes of resolving moral hazard, but to seek good and deter bad risks within an insurance pool (Newhouse 1996). By law, insurers in all four countries must accept all who seek to enroll, eliminating "direct selection" (denial of coverage for those anticipated to lose money) by insurers, but redirecting incentives to "indirect selection." Some enrollees are predictably unprofitable even with sophisticated risk-adjustment payment models, and so insurers have an incentive to deter

[7]Moreover, the Dutch Minister of Health can reclaim any overrun of the agreed-upon maximum total hospital expenditure growth by imposing a levy on each hospital in proportion to its revenue. Although this instrument has not been used so far, the possibility that it can be used might have incentivized insurers and hospitals to control total spending (Gajadien et al. 2023).

them.[8] In the United States, the means by which insurers can restrict access to the services valued by those whose anticipated costs exceed anticipated revenues are considerably greater than in the European countries. For example, access to services for children with serious mental health issues is poor in the United States, because families with this need are regarded as predictable "losers." In a typical market, competition pushes a seller to match product design to consumer preferences, but in the case of health insurance, when some consumers likely bring less revenue than they cost, this competitive incentive is turned on its head. Profit incentives can deter insurers from offering good, accessible quality of care for those with predictable losses. In the US Marketplaces, Geruso, Layton, and Prinz (2019) found that insurers set coverage in drug formularies to disfavor drugs used by enrollees with predictable losses within the risk-adjustment formula—as a profit-maximizing insurer would be predicted to do. Furthermore, consumers may not appreciate the degree of managed care they are buying when they choose a particular plan. Copayments and deductibles are easy enough to compare across plans; stringency of management is not.

Social Health Insurance Markets on Two Sides of the Atlantic

After contemplating similarities and differences among four social health insurance systems employing principles of regulated competition, it is natural to ask: which of these approaches do we regard as superior? Unsurprisingly, we are unable to answer this ultimate question. The reason is two-fold. First, the health outcomes of the four systems are hard to compare due to a lack of data and a universe of confounders. Second, the culture, history, and even the geography of the countries are different. Although all four systems pursue concepts of "fairness" and "efficiency" at a high level, the interpretation and weighing of these concepts varies across countries. What might be regarded superior in one country might be regarded second best in another. For the same reason, design and implementation of regulated competition is not a matter of one-size-fits-all.

In an earlier effort highlighting similarities and differences across national healthcare systems, van de Ven et al. (2013) compared the regulated health insurance markets in Belgium, Germany, Israel, the Netherlands and Switzerland. They evaluated the extent to which what they referred to as the "preconditions" for adequate functioning of these markets were in place. By identifying these preconditions and comparing countries in terms of the extent to which they are fulfilled, one gets an indication of the "maturity" of regulated health insurance markets. In a recent update, van de Ven et al. (2024) conclude that regulated competition is a

[8] Many papers in all four countries study "over and undercompensation" for groups of users to monitor such incentives. As examples, see Montz et al. (2016) for the United States; van Kleef, Eijkenaar, and van Vliet (2020) for the Netherlands; Schindler, Berndt, and Häckl (2026) for Germany; and Kauer, McGuire, and Beck (2020) for Switzerland.

work in progress in all five countries, an observation that could certainly include regulated competition health insurance markets in the United States.

Of course, regulated competition is not the only approach to financing health care. Other high-income countries pay for health care without introducing competing insurers as intermediaries between the government and consumers. In Canada, *socialized health insurance* run by the provinces pays private health care providers (for a Canadian-US comparison in some of the same spirit as our current paper, see Gruber 2022). In the United Kingdom, *socialized medicine* makes health care available directly to the population without a third-party payer. These partially or completely socialized systems can boast some notable accomplishments. The United Kingdom provides health care to its population at less than half the per capita cost of total US health care spending (OECD 2023). Administrative costs in Canada are about 20 percent of the $1,000+ per person in the United States (OECD 2023), and much less still on the other side of the Atlantic. Douven et al. (2022) finds that in the US Marketplaces, administrative costs make up 15 percent of the insurance premium, while for Germany, the Netherlands and Switzerland this is 5, 3, and 5 percent, respectively. Although it is not possible to make blanket statements on whether regulated competition is better or worse than other approaches for social health insurance policy, we can say that regulated competition has found a place at the national level in several European countries and the sectoral level in the United States and generates interest of researchers and policymakers around the world.

Supporting the functioning of regulated competition in all four countries are backroom operations in which premiums and other revenue sources are channeled to insurers enrolling a higher-risk population. The backroom is the realm of risk adjustment, risk sharing, and other payment system features not emphasized here but about which there has been extensive and on-going comparative research among countries employing regulated competition. Commonality in the forms of risk adjustment and statistical methods have facilitated much comparative empirical research (for example, McGuire, Schillo, and van Kleef 2020; van Kleef et al. 2024, and the references therein).

The lower profit orientation of health insurers in Europe raises an issue for research and policy that has not been explicitly addressed in the existing literature. If profit maximization is not the objective of insurers, then what is the objective? And what does this mean for how we design and evaluate incentives? For example, researchers study the effect of diagnosis-based and spending-based components of a payment system on incentives for cost control (Douven et al. 2025). The policy orientation of this literature has focused more on incentives, but less on the possibility that objectives of the targeted organization may differ. A smaller incentive might go a long way with a highly profit-oriented organization with broad scope of potential actions to increase profits, whereas a more powerful payment model might be necessary to get the attention of a less profit-oriented insurer. Generally, a fundamental premise of principal-agent analysis is that in order to design a contract for an agent (a health insurer), the principal (the regulator) needs to

be able to anticipate how the agent will respond to the contract. A conception of the objective of the health insurer seems necessary to formulate a reasonable expectation of how the insurer responds to regulation and incentives created by the payment system.

■ *We thank Konstantin Beck, Florian Buchner, Rudy Douven, Randy Ellis, Jacob Glazer, Joe Newhouse, Lennart Pirktl, Erik Schut, Maria Trottman, and Wynand van de Ven for comments on an earlier draft. This paper was presented at the 2025 meeting of the Risk Adjustment Network, September 11–13, 2025, at Stanford University, and we thank attendees for helpful discussion. Jeffrey Kling, Timothy Taylor, and Heidi Williams provided editorial guidance and helpful comments. We also thank Justin Li of the Belmont Hill School for excellent research assistance.*

References

Aron-Dine, Aviva, Liran Einav, and Amy Finkelstein. 2013. "The RAND Health Insurance Experiment, Three Decades Later." *Journal of Economic Perspectives* 27 (1): 197–222.

Arrow, Kenneth J. 1963. "Uncertainty and the Welfare Economics of Medical Care." *American Economic Review* 53 (5): 941–73.

ASPE (Assistant Secretary for Planning and Evaluation, Department of Health and Human Services). 2024. "Health Insurance Marketplaces: 10 Years of Affordable Private Plan Options." Issue Brief HP-2024–09.

Baicker, Katherine, and Amitabh Chandra. 2018. "Challenges in Understanding Differences in Health Care Spending between the United States and Other High-Income Countries." *Journal of the American Medical Association* 319 (10): 986–87.

Bauhoff, Sebastian. 2012. "Do Health Plans Risk-Select? An Audit Study on Germany's Social Health Insurance." *Journal of Public Economics* 96 (9–10): 750–59.

BAG (Bundesamt für Gesundheit). 2025. "Statistik der obligatorischen Krankenversicherung." Bundesamt für Gesundheit. https://www.bag.admin.ch/de/statistik-der-obligatorischen-krankenversicherung.

Baumol, William J. 1959. *Business Behavior, Value, and Growth.* MacMillian Publishing.

Boonen, Lieke H. H. M., Stéphanie A. van der Geest, Frederik T. Schut, and Marco Varkevisser. 2010. "Pharmaceutical Policy in the Netherlands: From Price Regulation towards Managed Competition." *Advances in Health Economics and Health Services Research* 22: 53–76.

Buchner, Florian, and Frederik T. Schut. 2026. "High-Risk Individuals in Voluntary Health Insurance Markets: The Elephant in the Room?" *Health Economics, Policy and Law* 21 (1): 40–53.

Carey, Colleen. 2017. "Technological Change and Risk Adjustment: Benefit Design Incentives in Medicare Part D." *American Economic Journal: Economic Policy* 9 (1): 38–73.

Cheng, Tsung-Mei, and Uwe E. Reinhardt. 2008. "Shepherding Major Health System Reforms: A Conversation with German Health Minister Ulla Schmidt." *Health Affairs* 27 (S1): 204–13.

Cutler, David M. 2002. "Equality, Efficiency, and Market Fundamentals: The Dynamics of International Medical-Care Reform." *Journal of Economic Literature* 40 (3): 881–906.

Cutler, David M., and Sarah J. Reber. 1998. "Paying for Health Insurance: The Trade-Off between Competition and Adverse Selection." *Quarterly Journal of Economics* 113 (2): 433–66.

Dafny, Leemore. 2019. "Does It Matter If Your Health Insurer Is For Profit? Effects of Ownership on Premiums, Medical Spending, and Insurance Coverage." *American Economic Journal: Economic Policy* 11 (1): 222–65.

Douven, Rudy, Shuli Brammli, Florian Buchner, Lukas Kauer, Richard van Kleef, Thomas McGuire, Francesco Paolucci, Marissa Reitsma, Sherri Rose, Christian P. R. Schmid, and Jürgen Wasem. 2025. "Power of Volume and Price Incentives in Health Plan Payment Models: A Country Comparison." Unpublished.

Douven, Rudy, Lukas Kauer, Sylvia Demme, Francesco Paolucci, Wynand van de Ven, Jürgen Wasem, and Xiaoxi Zhao. 2022. "Should Administrative Costs in Health Insurance Be Included in the Risk-Equalization? An Analysis of Five Countries." *European Journal of Health Economics* 23 (9): 1437–53.

Douven, Rudy C. H. M., and Frederik T. Schut. 2011. "Pricing Behavior of Nonprofit Insurers in a Weakly Competitive Social Health Insurance Market." *Journal of Health Economics* 30 (2): 439–49.

Einav, Liran, and Amy Finkelstein. 2011. "Selection in Insurance Markets: Theory and Empirics in Pictures." *Journal of Economic Perspectives* 25 (1): 115–38.

Einav, Liran, Amy Finkelstein, and Mark R. Cullen. 2010. "Estimating Welfare in Insurance Markets Using Variation in Prices." *Quarterly Journal of Economics* 125 (3): 877–921.

Einav, Liran, Amy Finkelstein, and Ray Fisman. 2023. *Risky Business: Why Insurance Markets Fail and What to Do about it*. Yale University Press.

Ellis, Randall P., Bruno Martins, and Sherri Rose. 2018. "Risk Adjustment for Health Plan Payment." In *Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets: Theory and Practice*, edited by Thomas G. McGuire and Richard C. van Kleef, 55–104. Academic Press.

Enthoven, Alain C. 1980. *Health Plan: The Only Practical Solution to the Soaring Cost of Medical Care*. Addison-Wesley Publishing.

Enthoven, Alain C. 2018. "Regulated Competition in Health Insurance Markets: Foreword by Alain Enthoven." In *Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets: Theory and Practice*, edited by Thomas G. McGuire and Richard C. van Kleef, xxi–xxviii. Academic Press.

Enthoven, Alain, and Richard Kronick. 1989. "A Consumer-Choice Health Plan for the 1990s: Universal Health Insurance in a System Designed to Promote Quality and Economy." *New England Journal of Medicine* 320 (2): 94–101.

Enthoven, Alain C., and Wynand P. M. M. van de Ven. 2007. "Going Dutch—Managed-Competition Health Insurance in the Netherlands." *New England Journal of Medicine* 357 (24): 2421–32.

Frank, Richard G., Jacob Glazer, and Thomas G. McGuire. 2000. "Measuring Adverse Selection in Managed Health Care." *Journal of Health Economics* 19 (6): 829–54.

FSIO (Federal Social Insurance Office). 2013. "History of Social Security in Switzerland." https://www.historyofsocialsecurity.ch/risk-history/illness (accessed February 6, 2026).

Gajadien, Chandeni C., Peter J. G. Dohmen, Frank Eijkenaar, Frederik T. Schut, Erik M. van Raaij, and Richard Heijink. 2023. "Financial Risk Allocation and Provider Incentives in Hospital–Insurer Contracts in the Netherlands." *European Journal of Health Economics* 24 (1): 125–38.

Geruso, Michael, Timothy Layton, and Daniel Prinz. 2019. "Screening in Contract Design: Evidence from the ACA Health Insurance Exchanges." *American Economic Journal: Economic Policy* 11 (2): 64–107.

GKV-Spitzenverband. 2025. "Die gesetzlichen Krankenkassen." https://www.gkv-spitzenverband.de/krankenversicherung/kv_grundprinzipien/alle_gesetzlichen_krankenkassen/alle_gesetzlichen_krankenkassen.jsp.

Glazer, Jacob, and Thomas G. McGuire. 2002. "Multiple Payers, Commonality, and Free-Riding in Health Care: Medicare and Private Payers." *Journal of Health Economics* 21 (6): 1049–69.

Glazer, Jacob, and Thomas G. McGuire. 2011. "Gold and Silver Health Plans: Accommodating Demand Heterogeneity in Managed Competition." *Journal of Health Economics* 30 (5): 1011–19.

Gruber, Jonathan. 2022. "Financing Health Care Delivery." *Annual Review of Financial Economics* 14: 209–29.

Handel, Ben, Igal Hendel, and Michael D. Whinston. 2015. "Equilibria in Health Insurance Exchanges: Adverse Selection versus Reclassification Risk." *Econometrica* 83 (4): 1261–313.

Henriquez, Josefa, Marica Iommi, Thomas G. McGuire, Emmanouil Mentzakis, and Francesco Paolucci. 2023. "Designing Feasible and Effective Health Plan Payments in Countries with Data Availability Constraints." *Journal of Risk and Insurance* 90 (1): 33–57.

Henriquez, Josefa, Thomas McGuire, Richard van Kleef, Francesco Paolucci, and Andrew Matthews.

2026. "Policy Innovation in Voluntary Hospital Private Health Insurance: Combining Risk Adjustment and Risk Sharing in Australia." *Health Policy* 167: 105584.

Jeurissen, Patrick, and Hans Maarse. 2021. *The Market Reform in Dutch Health Care: Results, Lessons and Prospects*. European Observatory on Health Systems and Policies.

Kaiser Family Foundation. 2025. "Employer-Sponsored Health Insurance 101." October 8. https://www.kff.org/health-costs/health-policy-101-employer-sponsored-health-insurance/?entry=table-of-contents-introduction.

Kannan, Sneha, and Zirui Song. 2025. "Variation in Hospital Salary Expenditures and Utilization Changes after Private Equity Acquisition, 2005–19." *Health Affairs* 44 (2): 206–14.

Kauer, Lukas, Thomas G. McGuire, and Konstantin Beck. 2020. "Extreme Under and Overcompensation in Morbidity-Based Health Plan Payments: The Case of Switzerland." *Health Policy* 124 (1): 61–68.

Kong, Edward, Mark Shepard, and Adrianna McIntyre. 2022. "Turnover in Zero-Premium Status among Health Insurance Marketplace Plans Available to Low-Income Enrollees." *JAMA Health Forum* 3 (4): e220674.

Layton, Timothy J., Ellen Montz, and Mark Shepard. 2018. "Health Plan Payment in US Marketplaces: Regulated Competition with a Weak Mandate." In *Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets: Theory and Practice*, edited by Thomas G. McGuire and Richard C. van Kleef, 491–522. Academic Press.

McDermott, Daniel, and Cynthia Cox. 2020. *Insurer Participation on the ACA Marketplaces, 2014–2021*. Kaiser Family Foundation.

McGuire, Thomas G. 2012. "Demand for Health Insurance." In *Handbook of Health Economics*, Vol. 2, edited by Mark V. Pauly, Thomas G. McGuire, and Pedro P. Barros, 317–96. North-Holland.

McGuire, Thomas G., Sonja Schillo, and Richard C. van Kleef. 2020. "Reinsurance, Repayments, and Risk Adjustment in Individual Health Insurance: Germany, the Netherlands and the US Marketplaces." *American Journal of Health Economics* 6 (1): 139–68.

McGuire, Thomas G., and Richard C. van Kleef, eds. 2018. *Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets: Theory and Practice*. Academic Press.

Montz, Ellen, Tim Layton, Alisa B. Busch, Randall P. Ellis, Sherri Rose, and Thomas G. McGuire. 2016. "Risk Adjustment Simulation: Health Plans May Have Incentives to Distort Mental Health and Substance Abuse Coverage." *Health Affairs* 35 (6): 1022–28.

Mulcahy, Andrew W., Daniel Schwam, and Susan L. Lovejoy. 2024. "International Prescription Drug Price Comparisons: Estimates Using 2022 Data." *RAND Health Quarterly* 11 (3): 5.

Newhouse, Joseph P. 1993. *Free-for-All: Health Insurance, Medical Costs, and Health Outcomes: The Results of the Health Insurance Experiment*. Harvard University Press.

Newhouse, Joseph P. 1996. "Reimbursing Health Plans and Health Providers: Efficiency in Production versus Selection." *Journal of Economic Literature* 34 (3): 1236–63.

Netherlands Authority for Consumers and Markets. 2017. *Toetredings—en groeidrempels op de zorgverzekeringsmarkt*. Netherlands Authority for Consumers and Markets.

OECD. 2023. *Health at a Glance 2023: OECD Indicators*. OECD Publishing. https://doi.org/10.1787/7a7afb35-en (accessed June 10, 2025).

OECD Data Explorer. 2025. *Health Expenditure and Financing*. https://www.oecd.org/en/data/datasets/oecd-DE.html (accessed June 10, 2025).

Ortaliza, Jared, and Cynthia Cox. 2024. "2024 Medical Loss Ratio Rebates." Kaiser Family Foundation, June 5. https://www.kff.org/private-insurance/medical-loss-ratio-rebates/.

Ortaliza, Jared, Matt McGoug, Kaitlyn Vu, Imani Telesford, Shameek Rakshit, Emma Wager, Lynne Cotter, and Cynthia Cox. 2025. "How Much and Why ACA Marketplace Premiums Are Going Up in 2026." Peterson–KFF Health System Tracker, August 6. https://www.kff.org/affordable-care-act/how-much-and-why-aca-marketplace-premiums-are-going-up-in-2026/.

Pauly, Mark V. 1968. "The Economics of Moral Hazard: Comment." *American Economic Review* 58 (3): 531–37.

Peterson Foundation. 2024. "How Does the Federal Government Subsidize Healthcare under the ACA and What Does It Cost?" December 6. https://www.pgpf.org/article/how-does-the-federal-government-subsidize-healthcare-under-the-aca-and-what-does-it-cost/.

Schindler Christian, J. A., Benjamin Berndt, and Dennis Häckl. 2026. "Pharmacy Cost Groups for the German Morbidity-Based Risk Compensation Scheme." *European Journal of Health Economics* 27 (1): 135–47.

Schut, Frederik T., Cornelia Henschke, and Zeynep Or. 2023. "Changing Roles of Health Insurers in

France, Germany, and the Netherlands: Any Lessons to Learn from Bismarckian Systems?" *Health Economics, Policy and Law* 18 (4): 362–76.

Shepard, Mark. 2022. "Hospital Network Competition and Adverse Selection: Evidence from the Massachusetts Health Insurance Exchange." *American Economic Review* 112 (2): 578–615.

Singh, Yashaswini, Zirui Song, Daniel Polsky, and Jane M. Zhu. 2025. "Increases in Physician Professional Fees in Private Equity-Owned Gastroenterology Practices." *Health Affairs* 44 (2): 215–23.

Stolper, Karel C. F., Lieke H. H. M. Boonen, Frederik T. Schut, and Marco Varkevisser. 2022. "Do Health Insurers Use Target Marketing as a Tool for Risk Selection? Evidence from the Netherlands." *Health Policy* 126 (2): 122–28.

Tevis, Delaney, Matt McGough, Juliette Cubanski, Matthew Rae, and Cynthia Cox. 2025. "How Do Healthcare Prices and Utilization in the United States Compare to Peer Nations?" Peterson–KFF Health System Tracker, September 4. https://www.kff.org/health-costs/how-do-healthcare-prices-and-use-in-the-u-s-compare-to-other-countries/.

Tobin, James. 1970. "On Limiting the Domain of Inequality." *Journal of Law and Economics* 13 (2): 263–77.

van Kleef, Richard C., Frank Eijkenaar, and René C. J. A. van Vliet. 2020. "Selection Incentives for Health Insurers in the Presence of Sophisticated Risk Adjustment." *Medical Care Research and Review* 77 (6): 584–95.

van Kleef, Richard C., Mieke Reuser, Thomas G. McGuire, John Armstrong, Konstantin Beck, Shuli Brammli-Greenberg, Randall P. Ellis, Francesco Paolucci, Erik Schokkaert, and Juergen Wasem. 2024. "Scope and Incentives for Risk Selection in Social Health Insurance Markets with Regulated Competition: A Conceptual Framework and International Comparison." *Medical Care Research and Review* 81 (3): 175–94.

van Kleef, Richard C., and René C. J. A. van Vliet. 2025. *Evaluatie risicoverevening 2024 and 2025.* Erasmus School of Health Policy and Management.

van de Ven, Wynand P. M. M., Konstantin Beck, Florian Buchner, Erik Schokkaert, Frederick T. Schut, Amir Shmueli, and Juergen Wasem. 2013. "Preconditions for Efficiency and Affordability in Competitive Healthcare Markets: Are They Fulfilled in Belgium, Germany, Israel, the Netherlands and Switzerland?" *Health Policy* 109 (3): 226–45.

van de Ven, Wynand P. M. M., Konstantin Beck, Florian Buchner, Erik Schokkaert, Frederick T. Schut, Amir Shmueli, and Juergen Wasem. 2024. "Preconditions for Efficiency and Affordability in Competitive Healthcare Markets: Are They Fulfilled in Belgium, Germany, Israel, the Netherlands and Switzerland? Ten Years Later." *Health Policy* 146: 105–09.

van de Ven, Wynand P. M. M., and Randall P. Ellis. 2000. "Risk Adjustment in Competitive Health Plan Markets." In *Handbook of Health Economics*, Vol. 1A, edited by Anthony J. Culyer and Joseph P. Newhouse, 755–845. Elsevier.

van de Ven, Wynand P. M. M., and Frederik T. Schut. 2008. "Universal Mandatory Health Insurance in the Netherlands: A Model for the United States?" *Health Affairs* 27 (3): 771–81.

Vektis. 2025. *Zorgthermometer 2025: Verzekerden in beeld.* Vektis Intelligence.

Ways and Means Committee Staff. 2019. *A Painful Pill to Swallow: U.S. vs. International Prescription Drug Prices.* US House of Representatives.

Zeckhauser, Richard. 1970. "Medical Insurance: A Case Study of the Tradeoff between Risk Spreading and Appropriate Incentives." *Journal of Economic Theory* 2 (1): 10–26.

Zorginstituut. 2026. *Zorgcijfersdatabank.* https://www.zorgcijfersdatabank.nl/ (accessed February 17, 2026).

Journal of Economic Perspectives—Volume 40, Number 2—Spring 2026—Pages 69–92

Anticompetitive Contracts Between Insurers and Providers in Health Care

Anna D. Sinaiko

In 2024, 213.1 million Americans (63.2 percent of the population) were covered by health insurance insured via a private (that is, employer-sponsored or direct-purchase) commercial insurance plan (Bunch and Ketema 2025). In the economics literature, the most common justification for health insurance is that by pooling people together into a health insurance plan and charging a premium in exchange for coverage of future health care costs, insurers protect consumers against large expenditures in the case of future illness. This effect reduces uncertainty in a consumer's future consumption, and because the average consumer is risk-averse, greater certainty makes consumers better off (Arrow 1963; Cutler and Zeckhauser 2000).

However, this essay will focus on a different way in which health insurers could benefit consumers: via the contracts between insurers and health care providers—including hospitals, physicians, and pharmacies—that set the terms for the medical services received by patients. These contracts can potentially be welfare-improving for consumers in two ways. First, although contracts vary, they generally include a fee schedule, achieved through insurer-provider bargaining, that specifies how providers will be paid and the prices that providers will receive for different services. These prices represent the total amount that a provider receives for each service, and in practice are made up of a combination of payments from the insurer and the consumer at the time the service is provided. This bargaining can lead to welfare gains for consumers because of asymmetric information between providers and

■ *Anna D. Sinaiko is Associate Professor of Health Economics and Policy, Harvard TH Chan School of Public Health, Boston, Massachusetts. Her email address is asinaiko@hsph.harvard.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20251471.

consumers—consumers do not know the types, quantities, or quality of services they need (McGuire 2000)—and because it avoids transaction costs that would be incurred if all patients had to negotiate prices themselves.

Second, insurer contracting can counter market power of health care providers by rewarding a provider with volume in exchange for price concessions. This outcome is achieved primarily through what is called "selective contracting," which occurs when an insurer only contracts with some of the providers in a geographic area, but not all, and then distinguishes those with contracts as "in-network" providers. Insurers give patients financial incentives to seek care from in-network providers, typically by charging higher cost-sharing or providing no coverage for services from out-of-network providers. Because of this threat of exclusion from the network and the associated loss of patient volume, health care providers will compete to be in-network, accepting lower prices and other terms favorable to the insurer. In this way, "selective contracting" can reduce spending by shifting utilization to lower-price providers, and can further improve welfare by creating additional pressure among in-network providers to lower their prices. As we will discuss, insurers can also use other aspects of benefit design to steer patients to lower-cost providers, with similar goals of lower spending and creating price pressure on higher-cost providers.

However, insurer-provider contract provisions do not always advance competition and consumer welfare. In some places, health care markets can have dominant provider systems, who may be able to negotiate terms with insurers that have the opposite effect, protecting themselves from competition.[1] There are four categories of anticompetitive contract terms. First, "anti-tiering" and "anti-steering" clauses restrict an insurer from encouraging consumers, whether through financial incentives or other means, to obtain health care services from other health systems that are competitors of the contracting provider system. Second, "all or nothing" clauses require an insurer to contract with all facilities in the health system if they want to include any facilities in the plan. Both types of clauses prevent insurers from selectively excluding or steering consumers away from higher-price providers, which would promote competition. Third, a "most favored nation" clause can set floors for negotiated prices, inhibiting price competition. Fourth, "gag clauses" prevent insurers from disclosing the health care prices paid to providers, along with other details about insurer-provider contracts, which inhibits the ability of consumers, or employer purchasers on their behalf, to shop for care based on prices. All four categories of contract provisions can allow dominant health provider systems to maintain and exercise monopoly power, which is almost always harmful to consumers. Understanding these contract provisions and the potential policy responses to limit their impact is thus critical to health care competition.

[1] There are other reasons why results of insurer-provider contracting in practice are not optimal, and two important reasons for inefficient outcomes are asymmetric information and selection in insurance markets. For a review of these market failures and their impact on insurer-provider contracting, see in this journal Einav and Finkelstein (2011).

In this essay, I define anticompetitive behavior not in a narrow legal sense, but in a broader economic sense—that is, as the exercise of monopoly or market power and including provider behaviors that lead to increased quantities of care, restrict quality, set prices above competitive levels, or create barriers to entry for competing firms. Although my focus here will be on market power of health care providers, the effect of insurer market power on competition can also be important, though more ambiguous; while insurer market power can be a response to counter provider market power and bargain for lower prices (Scheffler and Arnold 2017; Trish and Herring 2015), dominant insurers also face less pressure from consumers in the market and thus may not pass on these price concessions in the form of lower premiums. To focus this paper, I do not evaluate the anticompetitive effects from insurer market power except in cases when it augments provider anticompetitive behaviors by health care producers.

This paper begins by describing insurer strategies in insurer-provider contracts that are aimed at countering provider market power and improving efficiency, along with evidence on the effects of these strategies. I then turn to the anticompetitive clauses that powerful providers have sought to include in insurer-provider contracts to counter these insurer strategies, and discuss their anticompetitive effects. I conclude with a discussion of policy responses that can be used to address provider use of anticompetitive clauses and that can reduce the negative impacts of provider market power.

Insurer-Provider Contracting to Foster Health Care Competition

Contracting and Commercial Health Plan Design

The majority of health systems in high-income countries as well as public programs in the United States set administered prices for health care services; in contrast, in US commercial health insurance, the negotiations behind insurer-provider contracts determine both the forms and levels of provider payment. The predominant form of payment in the United States is fee-for-service, where providers receive a fee for every service—that is, every visit, test, and procedure by physicians, or inpatient stay in hospitals—that is delivered. Many health care providers have lists of fees they charge per service, known as "chargemasters," but in actuality, very few patients pay these fees. Instead, insurers and providers negotiate fee schedules that define the actual prices paid to providers for care for the patients enrolled in the insurers' plans. In health insurance claims, these prices are called the "allowed amount." Alternative forms of payment also common in the US health care system are risk-based "capitated" contracts where providers accept an overall payment for costs of care incurred by a patient for a predetermined period of time, or bundled payments where providers accept a predetermined payment for all services provided in an episode of care (say, all the separate services involved in a joint replacement surgery).

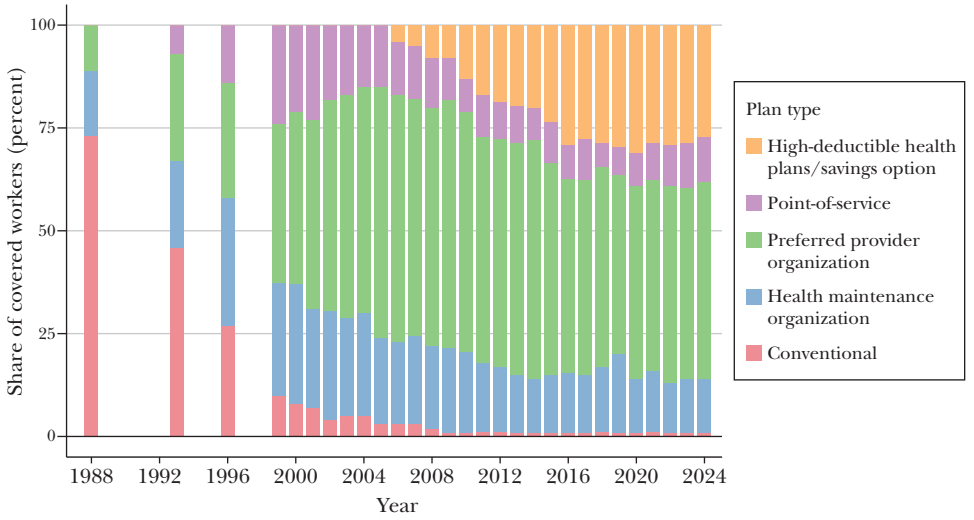
Commercial health plans vary in their design, and two dimensions related to insurers' price negotiations with providers are the breadth of their provider

network and degree of health insurance coverage for care received from providers who are out-of-network. A plan's provider network includes all providers who have contracted with the insurer to participate in that plan, and consumers receiving care from one of these providers will have all care covered aside from their plan's cost-sharing obligations (in the form of required deductibles, copayments, and coinsurance). Plans can have completely open networks where all providers in a geographic area are included; they can have broad networks where the vast majority, but not all providers in an area are included; or they can have narrow or limited networks where fewer than half of providers in an area are included in-network. In terms of coverage, some types of plans provide no coverage for care received from out-of-network providers, and these are typically health maintenance organization (HMO) plans. (These plans do usually include an exception process so that in special circumstances, out-of-network care may be covered by the insurer.) Other plans are less strict, and patients receive some coverage, though must pay higher cost-sharing, for care from out-of-network providers in comparison to in-network providers; preferred provider organizations (PPO) and point-of-service (POS) plans are in this category. Plans with open networks are generally known as conventional or indemnity plans.

Health maintenance organizations, preferred provider organizations, and point-of-service plans are typically known as "managed care." Managed care, a broad term, generally includes the use of financial incentives to physicians to limit utilization, efforts to restrict the services patients received through supply-side controls such as prior authorization and utilization review, and bargaining with networks of health care providers to obtain lower prices.

The distribution of consumer market share across plan type has changed significantly since the mid-1980s, increasing the importance of selective contracting between insurers and providers in the US market (as shown in Figure 1). Among consumers with employer-sponsored health insurance, having an insurance plan with no network restrictions went from being the most common plan type with 73 percent of all enrollments in the late 1980s to barely present in the market by the early 2000s. Health maintenance organizations represented as much as 30 percent of all plans in the late 1990s, but have gradually fallen to about 13 percent of plans by 2024. Preferred provider organizations have risen from about one-quarter of all plans in the mid-1990s to nearly half of all plans today. Point-of-service plans were uncounted in the 1988 data, rose to almost one-quarter of all plans in the mid-1990s, but since then have declined to more like 10 percent of these plans. Finally, high-deductible health plans with a savings option (HDHP/SO) were established via federal legislation in the mid-2000s and have grown in prevalence considerably since that time. In these plans consumers can save their own money in a tax-preferred account, which can be used to pay the high deductibles before insurance coverage kicks in; these plans typically are paired with broader networks (like that of a PPO), but could have narrower networks. HDHP/SO plans first appear in this data in 2006 and now represent more than one-quarter of all coverage.

Figure 1
Distribution of Health Plan Enrollment for Covered Workers, by Plan Type, 1988–2024



Source: Kaiser Family Foundation (KFF) Employer Health Benefits Survey, 2018–2024; Kaiser/HRET Survey of Employer-Sponsored Health Benefits, 1999–2017; KPMG Survey of Employer-Sponsored Health Benefits, 1993 and 1996; The Health Insurance Association of America (HIAA), 1988.
Note: Information was not obtained for POS plans in 1988 or for HDHP/SO plans until 2006. See the Survey Design and Methods section from the 2005 Kaiser/HRET Survey and the 2021 KFF Survey for a discussion of weighting changes.

The story underlying the changing prevalence of managed care, and in particular the decreased penetration of health maintenance organizations and increased penetration of preferred provider organization plans, is that during the early 2000s, consumers expressed strong dislike of HMO's narrow provider networks (which created relatively tight restrictions on their choice of provider) and of other utilization management practices (like requiring a referral from a primary care physician before allowing specialist visits). The resulting "managed care backlash" led to a rise in preferred provider organizations and point-of-service plans, which tend to have broader networks, less use of utilization management, and some partial coverage for out of network care (Levitt 2025). These trends have had consequences for insurers' ability to use selective contracting over time, and they have raised the importance of insurers being able to also use benefit design to steer consumers to lower-price providers within broader provider networks. I will now discuss in turn how selective contracting and benefit design play a role in fostering provider competition.

Selective Contracting and Provider Competition

"Selective contracting" is the term used to describe insurer practice of contracting with some, but not all, providers in a geographic area. In the presence

of selective contracting, provider competition can be thought of as a two-stage game, where in the first stage providers compete to be in the insurer's network and in the second stage providers chosen to be within the network compete for patients (Vistnes 2000). Patients in the second stage are insured and thus pay very little of the full price of their care if they remain in-network, but would pay substantially higher costs if they choose to receive out-of-network care.

In general, health care provider markets are consistent with monopolistic competition, primarily because asymmetric information implies that providers are considered "imperfect substitutes" by patients (for more detail on this literature, see McGuire 2000). However, health care providers have used asymmetric information in combination with their market power to determine quantities of care, restrict quality of care, and negotiate prices that are above the competitive level (McGuire 2000). Prices above the competitive level include profits, or "rents" in economic terms, and are inefficient. Thus, selective contracting between insurers and providers in the first-stage can be viewed as a method for creating price and quality competition.[2] Dranove, Shanley, and White (1993) terms this first-stage as "payer-driven competition," because in order to obtain patients, providers must respond to payers (another term for insurers) as well when setting prices.

The breadth of an insurer's provider network and the comprehensiveness of a plan's out-of-network coverage affects the ability of a health insurer to bargain for lower prices. Preferred provider organization plans, a form of managed care with broader provider networks and some insurance coverage for out-of-network care, effectively give consumers in these plans more choices among providers and weaker financial incentives to stay in-network for care. As a result, providers in a preferred provider organization network face more competition for consumers in the second stage. In contrast, health maintenance organization plans have narrower provider networks and nearly no coverage for care received from out-of-network providers, so in-network providers face less competition with other providers. Insurers are thus able to extract greater price concessions from health care providers in exchange for inclusion in their HMO plan networks, because those providers know that by being in the HMO network they will have higher volumes of patients.

Impacts of Selective Contracting

Managed care insurance plans, in particular health maintenance organizations, are very effective at bargaining for lower prices from health care providers.[3]

[2] The classic industrial organization model of health care describes that first providers determine their investments in quality, which are affected by the provider's expected degree of horizontal competition from other providers and the expected impact of these quality investments on consumer demand. Then, given these quality levels, providers negotiate with insurers to determine insurers' provider networks and the prices paid to providers. Gaynor, Ho, and Town (2015) provide an excellent description of this theory and the impact of prices on health care quality.

[3] All commercial health care prices in the United States are higher, however, than prices paid in public programs in the United States and by payers in other OECD countries (Anderson, Hussey, and Petrosyan 2019; Congressional Budget Office 2022).

In a paper by Cutler, McClellan, and Newhouse (2000), the authors start from the observation that managed care spending in 1994–1995 was 30–40 percent lower than indemnity-style plans. They decompose lower spending in managed care plans into a price component and a quantity/quality component. Using health insurance claims data from a very large firm in Massachusetts and statewide data on hospital inpatient admissions, the authors compared health care prices paid by traditional, indemnity-style health plans to those of managed care plans for patients with heart disease. The focus on heart disease strengthens this study design. The condition matters in quantitative terms because it is common and expensive to treat. But more important, the focus on heart disease reduces the threat of selection bias in their results, because it is not likely to be known by the patient in advance and thus unlikely to have influenced the patient's choice of health plan.

In this paper, the treatments for heart disease were categorized as medical management for the condition, cardiac catheterization, bypass surgery, and angioplasty (that is, using a catheter with a balloon tip to reduce obstruction in arteries). In all four categories, as shown in Table 1, the health maintenance organizations reimbursed only about 60 percent or less what indemnity insurance was paying for the same medical services. This study found minimal differences in the treatment rates by plan type. Thus, their aggregate finding that average reimbursement by HMOs was 40 percent lower than indemnity plans persisted whether the analysis used raw data (known as unadjusted results) or controlled for treatment type (reported in the column showing adjusted results). Moreover, this study did not show any differences in rates of consumers who had these heart disease treatments or differences in rates of consumers with adverse outcomes across plans. However, in other medical care contexts, managed care plans with narrow networks have led to decreased quantities of care, as well as lower-paid prices (Ma and McGuire 2002; Gruber and McKnight 2016).[4]

Today, indemnity plans have nearly disappeared from the commercial insurance market, and most consumers are enrolled in a health maintenance organization or a preferred provider organization plan. HMOs, which have narrower networks and stronger financial incentives for consumers to seek care in-network, have lower prices than PPO plans. Using data from Massachusetts from 2009 to 2011, Craig Ericson, and Starc (2021) find that HMOs pay prices about 5.4 percent lower than PPOs and that greater discounts are achieved by HMOs with more consumers (and more volume to steer to in-network providers). The magnitude of price difference, at 5.4 percent, is much smaller than observed in Cutler, McClellan, and Newhouse (2000). This is likely for a few reasons. First, the HMO prices in Craig, Ericson, and Starc are compared to those paid by PPO plans, which include selective contracting and thus exert some downward pressure on prices in comparison to the indemnity plans studied in Cutler, McClellan, and Newhouse (2000). Also, Craig, Ericson, and Starc analyze prices for a different set of medical services, including joint

[4]Similar results have been found in the Medicaid setting (Wallace 2023).

Table 1
Heart Attack (Acute Myocardial Infarction) Reimbursement by Plan Type

Plan	Average reimbursement (unadjusted)	Treatment Regimen				Average reimbursement (adjusted)
		Medical management	Cardiac catherization	Bypass surgery	Angioplasty	
Indemnity	$38,502	$26,601	$38,448	$97,347	$41,597	$39,410
HMO	$23,632	$16,318	$17,604	$55,826	$24,181	$22,836
HMO as percent of indemnity	61%	61%	46%	57%	58%	58%

Source: Cutler, McClellan, and Newhouse (2000).
Note: Reimbursement is within 90 days of the heart attack.

replacement procedures, childbirth, and imaging. Finally, it is likely that differences across insurers (in negotiations, size, and market power) also affects prices. The analysis of Craig et al. controls for the insurer, and their estimates represent a within-insurer effect of HMO contracts (relative to PPO) on prices.

Benefit Design and Competition Among In-Network Providers

An important driver of health plan design, by which I mean both a plan's provider network and the cost-sharing obligations facing consumers for care, is consumer demand for insurance plans, because insurers ultimately need to sell their plans to employer purchasers. In turn, these employer purchases face pressure to be competitive in labor markets and will seek to offer plan designs that align with employee preferences (Bundorf 2002; Tilipman 2022). As noted earlier, there was a "managed care backlash" in the early 2000s, which in turn led less restrictive managed care options like preferred provider organizations and point-of-service plan networks to become more popular—and also to weaken insurer ability to extract price concessions from the providers in their broader networks. As a result, wide differences in health care prices existed across providers within a single plan's network (Sinaiko, Kakani, and Rosenthal 2019). Moreover, the wide variation in health care prices across in-network providers seems only weakly correlated with health care quality but strongly related to provider market power (Hussey, Wertheimer, and Mehrotra 2013; Cooper, Scott Morton, and Shekita 2020).

However, in the typical US commercial health plans in the 1990s and early 2000s, consumers were not exposed to this provider price variation at the point of care.[5] Instead, insurers often used co-payments for care, which are typically fixed prices that are the same for providers of the same specialty or type in-network.

[5] Zeckhauser (1970) showed that in a competitive market, the optimal demand-side cost sharing scheme offered by insurers balances risk protection and moral hazard, by varying cost-sharing with demand-response for care. Optimal cost-sharing will be higher for care that is more price elastic, to reflect the less favorable risk protection/moral hazard tradeoff with more elastic demand.

Co-payments, because they do not vary across providers, thus give consumers no incentive to choose lower-price providers within-network. Recall that in the second stage of the two-part Vistnes (2000) game, in-network providers compete for patients. With fixed co-payments, competition among providers for consumers is mostly "non-price," and presumably based on observable differences including geographic location, customer service, and reputation.

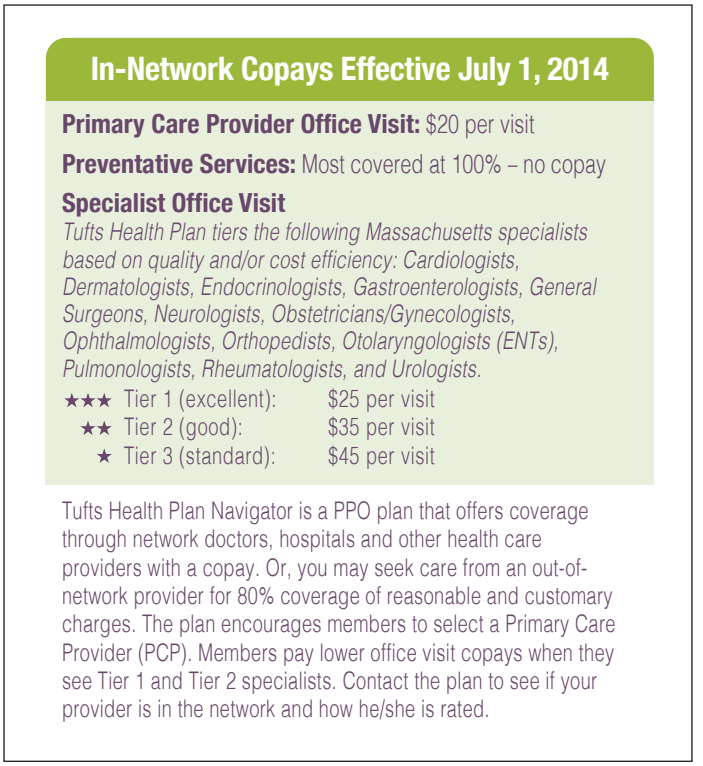
The first attempt by insurers to use benefit design to foster demand response to provider price differences among in-network providers was the use of high-deductible health plans. These plans include a deductible of $1000 or more for the vast majority of in-network care, although regulations exempt specified preventive care from the deductible. In 2023, nearly one in three patients insured through an employer was enrolled in a high-deductible health plan, up from fewer than one in 20 in 2006 (as shown earlier in Figure 1). These high-deductible health plans provide an incentive for consumers to "shop" based on price for their care, because consumers who receives care from a lower-price provider will spend less out-of-pocket when they are under their deductible.[6] However, this incentive only exists for consumers who do not expect their annual health care spending to exceed their deductible level. If consumers expect to exceed their deductible in the benefit period, then pre-deductible prices should not affect use of care (Brot-Goldberg et al. 2017).

A second form of benefit design stratifies providers who are in-network based on provider performance on cost and/or quality measures—and then assesses differential cost-sharing for providers in different strata. Examples of these are tiered network plans and reference-based pricing benefit design. In tiered networks, consumers pay lower cost-sharing when they choose a physician or hospital assigned to a preferred, or top-ranked, tier. Tiered physician network designs generally vary the co-payment for an office visit: for example, a patient might pay $25 for an office visit with a preferred (top-tier) physician, $35 for a visit with a middle-tier physician, and $45 for an office visit with a non-preferred (bottom tier) physician. Figure 2 reproduces the description of a tiered system from the Tufts Health Plan in Massachusetts in 2014. To make the financial incentive stronger, differentials in office visit co-payments between tiers can be larger: say, $30/$60/$90 across strata in the tiered network. Tiered hospital networks use differential deductible amounts, like $500 versus $1500 for admission to a preferred versus non-preferred hospital, to incentivize patients to choose preferred-tier hospitals. In effect, this benefit design creates price differences for in-network providers at the point of care, which works as an incentive to steer patients to the in-network providers who are lower cost.

Reference-based pricing works similarly to a tiered system but only applies to specific medical services. For services subject to reference-based pricing, a consumer who selects a provider that charges above the health plan determined "reference price" for a service must pay the required cost-sharing in addition to the difference

[6]High-deductible health plans are also based on economic theory which argues that among the various forms of cost-sharing, deductibles are optimal because they produce a reduction in wasteful care with the smallest possible financial harm to patients.

Figure 2
Sample Tiered Network Plan Cost Sharing Design



Source: Massachusetts Group Insurance Commission Benefit Decision Guide, Fiscal Year 2014.

between the reference price and the provider's price. To avoid higher cost-sharing, patients must seek care from providers where the price of care is below the reference price. The process can be made easier when insurance carriers publish lists of providers, or implement patient outreach and education programs, to help identify providers below the reference price.

Effects of Benefit Design on Consumer Behavior and Prices

High-deductible health plans, plans with tiered provider networks, and reference-based pricing benefit design all seek to steer consumer demand away from higher price providers but have had mixed success. For high-deductible health plans, the promise that consumers would shop for care based on price and foster competition within insurer networks has not been borne out. Consumers in high-deductible plans do use less health care and have lower spending than consumers in traditional plans without a high-deductible, but they have not been proven good at discriminating in their use of care based on value. Instead, they decrease use of

both high- and low-value services (Parente, Feldman, and Christianson 2004; Christianson, Parente, and Feldman 2004; Zhang et al. 2018; Brot-Goldberg et al. 2017), and they are no more likely than those in lower- or no-deductible plans to choose lower-cost providers (Zhang et al. 2018).

In contrast, tiering and steering network designs strategies have affected consumer choice of provider and in some instances, created downward pressure on provider prices (Mazurenko, Taylor, and Menachemi 2022). For example, physician copayment differentials of $10 across three tiers decreased market share of new patients for physicians in the worst (that is, highest-copay) tier by 11 percent (Sinaiko and Rosenthal 2014). These effects were concentrated among consumers who were older and sicker, and thus in expectation were likely to need higher quantities of care (Sinaiko 2016). Differential inpatient hospital deductibles with a span of $1250 from lowest to highest were found to decrease hospital volume among the least preferred (highest-deductible) hospitals by 7.9 percent and to increase volume among the lowest-deductible hospitals by 5.9 percent (Prager 2020).[7] Plans with tiered provider networks had on the order of 5 percent lower total medical spending than health plans requiring the same cost-sharing for all providers (Sinaiko, Landrum, and Chernew 2017; Ackley 2022), although some evidence suggests that these demand effects may fade over time (Prager et al. 2023). On the provider side, primary care clinics participating in the Minnesota State Employee Group Insurance Program for health benefits agreed to price reductions or alternate risk-sharing arrangements to move to (or remain in) a tier with lower cost-sharing (McDonald et al. 2021).

Reference-based pricing benefit design has been associated with consumer selection of lower-cost alternatives for several outpatient medical services, including colonoscopy, laboratory tests, cataract surgery, advanced imaging services, and non-emergent orthopedic surgery (Robinson and Brown 2013; Robinson, Brown, and Whaley 2015; Robinson, Whaley, and Brown 2016). From these steering effects—that is, consumers receiving care from price providers who generally charge lower prices—these programs achieve lower spending. Models extending the behavioral responses from across these reference-based pricing benefit design programs to the full set of services in the commercial market well-suited to this benefit design—that is, to nonemergency procedures that are scheduled in advance, that take place in markets with multiple providers, and for which some quality metrics are available to patients—estimate, as an upper bound, that savings on the order of 20 percent of spending on these services could be realized through these programs (Robinson, Brown, and Whaley 2017).

[7] A similar form of benefit design is three-tiered formularies for prescription drugs, in which drugs are allocated to tiers ranging from most cost-effective and least expensive (often including generic drugs) to more expensive and specialized drugs. This approach has been very effective at steering consumer demand to the lowest cost-sharing tier (Motheral and Fairman 2001). As in the case with provider market power, these formularies were established in part to counter the market power of pharmaceutical companies.

In some settings, benefit design using reference-based pricing also led to decreases in provider prices. One study looked at a reference-based pricing benefit design for non-emergent hip and knee replacement surgeries in California (Robinson and Brown 2013). As Figure 3 illustrates, the study included two groups: one group of patients in health plans with reference-based pricing, and a comparison group of patients in health plans offered by the same carrier but without reference-based pricing. Both patients had access to the same California hospitals for their hip or knee replacement surgeries, and these hospitals were divided into two groups based on price. The low-price hospitals offered hip and knee replacement at a price below $30,000, which was the reference price at the time when reference-based pricing was implemented. The high-price hospitals had prices for these procedures above $30,000. The reference-based pricing aimed to steer consumers to lower price providers, and that was what was observed. The share of patients in a plan with reference-based pricing choosing the high-priced facilities dropped dramatically when the program was implemented, as shown in the top panel. Moreover, the high-price hospitals reacted by lowering their prices for these orthopedic procedures towards the level of the reference price, as shown in the bottom panel.

Another program that assessed differential cost-sharing for laboratory tests in New Hampshire led to lower provider prices. This program required zero cost sharing at preferred providers and a deductible for non-preferred providers. Analysis of the program's impact found that program provider lab prices fell 12–15 percent, and these price decreases accounted for about half the magnitude of program savings (Ackley 2025).

Provider Response and Anticompetitive Contract Provisions

It is reasonable to expect health care providers to attempt to respond to any loss of revenue from lower prices or, when consumers are steered to other lower price providers, their lost market share. For example, providers could shift their service provision towards higher- versus lower-margin services. Because health insurers pay different levels of prices for care, with commercial insurance reimbursement the highest and Medicaid the lowest, providers might also seek to increase the proportion of commercially insured patients and/or decrease access for less well reimbursed Medicaid patients (Alexander and Schnell 2024). More concerning are provider efforts to increase market power or to use their market power in anticompetitive ways.

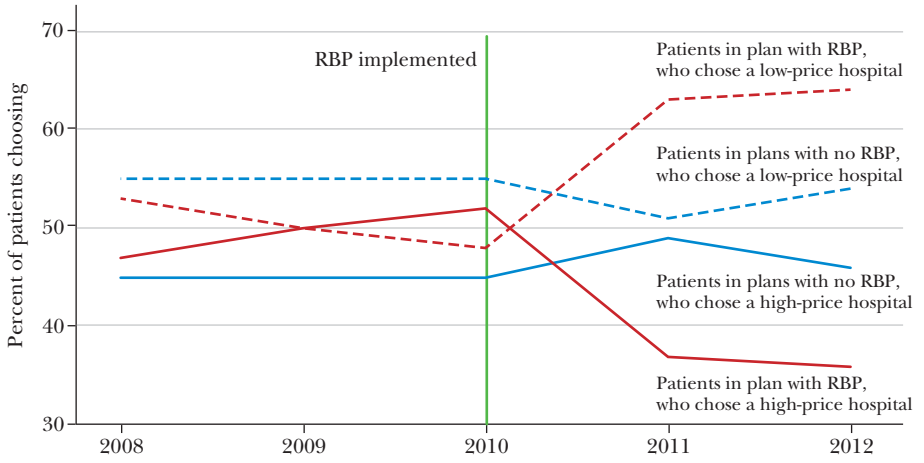
Provider Market Power and Price Increases

Provider market power stems in large part from consolidation, including mergers, acquisitions, joint ventures, affiliations, or any other material change in ownership. Consolidations can be along the horizontal chain of production: say, horizontal mergers between hospitals, or between two medical groups. For

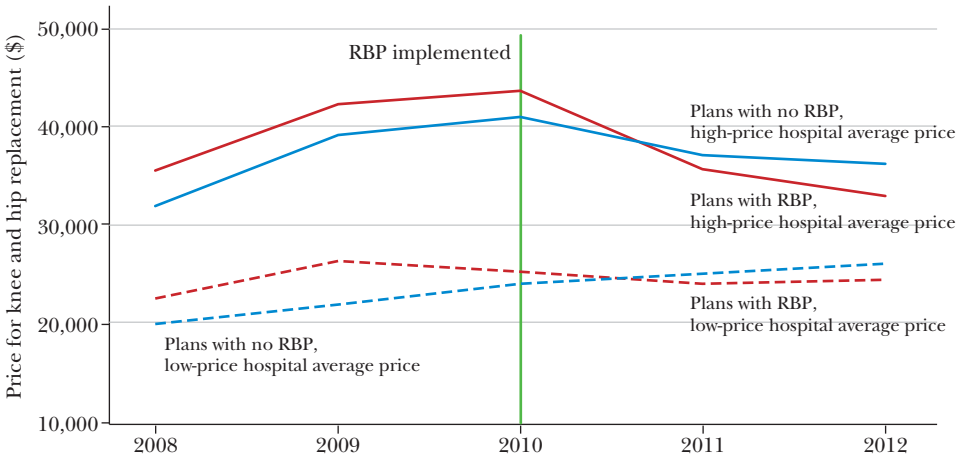
Figure 3

Patient Choices and Health Care Prices with Reference-Based Pricing Benefit Design

Panel A. Patients choosing high-price or low-price California hospitals for knee or hip replacement surgery, 2008–2012



Panel B. Prices paid for knee and hip replacement surgery in high-price and low-price California hospitals, 2008–2012



Source: Robinson and Brown (2013).

Note: RBP = Reference-based pricing; California hospitals providing hip or knee replacement surgeries were divided into two groups based on price paid for this procedure at the start of the reference-based pricing program. Low-price hospitals were those that offered hip and knee replacement at a price below $30,000 (which was the reference-price). The high-price hospitals had prices for these procedures above $30,000.

example, US hospital markets are highly concentrated and have become increasingly concentrated over time. One analysis of changes in hospital concentration from the 1990s to 2006 finds a 900 point increase in the population-weighted

average hospital Herfindahl-Hirschman Index (HHI) at the metropolitan statistical area level to 3261 (Gaynor, Ho, and Town 2015), which is well above the Federal Trade Commission's definition of highly concentrated markets (HHI > 1800) (US Department of Justice 2024).[8] More recently, analyses of American Hospital Association data from 2013, 2017, and 2021 estimated a similar weighted average hospital HHI by metropolitan statistical area, reporting that 97 percent of hospital markets were highly concentrated in 2013 and by 2021, the average hospital HHI was 4062 (Guardado 2024). Due to changes in the number and delineations between metropolitan statistical areas over these decades, these HHI estimates should be compared with caution, but qualitatively, hospital concentration has clearly continued to increase. Vertical integration in health care has also increased, although this trend is more recent; 47 percent of physicians were employed by or affiliated with hospital systems in 2024, up from less than 30 percent in 2012 (US Government Accountability Office 2025).

A first-order consequence of increasing health care provider market power is that powerful provider systems can use their market power to negotiate higher prices from insurers for all providers in the system. When insurers need to include these large, dominant systems in their networks to meet consumer demand, they will agree to pay higher prices. The weaker the insurer's market power, the more the insurer will concede. These anticompetitive price effects also arise for provider consolidations across geographic markets. This is because large health provider systems can jointly contract with insurers for all hospitals at once. Vistnes (2000) writes that "there can be a type of inter-hospital linkage that makes a health plan's bargaining position relative to one hospital depend on whether the plan can contract with a second hospital," regardless of whether that hospital is in the same geographic market or not. One of the earliest examples occurred in California in 1999–2003, where joint contracting was observed for health care systems that include hospitals and physicians across multiple geographic area, in which systems with the strongest bargaining positions in some markets extended that leverage to negotiate high rates elsewhere (Melnick and Keeler 2007). Hospital horizontal consolidation has resulted in higher prices, and these effects are larger in concentrated markets; cross-market horizontal mergers have increased health care prices, though to a lesser degree (Dafny, Ho, and Lee 2019). Vertical integration in health care has also resulted in higher health care prices; physician consolidation with hospital systems leads to higher prices for physician services (US Government Accountability Office 2025).

Anticompetitive Contract Provisions

The same market power that providers use to bargain for higher prices can be used to insist on the inclusion of anticompetitive clauses in contracts. These clauses can allow providers to protect and/or extend the high prices achieved through

[8]The Herfindahl-Hirschman Index is calculated as the sum of the market shares of the firms in a given market: thus, a monopoly provider with 100 percent of a market would have an HHI of 10,000.

market power and dampen insurer efforts to foster competition. I will focus on four types of clauses: all or nothing, anti-tiering or anti-steering, gag, and most favored nation.

An *"all-or-nothing" clause* with a health system requires an insurer to contract with all facilities in the health system if they want to include any facilities in the plan. Related are tying arrangements, which condition the sale of one product (in this case, access to part of a health system network) to the insurer's agreement to purchase a second product (access to the rest of the network).[9] These exclusive contracting clauses prevent insurers from selectively contracting with some but not all hospitals or medical groups in a health system. These provisions are often used by providers when a health system includes a "must-have" facility, as the insurer will agree to unfavorable terms in order to include the "must-have" facility in their network. Vistnes (2000) writes: "Even though a health plan may be able to continue marketing its plan to employers when they have one or two important 'holes' in their provider network, at some point a plan may have so many holes in its network that employers will be unwilling to offer that plan to their employees." All-or-nothing clauses protect providers from having to make price concessions or experiencing lost market share to competitors that insurers otherwise could achieve through selective contracting and managed care network design.

It is difficult to measure the empirical impact of all-or-nothing clauses because, due to confidentiality protections of insurer-provider contracts, we cannot observe which contracts employ these clauses. One of the best sources to understand the use of these contract provisions and their consequences is antitrust activity, because this activity can make formerly confidential contracts public. Analysis of confidential contracts made public through antitrust litigation against Sutter Health, a dominant health system with 24 state-licensed hospitals throughout Northern California, found that Sutter's use of all-or-nothing clauses led to health care prices approximately 30 percent above prices charged by other, non-Sutter hospitals in a control group. In addition, whereas prior to the use of these clauses Sutter hospital prices were higher in less competitive markets relative to more competitive markets, as one might expect, once these clauses were implemented the relationship between prices and a hospital market concentration went away. Melnick and Fonkych (2024) write: "These findings suggest that Sutter hospitals were no longer held in check by local market competition, and Sutter hospitals in more competitive markets could raise their prices to levels [even] above prices in less competitive markets."

Anti-tiering and anti-steering clauses rest at the intersection of health insurer network design and benefit design. An anti-steering clause restricts an insurer from encouraging consumers, whether through financial incentives or other means, to obtain health care services from other health systems that are competitors of the contracting provider system. An anti-tiering clause can limit an insurer's ability to

[9] A similar concept appears in regulation in the form of "Any Willing Provider" laws. Another version of this is called "exclusive dealing provisions," which require an insurer to only contract with one supplier of health care services and none of their competitors (Gudiksen et al. 2020).

implement a tiered provider network in their health plan, or alternatively, require that an insurer place all physicians, hospitals, and other facilities associated with a health system in the most favorable tier of providers or at the lowest cost-sharing rate. Both of these types of clauses prevent the insurer from giving consumers incentives, such as lower co-payments or deductible payments, that would steer them away from that health system.

As was the case with all-or-nothing clauses, anti-tiering and anti-steering clauses are difficult to study because, due to confidentiality provisions, most insurer-provider contracts are not transparent. Again, however, evidence is available from settlements of antitrust cases. For example, in 2016 the US Department of Justice and North Carolina's Attorney General's office filed suit against Atrium (formerly Carolinas Healthcare System), a dominant North Carolina hospital system, alleging their use of anti-tiering and anti-steering clauses violated the Sherman Antitrust Act. The court denied Atrium's motion to dismiss the case, citing the Department of Justice arguments that Atrium uses its market power as a "must have" health care system in the market to include steering restrictions in the contracts, and that the effects of steering restrictions include increased prices, reduced numbers of health plans, and reduced quality than would be possible if insurers could steer freely.[10] The case was then settled out of court, with the settlement prohibiting Atrium's use or enforcement of anti-steering or anti-tiering provisions in its contracts with its insurers (Gu 2020).

Gag clauses prevent insurers and providers from disclosing the negotiated fee schedule of their health care prices, along with other details about insurer-provider contracts. These clauses can be anticompetitive, because this information could be used by consumers to shop for care by employer purchasers and plan sponsors to develop initiatives to foster competition such as tiered or limited network designs, and by regulators to determine in which markets health care competition is most threatened. However, gag clauses seem to have only minor effects on direct consumer choices. Starting in 2010, several insurers, employers, and states began to give consumers access to health care price transparency tools—the best of which displayed a consumer's estimate of their personalized out-of-pocket costs at a specific provider. However, even among consumers with high-deductible insurance plans, only a small minority of patients use price transparency tools to obtain price information (Brot-Goldberg et al. 2017; Sinaiko and Rosenthal 2016; Desai et al. 2016). In most settings, the availability of out-of-pocket price estimates has generally not led patients to choose lower-priced providers.[11] The likelihood that data on health care prices could be used by employers and plan sponsors to increase

[10] *United States of America and the State of North Carolina v. The Charlotte-Mecklenburg Hospital Authority*, No 3:16-cv-00311, Document 92 (W.D.N.C 2018).

[11] An exception to this statement is advanced imaging services, where price transparency has led consumers to choose lower price imaging providers and some providers to lower their prices for imaging services (Desai et al. 2017; Sinaiko, Joynt, and Rosenthal 2016; Brown 2019).

competition is much higher. This information forms the basis of reference-based pricing benefit design, for example.

Most-favored-nation clauses take their name from similar provisions in international trade agreements: they are agreements between buyers and sellers that ensure that a buyer has the same terms (usually prices) as their "most favored" buyer. These clauses are not inherently anticompetitive; in settings where neither buyers nor sellers have market power, these clauses can prevent price discrimination. The risk of anticompetitive effects arises when the buyers and sellers, or in a health care market the insurers and providers, have market power. While a large literature has focused on the anticompetitive effects of most-favored-nation clauses in markets with a dominant insurer,[12] most relevant to this paper are the outcomes from most-favored-nation clauses in markets that have both a dominant insurer and a dominant provider. In these contexts, most-favored-nation clauses are not always used by insurers to lower their own costs. Instead, insurers can offer higher rates to a dominant provider in exchange for a most-favored-nation agreement, because the most-favored-nation clause ensures that the insurer's competitors must also pay the same or higher rates for access to the dominant provider (Salop and Scott Morton 2013). This practice by the insurer to inflict a high costs on itself in order to create high costs for other insurers maintains high provider prices and is anticompetitive.

Policy Responses

The lack of transparency in insurer-provider contracts makes it difficult for antitrust regulators to observe when potentially anticompetitive clauses are in place, and case-by-case adjudication is resource-intensive and inefficient. However, there are other options for policy that address use of anticompetitive provisions in contracts between health insurance companies and health care providers.

Ban or Restrict Anticompetitive Contract Clauses

States or the federal government could ban or restrict the use of anticompetitive contract provisions in insurer-provider contracts: for example, legislation could prohibit some or all of these clauses in any contracts where a health care provider, health care insurer, or health care administrator is one of the parties. Model state-level legislation developed by the National Academy for State Health

[12] Although the focus of this paper is on actions by health care providers, most-favored-nation clauses can also be barriers to entry in the insurer market. Because of advantages held by existing, dominant insurers, including brand and reputation, entrants usually need something to win market share (and overcome any switching costs for employers or consumers). A pro-competitive strategy would have new insurers negotiate lower prices with providers, so they can offer lower-premium health plans. However, when most-favored-nation clauses are in place, providers would be less likely to agree to lower prices with a potential entrant insurer, because doing so would mean having to lower prices for the incumbent insurer with the most-favored-nation clause. Thus, it becomes impossible for an entering insurer to gain a cost advantage over an existing insurer (Gudiksen et al. 2020).

Policy (NASHP) proposes that these bans (along with an exception process) be enforced by either the state insurance commissioner or attorney general (National Academy for State Health Policy 2021).

Several instances of such bans already exist. As of 2020, 20 states implemented bans of most-favored-nation clauses, especially in markets with a dominant insurer, which have slowed hospital price growth (Arnold et al. 2022). The US Congress prohibited gag clauses in insurer-provider as part of the in the Consolidated Appropriations Act of 2021; contracts can no longer include clauses that restrict insurers from sharing provider-specific price and quality data with consumers, plan sponsors, or referring providers. State-level bans of anti-tiering, anti-steering, and all-or-nothing clauses have been less common. Massachusetts banned the use of all-or-nothing, anti-tiering, and anti-steering clauses in insurer-provider contracts as of 2012. In 2021 the state of Nevada (S.B. 329), and in 2023 both Connecticut (H.B. 6669) and Texas (H.R 711), passed legislation banning the use of anti-tiering, anti-steering, and all-or-nothing clauses.

Policies That Complement Bans of Anticompetitive Contract Clauses

Simply banning these clauses may not generate competition, but additional policies can complement these bans. In Massachusetts, legislators paired the 2012 ban of anti-tiering and anti-steering clauses with a requirement that all insurers with at least 5,000 enrollees offer one tiered- or limited-network plan in the commercial market. In the following five years, penetration of tiered-network plans in the employer-sponsored insurance market in Massachusetts ranged from 19.2 percent to 28.3 percent, reaching its highest level in 2016. Overall penetration was driven by take-up in the large employer market, while penetration in the small and mid-sized employer market ranged from 10 to 15 percent (Sinaiko et al. 2024).

Of course, health plan penetration in a market also depends on employer decisions to offer and consumer decisions to enroll in plans, which may vary with market conditions. The Massachusetts experience suggests that provider consolidation may limit the ability to implement tiered-network health plans. Penetration of these plans in the employer-sponsored market is negatively associated with physician market concentration and with the percentage of physicians in a vertical relationship with a large health system in a geographic area (Sinaiko et al. 2024). The presence of large medical groups and additional health systems within local markets may play an important role in the feasibility of tiered-network health plans in the market.

Required disclosure of provider-negotiated prices is another example of a complementary policy, this time complementing the federal prohibition on gag clauses. Two federal rules, known as the "Hospital Price Transparency Rule" and the "Transparency in Coverage" rule, require hospitals starting in 2021 and insurers in 2022 to release complete lists of fee schedules and negotiated prices for all medical services. Both rules also include components that release price information in formats more suited for consumer comparison shopping. If employer purchasers or insurers use this information to modify health insurance benefit design, select higher-value

insurance carriers, or negotiate lower prices with health systems, these rules can have pro-competitive effects. However, in markets with a few large health systems with substantial market power, market-wide price transparency could facilitate tacit coordination or other anticompetitive behavior that leads prices to increase. Disclosure of prices may also discourage price concessions that providers were willing to make when prices were confidential—and there was less risk that other insurers or purchasers could learn about them and demand similar rates.

Since these rules went into effect, using these price data has proven difficult. The datasets are challengingly large: the full set of files released by insurers each month in 2023 include over one trillion price observations and are over one petabyte of data (Whaley et al. 2025). In addition, the insurer-released files include unlikely rates or multiple rates for the same service at the same provider, use different rate structures, and do not include unique provider identifiers (Claxton, Cotter, and Rakshit 2025). Hospitals face only low monetary penalties for noncompliance, and a 2024 Office of the Inspector General audit estimated that 46 percent of hospitals that were required to comply with the Hospital Price Transparency Rule were noncompliant (Office of Inspector General, US Department of Health and Human Services 2024). Nonetheless, studies using these data thus far have demonstrated wide variation in commercial prices within and across insurers and geographic areas (Whaley et al. 2025; Chartock, Simon, and Whaley 2023).

"Surprise billing laws" are another strategy to complement selective contracting and foster competition. Surprise billing exists because hospitals and physicians often negotiate contracts with insurers independently, and as a result, the network status of physicians working in facilities hospitals can vary. For example, a patient might choose an emergency room in their plan's provider network and later find out that one of the clinicians who was working in the emergency room and who treated them was out-of-network. Consumers also have little ability to choose the anesthesiologists, pathologists, or radiologists who participate in episodes of care initiated with other physicians, but who may be out-of-network. These scenarios can result in large and unexpected out-of-pocket costs for the consumer, because care received from an out-of-network provider is either not covered, or is less comprehensively covered, than in-network care. Surprise billing laws eliminate these scenarios. These laws stipulate that patients who saw an out-of-network physician within an in-network facility would pay cost-sharing based on the insurers negotiated rates with in-network physicians, and physicians were prohibited from charging patients anything additional (known as "balance billing").

Relevant to health care competition, these laws have the joint effect of reducing provider incentives to stay out-of-network and reducing provider leverage to negotiate higher rates in-network. In the context of an emergency department setting, Cooper, Scott Morton, and Shekita (2020) articulate the theory that in the absence of surprise billing laws, emergency department physicians are able to negotiate higher out-of-network rates with insurers, while working at in-network facilities, without risking much loss in patient volume. With surprise billing laws in place, this is no longer the case, and an empirical test following the state of New York's elimination

of emergency department surprise bills showed that out-of-network billing fell by 88 percent and was accompanied by a 15 percent reduction in in-network physician. Garmon et al. (2024) find substantial heterogeneity in these price effects across physician specialties and across states.

Antitrust Reform

Antitrust enforcement remains an important component of efforts to protect competition in health care markets, and reform to antitrust review is an additional policy response. The legal standard for anticompetitive conduct is a substantial lessening of competition that results in consumer harm (Abbott 2021), which is narrower than the economic definition which considers any exercise of market power anticompetitive. By extension, the traditional scope for anticompetitive harms could be expanded to include consumer non-price effects like limited choice or reductions in quality, as well as harms to others (perhaps through labor market effects) (US Federal Trade Commission 2022; US Department of Justice and Federal Trade Commission 2023). Broadening the legal standard to include these harms may increase the effectiveness of antitrust in countering anticompetitive contract provisions used by providers with market power.

Conclusion

In the US health care market, where third-party insurance plays such a dominant role, competition is not primarily driven by direct interactions between the buyers and sellers of services (that is, between patients and individual health care providers). Instead, competition in the health care market is primarily driven by the interactions between large market players: on the demand side, the insurers who are often acting on behalf of employers and patients, and on the supply side, the provider organizations who are often acting on behalf of doctors, nurses, and other clinicians, as well as hospitals and clinics. In this setting, competition does not follow the textbook model in which demand and supply for a given good or service is intermediated by price, but instead by occurs through complex contracts that describe interlinked prices and networks over a very wide array of goods and services. If competitive forces are to exert pressure for lower costs and greater quality in health care markets, scrutinizing the details of these contracts is of central importance.

Several practices that are common in insurer-provider contracting have the potential to drive competition in health care. Insurers can continue to innovate in their use of networks and benefit design to strengthen consumer incentives to select higher-value providers and to create downward pressure on health care prices. However, in settings where providers have market power, providers may use their dominant position to negotiate contract terms that inhibit competition. Thus, efforts to monitor anticompetitive behaviors in contracting, both those that are known and those that emerge over time, are important. Assessing these practices

and enacting policy and market responses to limit the effects of anticompetitive contract provisions will thus be critical to support health care competition.

■ *I am grateful to Tom McGuire, Elena Prager, and Meredith Rosenthal who provided helpful comments on an earlier draft, and to Hailey Clark for excellent research assistance.*

References

Abbott, Alden. 2021. *US Antitrust Laws: A Primer.* Mercatus Center, George Mason University Policy Brief, March. https://www.mercatus.org/research/policy-briefs/us-antitrust-laws-primer.

Ackley, Calvin A. 2022. "Tiered Cost Sharing and Health Care Demand." *Journal of Health Economics* 85: 102663.

Ackley, Calvin A. 2025. "The Impact of Preferred Provider Incentives on Demand and Negotiated Prices." *American Journal of Health Economics* 11 (4): 668–96.

Alexander, Diane, and Molly Schnell. 2024. "The Impacts of Physician Payments on Patient Access, Use, and Health." *American Economic Journal: Applied Economics* 16 (3): 142–77.

Anderson, Gerard F., Peter Hussey, and Varduhi Petrosyan. 2019. "It's Still the Prices, Stupid: Why the US Spends So Much on Health Care, and a Tribute to Uwe Reinhardt." *Health Affairs* 38 (1): 87–95.

Arnold, Daniel R., Katherine L. Gudiksen, Jaime S. King, Brent D. Fulton, and Richard M. Scheffler. 2022. "Do State Bans of Most-Favored-Nation Contract Clauses Restrain Price Growth? Evidence from Hospital Prices." *Milbank Quarterly* 100 (2): 589–615.

Arrow, Kenneth J. 1963. "Uncertainty and the Welfare Economics of Medical Care." *American Economic Review* 53 (5): 941–73.

Brot, Zarek, Zack Cooper, Stuart V. Craig, and Lev Klarnet. 2024. "Is There Too Little Antitrust Enforcement in the US Hospital Sector?" *American Economic Review: Insights* 6 (4): 526–42.

Brot-Goldberg, Zarek C., Amitabh Chandra, Benjamin R. Handel, and Jonathan T. Kolstad. 2017. "What Does a Deductible Do? The Impact of Cost-Sharing on Health Care Prices, Quantities, and Spending Dynamics." *Quarterly Journal of Economics* 132 (3): 1261–318.

Brown, Zach Y. 2019. "Equilibrium Effects of Health Care Price Information." *Review of Economics and Statistics* 101 (4): 699–712.

Bunch, Lisa N., and Halelujha Ketema. 2025. *Table 1. Number of People by Health Insurance Coverage Status and Type: 2023 to 2024. Nos. P60-288.* Current Population Reports. US Census Bureau. https://www2.census.gov/programs-surveys/demo/tables/p60/288/table1.xlsx (accessed October 9, 2025).

Bundorf, M. Kate. 2002. "Employee Demand for Health Insurance and Employer Health Plan Choices." *Journal of Health Economics* 21 (1): 65–88.

Chartock, Benjamin L., Kosali Simon, and Christopher M. Whaley. 2023. "Transparency in Coverage Data and Variation in Prices for Common Health Care Services." *JAMA Health Forum* 4 (10): e233663.

Christianson, Jon B., Stephen T. Parente, and Roger Feldman. 2004. "Consumer Experiences in a Consumer-Driven Health Plan." *Health Services Research* 39 (4p2): 1123–40.

Claxton, Gary, Lynne Cotter, and Shameek Rakshit. 2025. "Challenges with Effective Price Transparency Analyses." Peterson-KFF Health System Tracker, February 25. https://www.healthsystemtracker.org/brief/challenges-with-effective-price-transparency-analyses/.

Congressional Budget Office. 2022. *The Prices That Commercial Health Insurers and Medicare Pay for Hospitals' and Physicians' Services.* Congressional Budget Office.

Cooper, Zack, Fiona Scott Morton, and Nathan Shekita. 2020. "Surprise! Out-of-Network Billing for

Emergency Care in the United States." *Journal of Political Economy* 128 (9): 3626–77.

Craig, Stuart V., Keith Marzilli Ericson, and Amanda Starc. 2021. "How Important Is Price Variation between Health Insurers?" *Journal of Health Economics* 77: 102423.

Cutler, David M., Mark McClellan, and Joseph P. Newhouse. 2000. "How Does Managed Care Do It?" *RAND Journal of Economics* 31 (3): 526–48.

Cutler, David M., and Richard J. Zeckhauser. 2000. "The Anatomy of Health Insurance." In *Handbook of Health Economics*, Vol. 1A, edited by Anthony J. Culyer and Joseph P. Newhouse, 563–643. Elsevier.

Dafny, Leemore, Kate Ho, and Robin S. Lee. 2019. "The Price Effects of Cross-market Mergers: Theory and Evidence from the Hospital Industry." *RAND Journal of Economics* 50 (2): 286–325.

Desai, Sunita, Laura A. Hatfield, Andrew L. Hicks, Michael E. Chernew, and Ateev Mehrotra. 2016. "Association between Availability of a Price Transparency Tool and Outpatient Spending." *JAMA* 315 (17): 1874–81.

Desai, Sunita, Laura A. Hatfield, Andrew L. Hicks, Anna D. Sinaiko, Michael E. Chernew, David Cowling, Santosh Gautam, Sze-jung Wu, and Ateev Mehrotra. 2017. "Offering a Price Transparency Tool Did Not Reduce Overall Spending among California Public Employees and Retirees." *Health Affairs* 36 (8): 1401–07.

Dranove, David, Mark Shanley, and William D. White. 1993. "Price and Concentration in Hospital Markets: The Switch from Patient-Driven to Payer-Driven Competition." *Journal of Law and Economics* 36 (1): 179–204.

Einav, Liran, and Amy Finkelstein. 2011. "Selection in Insurance Markets: Theory and Empirics in Pictures." *Journal of Economic Perspectives* 25 (1): 115–38.

Garmon, Christopher, Yiting Li, Sheldon M. Retchin, and Wendy Yi Xu. 2024. "The Impact of Surprise Billing Laws on Hospital-Based Physician Prices and Network Participation." *Health Economics, Policy and Law*. https://doi.org/10.1017/S1744133124000239.

Gaynor, Martin, Kate Ho, and Robert J. Town. 2015. "The Industrial Organization of Health-Care Markets." *Journal of Economic Literature* 53 (2): 235–84.

Gruber, Jonathan, and Robin McKnight. 2016. "Controlling Health Care Costs through Limited Network Insurance Plans: Evidence from Massachusetts State Employees." *American Economic Journal: Economic Policy* 8 (2): 219–50.

Gu, Amy Y. 2020. "[Case Brief] Atrium Health Settlement Encourages Enforcement of Anti-Tiering/Anti-Steering Clauses in Healthcare Contracts." Source on HealthCare Price and Competition, November 16. https://sourceonhealthcare.org/case-brief-atrium-health-settlement-encourages-enforcement-of-anti-tiering-anti-steering-clauses-in-healthcare-contracts/.

Guardado, José R. 2024. "Policy Research Perspectives: Competition in Hospital Markets, 2013–2021." American Medical Association. https://www.ama-assn.org/system/files/prp-competition-in-hospital-markets.pdf.

Gudiksen, Katherine L., Alexandra D. Montague, Jaime S. King, Amy Y. Gu, Brent D. Fulton, and Thomas L. Greaney. 2020. *Preventing Anticompetitive Contracting Practices in Healthcare Markets*. Source on Healthcare Price and Competition, University of California Hastings College of Law.

Hussey, Peter S., Samuel Wertheimer, and Ateev Mehrotra. 2013. "The Association between Health Care Quality and Cost: A Systematic Review." *Annals of Internal Medicine* 158 (1): 27–34.

Jaromin, Sarah. 2024. *The Evolving Landscape of State Health Care Transaction Laws*. National Conference of State Legislatures.

Levitt, Larry. 2025. "A Backlash against Health Insurers, Redux." *JAMA Health Forum* 6 (3): e250954.

Ma, Ching-To Albert, and Thomas G. McGuire. 2002. "Network Incentives in Managed Health Care." *Journal of Economics and Management Strategy* 11 (1): 1–35.

Mazurenko, Olena, Heather L. Taylor, and Nir Menachemi. 2022. "The Impact of Narrow and Tiered Networks on Costs, Access, Quality, and Patient Steering: A Systematic Review." *Medical Care Research and Review* 79 (5): 607–17.

McDonald, Tim, Katie M. White, Tsan-Yao Huang, Christopher M. Whaley, and Bryan Dowd. 2021. "Clinic Price Reductions in a Tiered Total Cost Benefit Design." *American Journal of Managed Care* 27 (9): e316–21.

McGuire, Thomas G. 2000. "Physician Agency." In *Handbook of Health Economics*, Vol. 1A, edited by Anthony J. Culyer and Joseph P. Newhouse, 461–536. Elsevier.

Melnick, Glenn, and Katya Fonkych. 2024. "Are 'All or Nothing' Contracts by Hospital Systems Anti-ccompetitive?—Evidence from a Recent Antitrust Lawsuit." *Journal of Hospital Management and Health Policy* 8: 21.

Melnick, Glenn, and Emmett Keeler. 2007. "The Effects of Multi-hospital Systems on Hospital Prices." *Journal of Health Economics* 26 (2): 400–413.

Motheral, Brenda, and Kathleen A. Fairman. 2001. "Effect of a Three-Tier Prescription Copay on Pharmaceutical and Other Medical Utilization." *Medical Care* 39 (12): 1293–304.

National Academy for State Health Policy. 2021. "NASHP Model Act to Address Anticompetitive Terms in Health Insurance Contracts." National Academy for State Health Policy (blog), April 8. https://nashp.org/nashp-model-act-to-address-anticompetitive-terms-in-health-insurance-contracts/.

Office of Inspector General, US Department of Health and Human Services. 2024. *Not All Selected Hospitals Complied with the Hospital Price Transparency Rule.* Audit A-07-22-06108. Office of Inspector General, US Department of Health and Human Services.

Parente, Stephen T., Roger Feldman, and Jon B. Christianson. 2004. "Evaluation of the Effect of a Consumer-Driven Health Plan on Medical Care Expenditures and Utilization." *Health Services Research* 39 (4p2): 1189–210.

Prager, Elena. 2020. "Healthcare Demand under Simple Prices: Evidence from Tiered Hospital Networks." *American Economic Journal: Applied Economics* 12 (4): 196–223.

Prager, Elena, Vilsa E. Curto, Alexa Magyari, Marema Gaye, and Anna D. Sinaiko. 2023. "Tiered Physician Network Plans and Patient Choices of Specialist Physicians." *JAMA Network Open* 6 (11): e2341836.

Robinson, James C., and Timothy T. Brown. 2013. "Increases in Consumer Cost Sharing Redirect Patient Volumes and Reduce Hospital Prices for Orthopedic Surgery." *Health Affairs* 32 (8): 1392–97.

Robinson, James C., Timothy Brown, and Christopher Whaley. 2015. "Reference-Based Benefit Design Changes Consumers' Choices and Employers' Payments for Ambulatory Surgery." *Health Affairs* 34 (3): 415–22.

Robinson, James C., Timothy T. Brown, and Christopher Whaley. 2017. "Reference Pricing Changes the 'Choice Architecture' of Health Care for Consumers." *Health Affairs* 36 (3): 524–30.

Robinson, James C., Christopher Whaley, and Timothy T. Brown. 2016. "Reference Pricing, Consumer Cost-Sharing, and Insurer Spending for Advanced Imaging Tests." *Medical Care* 54 (12): 1050–55.

Salop, Steven C., and Fiona Scott Morton. 2013. "Developing an Administrable MFN Enforcement Policy." *Antitrust* 27 (2): 15–19.

Scheffler, Richard M., and Daniel R. Arnold. 2017. "Insurer Market Power Lowers Prices In Numerous Concentrated Provider Markets." *Health Affairs* 36 (9): 1539–46.

Sinaiko, Anna D. 2016. "Variations in Patient Response to Tiered Physician Networks." *American Journal of Managed Care* 22 (6): 420–25.

Sinaiko, Anna D., Vilsa E. Curto, Elizabeth Bambury, Mark J. Soto, and Meredith B. Rosenthal. 2024. "Variation in Tiered Network Health Plan Penetration and Local Provider Market Characteristics." *Health Services Research* 59 (4): e14223.

Sinaiko, Anna D., Karen E. Joynt, and Meredith B. Rosenthal. 2016. "Association between Viewing Health Care Price Information and Choice of Health Care Facility." *JAMA Internal Medicine* 176 (12): 1868–70.

Sinaiko, Anna D., Pragya Kakani, and Meredith B. Rosenthal. 2019. "Marketwide Price Transparency Suggests Significant Opportunities for Value-Based Purchasing." *Health Affairs* 38 (9): 1514–22.

Sinaiko, Anna D., Mary Beth Landrum, and Michael E. Chernew. 2017. "Enrollment in a Health Plan with a Tiered Provider Network Decreased Medical Spending by 5 Percent." *Health Affairs* 36 (5): 870–75.

Sinaiko, Anna D., and Meredith B. Rosenthal. 2014. "The Impact of Tiered Physician Networks on Patient Choices." *Health Services Research* 49 (4): 1348–63.

Sinaiko, Anna D., and Meredith B. Rosenthal. 2016. "Examining a Health Care Price Transparency Tool: Who Uses It, and How They Shop for Care." *Health Affairs* 35 (4): 662–70.

Tilipman, Nicholas. 2022. "Employer Incentives and Distortions in Health Insurance Design: Implications for Welfare and Costs." *American Economic Review* 112 (3): 998–1037.

Trish, Erin E., and Bradley J. Herring. 2015. "How Do Health Insurer Market Concentration and Bargaining Power with Hospitals Affect Health Insurance Premiums?" *Journal of Health Economics* 42: 104–14.

US Department of Justice. 2024. "Antitrust Division—Herfindahl-Hirschman Index." January 17. https://www.justice.gov/atr/herfindahl-hirschman-index.

US Department of Justice and Federal Trade Commission. 2023. *Merger Guidelines.* US Department of Justice; Federal Trade Commission.

US Federal Trade Commission. 2022. *Policy Statement Regarding the Scope of Unfair Methods of Competition*

Under Section 5 of the Federal Trade Commission Act. Commission File P2221202. Federal Trade Commission.

US Government Accountability Office. 2025. *Health Care Consolidation: Published Estimates of the Extent and Effects of Physician Consolidation*. GAO-25-107450.

Vistnes, Gregory. 2000. "Hospitals, Mergers, and Two-Stage Competition." *Antitrust Law Journal* 67 (3): 671–92.

Wallace, Jacob. 2023. "What Does a Provider Network Do? Evidence from Random Assignment in Medicaid Managed Care." *American Economic Journal: Economic Policy* 15 (1): 473–509.

Whaley, Christopher, Nandita Radhakrishnan, Michael Richards, Kosali Simon, and Benjamin Chartock. 2025. "Understanding Health Care Price Variation: Evidence from Transparency-in-Coverage Data." *Health Affairs Scholar* 3 (2): qxaf011.

Zeckhauser, Richard. 1970. "Medical Insurance: A Case Study of the Tradeoff between Risk Spreading and Appropriate Incentives." *Journal of Economic Theory* 2 (1): 10–26.

Zhang, Xinke, Amelia Haviland, Ateev Mehrotra, Peter Huckfeldt, Zachary Wagner, and Neeraj Sood. 2018. "Does Enrollment in High-Deductible Health Plans Encourage Price Shopping?" *Health Services Research* 53 (S1): 2718–34.

Journal of Economic Perspectives—Volume 40, Number 2—Spring 2026—Pages 93–116

The Current Era of Health Care Consolidation

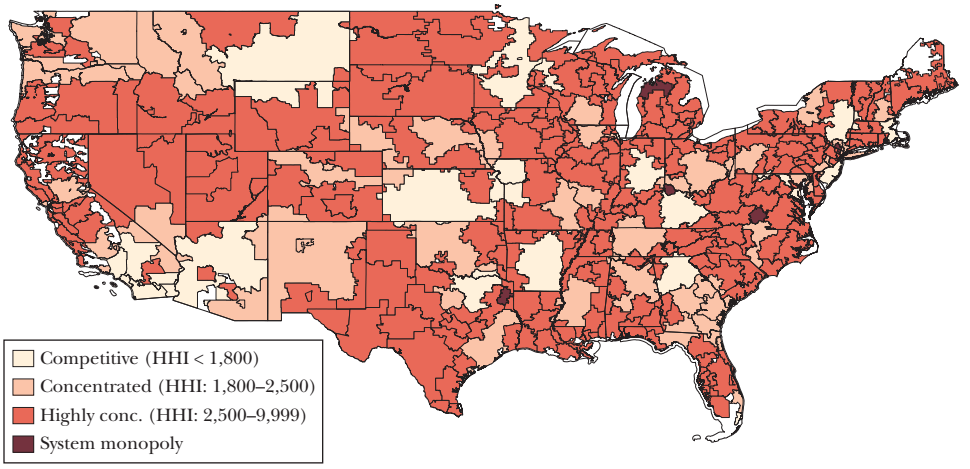
Michael R. Richards and Christopher M. Whaley

Consolidation in the last few decades has reshaped the organization and structure of US health care markets, among both providers and insurers. Nearly all US hospital and insurer markets exceed established regulatory thresholds for competitive markets, and over half of physicians are now employed by a hospital or health system, which can increase spending for patients, payers, and taxpayers. As one example, Figure 1 shows that market concentration is the norm in the hospital industry. Increased supply-side concentration can alter the balance of negotiations between providers and insurers, and unsurprisingly, prices for patients with commercial insurance are approximately 2.5 times the prices paid by those with public insurance. High and variable prices demonstrate a minimal link with higher quality, and the United States leads peer nations in health care spending. Over 40 percent of Americans *with insurance* report difficulty managing health care expenses, and a nearly equal number carry some level of medical debt (Sparks et al. 2026). Health care consolidation trends therefore sit uncomfortably in the midst of these consumer realities and ongoing national debates over an "affordability crisis" (Saad and Brenan 2025).

■ *Michael R. Richards is Professor and Director of the Sloan Program in Health Administration, both at the Cornell Brooks School of Public Policy, Ithaca, New York. Christopher M. Whaley is Associate Professor of Health Services, Policy and Practice, and Associate Director of the Center for Advancing Health Policy through Research, both at the Brown University School of Public Health, Providence, Rhode Island. Their email addresses are michael.richards@cornell.edu and christopher_whaley@brown.edu.* *The authors have received financial support from interested parties.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20251472.

Figure 1
Market Concentration of the US Hospital Industry



Source: Analysis of 2024 American Hospital Association data.
Note: Markets are defined using Hospital Referral Regions (HRRs), and Herfindahl–Hirschman Indexes (HHIs) are constructed using each hospital's number of beds. HHIs are categorized according to market concentration categories used by the Antitrust Division of the US Department of Justice and the Federal Trade Commission: that is, Competitive, HHI $<$ 1,800; Concentrated, HHI between 1,800 and 2,500; Highly Concentrated, HHI 2,500 to $<$ 10,000; and Monopoly, HHI of 10,000).

In this paper, we begin by discussing three waves of health care industry consolidation since 2000: (1) horizontal integration, in which hospitals form health systems or health insurers merge to form larger, national brands; (2) a higher level of mergers and acquisitions activity among a new slate of typically smaller firms like physician practices, dialysis clinics, ambulatory surgery centers, and diagnostic testing centers, as well as new organizational combinations like vertical integration arrangements across health care industries; and (3) private equity and insurers as new financing sources driving health care market consolidation.

As we will discuss, these newer forms of health industry consolidation were often justified as attempts to take advantage of economies of scale, to create negotiating leverage to hold down price increases, or to provide consumers with easier access to an array of care. Medical service prices for commercial payers, in particular, are the result of negotiations between large insurers and large providers, with both sides searching for profitable opportunities to leverage their market power—perhaps to the detriment of the average consumer or taxpayer. The new entities have also proven capable of maximizing the financial benefits from existing reimbursement mechanisms and policies, including in ways that create arbitrage opportunities within public and private payer market segments. They can additionally reshape competition among local provider markets to generate higher industry profits while sacrificing efficiency and access for the overarching health care system.

A source, as well as a manifestation, of these various incentives can be seen in the wide variation in reimbursement rates for medical services within the US health care system. Substantial health care price variation exists between sources of insurance coverage—for example, commercial, Medicare, and Medicaid—and both within and across geographies. For a given payer, payment rates can also differ between settings where the care is provided for the exact same procedure, such as a physician's office versus a hospital outpatient department. Looking across payers, commercial insurers' negotiated rates for hospital care average approximately 250 percent of what Medicare pays (Whaley et al. 2024), and commercial prices commonly vary by nearly an order of magnitude among frequent and commoditized services, such as routine magnetic resonance imaging (MRI) exams (Cooper et al. 2019). These high provider prices are not tied to higher care quality or greater access for patients (Cooper et al. 2022; Beaulieu et al. 2020; Crespin and Whaley 2023). These patterns were not always the case. Commercial insurance and Medicare payment rates were once similar, and as recently as 1996, commercial insurers paid hospital rates that averaged just 106 percent of Medicare's corresponding reimbursement levels (Selden et al. 2015).

Wide variation in prices encourages the restructuring of health care markets to take advantage of higher payments, while consumers may have only limited or no awareness of these price differentials due to health insurance coverage. For example, employers who choose health insurers on behalf of their employees often lack the ability or incentives to navigate a complex market and to balance the competing priorities of broad coverage and cost containment. Employers also pass the costs of health expenses to workers in the form of wage reductions, which weakens their own incentives to engage in costly and time-consuming work to reduce health spending (Baicker and Chandra 2006; Arnold and Whaley 2020; Brot-Goldberg et al. 2024). Patients and providers can therefore be less responsive to differences in prices—allowing striking reimbursement gaps to persist when market forces would otherwise narrow or close such differences (Rochet and Tirole 2006; Bardey and Rochet 2010). Indeed, when patients face full marginal prices, their provider choices dramatically change and favor lower-priced providers (Whaley, Guo, and Brown 2017; Prager 2020), and providers likewise charge lower prices (Whaley and Brown 2018; Wang et al. 2023). Many insurers and employers have consequently designed and implemented "tiered" networks that lower patients' cost-sharing amounts when they choose care from lower-priced providers or outright restrict access to higher-priced and lower-quality providers (Gruber and McKnight 2016; Sinaiko, Landrum, and Chernew 2017; Dowd, Huang, and McDonald 2021; Whaley et al. 2021b). These targeted benefit design innovations have proven to be much more effective in encouraging the use of efficient providers than broader and nontargeted policies, such as high-deductible health plans (Brot-Goldberg et al. 2017).

These existing market dynamics and the overall state of US health care did not arrive suddenly, but instead reflect a long history of events and policy choices. Earlier forms of health sector consolidation precipitated subsequent and countervailing

consolidation strategies, with the latter bringing their own unique challenges to health care market functioning. Regulators and policymakers therefore have to grapple with the reality that their efforts to curb consolidation are often inherently reactive. Moreover, the existing regulatory playbook for tempering overly aggressive merger and acquisition activity may prove ill-equipped for emergent consolidation waves, especially as the resulting tie-ups and underlying financial flows grow in their complexity and creativity. With this consideration in mind, we conclude with an overview of a variety of recent policy levers, particularly at the state level, that seek to modify some of these emerging market dynamics to improve the efficiency of the US health care system. However, this menu of policies involves a mix of trade-offs related to the dual aims of ameliorating the adverse consequences of modern consolidation, while avoiding excessive market interference via the blunt instruments of regulation.

Forms and Prevalence of US Provider Consolidation

The recent weakening of US health care market competition has largely occurred through three distinct "waves" of consolidation. Each wave and related forms of consolidation impact certain aspects of the insurance and care delivery segments.

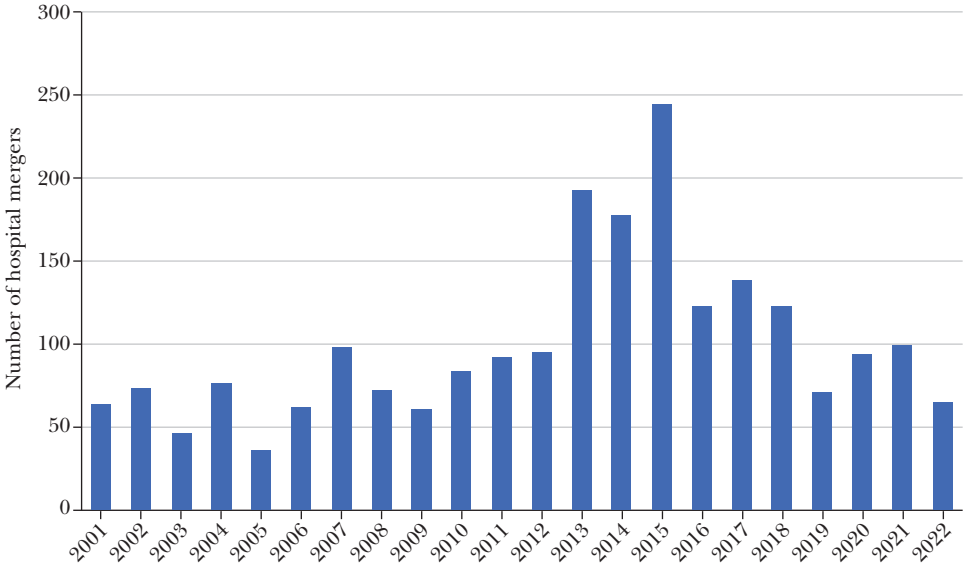
Hospital and Insurer Horizontal Consolidation

The first wave of modern consolidation in US health care involved a self-reinforcing combination of horizontal integration: hospitals combining to form health systems and insurers combining to form multi-state insurance conglomerates. In the 1990s, the adoption of managed care approaches was a response to the pressures insurers were facing from a combination of increased consumer sensitivity to insurance rates and employer demand for customized insurance networks (Robinson 1999). Insurers responded to these demands by attempting to place downward pressure on provider payment rates through tougher negotiations for more favorable rates and by shifting patient volume to preferred providers. The care delivery suppliers, namely hospitals, responded to these pressures by rapidly consolidating themselves to form larger entities with often wider geographic footprints and stronger negotiation leverage with the insurance companies.

Figure 2 presents the approximate number of US hospital mergers between 2001 and 2022. These many hospital combinations have led to large and sprawling systems, with the largest ten hospital systems accounting for 22 percent of all US hospital beds, as shown in Table 1. Moreover, as a result of these mergers, nearly all US hospital markets exceed the definitions used by the US Department of Justice, Federal Trade Commission, and other regulators to measure "highly concentrated" market competitiveness (Fulton 2017), as shown in Figure 1 above.

A large body of evidence links increased hospital mergers to higher prices, with post-merger price increases ranging from 6 percent to over 20 percent

Figure 2
Annual Number of US Hospital Mergers



Source: Analysis of American Hospital Association data.

Table 1
Ten Largest US Hospital Systems (by Beds)

Health system	Number of beds	Share of US beds
HCA Healthcare	37,670	5.7%
CommonSpirit Health	17,238	2.6%
Ascension Health	15,238	2.3%
Trinity Health MI	15,069	2.3%
Tenet Healthcare	12,587	1.9%
Advocate Health	10,573	1.6%
Kaiser Permanente	9,337	1.4%
Providence	9,184	1.4%
Community Health Systems	9,090	1.4%
AdventHealth	8,834	1.3%

Source: Analysis of American Hospital Association data.

(Gaynor, Ho, and Town 2015; Cooper et al. 2019; Liu et al. 2022; Andreyeva et al. 2024). However, care quality and patient access do not appear to improve following hospital consolidation, and in many cases, exhibit deterioration (Gaynor 2006; Beaulieu et al. 2020). More recent evidence also finds that following a merger, hospitals shift resources to higher-margin commercially insured patients

at the expense of Medicare and Medicaid patients, particularly with respect to lower-margin services such as labor and delivery care (Desai et al. 2023; Arnold, Radhakrishnan, and Whaley 2025; Dranove, Gaynor, and Geddes 2025; Oh and Whaley 2026).

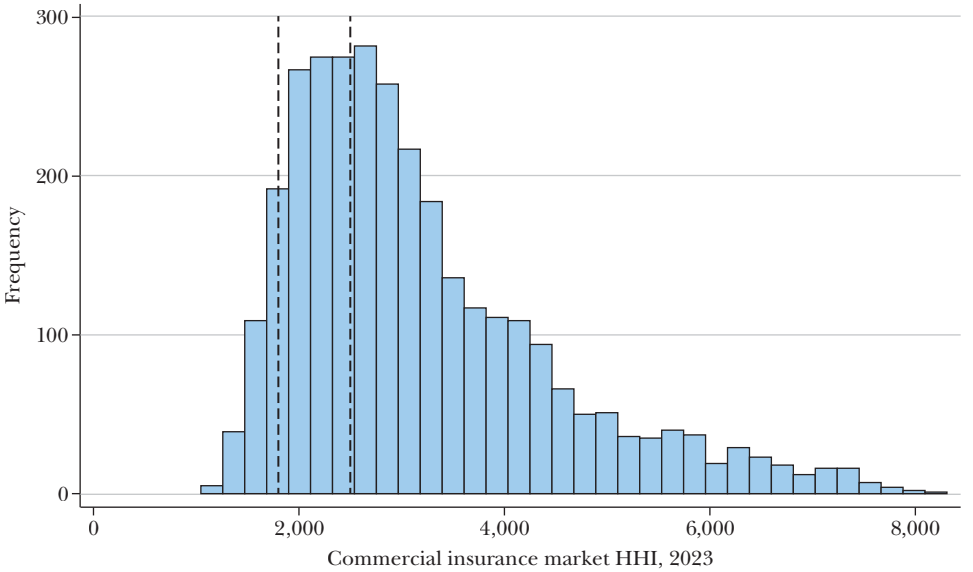
As hospitals grew into multi-state conglomerates, the incentives for insurers to combine for similar market reach increased as well, especially as many large insurers expanded their "third party administrator" lines of business. These third-party administrator contracts involve self-insured employers—that is, employers that form their own risk pools and ultimately bear the financial risks tied to their employees' medical spending—as well as traditional insurers that handle provider network formation, price negotiations, and provider payments on behalf of the employer but do not bear the actual financial risk related to the workers' medical spending. Successful third-party administrator offerings typically need to accommodate large firms with a regional presence, or even a national one. When shopping for a third-party administrator contract for their employer-sponsored health insurance plans, these companies would often require insurance coverage across multiple markets where their employees work.

But perhaps the most critical driver of health insurer consolidation over time has been directly connected to the evolutions in provider-specific market structures, as the insurers aimed to create countervailing negotiation leverage with hospitals and other providers when bargaining over service payment rates (Barrette, Gowrisankaran, and Town 2022). Larger market presence for insurers also expands their ability to construct provider networks with more financially favorable rates. Indeed, this type of market power was explicitly stated as the basis for recently proposed company mergers among some of the largest national health insurers, although this particular merger was blocked by antitrust authorities and the courts.[1]

Likewise, the US market for health insurance commonly exceeds highly concentrated thresholds. Again using the Herfindahl–Hirschman Index guidelines from the US Department of Justice and the Federal Trade Commission, only 7 percent of US counties have a competitive commercial health insurance market, while 26 percent have moderately concentrated ($1{,}500 < \text{HHI} < 2{,}500$) and 65 percent have highly concentrated markets ($\text{HHI} > 2{,}500$), as shown in Figure 3. The extreme consolidation of health insurance in the United States has given rise to the "BUCAH" terminology, which refers to the five largest insurers: Blue Cross Blue Shield, UnitedHealthcare, Cigna, Aetna, and Humana. Collectively, the BUCAHs account for approximately 80 percent of US commercial health insurance, and in the Medicare Advantage program (Medicare "Part C"), where the government finances a private health insurance policy for enrollees, UnitedHealth and Humana alone account for nearly half of all Medicare Advantage enrollment (Ochieng et al. 2025).

[1] *United States of America v. Anthem, Inc.*, Civil Action No. 16-1493 (2017).

Figure 3
Distribution of US Commercial Insurance Market Concentration



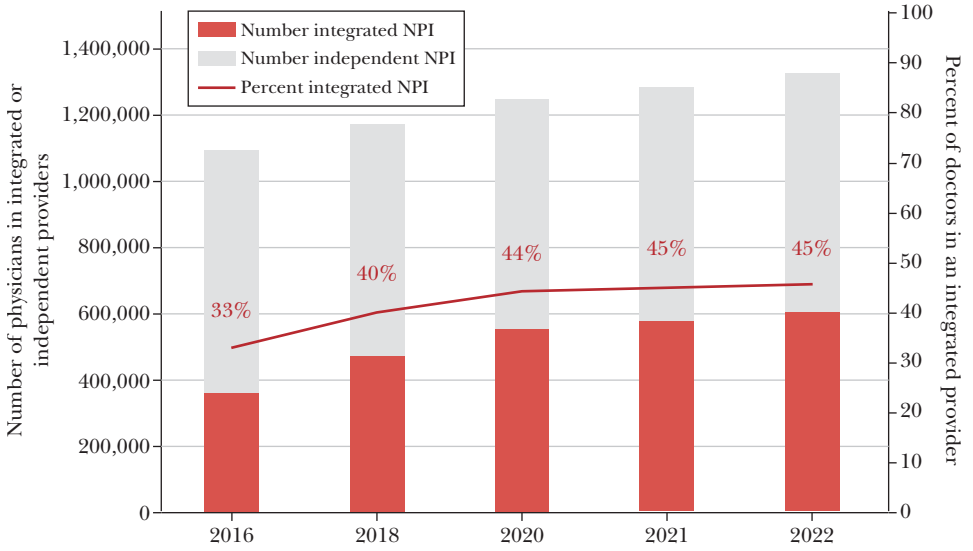
Source: Analysis of Clarivate Insurance enrollment data.
Note: For the horizontal axis, the market size here is defined as US counties. The total number of markets is 3,132 US counties. For references, the Herfindahl–Hirschman Index levels used as guidelines by the Antitrust Division of the US Department of Justice and the Federal Trade Commission are: Competitive, HHI $<$ 1,800; Concentrated, HHI between 1,800 and 2,500; Highly Concentrated, HHI 2,500 to $<$ 10,000; and Monopoly, HHI of 10,000.

Despite the movement toward concentrating purchasing power for medical services, the actual success of insurers at limiting hospital price rises, much less reducing prices, has limited empirical support to date (Barrette, Gowrisankaran, and Town 2022). Additionally, the same market power that can improve health insurers' negotiation leverage also reduces the competitive incentives to run efficient plans and pass savings onto consumers in the form of lower premiums (Dafny 2010). As a result, the long-running streak of hospital mergers has consistently placed upward pressure on commercial prices and only occasionally been restrained by regulatory interference.

Hospital/Health System–Physician Vertical Integration

Following the wave of horizontal integration among hospitals and health insurers, the second wave of modern health care market consolidation has been more localized and involved greater crossing of industry boundaries. In recent years, hospitals and health systems have aggressively acquired a growing share of physician practices. As highlighted in Figure 4, the share of US physicians employed by a hospital or health system increased by approximately 40 percent over the 2016

Figure 4
Changes in Share of US Physicians Vertically Integrated with a Hospital or Health System



Source: Analysis of Compendium of US Health Systems and Medicare Fee-for-Service claims data from the Agency for Health Care Research and Quality (AHRQ).
Note: This figure shows the number of US physicians that are independent. NPI stands for National Provider Identifier, a government-issued number required for all US physicians and health care providers.

to 2022 period. The American Medical Association estimates that over half of US physicians are now vertically integrated—and hence are employees, rather than owners of their own medical practice firms (Kane 2025). This trend is also ubiquitous, though not uniform, across all physician specialties (Nikpay, Richards, and Penson 2018; Whaley et al. 2021a).

The rapid formation of vertically integrated physician groups is largely due to financial incentives for exercising greater control over physician referrals, especially when these referral decisions impact the flow of patients and revenues to hospital settings. For example, the 340B program, which creates substantial drug acquisition discounts for hospitals, can motivate greater hospital-physician vertical integration. Profit margins for oncology and other provider-administered drugs at 340B hospitals can be nearly 700 percent, which creates strong incentives to capture referrals from oncologists and related specialist physicians (Robinson, Whaley, and Dhruva 2024). Physicians can benefit from vertical integration as well. Beyond avoiding the challenges of being a small business owner, vertically integrating with a health system also allows for a physician's service prices to be negotiated by the owning health system. An overarching health system is much more likely to wield greater negotiation leverage vis-à-vis health insurers when

compared to an individual physician or small physician group, which would typically lack market dominance and accompanying bargaining power (Peters 2014; Liu et al. 2022; US Government Accountability Office 2025; Cooper et al. 2025; Baker et al. 2014).

New Financing for Health Sector Integration

The US health care system is now undergoing its third wave of modern consolidation, including new financial actors and even more mixing and matching of previously separate industries within the health care ecosystem. First, various financial services firms, led by private equity firms, have rapidly expanded into health care markets. These changes in ownership arrangements are less familiar and less observable to health care consumers and policymakers than the previous rounds of mergers. These investment firms also possess the financial capital to potentially underwrite consolidation at a scale and speed that health care providers could not accomplish themselves.

Private equity, specifically, has been present in health care markets for several decades, making deals that have sometimes surfaced in the popular press. For example, the 2006 acquisition of the for-profit HCA hospital chain represented the largest leveraged buy-out in US financial history up to that point (Richards and Whaley 2024). However, private equity's overall *growth* in health care markets has attracted greater scrutiny. Over the 2013 to 2023 period, private equity firms have deployed nearly $800 billion to acquire a variety of health care provider companies within their investment portfolios—ranging from hospitals, physician practices, nursing homes, and surgical centers (Singh et al. 2026).

This surge in private equity investment activity has spurred a contemporary literature focused on its downstream consequences for providers and patients. Across several care delivery settings, private equity acquisitions have been linked to increases in acquired providers' charges and transaction prices (Singh et al. 2022; La Forgia et al. 2022; Liu 2022; Lin et al. 2023). Price increases are particularly notable when private equity investors are able to "roll-up" (that is, purchase and combine) previously fragmented provider firms within a market—and thereby strengthen negotiation leverage with commercial payers (Asil et al. 2024). The evidence exploring effects of private equity on other dimensions relevant to patient welfare and the health care delivery efficiencies is more mixed. Some studies find decreases in quality, especially within private equity–acquired nursing homes (Gupta et al. 2024); however, other studies find no impact on quality or treatment style, with some showing quality improvements in certain clinical contexts (Gao, Kim, and Sevilir 2025; Lin et al. 2023; La Forgia and Bodner 2025). In studies of the private equity influence on clinician labor markets, some have found decreases in physician employment following private equity ownership events (Berquist, Klarnet, and Dafny 2025; Singh et al. 2025a). However, in other business contexts, private equity firms seemingly "tie" physicians to facilities by encouraging individual physician equity stakes, alongside the private equity investments into these same facilities. Doing so gives the affected physicians direct exposure to the facility's overall

financial performance—akin to stock option packages aiming to align incentives for managers and executives with those of the broader organization in a host of other industries (Lin et al. 2023). The lack of clear patterns of studies examining the impacts of private equity acquisition likely reflects the non-random set of firms acquired by private equity, which can range from failing companies to those with substantial, yet unrealized, financial upside.

Private equity firms typically seek investments for a medium-term but limited duration (say, three to seven years), at the end of which they divest from the acquisition and return the proceeds to the initial investors in the fund. For sufficiently large health care businesses, the private equity liquidation event could be precipitated by "taking the business public," which refers to listing the firm on a public stock exchange for an initial public offering. However, for many health care firms, the public equity fund divesture strategy is predicated on a secondary buyer—often including hospitals and health systems, which can risk further competition erosion in the local health care market (Lin et al. 2023; Singh et al. 2025b).

Another emerging source of funding for industry consolidation comes from insurance companies following a strategy of acquiring health care providers directly, sometimes called the "payvider" structure. These acquisitions have expanded greatly in scale and scope in the past few years. As a prominent example, UnitedHealth claims that its subsidiary Optum directly employs or manages thousands of physicians, as well as a multitude of other provider types (home health, hospice, infusion clinics, and surgery centers).[2] Moreover, insurer-led acquisitions often expand beyond physician practices to include other care delivery settings that result in the insurer owning or managing a "platform" of providers across the spectrum of patient care. Insurer-provider tie-ups can lack transparency and are too recent to have garnered a rich set of empirical studies. That said, the rapid and expansive foray of insurers into acquiring clinicians—and in this way owning both sides of the negotiation table over price and structure of health care provision—has already attracted substantial policymaker and regulatory attention.

The effects that these payvider health care mergers and acquisition combinations have on the efficiency and performance of the US health care system are ambiguous. One potential driver of the payvider trend can be a desire to succeed in situations where insurers are receiving flat or capitated annual payments and taking on the risk of care: for example, the Medicare Advantage market and other "value-based care" models. From this view, vertical integration could allow for improved care coordination and alignment between insurers who pay for care and the clinicians who direct care and shape the overall medical spend. Vertically integrated insurers often point to the Kaiser Permanente health maintenance organization model as a leading example of how payer-provider vertical integration can achieve improved

[2]The number of directly employed or managed clinicians is unclear. While United has claimed approximately 90,000 in investor-facing statements, studies using hand-collected and administrative data find far fewer clinicians affiliated with UnitedHealth or its main subsidiary, Optum (Lake et al. 2025; Marr, Whaley, and Zhao 2026).

patient outcomes and lower costs. Capitated payments from public funds to manage insurance benefits now account for approximately 25 percent of private insurer revenue—primarily through the Medicare and Medicaid public insurance segments (Graham 2025). These business lines offer insurers high-powered incentives to engage in cost control in order to maximize profits under these government insurance contracts.

Likewise, these payvider approaches can improve the ability of the acquisition targets—that is, physician groups and other providers—to participate in various value-based payment models that are increasingly promoted by Medicare and other payers (Rooke-Ley and Ryan 2025). An estimated 83 percent of Medicare beneficiaries receive care from a physician reimbursed through some form of capitated payment or quality-based payment model.[3] Importantly, insurers often have the data-monitoring capabilities to track and manage patient costs better across multiple modes and settings of care when compared to any one physician group or other small provider firm, who may just see the patient for a fraction of total care needs. Given that these payment models explicitly place providers "at risk" for patient spending and quality outcomes, it is not surprising that some are looking more to insurers to help manage and spread this risk.

Whatever the merits of these arguments, the actual evidence on integrated insurer-provider delivery systems is limited. The older forms of integrated delivery systems, like the Kaiser Permanente health maintenance organization, also differ from the latest evolutions of insurer-physician vertical integration in terms of patient scope and potential market implications. Kaiser, as well as other fully integrated plans, will commonly only treat patients holding Kaiser insurance and utilize very restrictive provider options—meaning, patients with Kaiser insurance know to use only Kaiser providers. In contrast, physicians acquired by United/Optum or other insurer entities commonly treat patients with multiple forms of insurance, most of which are direct competitors with the owning insurer's products. This "open" competitive dynamic has motivated some concern that insurer-acquired physicians may change care patterns or reduce access for patients with a rival insurance plan. Payvider practices could also raise prices for competing insurers as a way of both extracting revenue from rivals in the short run and pushing up their premiums over the longer run, thus inflicting competitive damage twice over.

Policy Incentives for Consolidation: The Case of Site-of-Care Payment Differentials

Along with the self-reinforcing cycle of health insurer and provider consolidations discussed earlier, certain policy choices have implicitly provided incentives for these forms of consolidation in the health care sector. For example, "site-of-care"

[3] See the website of the Health Care Payment Learning & Action Network (HCPLAN) at https://hcp-lan.org/2024-infographic/.

payment differentials arise when reimbursement for a given service depends on the site in which a procedure or service is performed. Such payment differentials occur for many services that can be performed in multiple sites of care, including outpatient surgical/procedural services, diagnostic testing procedures, imaging services, and clinician administered drugs, which are often reimbursed at a higher rate if provided in a hospital rather than in a doctor's office or a specialized center. By statute, Medicare typically reimburses at a higher rate for identical services and procedures performed in a hospital setting than in a nonhospital setting—say, an ambulatory surgery center or stand-alone imaging center.

For example, Medicare pays $1,088 for a standard colonoscopy performed in a hospital setting but only $666 when performed in a nonhospital ambulatory surgery center.[4] Figure 5 presents the range of Medicare payment differentials between ambulatory surgical centers and hospital outpatient departments for 3,410 common outpatient procedures that are performed in both settings. In these data, Medicare payment rates are often half when the same outpatient procedure is performed in a nonhospital setting like an ambulatory surgical center. Importantly, the "professional fee" paid to the physician is the same, regardless of site-of-care. The Medicare payment differential only applies to the "facility fee" that is directed to the hospital or nonhospital facility like an ambulatory surgery center, diagnostic/imaging center, or infusion center.

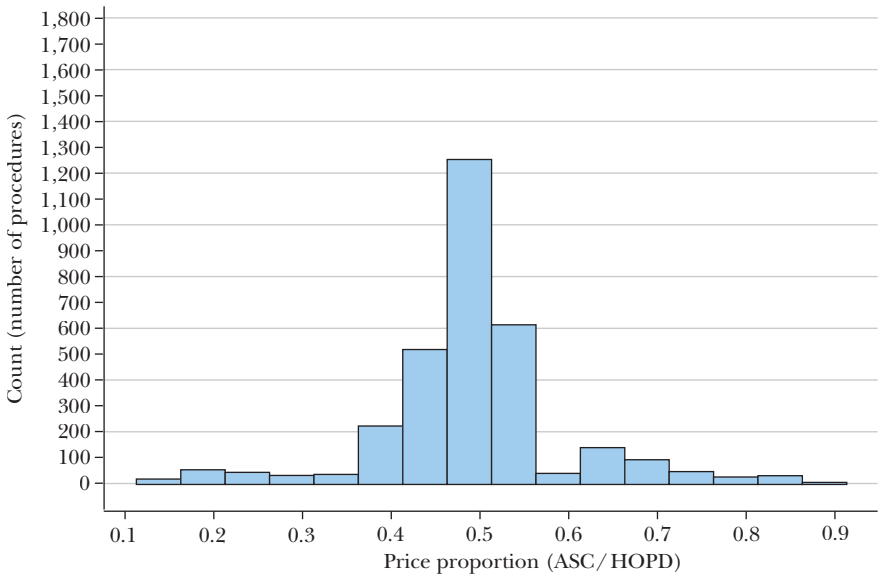
These payment rate gaps are not linked to differences in clinical approach or performance across sites of care. For instance, the "focused factory" nature of ambulatory surgery centers and other nonhospital providers improves quality and patient experience (Casalino, Devers, and Brewster 2003; Munnich and Parente 2018; Aouad, Brown, and Whaley 2019). Instead, hospitals argue that these payment differentials are necessary to account for higher fixed costs of hospital operation, or for higher illness burden or care complexity among patients receiving care at hospitals (American Hospital Association 2025). Yet, little causal evidence exists to support the latter claim. In one study examining changes in use of ambulatory surgery centers following physician ownership investments in the center, patients who shifted to the ambulatory surgery center setting were observably similar to patients who remained in hospital settings (Munnich et al. 2026). Across several dimensions of patient characteristics, including patient age, sex, risk score, race, and income proxies, patients shifted to ambulatory surgery centers were the same as patients in hospital outpatient departments. This result leaves hospitals' relatively high cost structures as the main justification for maintaining the site-of-care differentials for reimbursement.

Site-of-care payment differentials in Medicare also have at least two important market implications for the broader health care system. First, despite the existence of a separate negotiation process to set prices, many commercial payers follow Medicare payment methodologies and are indirectly impacted by Medicare

[4]For details, see the Medicare.gov website at https://www.medicare.gov/procedure-price-lookup/cost/45378 (accessed January 22, 2026).

Figure 5

Medicare Payment Differences Between Services Performed in Ambulatory Surgical Centers and Hospital Outpatient Departments



Source: Analysis of Medicare outpatient price lookup data, https://www.medicare.gov/procedure-price-lookup/.

Note: The horizontal axis shows the proportion of the facility fee in an ambulatory surgical center (ASC) in proportion to a hospital outpatient department (HOPD). The vertical axis shows the number of procedures with this proportion, based on a sample of 3,410 common outpatient procedures.

payment policies (Clemens and Gottlieb 2017; Geruso and Richards 2022; Richards, Seward, and Whaley 2021). Commercial health insurance payments are even negotiated to be a multiple of Medicare rates in some contexts (Weber et al. 2021). As a result, site-of-care payment differentials embedded in the Medicare program can feed into commercial pricing for the same services and generate an important externality leading to higher and more variable spending among the commercial market segment. Indeed, for outpatient surgical services, ambulatory surgery center versus hospital outpatient department price differences are *larger* among commercial plans than within Medicare (Whaley et al. 2024).

Second, the large difference in payments creates an arbitrage opportunity that implicitly incentivizes consolidation. Most services with noticeable site-of-care payment differentials require patients to be referred by a physician. Thus, directing the necessary physician referrals to hospitals and health systems can be a way for these organizations to exploit the opportunities for higher payment for otherwise identical care. Within the Medicare program, anti-kickback provisions prohibit physicians being paid for referrals, and the federal Physician Self-Referral Law, often called the Stark law, limits physicians' ability to refer patients to sites in which

they have a financial interest.[5] However, these restrictions can be circumvented through vertical integration between physicians and the acquiring hospitals or health systems.

As discussed above, the share of physicians employed by a hospital or health system has more than doubled since 2010. Several studies find that following physician-hospital integration, referrals shift towards the acquiring hospital (Whaley et al. 2021c; Richards, Seward, and Whaley 2022; Whaley and Zhao 2024; Cooper et al. 2025). Additionally, Whaley and Zhao (2024) document how changes to referral patterns following vertical integration may harm patient welfare through increased travel distances and adverse quality impacts. Other studies find increases in overall spending, but no gain or loss in quality outcomes post-integration (Koch, Wendling, and Wilson 2017, 2021; Lin et al. 2025). This form of vertical integration can also foster labor market monopsony power—placing downward pressure on health care worker wages (Whaley et al. 2021a). Finally, the subsequent changes in market structure also correspond to downstream shifts in baragaining leverage with commercial insurers and higher prices for physician and hospital services following hospital-physician vertical integration (Baker, Bundorf, and Kessler 2014; Capps, Dranove, and Ody 2018; Dranove and Ody 2019; Lin, McCarthy, and Richards 2021; Cooper et al. 2025).

Policy Incentives for Consolidation: The Case of the Hart-Scott-Rodino Rule

Many factors have contributed to the expansion of private equity into US health care, but one potential policy contributor relates directly to the Hart-Scott-Rodino rule for firms considering mergers and acquisition deals. This rule requires that firms file a pre-merger report with antitrust authorities when considering any deals above certain threshold size, set at $133.9 million in 2026. With scarce resources, instructing regulators to focus their attention on presumably more consequential consolidation events is reasonable. However, the share of deals falling below the statutory Hart-Scott-Rodino threshold is likely to vary across sectors of the economy. For health care in particular, the plurality of providers—physician practices, independent pharmacies, dialysis centers, imaging centers, ambulatory surgery centers, and others—are going to have valuations and purchase prices well below the Hart-Scott-Rodino threshold (Wollmann 2020; Eliason et al. 2020; Asil et al. 2024).

These antitrust policy realities create an environment where a private equity investment group can establish a platform provider practice—for example, an anesthesia provider network—and grow it substantially without ever crossing the Hart-Scott-Rodino threshold. The overall valuation of the growing provider platform

[5] One exception to these Stark Law restrictions is ambulatory surgery centers, which commonly have physician owners. To comply with this Stark Law exception, ambulatory surgery centers are required to report their ownership structure to the Centers for Medicare Centers & Medicaid Services (CMS), including any physician owners with at least a 5 percent ownership stake.

can become quite high, but each individual "tuck-in" acquisition is relatively small in dollar terms. In this way, provider consolidation that permanently transforms a local health care market can happen stealthily and quickly.

Policy Incentives for Consolidation: The Case of Insurer Regulations

Federal and state regulations and payment policies for private insurers are potential drivers of greater insurer-provider integration. First, federal regulations regarding minimum "medical loss ratios" require that insurers devote at least 80 to 85 percent of their premium dollar to medical and care quality improvement expenses for relevant (regulated) plans (Cicala, Lieber, and Marone 2019; Zhao 2021; Chen, Grabowski, and Trish 2025). These regulations were born out of the Patient Protection and Affordable Care Act of 2010 and touted as restraints on insurers' profits. Second, state-level rate review and approval regulations require insurers to justify and gain approval for health insurance premium increases. However, medical loss ratio and rate review requirements can also create perverse incentives, including a form of "self-dealing" in which insurers recapture some of those lost premium dollars by paying providers that they also happen to own. In other words, acquiring insurers can redirect and even increase payment rates to the providers they own as a form of intra-company transfer, which can make medical-loss-ratio requirements and state restrictions on premium increases less binding. As an example, approximately 30 percent of UnitedHealthcare insurer expenses are flowing to its provider assets or other subsidiaries of the United Health Group.

Another implicit incentive for the payvider consolidations described earlier comes from the Medicare Advantage program structure, the Medicare Part C provision in which the federal government pays participating private insurers to administer health insurance coverage for enrollees—now totaling nearly $500 billion in annual spending. These fixed ("per member per month") payments to Medicare Advantage insurers are adjusted based on patient characteristics, including demographics (age and sex) as well as the expected patient illness burden. The latter factor is accomplished through risk adjustment. While Medicare Advantage plans are paid based on patient risk, the clinicians are ultimately responsible for coding and documenting patient illness. These realities create a perverse financial incentive for insurers to influence how clinicians code patient risk, with the underlying goal to "upcode" severity in order to trigger higher risk-adjustment payments for a given enrollee. Vertical integration between insurers and providers can be a way to align incentives between Medicare Advantage plans and the relevant clinicians to ensure higher risk-based payments.[6] While research on this subject is just taking shape, studies have found evidence of extensive patient risk score gaming, and

[6]These incentives for higher risk score payments can also occur in the absence of formal acquisition or integration. Many Medicare Advantage plans pay providers based on "sub-capitation" arrangements, in

more specifically, risk score increases of up to 30 percent following vertical integration or other alignment between physician practices and Medicare Advantage plans (Geruso and Layton 2020; Meyers et al. 2025; Marr, Whaley, and Zhao 2026). These increases in *documented* patient illness seem to occur in the absence of other measurable changes in patient health status or care utilization, suggesting that they reflect increased coding intensity rather than truly higher disease burden (Meyers et al. 2025).

Options to Address Modern Health Care Market Consolidation

A range of policy options could be used to address consolidation in the health care sector: increased oversight and regulation of health care markets, changes to payment policies that favor consolidation, as well as more expansive regulation of provider payment and market conduct.

Increased Transparency of Market Structure and Commercial Prices

Data on health care organizations and prices are mostly complex, nontransparent, and unavailable. Transparency policies seek to increase available information on insurer and provider organizational structures and rates negotiated between insurers and providers.

Ownership patterns in the health care sector often seem constructed to offer a shield from regulatory scrutiny. While provider ownership should be reported to the Centers for Medicare & Medicaid Services through the Provider Enrollment, Chain and Ownership System (PECOS) data, reporting is often incomplete. In particular, PECOS data do not capture "management services only" arrangements, joint ventures, and other "arm's-length" arrangements prevalent in emergent forms of consolidation, including private equity or insurer-led deals (US Government Accountability Office 2025). Limitations with PECOS are particularly relevant when studying insurer-backed consolidation. For example, the nation's largest insurer, UnitedHealth, claims to employ or manage approximately 90,000 US physicians, but only approximately 5,000 can be identified in the PECOS data (Marr, Whaley, and Zhao 2026).

The opaqueness of these transactions limits the ability of researchers and regulators to monitor market conduct and the impact of these acquisitions on market structure, prices, and quality of care. Researchers studying consolidation are often required to use proprietary industry data (for example, American Hospital Association data to study hospital mergers or SK&A data to study hospital-physician vertical integration) or financial services data (for example, from firms like Pitchbook to study private equity transactions).

---

which the patient risk score is used to adjust payments to providers. These "sub-cap" payment models can also create incentives to increase risk scores.

Both federal and state policies have sought to expand transparency of changes in provider (particularly physician) ownership and direct management. At the federal level, the Agency for Health Care Research and Quality has constructed measures of hospital system horizontal and vertical consolidation over the 2016 to 2023 period. Relatedly, the Centers for Medicare & Medicaid Services has published extracted data on changes in ownership for hospitals and nursing homes. Notably, these data track ownership by financial services companies, including private equity firms. These resources are designed to inform the efforts of regulators and researchers to monitor the impacts of provider consolidation and financialization on patient access to care, spending, and quality.

At the state level, building upon model legislation produced by the National Academy for State Health Policy, Oregon now requires notification and approval for any change in direct ownership or management of any health care provider with over $10 million in revenue. This process was recently used to review the proposed merger between two Portland-area health systems, Oregon Health & Science University and Legacy Health. Following concerns from Oregon regulators about increased prices and reduced competition, the two parties dropped merger plans. Several states, including California, Massachusetts, and New Mexico, have implemented similar regulatory requirements. Since 2000, 16 states have passed laws that regulate or provide oversight to health care market consolidation.[7]

Transparency initiatives have also focused on increasing the health care price and quality information available for patients, with the hope that competitive pressures from increased consumer choice will, as in other markets, lead to reduced prices and increased provider quality. However, existing studies show only a minimal impact of consumer-focused tools. Even high-powered consumer-facing price transparency tools find minimal consumer engagement and no overall effect on spending (Whaley et al. 2014; Desai et al. 2016, 2017; Brot-Goldberg et al. 2017; Saad and Brenan 2025). A plausible reason behind such findings is that, as economists have long appreciated, health care markets have extraordinary information asymmetries between patient and physician (Arrow 1963). Additionally, even most "shoppable" services in US health care markets require physician referrals. Most patients are unlikely to push back against physician recommendations for where any "downstream" services should occur—and consolidated physicians may have strong financial incentives to retain these referrals (Chernew et al. 2021). As a result, current price transparency efforts have had only modest benefits for individual patients, who, even if they find actual prices, must still navigate a complex medical and billing system to determine true out-of-pocket costs.

At present, commercial insurance prices are obfuscated, limiting the ability of insurers and employers to design networks and of regulators from monitoring market conduct. However, recent federal policies require all US hospitals to post

[7]For details, see "2024 State Legislation to Lower Health System Costs" at the website of the National Academy for State Health Care Policy (NASHPP), https://nashp.org/state-tracker/state-legislative-action-to-lower-health-system-costs/.

negotiated rates for common services and insurers to post their entire set of negotiated prices. While these data are too unwieldly for routine consumer use, innovators and entrepreneurs have used them to craft benefit design programs steering patients to lower-priced providers. Naturally, providers have tried to use these data to negotiate higher rates, while insurers have simultaneously used them to push for contract discounts (Goldstein 2024). Yet, without improved financial incentives, patients are unlikely to leverage new data to "price shop" for medical services in the same ways observed in other markets.

Payment Policies

As means of both limiting the adverse consequences of existing levels of consolidation and disincentivizing future consolidation, several federal, state, and purchaser policies seek to change payment policies. A particular focus in this area is to limit or eliminate site-of-care payment differentials. The Medicare Payment Advisory Commission, which advises Congress on Medicare policy, has advocated for site-neutral payments (MedPAC 2023). Under their proposal, Medicare would establish a set reimbursement rate for selected services that can be safely and effectively be performed in nonhospital settings (including ambulatory surgery centers and physician offices). More recently, federal Medicare policies have ended payment differentials for clinician drug administration, but proposed policies call for a complete elimination of site-of-care payment differentials. The Congressional Budget Office estimates that full site-neutrality could reduce federal spending by approximately $150 billion over ten years. Commercial insurers could also follow Medicare's lead, and more broadly, site-neutrality would reduce the incentives for health systems to own and acquire physician practices—thus promoting competition in physician service markets. Moreover, several emerging state initiatives seek to expand site-neutrality to commercial markets within their boundaries. For example, Indiana recently prohibited additional facility fees for clinician visits, and in Indiana as well as New York, proposed legislation seeks to adopt MedPAC recommendations on site-neutrality and apply them to commercial markets within their states. Of course, these proposed policy shifts have garnered substantial opposition from the hospital industry, with the American Hospital Association warning of large adverse financial effects on hospitals.

Antitrust Policies

The 2023 Merger Guidelines adopted by the US Department of Justice and the Federal Trade Commission reduce the thresholds under which a horizontal merger would receive heightened regulatory scrutiny, from a Herfindahl–Hirschman Index of 2,500 down to 1,800 (the level prevailing before 2010). Under these revised guidelines, nearly all US hospital markets meet the "highly concentrated" definition. Other proposed antitrust changes include a lower threshold for Hart-Scott-Rodino merger notification in the case of a series of transactions that cumulatively generate market power. Lowering thresholds both expands the sets of consolidation events requiring notification and allows for review of newer forms of consolidation, such as

vertical acquisitions of physician practices and private equity "roll-ups" of physician groups and other nonhospital providers (Asil et al. 2024).

To address emerging vertical and "platform" consolidation more explicitly, recently proposed federal legislation would ban insurers from owning physician groups (Ross 2025). But at present, the consumer welfare effects of insurer-led vertical integration remain empirically unclear. Other sometimes discussed policies include capping payment rates to health care providers in markets with high market shares or those who operate in concentrated markets (Pany, Chernew, and Dafny 2021). These proposed policies recognize that decades of consolidation are not likely to be undone, and so they instead seek to address the pricing-specific consequences of accumulated market power directly.

More Expansive Regulatory Policies

A diverse set of states has decided that the adverse consequences of high commercial prices for health care providers, driven in part by industry consolidation, is beyond modest reform and requires heavy-handed intervention. In states as far-flung as Indiana, Montana, Oregon, Vermont, and Washington, state-imposed regulations are limiting the ability of providers to charge above a certain multiple of Medicare prices for the relevant medical service (Hostert et al. 2025). Similar rules are under consideration in Colorado, Nevada, and New Mexico. States have largely elected to set prices by using "reference-based pricing" as a percentage of Medicare rates because rates are administratively determined to approximate the break-even point for efficient providers and are publicly available (MedPAC 2023). Enacted Medicare reference rates range from 200 percent of Medicare in Oregon to a proposed 265 percent of Medicare rates in Indiana.

Conclusion

The US health care system faces well-known challenges. The United States spends nearly 20 percent of gross domestic product on health care services, which is nearly double peer countries like Australia, Austria, Belgium, Canada, France, Germany, Japan, Netherlands, Sweden, Switzerland, and the United Kingdom (Telesford et al. 2025). Higher US health care spending is largely due to differences in per-unit prices rather than differences in quantities. For example, patients in the United States receive approximately half the number of primary care visits, 35 percent fewer hospital discharges, and the lowest per-capita number of knee and hip joint replacement surgeries relative to peer nations (Tevis et al. 2025). Yet, the prices paid by both public and private insurance systems in the United States far exceed those of comparators, including 50 percent higher prices for joint replacement surgeries covered by public insurance and almost double the price for those same services covered by private insurance. Despite higher spending, the United States has minimal differences in overall health outcomes, and even worse outcomes on key measures like mortality, where its relative performance has declined in recent years.

One of the hopes behind the ongoing consolidation trends that have been restructuring United States health care markets for years is that these new combinations and organizations would push back against these issues, helping to hold down health care prices and improve quality. But so far, the record of consolidation in achieving these goals has proven mixed. Instead, the outcome has often been increased markups passed along to consumers who, due to insurance coverage, lack the price sensitivity to exert market discipline onto provider prices. These newer forms of consolidation may also limit patient choices and enable regulatory and policy avoidance as well as gaming, which undermine intentions to restrain spending and the overuse of low-value care in order to promote more efficient health care delivery.

■ *We thank Daniel Arnold, Qi Lei, and Corrie Mook for their helpful support.*

References

American Hospital Association. 2025. "Fact Sheet: Medicare Hospital Outpatient Site-Neutral Payment Policies." American Hospital Association. https://www.aha.org/system/files/media/file/2023/03/fact-sheet-medicare-hospital-outpatient-site-neutral-payment-policies-march-2023.pdf.

Andreyeva, Elena, Atul Gupta, Catherine Ishitani, Malgorzata Sylwestrzak, and Benjamin Ukert. 2024. "The Corporatization of Independent Hospitals." *Journal of Political Economy Microeconomics* 2 (3): 602–63.

Aouad, Marion, Timothy T. Brown, and Christopher M. Whaley. 2019. "Reference Pricing: The Case of Screening Colonoscopies." *Journal of Health Economics* 65: 246–59.

Arnold, Daniel, Nandita Radhakrishnan, and Christopher Whaley. 2025. "Foisted: The Spillover Effects of Hospital Mergers on Costs and Utilization." Preprint, SSRN. http://dx.doi.org/10.2139/ssrn.5265291.

Arnold, Daniel, and Christopher Whaley. 2020. "Who Pays for Health Care Costs?: The Effects of Health Care Prices on Wages." Preprint, SSRN. http://dx.doi.org/10.2139/ssrn.4959256.

Arrow, Kenneth J. 1963. "Uncertainty and the Welfare Economics of Medical Care." *American Economic Review* 53 (5): 941–73.

Asil, Aslihan, Paulo Ramos, Amanda Starc, and Thomas G. Wollmann. 2024. "Painful Bargaining: Evidence from Anesthesia Rollups." NBER Working Paper 33217.

Baicker, Katherine, and Amitabh Chandra. 2006. "The Labor Market Effects of Rising Health Insurance Premiums." *Journal of Labor Economics* 24 (3): 609–34.

Baker, Laurence C., M. Kate Bundorf, and Daniel P. Kessler. 2014. "Vertical Integration: Hospital Ownership of Physician Practices Is Associated with Higher Prices and Spending." *Health Affairs* 33 (5): 756–63.

Bardey, David, and Jean-Charles Rochet. 2010. "Competition Among Health Plans: A Two-Sided Market Approach." *Journal of Economics & Management Strategy* 19 (2): 435–51.

Barrette, Eric, Gautam Gowrisankaran, and Robert Town. 2022. "Countervailing Market Power and Hospital Competition." *Review of Economics and Statistics* 104 (6): 1351–60.

Beaulieu, Nancy D., Leemore S. Dafny, Bruce E. Landon, Jesse B. Dalton, Ifedayo Kuye, and J. Michael McWilliams. 2020. "Changes in Quality of Care after Hospital Mergers and Acquisitions." *New England Journal of Medicine* 382 (1): 51–59.

Berquist, Victoria, Lev Klarnet, and Leemore Dafny. 2025. "Sale of Private Equity-Owned Physician Practices and Physician Turnover." *JAMA Health Forum* 6 (2): e245376.

Brot-Goldberg, Zarek C., Amitabh Chandra, Benjamin R. Handel, and Jonathan T. Kolstad. 2017. "What Does a Deductible Do? The Impact of Cost-Sharing on Health Care Prices, Quantities, and Spending Dynamics." *Quarterly Journal of Economics* 132 (3): 1261–318.

Brot-Goldberg, Zarek, Zack Cooper, Stuart V. Craig, Lev R. Klarnet, Ithai Lurie, and Corbin L. Miller. 2024. "Who Pays for Rising Health Care Prices? Evidence from Hospital Mergers." NBER Working Paper 32613.

Capps, Cory, David Dranove, and Christopher Ody. 2018. "The Effect of Hospital Acquisitions of Physician Practices on Prices and Spending." *Journal of Health Economics* 59: 139–52.

Casalino, Lawrence P., Kelly J. Devers, and Linda R. Brewster. 2003. "Focused Factories? Physician-Owned Specialty Facilities." *Health Affairs* 22 (6): 56–67.

Chen, Amanda C., David C. Grabowski, and Erin Trish. 2025. "Medical Loss Ratio's Role in the Large Group Insurer Market." *American Journal of Managed Care* 31 (1): 33–36.

Chernew, Michael, Zack Cooper, Eugene Larsen Hallock, and Fiona Scott Morton. 2021. "Physician Agency, Consumerism, and the Consumption of Lower-Limb MRI Scans." *Journal of Health Economics* 76: 102427.

Cicala, Steve, Ethan M. J. Lieber, and Victoria Marone. 2019. "Regulating Markups in US Health Insurance." *American Economic Journal: Applied Economics* 11 (4): 71–104.

Clemens, Jeffrey, and Joshua D. Gottlieb. 2017. "In the Shadow of a Giant: Medicare's Influence on Private Physician Payments." *Journal of Political Economy* 125 (1): 1–39.

Cooper, Zack, Stuart V. Craig, Aristotelis Epanomeritakis, Matthew Grennan, Joseph R. Martinez, Fiona Scott Morton, and Ashley T. Swanson. 2025. "Are Hospital Acquisitions of Physician Practices Anticompetitive?" NBER Working Paper 34039.

Cooper, Zack, Stuart V. Craig, Martin Gaynor, and John Van Reenen. 2019. "The Price Ain't Right? Hospital Prices and Health Spending on the Privately Insured." *Quarterly Journal of Economics* 134 (1): 51–107.

Cooper, Zack, Joseph J. Doyle Jr., John A. Graves, and Jonathan Gruber. 2022. "Do Higher-Priced Hospitals Deliver Higher-Quality Care?" NBER Working Paper 29809.

Crespin, Daniel J., and Christopher Whaley. 2023. "The Effect of Hospital Discharge Price Increases on Publicly Reported Measures of Quality." *Health Services Research* 58 (1): 91–100.

Dafny, Leemore S. 2010. "Are Health Insurance Markets Competitive?" *American Economic Review* 100 (4): 1399–431.

Desai, Sunita, Laura A. Hatfield, Andrew L. Hicks, Michael E. Chernew, and Ateev Mehrotra. 2016. "Association between Availability of a Price Transparency Tool and Outpatient Spending." *JAMA* 315 (17): 1874–81.

Desai, Sunita, Laura A. Hatfield, Andrew L. Hicks, Anna D. Sinaiko, Michael E. Chernew, David Cowling, Santosh Gautam, Sze-jung Wu, and Ateev Mehrotra. 2017. "Offering a Price Transparency Tool Did Not Reduce Overall Spending among California Public Employees and Retirees." *Health Affairs* 36 (8): 1401–07.

Desai, Sunita M., Prianca Padmanabhan, Alan Z. Chen, Ashley Lewis, and Sherry A. Glied. 2023. "Hospital Concentration and Low-Income Populations: Evidence from New York State Medicaid." *Journal of Health Economics* 90: 102770.

Dowd, Bryan E., Tsan-yao Huang, and Tim McDonald. 2021. "Tiered Cost-Sharing for Primary Care Gatekeeper Clinics." *American Journal of Health Economics* 7 (3): 306–32.

Dranove, David, Martin Gaynor, and Eilidh Geddes. 2025. "Expecting Harm? The Impact of Rural Hospital Acquisitions on Maternal Health Care." NBER Working Paper 34159.

Dranove, David, and Christopher Ody. 2019. "Employed for Higher Pay? How Medicare Payment Rules Affect Hospital Employment of Physicians." *American Economic Journal: Economic Policy* 11 (4): 249–71.

Eliason, Paul J., Benjamin Heebsh, Ryan C. McDevitt, and James W. Roberts. 2020. "How Acquisitions Affect Firm Behavior and Performance: Evidence from the Dialysis Industry." *Quarterly Journal of Economics* 135 (1): 221–67.

Fulton, Brent D. 2017. "Health Care Market Concentration Trends in the United States: Evidence and Policy Responses." *Health Affairs* 36 (9): 1530–38.

Gao, Janet, Yongseok Kim, and Merih Sevilir. 2025. "Private Equity in the Hospital Industry." *Journal of Financial Economics* 171: 104107.

Gaynor, Martin. 2006. "What Do We Know about Competition and Quality in Health Care Markets?" NBER Working Paper 12301.

Gaynor, Martin, Kate Ho, and Robert J. Town. 2015. "The Industrial Organization of Health-Care Markets." *Journal of Economic Literature* 53 (2): 235–84.

Geruso, Michael, and Timothy Layton. 2020. "Upcoding: Evidence from Medicare on Squishy Risk Adjustment." *Journal of Political Economy* 128 (3): 984–1026.

Geruso, Michael, and Michael R. Richards. 2022. "Trading Spaces: Medicare's Regulatory Spillovers on Treatment Setting for Non-Medicare Patients." *Journal of Health Economics* 84: 102624.

Goldstein, Joseph. 2024. "Hospital vs. Insurer Dispute Could Force Thousands to Switch Doctors." *New York Times*, March 14. https://www.nytimes.com/2024/03/14/nyregion/united-healthcare-mount-sinai-insurance.html.

Graham, John R. 2025. "Government Accounts for More Than One Quarter of 'Private' Health Insurance Spending." Paragon Health Institute, January 29. https://paragoninstitute.org/paragon-pic/government-accounts-for-more-than-one-quarter-of-private-health-insurance-spending/.

Gruber, Jonathan, and Robin McKnight. 2016. "Controlling Health Care Costs through Limited Network Insurance Plans: Evidence from Massachusetts State Employees." *American Economic Journal: Economic Policy* 8 (2): 219–50.

Gupta, Atul, Sabrina T. Howell, Constantine Yannelis, and Abhinav Gupta. 2024. "Owner Incentives and Performance in Healthcare: Private Equity Investment in Nursing Homes." *Review of Financial Studies* 37 (4): 1029–77.

Hostert, Nathan, Roslyn Murray, Christopher M. Whaley, and Erin C. Fuse Brown. 2025. "How States Are Using Hospital Price Caps To Save Money." *Health Affairs Forefront*. https://doi.org/10.1377/forefront.20251008.247769.

Kane, Carol K. 2025. *Physician Practice Characteristics in 2024: Private Practices Account for Less Than Half of Physicians in Most Specialties*. American Medical Association.

Koch, Thomas G., Brett W. Wendling, and Nathan E. Wilson. 2017. "How Vertical Integration Affects the Quantity and Cost of Care for Medicare Beneficiaries." *Journal of Health Economics* 52: 19–32.

Koch, Thomas G., Brett W. Wendling, and Nathan E. Wilson. 2021. "The Effects of Physician and Hospital Integration on Medicare Beneficiaries' Health Outcomes." *Review of Economics and Statistics* 103 (4): 725–39.

La Forgia, Ambar, and Julia Bodner. 2025. "Getting Down to Business: Chain Ownership and Fertility Clinic Performance." *Management Science* 71 (6): 5022–44.

La Forgia, Ambar, Amelia M. Bond, Robert Tyler Braun, Leah Z. Yao, Klaus Kjaer, Manyao Zhang, and Lawrence P. Casalino. 2022. "Association of Physician Management Companies and Private Equity Investment with Commercial Health Care Prices Paid to Anesthesia Practitioners." *JAMA Internal Medicine* 182 (4): 396–404.

Lake, Derek T., Lawrence Casalino, Michael Richards, Sean Nicholson, Rahul Fernandez, Brendan O'Connell, Manyao Zhang, and Robert Tyler Braun. 2025. "The Impact of Health Insurer Acquisitions of Physician Practices on Prices and Patient Visits." *Health Services Research* 61 (2): e70025.

Lin, Haizhen, Ian M. McCarthy, and Michael Richards. 2021. "Hospital Pricing Following Integration with Physician Practices." *Journal of Health Economics* 77: 102444.

Lin, Haizhen, Ian M. McCarthy, Michael Richards, and Christopher Whaley. 2025. "Owning the Agent: Hospital Influence on Physician Behaviors." *Journal of Human Resources*. https://doi.org/10.3368/jhr.0723-13024R2.

Lin, Haizhen, Elizabeth L. Munnich, Michael R. Richards, Christopher M. Whaley, and Xiaoxi Zhao. 2023. "Private Equity and Healthcare Firm Behavior: Evidence from Ambulatory Surgery Centers." *Journal of Health Economics* 91: 102801.

Liu, Jodi L., Zachary M. Levinson, Annetta Zhou, Xiaoxi Zhao, PhuongGiang Nguyen, and Nabeel Qureshi. 2022. *Environmental Scan on Consolidation Trends and Impacts in Health Care Markets*. RAND Corporation.

Liu, Tong. 2022. "Bargaining with Private Equity: Implications for Hospital Prices and Patient Welfare." Preprint, SSRN. https://doi.org/10.2139/ssrn.3896410.

Marr, Jeffrey, Christopher Whaley, and Xiaoxi Zhao. 2026. "From Payer to Provider: Impacts of Insurer Vertical Expansion into Physician Services." Unpublished.

MedPAC (Medicare Payment Advisory Commission). 2023. *Medicare and the Health Care Delivery System, Report to the Congress*. MedPAC.

Meyers, David, Jay Shroff, Yashaswini Singh, and Christopher Whaley. 2025. "Corporate Practice of

Medicine: Vertical Alignment and Medicare Advantage Risk Coding." Unpublished.

Munnich, Elizabeth L., and Stephen T. Parente. 2018. "Returns to Specialization: Evidence from the Outpatient Surgery Market." *Journal of Health Economics* 57: 147–67.

Munnich, Elizabeth L., Michael R. Richards, Christopher M. Whaley, and Xiaoxi Zhao. 2026. "Raising the Stakes: Physician Facility Investments and Provider Agency." *American Economic Review* 116 (2): 502–34.

Nikpay, Sayeh S., Michael R. Richards, and David Penson. 2018. "Hospital-Physician Consolidation Accelerated in the Past Decade in Cardiology, Oncology." *Health Affairs* 37 (7): 1123–27.

Ochieng, Nancy, Meredith Freed, Jeannie Fuglesten Biniek, Anthony Damico, and Tricia Neuman. 2025. "Medicare Advantage in 2025: Enrollment Update and Key Trends." KFF, July 28. https://www.kff.org/medicare/medicare-advantage-enrollment-update-and-key-trends/.

Oh, Hyesung, and Christopher Whaley. 2026. "Delivering Amidst Consolidation: Implications of Hospital Mergers on Newborn and Maternal Outcomes in the US Medicaid Program." Preprint, SSRN. https://doi.org/10.2139/ssrn.5804282.

Pany, Maximilian J., Michael E. Chernew, and Leemore S. Dafny. 2021. "Regulating Hospital Prices Based on Market Concentration Is Likely to Leave High-Price Hospitals Unaffected." *Health Affairs* 40 (9): 1386–94.

Peters, Craig T. 2014. "Bargaining Power and the Effects of Joint Negotiation: The 'Recapture Effect.'" US Department of Justice Economic Analysis Group Discussion Paper 14-3.

Prager, Elena. 2020. "Healthcare Demand under Simple Prices: Evidence from Tiered Hospital Networks." *American Economic Journal: Applied Economics* 12 (4): 196–223.

Richards, Michael R., Jonathan A. Seward, and Christopher M. Whaley. 2021. "Removing Medicare's Outpatient Ban and Medicare and Private Surgical Trends." *American Journal of Managed Care* 27 (3): 104–08.

Richards, Michael R., Jonathan A. Seward, and Christopher M. Whaley. 2022. "Treatment Consolidation after Vertical Integration: Evidence from Outpatient Procedure Markets." *Journal of Health Economics* 81: 102569.

Richards, Michael R., and Christopher M. Whaley. 2024. "Hospital Behavior over the Private Equity Life Cycle." *Journal of Health Economics* 97: 102902.

Robinson, James C. 1999. "The Future of Managed Care Organization." *Health Affairs* 18 (2): 7–24.

Robinson, James C., Christopher Whaley, and Sanket S. Dhruva. 2024. "Hospital Prices for Physician-Administered Drugs for Patients with Private Insurance." *New England Journal of Medicine* 390 (4): 338–45.

Rochet, Jean-Charles, and Jean Tirole. 2006. "Two-sided markets: a progress report." *RAND Journal of Economics* 37 (3): 645–67.

Rooke-Ley, Hayden, and Andrew M. Ryan. 2025. "A New Medicare Agenda—Moving beyond Value-Based Payment and the Managed Care Paradigm." *JAMA* 333 (14): 1203–04.

Ross, Casey. 2025. "Democrats Introduce Bill to Break Up Health Care Conglomerates Like UnitedHealth Group." *STAT*, September 17. https://www.statnews.com/2025/09/17/unitedhealth-breakup-targeted-new-proposed-law-pat-ryan/.

Saad, Lydia, and Megan Brenan. 2025. "Cost Leads Americans' Top-of-Mind Healthcare Concerns." *Gallup*, December 15. https://news.gallup.com/poll/699770/cost-leads-americans-top-mind-healthcare-concerns.aspx.

Selden, Thomas M., Zeynal Karaca, Patricia Keenan, Chapin White, and Richard Kronick. 2015. "The Growing Difference between Public and Private Payment Rates for Inpatient Hospital Care." *Health Affairs* 34 (12): 2147–50.

Sinaiko, Anna D., Mary Beth Landrum, and Michael E. Chernew. 2017. "Enrollment in a Health Plan with a Tiered Provider Network Decreased Medical Spending by 5 Percent." *Health Affairs* 36 (5): 870–75.

Singh, Yashaswini, Geronimo Bejarano Cardenas, Hamid Torabzadeh, Durga Borkar, and Christopher M. Whaley. 2025a. "Physician Turnover Increased in Private Equity–Acquired Physician Practices." *Health Affairs* 44 (3): 280–87.

Singh, Yashaswini, Nandita Radhakrishnan, Loren Adler, and Christopher Whaley. 2025b. "Growth of Private Equity and Hospital Consolidation in Primary Care and Price Implications." *JAMA Health Forum* 6 (1): e244935.

Singh, Yashaswini, Megha Reddy, Irene Papanicolas, and Richard Scheffler. 2026. "Private Equity Investments in Health Care in OECD Countries: An Exploratory Analysis." *Health Economics, Policy and Law*. https://doi.org/10.1017/S1744133125100352.

Singh, Yashaswini, Zirui Song, Daniel Polsky, Joseph D. Bruch, and Jane M. Zhu. 2022. "Association of Private Equity Acquisition of Physician Practices with Changes in Health Care Spending and Utilization." *JAMA Health Forum* 3 (9): e222886.

Sparks, Grace, Lunna Lopes, Alex Montero, Marley Presiado, and Liz Hamel. 2026. "Americans' Challenges with Health Care Costs." KFF, January 29. https://www.kff.org/health-costs/americans-challenges-with-health-care-costs/.

Telesford, Imani, Lynne Cotter, Emma Wager, and Cynthia Cox. 2026. "How Does Health Spending in the U.S. Compare to Other Countries?" Peterson-KFF Health System Tracker, March 11. https://www.healthsystemtracker.org/chart-collection/health-spending-u-s-compare-countries/.

Tevis, Delaney, Matthew McGough, Juliette Cubanski, Matthew Rae, and Cynthia Cox. 2025. "How Do Healthcare Prices and Utilization in the United States Compare to Peer Nations?" Peterson-KFF Health System Tracker, September 4. https://www.healthsystemtracker.org/chart-collection/how-do-healthcare-prices-and-use-in-the-u-s-compare-to-other-countries/.

US Government Accountability Office. 2025. *Health Care Consolidation: Published Estimates of the Extent and Effects of Physician Consolidation*. GAO.

Wang, Yang, Mark Katz Meiselbach, John S. Cox, Gerard F. Anderson, and Ge Bai. 2023. "The Relationships among Cash Prices, Negotiated Rates, and Chargemaster Prices for Shoppable Hospital Services." *Health Affairs* 42 (4): 516–25.

Weber, Ellerie, Eric Floyd, Youngran Kim, and Chapin White. 2021. "Peering behind the Veil: Trends in Types of Contracts between Private Health Plans and Hospitals." *Medical Care Research and Review* 78 (3): 260–72.

Whaley, Christopher M., Daniel R. Arnold, Nate Gross, and Anupam B. Jena. 2021a. "Physician Compensation in Physician-Owned and Hospital-Owned Practices." *Health Affairs* 40 (12): 1865–74.

Whaley, Christopher M., and Timothy T. Brown. 2018. "Firm Responses to Targeted Consumer Incentives: Evidence from Reference Pricing for Surgical Services." *Journal of Health Economics* 61: 111–33.

Whaley, Christopher M., Christoph Dankert, Michael R. Richards, and Dena Bravata. 2021b. "An Employer-Provider Direct Payment Program Is Associated with Lower Episode Costs." *Health Affairs* 40 (3): 445–52.

Whaley, Christopher M., Chaoran Guo, and Timothy T. Brown. 2017. "The Moral Hazard Effects of Consumer Responses to Targeted Cost-Sharing." *Journal of Health Economics* 56: 201–21.

Whaley, Christopher M., Rose Kerber, Daniel Wang, Aaron Kofner, and Brian Briscombe. 2024. *Prices Paid to Hospitals by Private Health Plans: Findings from Round 5.1 of an Employer-Led Transparency Initiative*. RAND Corporation.

Whaley, Christopher, Jennifer Schneider Chafen, Sophie Pinkard, Gabriella Kellerman, Dena Bravata, Robert Kocher, and Neeraj Sood. 2014. "Association between Availability of Health Service Prices and Payments for These Services." *JAMA* 312 (16): 1670–76.

Whaley, Christopher M., and Xiaoxi Zhao. 2024. "The Effects of Physician Vertical Integration on Referral Patterns, Patient Welfare, and Market Dynamics." *Journal of Public Economics* 238: 105175.

Whaley, Christopher M., Xiaoxi Zhao, Michael R. Richards, and Cheryl L. Damberg. 2021c. "Higher Medicare Spending on Imaging and Lab Services after Primary Care Physician Group Vertical Integration." *Health Affairs* 40 (5): 702–09.

Wollmann, Thomas G. 2020. "How to Get Away with Merger: Stealth Consolidation and Its Effects on US Healthcare." NBER Working Paper 27274.

Zhao, Xiaoxi. 2021. "Medical Loss Ratio Regulation and Insurer Pricing." Unpublished.

Journal of Economic Perspectives—Volume 40, Number 2—Spring 2026—Pages 117–142

Physician Competition: Entry and Substitution

Joshua D. Gottlieb and Sean Nicholson

Since at least Milton Friedman's classic dissertation (Friedman and Kuznets 1945), economists have argued that regulation creates barriers to entry in the health professions, giving physicians market power. When a market is highly regulated, incumbents have an easier time restricting the competitive fringe and maintaining a collusive market structure (Schmalensee 1989; Ellickson 2015). In recent years, these barriers to entry in the health professions have been eroded by competing occupations and perhaps new technologies that can substitute for high-priced physician labor. This article offers key facts and frameworks for thinking about competition among physicians and other medical professionals.

We think about physician competition in two stages. First, potential physicians compete to enter medical school and then many of them compete to enter higher-paid specialties. Within the market for physicians, tight regulatory caps on medical-school seats and residency slots—especially for high-paying specialties—continue to ration entry, enable high returns for those who gain these slots, and steer the most academically accomplished trainees toward lucrative fields.

■ *Joshua D. Gottlieb is Professor at the Harris School of Public Policy, University of Chicago, Chicago, Illinois. Sean Nicholson is Professor of Economics and Public Policy, Cornell University, Ithaca, New York. Gottlieb and Nicholson are both Research Associates, National Bureau of Economic Research, Cambridge, Massachusetts. Their email addresses are jgottlieb@uchicago.edu and sean.nicholson@cornell.edu.* *The authors have received financial support from interested parties.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20251473.

Second, fully trained physicians compete with each other and—as we will emphasize—with other health care occupations in the market for patients. Accounting for health care providers who can substitute for medical doctors, barriers have eroded even without major expansion in medical schools. These substitutes—such as nurse practitioners (NPs), physician assistants, certified registered nurse anesthetists, doctors of osteopathic medicine, foreign-trained physicians, telemedicine, and even emerging artificial intelligence tools—can expand capacity and mitigate the oft-predicted physician shortage.

The physician profession before 1980 had all of the characteristics that Stigler (1971) argued are favorable for regulatory capture: a large occupation with high income absent licensing; stable occupational membership; operating in a local/state rather than a national market; and where all consumers are customers but they are not aggregated (for example, into large health insurance plans that purchase physician services), so lack strong incentives to oppose entry barriers. As Stigler (1971, p. 14) says, "[A] large occupation serving everyone will encounter no organized opposition."

But since the late 1980s, conditions for regulatory capture have become less favorable, and physicians have lost some control over potential competitors' entry. Health care consumers and their agents have stronger incentives to oppose restricting the supply of physician services, now that medical spending constitutes 18 percent of GDP, and they have greater ability to do so because most consumers buy care through a private health insurer that negotiates physician prices on their behalf. The three largest private insurers now cover 122 million people, so the opposition is much more organized than it used to be, in Stigler's (1971) parlance. Furthermore, Medicaid is now the second largest state expenditure after K–12 education (National Association of State Budget Officers 2025). Even though Medicaid physician prices are usually set rather than negotiated, states have incentives to ensure enough physicians (and physician substitutes) are willing to accept low prices to treat low-income patients. There has been moderate growth of medical school and residency program positions since 2000, as well as substantial growth of mid-level health care practitioners who are increasingly allowed to compete with physicians, especially in providing primary care.

To understand competition in the market for physicians, we must therefore go beyond measuring existing levels of concentration through standard concentration metrics such as the Herfindahl-Hirschman Index (HHI). An effective pro-competition policy in the physician market would have two prongs. It would encompass, first, the upstream gatekeeping institutions (accreditation bodies, residency review committees, and state licensure boards) that ration physician entry and, second, the downstream scope-of-practice and technology rules that determine how easily other professionals can substitute for (or complement) physicians in treating patients.

This essay explains the institutions, describes empirical trends, and offers conceptual frameworks to understand competition in physician markets. We emphasize the margins of entry to become a physician and substitution between physicians and other medical practitioners. We do not cover other medical professionals, such

as nurses, except to the extent that they obtain more training to become mid-levels. Compared with physicians, it is easier to enter and faster to complete training in other health care occupations like nursing. Understanding the dynamics of that labor market, and in particular what frictions might impede nurse training, is an important topic for future work.

Preliminary Facts

From 1980 to 2025, the US population grew by 50 percent (World Bank 2025). The population above age 64 grew by 140 percent, while that above age 80 nearly tripled; these groups use a disproportionate share of health care and might thus be a better proxy for demand than total population. Over the same 1980–2025 period, the number of first-year positions in traditional US medical schools that award the Doctor of Medicine (MD) degree increased by only 34 percent. This smaller increase is unlikely to reflect a lack of students interested in becoming doctors; in 2025, there were 2.3 times as many applicants to US MD programs as available positions. This ratio has exceeded 2 every year since 2003.

The relatively small flow of entering physicians shows up in the stock: the United States has 2.7 practicing physicians per 1,000 population versus an average of 3.8 for OECD countries. The average annual growth rate of physicians per capita in the United States between 2000 and 2022 (0.8 percent) is about one-half of the average growth rate for OECD countries as a whole (1.5 percent).

Many organizations and reports have pointed out these patterns and expressed concern about their implications. For example, the American Association of Medical Colleges (2021) predicts a shortage of up to 124,000 physicians in 2034. In 15 large cities, it now takes an average of 31 days to schedule a physician appointment, an increase from 26 days in 2022 and 21 days in 2004 (AMN Healthcare 2025). In Boston, the average wait to schedule an appointment is 65 days across six medical specialties, with average waits of over 80 days for a dermatology or obstetrics/gynecology appointment.

Economists who study competition in labor markets often measure concentration for specific workers or services within a geographic market. The Herfindahl-Hirschman Index, one common way to measure market concentration, is 10,000 times the probability that two randomly drawn purchases in a market are from the same firm. In the average geographic market, the HHI of physician practices has increased over the past 25 years (Fulton 2017; Gaynor 2018; Gaynor, Ho, and Town 2015), and this is associated with increased private health insurance prices (Sun and Baker 2015; Dunn and Shapiro 2014; Baker et al. 2014; Koch and Ulrick 2021; Gaynor 2018; Clemens and Gottlieb 2017; Hausman and Lavetti 2021). Fulton (2017) reports that the average HHI of primary care physician practices increased from about 1850 to 2400 between 2010 and 2016.

In 1983, 41 percent of physicians were in solo practice. Now, only 12 percent of physicians are in solo practice, 55 percent are employed by a health system, and

4 percent are employed by private health insurers such as UnitedHealthcare/Optum (American Medical Association 2025; Physicians Advocacy Institute 2024; Adler et al. 2025).

Conventional measures of market concentration for physicians seem increasingly limited. Patients increasingly travel to physicians with more experience, more appropriate expertise, and rare specializations and capital equipment (Dingel et al. 2023). In addition, telemedicine enables physicians to treat patients far away. Dahlstrand (2025) shows how online health care can enable better matches of physicians and their patients. Other methods of outsourcing certain elements of health care include remote reading of radiology images, centralized pathology centers for analyzing patient samples, and even remote monitoring of intensive care unit patients. All these mechanisms for matching patients and care—potentially across long distances—render measurements of physician competition within a particular geographic location incomplete. Later in the paper, we will also discuss the rise of mid-level health care professionals, such as nurse practitioners, as another force shifting the competitive environment for physicians.

The Physician Education Pipeline

After graduating from medical school, newly minted physicians must receive residency training at an accredited residency program in the United States (or, in some cases, Canada) to practice medicine.[1] Therefore, the market for medical residents essentially determines the flow of new physicians in each specialty practicing in the United States.

Almost all first-year residency positions are allocated through the National Resident Matching Program (NRMP), referred to as "the Match." After interviewing candidates in autumn, each residency program ranks applicants; each applicant ranks residency positions; and a computer algorithm makes binding assignments that economic theory shows to be stable and Pareto-optimal for applicants (Roth and Peranson 1999). These results are revealed to much fanfare on "Match Day" in March. A residency program is a combination of a hospital and a specialty, such as "pediatrics at Massachusetts General Hospital." Primary care residencies usually run for three years, while other residencies can take four or five years, and some particularly specialized ones even longer. Some graduating residents subsequently complete a fellowship to further specialize; for example, a specialization in cardiology requires a three-year fellowship following completion of an internal medicine residency. Even further sub-specialization is possible,

[1] Until 2020, osteopathic (DO) graduates could match to residency positions through a separate system administered by the American Osteopathic Association. Between 2014 and 2020, the two accreditation systems merged, and all residency programs and applicants now participate in a single Match (Almarzooq et al. 2021).

for instance with a two-year electrophysiology fellowship following a cardiology fellowship.

From Predicted Surplus to Predicted Shortage

In determining the quantity of students who will enter the physician education pipeline, medical schools, residency programs, and the organizations that regulate them appear responsive to forecasts of future physician supply. In 1976, Congress asked the Graduate Medical Education National Advisory Committee (GMENAC) to estimate the number and specialty mix of physicians required to meet the nation's health care needs (American Academy of Pediatrics 1981). The resulting 1981 report (McNutt 1981) predicted a surplus of 145,000 physicians by 2000, or 23 percent of the projected workforce, and recommended restricting enrollment in US medical schools and the flow of immigrating international medical school graduates. Congress responded by eliminating medical school subsidies. This had the intended effect; the number of students graduating from those schools essentially remained constant until the mid-2000s.

In 2005 the Council on Graduate Medical Education, the successor to GMENAC, updated its forecast model, predicted a shortage of 85,000 physicians by 2020, and recommended that US medical schools expand enrollment. States and specialty societies concurred. Fourteen states issued reports in the 2000s concluding that there was, or soon would be, a shortage of physicians (Iglehart 2008), and in 2006 the Association of American Medical Colleges (AAMC) recommended a 30 percent increase in MD training capacity.

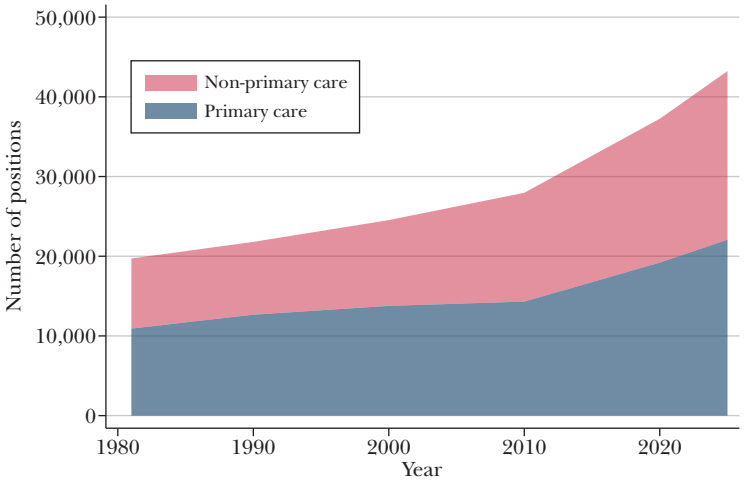
Slots and Applicants

These general patterns are illustrated in Figure 1, which shows the number of first-year residency positions offered over the last 45 years. In 1981, a total of 19,700 first-year residency positions were available to medical school graduates in 22 different specialties, with 55 percent of slots in primary care specialties. Given the earlier predictions of physician oversupply, total residency positions grew by only 42 percent between 1981 and 2010, but after that expanded more rapidly.

Residency applicants grew much more quickly than available positions, as shown in Figure 2. Panel A depicts the number of residency applicants in 1981, 1990, 2000, 2010, 2020, and 2025 by type of applicant, grouped by four medical school categories: (1) a conventional US-based medical school that grants a Doctor of Medicine; (2) a US-based medical school that grants a Doctor of Osteopathic Medicine (DO) degree, which offers similar preparation as an MD but with a different emphasis; (3) US citizens who graduated from a non-US medical school; and (4) foreign citizens who graduated from a non-US medical school.

In 1981, graduates of US-based Doctor of Medicine–granting medical schools comprised 81 percent of all residency applicants. That year, there were 2,600 fewer applicants than available first-year residency positions, and panel B of Figure 2 shows that 93 percent of MD graduates in the Match successfully matched. At that time, match rates for other applicant types were below 68 percent because

Figure 1
Residency Positions Offered



Source: Authors' calculations using data from the National Resident Matching Program (2025a) and historical osteopathic residency data (Cummings 1990; Fusco and Wachtler 1992; Obradovic, Bronersky, and Winslow-Falbo 2002; National Matching Services Inc., 2010).
Note: Before 2020, MD and DO residency programs operated separate match systems; the figure combines positions from both.

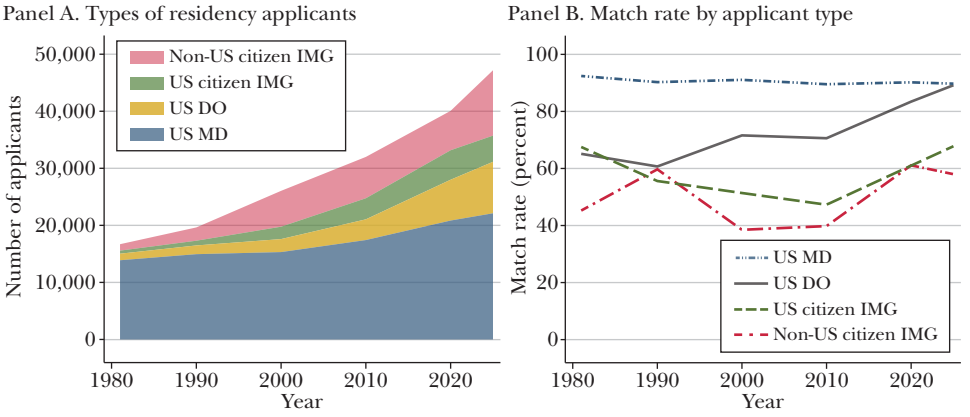
MD graduates were generally perceived by residency programs to be the most qualified.[2]

As the forecasts for future physicians shifted from surplus to scarcity, the number of applicants to the Match almost doubled between 2000 and 2025. The number of Doctor of Medicine–granting medical schools rose from 125 in 2002 to 155 schools today and existing schools increased their class sizes. But the real growth, as depicted in Figure 2, panel A, was in the other three applicant types. The number of Doctor of Osteopathic Medicine–granting medical school programs has more than doubled since 2002 from 19 to 43, and DO graduates now represent 19 percent of Match applicants.

Graduates from international medical schools are generally required to complete a US-based residency before they can be licensed, even if they have already completed a residency program or have practiced internationally. (As an illustration of how the landscape is changing in response to physician scarcity, twelve states have recently relaxed licensing laws by allowing foreign-trained physicians to practice without completing a US residency.) We would expect non-US (citizen) international medical graduates to be particularly responsive to

[2] Unmatched applicants and unfilled positions can coexist because some applicants list a limited set of programs and some programs choose not to list all interested applicants. Some positions that are unfilled in the Match are subsequently filled in the post-Match "scramble."

Figure 2
Applicants and Matches



Source: Authors' calculations using data from the National Resident Matching Program (2025a), the American Association of Colleges of Osteopathic Medicine (American Association of Colleges of Osteopathic Medicine 2024), and historical osteopathic residency data (Cummings 1990; Fusco and Wachtler 1992; Obradovic, Bronersky, and Winslow-Falbo 2002; National Matching Services Inc., 2010).
Note: In panel A, the "DO" series shows the number of graduates from DO schools, regardless of whether they entered an osteopathic residency or the NRMP. All other series show applicants participating in the NRMP. (IMG participation in the osteopathic residency match was negligible.) Match shares in panel B are based exclusively on NRMP applicants and matches.

US market conditions and policies because they already have a medical degree and could practice in their home country—although immigration policies can affect that mobility (Lo Sasso 2021). From 2011 to 2016, non-US international medical graduates Match applicants grew by 2.3 percent per year, on average. Those applicant numbers then decreased by 1.9 percent per year between 2016 and 2020, and increased by 10.7 percent per year between 2020 and 2025. This is consistent with different expectations regarding immigration policies during the first Trump administration and the subsequent Biden administration, and with evidence on how the Affordable Care Act and other policies would affect physician earnings. Because US immigration law provides incentives for non-US international medical graduates to practice in medically underserved areas (Hailat et al. 2025), they fill an important role in alleviating problems with access to medical care.

As the number of residency positions available in the Match surged around 2010 (as shown earlier in Figure 1), US-based Doctor of Osteopathic Medicine schools and international medical schools have had stronger incentives to open and expand. These two categories of schools also compete for talent. DO programs appear to have the advantage because they have attracted US citizens who might otherwise have attended international programs. Concretely, DO graduates have grown faster than US (citizen) international medical graduates in Figure 2, panel A. This revealed preference lines up with the much higher match rates panel B shows for DO graduates than US international medical graduates.

The match rates for all three non-US Doctor of Medicine groups rose sharply between 2010 and 2025 to the point where Doctor of Osteopathic Medicine graduates are now just as likely to match as MD graduates (Figure 2, panel B). By 2010, US-based MD graduates represented only 57 percent of National Resident Matching Program Match applicants. As the ratio of applicants-to-positions rose well above one, fewer than 50 percent of graduates of international medical schools were able to match. These three non-US MD types of medical graduates—US-based DOs, US citizens who graduated from foreign medical schools, and foreign citizens who graduated from foreign medical schools—tend to focus on specialties with lower average salaries where a residency match is easier.

Financial Support for Medical Residents

Residents are financially beneficial for hospitals thanks to federal government subsidies. Almost all residents from the mid-1980s through the 1990s were eligible for "direct graduate medical education" and "indirect medical education" payments supported through Medicare reimbursements, and about 74 percent of them today are eligible (Congressional Research Service 2025; Accreditation Council for Graduate Medical Education 2024). Through the "direct graduate medical education" payments, available since 1985, Medicare has paid hospitals to cover its allocation of the direct costs (for example, resident and faculty salaries) incurred to train an eligible resident, where share is defined as Medicare's proportion of a hospital's inpatient days. Since 1983, Medicare has also paid teaching hospitals an "indirect medical education" supplemental payment for every Medicare patient admitted to cover the estimated "indirect" costs of training eligible residents, such as their relatively lower productivity. A teaching hospital with 100 eligible residents and 400 beds, for example, currently receives an extra 13 percent for each Medicare patient it admits relative to a non-teaching hospital, with the premium based on a hospital's resident-to-bed ratio. In 2020, the average direct and indirect Medicare payments per eligible resident were $51,000 and $119,000, respectively.

Residents acquire general rather than specific human capital; residency training leads to a lifetime of physician earnings, whether practicing in the same hospital where a resident trains or elsewhere. As a result, residents should pay for these general training costs by accepting a salary below their marginal revenue product (Becker 1964).[3] Several studies confirm that residents' salaries, which currently average $67,000 for a first-year resident, are indeed considerably below the value they provide to a hospital (Todd et al. 2004; Green and Johnson 1995; Thorpe 1990), especially because residents often work close to the maximum allowed 80 hours per week. With these rules in place, if the Medicare supplemental payments fully cover a hospital's own cost of training a resident, hospitals can make money on residents. In 1998, Medicare capped the number of eligible residents

[3] See Nicholson (2002) for a more detailed explanation of the economic rationale for the Medicare Direct Medical Education (DME) and Indirect Medical Education (IME) programs.

at a hospital's 1996 level, so most residency positions added since 1998 do not generate incremental supplemental payments. Nevertheless, residents can help generate patient care revenue exceeding the hospital's cost of employing them, at least in general surgery (Richards, Seward, and Whaley 2025).

A Conceptual Framework for Choosing Specialties

To understand physicians' specialization choices, we summarize a conceptual framework from Gottlieb et al. (2023b). This stylized model offers a framework for understanding the relationships between physicians' skill level, incomes, specialty choice, and training difficulty. To simplify the analysis, the model assumes there are just two tracks for physicians: generalist and specialist. Training to become a specialist is long and challenging—but is less arduous for more-skilled physicians.

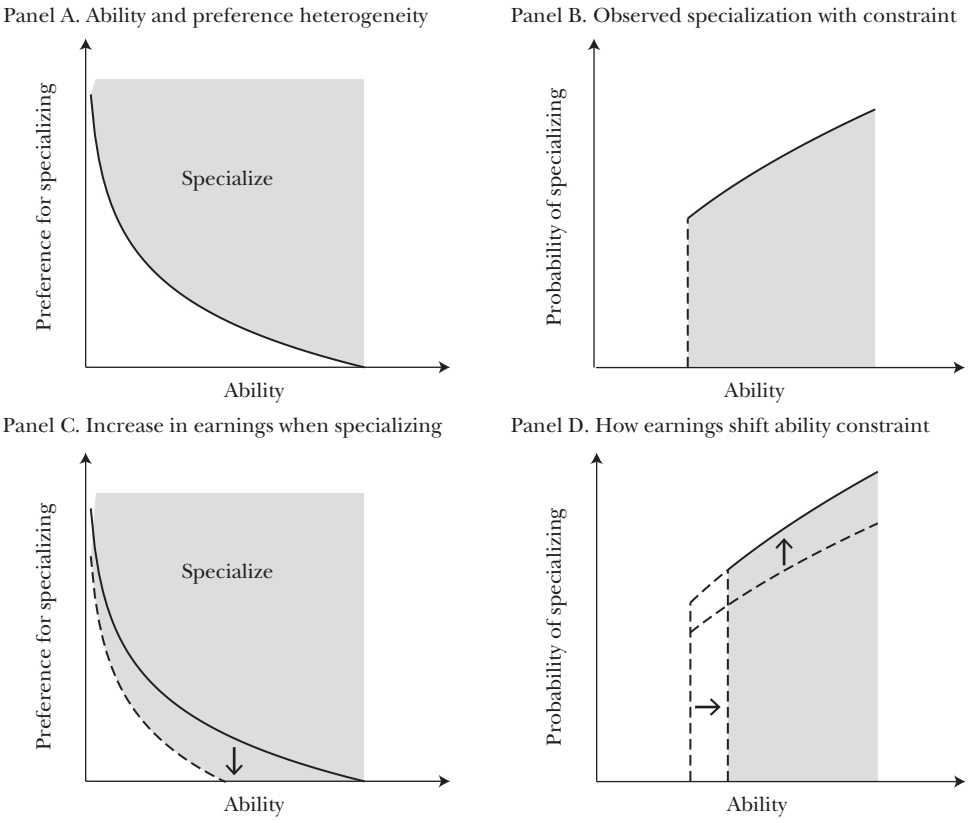
Consider what happens in this model if pay for specialists rises relative to generalists: more people find the long and arduous training worthwhile and would prefer to specialize. We expect to see an increase in the number of highly skilled physicians choosing that path. So far, such a model could apply to many industries. The distinguishing feature of the physician market is the cap on entry in a particular specialty. As a specialty's earnings increase, and more-skilled physicians become interested in that specialty, others must be squeezed out to maintain the entry cap. The specialty can become more selective in two ways: increasing the minimum skill level required to enter the specialty or making training harder.

The model features an equilibrium that balances the specialty's earnings, number of trainees permitted, difficulty of training, and minimum skill level. Figure 3 illustrates the patterns that emerge. As more-skilled physicians enter the higher-paid specialty, they displace the least-skilled physicians who would otherwise choose this specialty. With a fixed population of graduates choosing specialties, and fixed training capacity, Gottlieb et al. (2025a) find a positive supply response of the higher-scoring graduates—but a negative entry response for the lower-score graduates who are displaced.

Once specialties are chosen, physicians and non-physicians enter the market and compete for patients. While these two types of labor may be complementary within a firm, we abstract away from this and focus on the aggregate labor market, where physicians and non-physicians are likely substitutes in producing care. The stricter the entry barriers for any specialty or type of care, the higher the wages we would expect. These equilibrium wages then feed back into the initial specialty choice decision described above.

Given the many simplifications in this model, an important direction for future work is to generalize it. In reality, there are multiple specialties, heterogeneous training programs within a specialty, and multiple skill dimensions. The simple version presented here has one skill dimension, which drives both individual training preferences and residency applications. This approach has analytical advantages, allowing Gottlieb et al. (2023b) to obtain analytical solutions for ability

Figure 3
Conceptual Framework



Source: This figure illustrates the conceptual framework from Gottlieb et al. (2023b).
Note: Panel A shows the physicians who would prefer to specialize: those with higher ability (who thus find the training and other entry barriers more manageable) and those with a higher idiosyncratic preference for specialization, illustrated in the shaded area. Panel B illustrates entry restrictions imposing an ability cutoff limiting specialization to those who match for a residency position. The shaded area illustrates how many physicians end up specializing at each ability level. Panels C and D and illustrate what happens as a specialty becomes more attractive. As specialization becomes more desirable (say, due to higher earnings), the shaded area in panel C expands. In panel D, this is illustrated by the probability increasing at any given ability level (the arrow pointing upwards). If there is a restriction on the number who can specialize, the ability threshold to match will increase. This is illustrated by the arrow showing the ability threshold shifting to the right.

distributions and application thresholds. But in practice, skills and admissions are likely multi-dimensional. Residency programs primarily select based on test scores, medical school performance, and personal statements; these may have little or no relationship to clinical performance or patients' preferences (Lipman et al. 2023). Alexander and Schnell (2026) show that physicians with heterogeneous medical styles enter over the business cycle, and that test scores capture little if any of these

differences. The next section discusses how key elements of the model fit the context of actual specialty choices by physicians.

Barriers to Entry in Choice of Medical Specialties

Compared to primary care, the estimated returns to other specialties were strikingly high between 1951 and 1998—and increasing over time (Nicholson 2008). Between 1987 and 1998, the rate of return in radiology (relative to family practice) ranged from 47 percent to 105 percent. These returns come from a combination of Medicare payment policies, work hours, and training length (Gottlieb et al. 2025a).

How can we determine whether these differences in returns reflect entry barriers, selection, or compensating differentials? The persistence of these high rates of return, combined with a persistent excess of applicants relative to available positions in lucrative specialties, is consistent with the view that entry barriers are a key constraint on physicians' opportunity to specialize. Most visibly, there was much more competition among applicants to secure a non-primary care versus a primary care position in the 1980s, a phenomenon that has increased since then. The unmatched rates for Doctor of Medicine graduates in the three primary care specialties in 1985, the earliest year for which data are available, ranged from 1.6 percent to 4.5 percent, versus 11.0 percent to 16.9 percent in four desirable, high income non-primary care specialties.[4] In 1985, pediatricians and family practitioners had mean incomes of about $78,000 versus $140,000 to $200,000 in anesthesiology, radiology, general surgery, and orthopedic surgery.[5]

One might expect medical students who want to enter the desirable non-primary care specialties with high rates of return to bid down the residents' salaries in those specialties, thereby encouraging teaching hospitals to add more positions—especially before 1998 when the federal resident subsidies were not capped. But this has not happened, and first-year residents were, and still are, almost always paid the same amount by a hospital, regardless of specialty.

Nicholson (2003) discusses two possible rigidities in the market for medical residents which maintain high non-primary-care incomes: cartel behavior by professional associations and wage rigidity. Private organizations that consist primarily of physicians, called Residency Review Committees, restrict the flow of new physicians to non-primary care specialties. The Accreditation Council for Graduate Medical Education (ACGME) is a private organization responsible for overseeing residency training. The ACGME sets overall policies and allows each of the 26 separate specialty-specific Residency Review Committees to review and accredit residency programs in its specialty.

[4]The unmatched rates that we report are the percentage of applicants who rank programs in a single specialty in the Match, which is the most common strategy, and do not receive an assignment.

[5]Incomes reported in this paper are after all practice expenses, including malpractice insurance. The orthopedic surgery income above is from 1986 because it was not available for 1985.

A teaching hospital that wants to open a new residency program or increase the number of residents in their existing program must receive permission from a Residency Review Committee. In most states, medical students must attend a residency program that has been certified by the Accreditation Council for Graduate Medical Education in order to be eligible to take the licensing exam, and thus to practice in that state. Attending an ACGME-certified program is also usually required when physicians seek employment and to obtain admitting privileges at a hospital. Thus, the Residency Review Committees essentially control the flow of physicians into a specialty.

The other possible explanation is that teaching hospitals may not be willing or able to adjust residents' wages to allow the market to clear. The Accreditation Council of Graduate Medical Education (1996) used to require teaching hospitals to pay all residents the same wage, regardless of specialty. But the Federal Trade Commission interpreted the ACGME's policy on residents' wages as a restraint of trade, so the ACGME softened the policy language. Even so, the ACGME still requires that residents be paid an undefined amount—presumably positive—which might prevent the wage from adjusting to clear the excess supply of residents to certain non-primary care specialties.

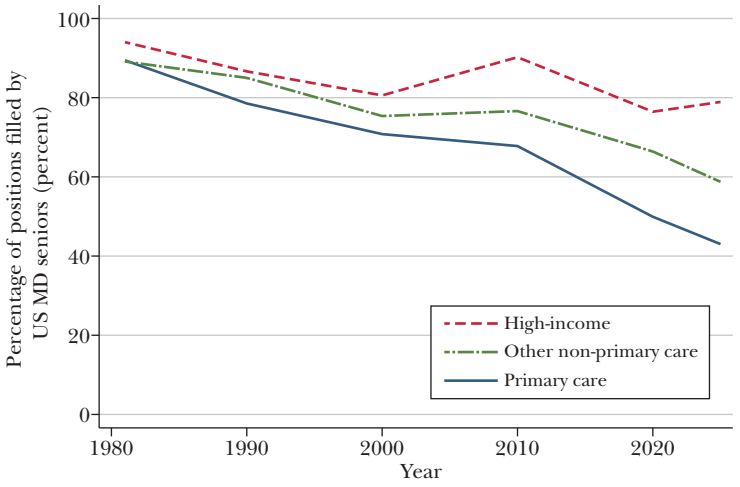
Over the past 30 years, US physicians who train and practice in non-primary care specialties—which usually require four or more years of post-medical school training—have earned substantially more than primary care physicians in family medicine, pediatrics, and general internal medicine, which generally require three years of post-medical school training. This persistent difference, along with barriers to entering non-primary care specialties, directs talent to the high-income specialties. In 2024, the average earnings in non-primary care specialties ranged from $342,000 in psychiatry to $680,000 in orthopedic surgery, compared with $265,000 to $326,000 in the three primary care specialties (Doximity 2025).

Figure 4 depicts the percentage of residency positions filled by US Doctor of Medicine graduates divided into three groups by specialty: primary care, non-primary care specialties with relatively low income, and high-income non-primary care specialties. We define a high-income specialty as one with a mean income greater than $150,000 in 1988 (in nominal dollars) and greater than $475,000 in 2023.[6]

In 1981, when there were relatively few non-MDs in the Match, there was little difference in the residents' medical school type across the three specialty groups. As residency positions expanded, the market has stratified. Non-US MD applicants are less likely to apply for or be accepted by the high-income specialty residency programs. This is also true, though to a lesser extent, with the lower-earning non-primary care specialties. Among Doctors of Osteopathic Medicine in 2025, for example, 14.5 percent ranked one of the six high-income specialties as their first

[6]The high-income non-primary care specialties in this classification are orthopedic surgery, dermatology, plastic surgery, otolaryngology (ENT), anesthesiology, and radiology. The other non-primary care specialties are obstetrics/gynecology, general surgery, psychiatry, emergency medicine, and pathology.

Figure 4
Share of Matches by Specialty Income



Source: Authors' calculations using data from the National Resident Matching Program.
Note: The figure shows the share of first-year residency positions filled by US Doctor of Medicine (MD) seniors in each specialty group. High-income specialties are those with mean physician income exceeding $150,000 in 1988 (in nominal dollars) and $475,000 in 2023: orthopedic surgery, dermatology, plastic surgery, otolaryngology, anesthesiology, and radiology. Other non-primary care specialties are obstetrics/gynecology, general surgery, psychiatry, emergency medicine, and pathology. Primary care specialties are family medicine, pediatrics, and internal medicine.

choice in the Match, versus 25.6 percent for MDs. Conditional on ranking one of those specialties as their first choice, 66 percent of DOs and 81 percent of MDs successfully matched. In 2025, 79 percent of the high-income residency positions were filled by US MDs, although they constituted only 47 percent of the total Match applicants. Conversely, a minority of first-year primary care residents are now US MDs, down from 89 percent in 1981.

Because the likelihood of matching, and matching in a desirable specialty, depends on the type of medical school one attends, prospective physicians compete to enter a type of medical school that affords the best opportunity to enter the profession, and to enter a desirable specialty if that is of interest. There have always been at least twice as many applicants as available positions in US-based Doctor of Medicine programs, and likewise for US-based Doctors of Osteopathic Medicine schools because the data were first available in 2010. The medical school market also stratifies by ability and educational effort. The mean Medical College Admission Test (MCAT) score among MD matriculants in 2024 was 512 (the 83rd percentile among those who took the MCAT exam in 2024), substantially higher than for matriculants to DO programs (503 in 2024, or the 56th percentile among test takers).

Similar patterns exist when comparing different MD-granting medical schools. Students from top-ranked schools do not commonly enter primary care. When

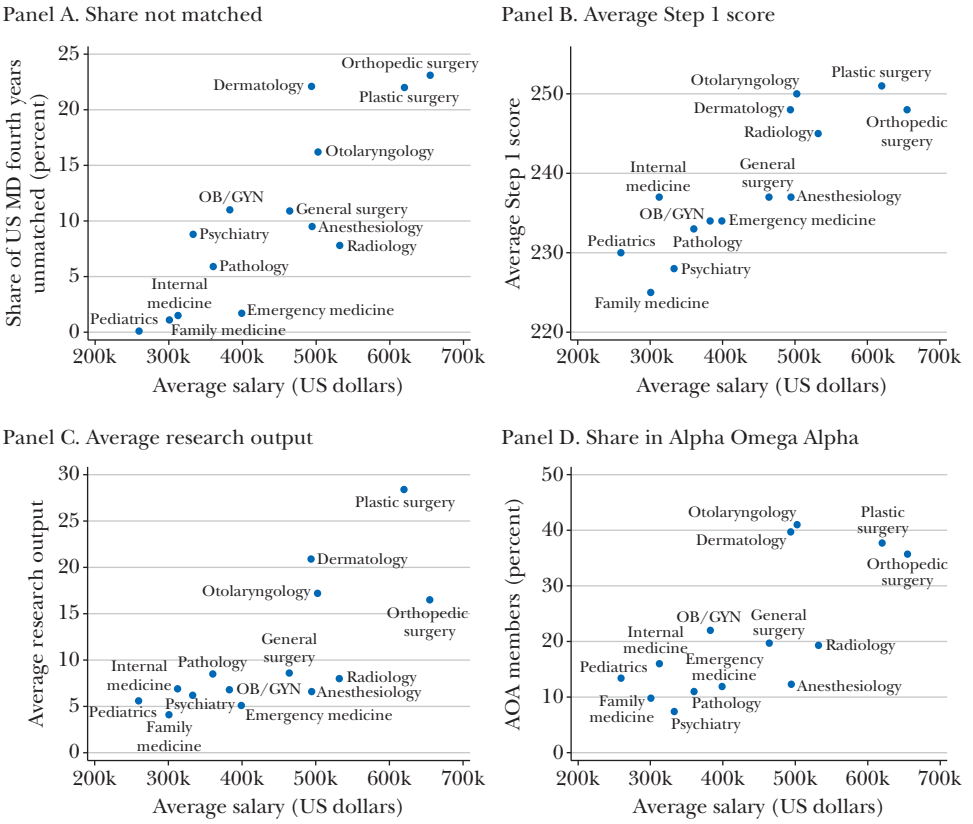
New York University eliminated medical school tuition, supposedly to enable students to practice primary care without the burden of tuition debt, its applicant quality spiked—and these more-competitive students were no more likely to enter primary care (Horowitch 2024). Fourteen percent of NYU's 2024 medical school graduates entered a primary care residency program, well below the 30 percent rate of physicians currently practicing in those specialties (Emanuel and Guido 2024). This could reflect the high returns, and high barriers, to entering selective specialties; students with the greatest financial resources are most likely to be able to take a year off from medical school to conduct research, thereby expanding their specialty choice set. The differences in perceived opportunity cost of this research time may not change when tuition is free. In a statistical sense, greater access to higher-paid specialties explains around 80 percent of the earnings gain from attending a top-ranked medical school (Gottlieb et al. 2025a, Table E.2), so an NYU student's opportunity cost of pursuing primary care is high, regardless of tuition.

Once in medical school, students compete to develop credentials and skills that will provide them with a broad specialty choice set. In Figure 5, we use data from US Doctor of Medicine graduates to depict how positions in the most desirable specialties are allocated. Panel A shows that the unmatched rate in a specialty in 2024 was highly correlated with the mean income of physicians practicing in that specialty in 2023. Over 20 percent of US MDs in 2024 who ranked dermatology, orthopedic surgery, or plastic surgery residency programs as their only choices in the Match did not obtain a match and had to scramble for a position (likely in a different specialty) following the Match, or take a year off and enter the Match the following year—a costly outcome in a high-income profession. The students who successfully enter desirable, high-income specialties are those who were aware of this risk, so these unmatched rates certainly understate the probability that the average US MD applicant would be able to successfully enter these specialties.

How does a US MD graduate increase the chance of entering an attractive specialty? Looking at panels B, C, and D, the answer appears to be excelling in courses and standardized tests while in medical school and investing substantial time conducting and presenting research. The Step 1 exam, which all students take after the second year of medical school, was scored through 2021, although now it is pass/fail. When the exam was scored, applicants with higher scores were much more likely to obtain a match in competitive specialties, as shown in panel B. Likewise, there is a strong positive correlation between the research credentials (panel C) and overall academic credentials (panel D) of students who successfully match in a specialty and that specialty's income. Each medical school may elect up to 20 percent of its graduating class to be members of Alpha Omega Alpha based on "high quality patient care, leadership, service, and scholarship."

Evidence from international physicians and international medical graduates bolsters this story (Gottlieb et al. 2025a; Buehler et al. 2026). US graduates comprise a larger share of matches in higher-paid specialties, suggesting that international graduates face disadvantages when trying to match in competitive residencies. When practicing, physicians in multiple developed countries are generally in

Figure 5
Specialty Earnings and Match Characteristics



Source: Authors' calculations using data from the National Resident Matching Program (2025a).
Note: Panel A shows the unmatched rate among US MD seniors who rank a single specialty in the Match (2024), plotted against the specialty's mean income in 2023 (from Doximity 2025). Panel B shows the Mean Step 1 score in 2022 among US MD graduates who successfully matched by specialty, plotted against the specialty's mean income in 2022 (from Doximity 2025). Panel C shows the mean number of presentations, abstracts, and publications among US MD graduates who successfully matched in each specialty in 2024. Panel D shows the share of students in AOA honor society among US MD graduates who matched in each specialty in 2024.

the top few percentiles of their countries' respective earnings distributions. This pattern could reflect absolute talent requirements to become a physician—causing salaries to be set high enough to induce talented people to choose this career—or inequality spillovers, or both. Even so, US physicians are more likely to be in the top percentile and in the top decile than physicians in Canada, the Netherlands, or Sweden. Combined with the higher top incomes in the United States, this makes it even more attractive to be a physician in the United States, despite the challenge of obtaining a residency.

Another factor pushing physicians to specialize may be the ability to compete for top-paying patients. Gottlieb et al. (2023a) emphasize the value of providing a premium service in an environment with growing income inequality. As income inequality increases, occupations providing services to these unequal consumers can themselves become more unequal. A physician's ability to effectively treat some high-income patients—and charge them a premium for it—can make that physician a high earner in turn. Thus, regardless of any other changes in the market for physicians, greater inequality of incomes might allow some top-earning physicians to retain their high earnings and elite status in the future.

Mid-Level Health Care Providers as Substitutes for Physicians

Even if the quantity of physicians in the United States has failed to keep up with demand, the health care industry has shown an ability to substitute. In pharmaceuticals, for example, when semaglutide demand exceeded Novo Nordisk's manufacturing capability, compound pharmacies arose to fill the gap (Mattingly and Conti 2025).[7] When baby formula was in short supply in 2022, some parents turned to imports while others shared breast milk (Pearson 2022). When demand for physicians increased, and thousands of medics were being discharged after serving in the Vietnam War, the physician assistant profession was institutionalized and expanded across states (Carter 2001; Cawley, Cawthon, and Hooker 2012).

But potential substitutes for some of the care provided by physicians—nurse practitioners, physician assistants, certified registered nurse anesthetists, and certified nurse midwives—are also regulated. Indeed, Starr (1982) argued that physicians have traditionally controlled not just the institutions for training doctors, but also the professions that are potential (partial) substitutes for doctors: "In industry, despite the resistance of artisans, the dictates of the market broke up the work of skilled craftsmen into low-skill—and consequently cheaper—labor. In medicine, physicians maintained the integrity of their craft and control of the division of labor. While medicine itself became highly specialized, the division of labor among physicians was negotiated by doctors themselves instead of being hierarchically imposed upon them by owners, managers, or engineers."

There is a large cost difference between physicians and so-called "mid-level practitioners" with advanced training who can increasingly substitute for physicians, which creates incentives for firms to treat patients using these lower-cost workers. For example, anesthesiologists earned $523,000 on average in 2024 versus $232,000 for certified registered nurse anesthetists; nurse practitioners and physician assistants earned $129,000 and $133,000 on average, respectively, less than one-half of what primary care physicians earn (Doximity 2025; Bureau of Labor Statistics 2024).

[7]Just as with the need to change health care licensing, discussed below, the Food and Drug Administration (FDA) had to designate a shortage for compound pharmacies to be allowed to offer this substitute.

Expansion of Mid-Level Practitioners

States have imposed restrictions on the care that can be provided by nurse practitioners and other mid-level health care professionals. However, recent regulatory changes, often driven by state-level policy initiatives, have relaxed these restrictions. For example, to practice as a NP, a registered nurse—who would have already graduated with an undergraduate nursing degree—must complete an accredited two-year NP masters (or doctoral) program, pass a national exam, and obtain a license from a state board of nursing. Since 1984, 27 states have passed liberal NP "scope of practice" laws, which allow NPs to diagnose and treat patients independently, including ordering and interpreting tests and writing prescriptions (American Association of Nurse Practitioners 2025). In the other states, a NP can only perform tasks under the supervision of a physician or other health care provider.[8] The growth of NPs and other mid-level providers, even in states with restrictive NP scope-of-practice laws, should increase competition among physicians because a physician practice can increase their physicians' income by using lower-cost inputs.

Figure 6 compares key aspects of new Doctors of Medicine and Doctors of Osteopathic Medicine with two prominent examples of mid-level health care providers: nurse practitioners, and certified registered nurse anesthetists (CRNAs). As panel A shows, training of non-MD providers—and especially of NPs—has expanded much more rapidly than that of physicians. Panel B illustrates the shifting provision of three specific types of care, as observed in private insurance claims data (Maretive 2021), with rapid growth in the share of anesthesia procedures provided by CRNAs and an expansion of both primary care and overall professional services provided by NPs. Patel et al. (2022, 2023) show similar patterns in Medicare data. A more detailed breakdown (not shown here) for CRNAs suggests that in 2005, CRNAs were focused on low-severity patients, but by 2021, this was no longer true. However, it remained true that rural areas consistently rely on CRNAs more than urban areas (US Census Bureau, Geography Division 2024; National Resident Matching Program 2025b).

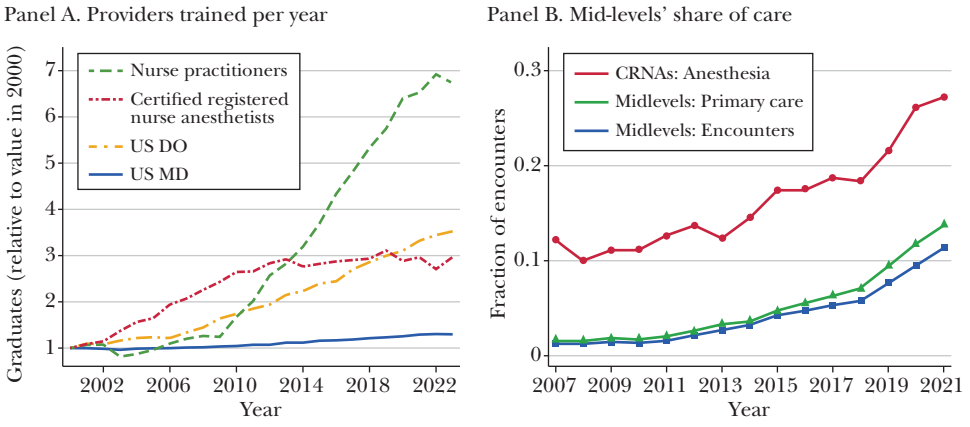
How Mid-Level Professions Fill the Skill Gap

If this broader set of mid-level health care occupations is substituting for hard-to-find physicians, in a context of rising demand for care, we would expect their compensation to reflect this high demand. Indeed, Gottlieb et al. (2025b) find that employment and earnings have grown faster in all clinical occupations than outside of health care. Among these occupations, mid-levels have the fastest employment growth, averaging nearly three times physicians.

Moving beyond averages, Figure 7 shows the earnings distribution for four categories of occupations in 1980 and in 2024. In 2024, physicians' earnings are more spread

[8]Most health insurers reimburse more for a given episode of care if the nurse practitioner is working with a physician than if the nurse practitioner is working independently. For example, a NP working with a physician might be paid at the same rate as the physician for providing a given service, but a NP working independently might be paid only 85 percent of that amount.

Figure 6
Training Rates of Physicians and Their Substitutes



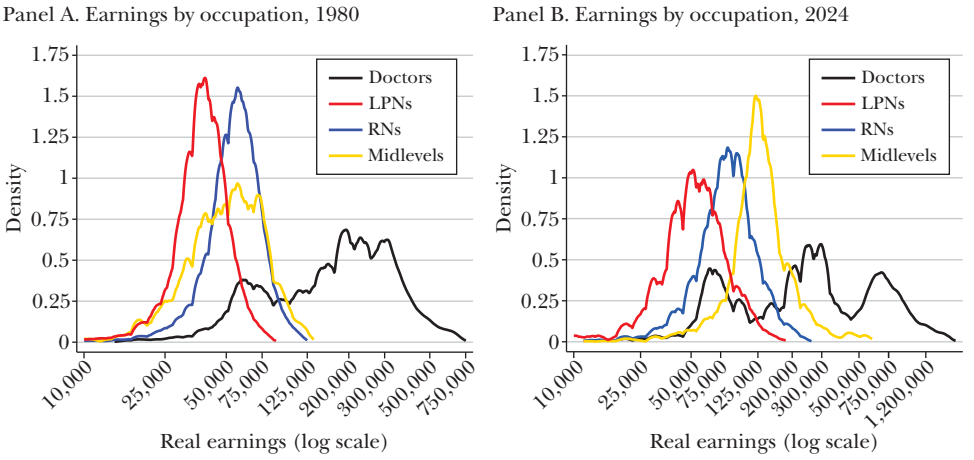
Source: Panel A is based on authors' calculations of data from the Department of Education's Integrated Postsecondary Education Data System (IPEDS) (US Department of Education, National Center for Education Statistics 2024). Panel B is reproduced from Gottlieb et al.'s (2025b, Figure C.8) calculations based on MarketScan private insurance claims data (Merative 2021).
Note: Panel A reports the number of students graduating in each category listed (MD = Medical Doctor; DO = Doctor of Osteopathic Medicine) relative to the value in 2000. Panel B shows CRNAs' share of anesthesia claims, nurse practitioners' share of office visits, and nurse practitioners' share of broader professional claims.

out than in 1980, with much of the distribution in the range of eight to ten times the modal earnings for registered nurses, compared with four to six times the earnings of registered nurses in 1980. The earnings of mid-level practitioners slot neatly between the distribution of registered nurses and that of physicians. These facts suggest that an increase in demand is the central force driving health care employment. But the sluggish growth in physician numbers, combined with their compensation, indicates that supply restrictions may be particularly relevant in that market.

Geographic variation shows evidence of mid-level providers substituting for physicians in response to these supply constraints. Physicians seem to prefer living in high-amenity areas (Lee 2010). Gottlieb et al. (2025b, Figure C.7) treat the area's share of college graduates as a proxy for these amenities. They show that the number of physicians per capita increases much faster in areas that are becoming more educated overall; conversely, growth in nurse practitioners per capita is faster in areas with a lower share of college graduates. Moreover, areas with more college graduates have a lower ratio of NPs per physician, but the wage ratio of NPs to physicians is higher in these same areas.

Putting these facts together enables an inference about the elasticity of substitution between nurse practitioners and physicians. Suppose the local college graduate share shifts the ratio of NPs to physicians in a given area, but does not otherwise shift technology or relative demand for nurse NPs and physicians. Under this assumption,

Figure 7
Earnings Distributions in 1980 and 2024 by Occupation



Source: Gottlieb et al. (2025b, Appendix Figure C.1), updated using IPUMS (Ruggles et al. 2025).
Note: Panels A and B and show the earnings distribution by health care occupation in 1980 and 2024, respectively. LPN stands for "licensed practical nurse"; RN stands for "registered nurse." Data come from IPUMS files for 1980 Decennial Census and the 2024 American Community Survey. Wages are inflation-adjusted to 2024 dollars using the CPI-U (US Bureau of Labor Statistics 2025). We impute physicians' incomes above the public-use censoring threshold using a Pareto distribution (US Census Bureau 2025). We take the (inverse) shape parameters from Gottlieb et al. (2023a), applying their 1980 parameter to our data in 1980 and their 2012 parameter to our 2024 data. We use the censoring threshold for each year and state from IPUMS as the scale parameter. We assign censored incomes by randomly drawing from a Pareto distribution with these parameters.

we can compute the elasticity of substitution between the labor of NPs and physicians. Because the same change in college graduate share associated with relative log wages of NPs increasing by 1.1 is associated with relative log quantity of nurse practitioner declining by 2.25, the implied elasticity of substitution is around 2.

Implications for Physician Competition

The rise in employment, scope-of-practice, and earnings for mid-level health care providers can help to explain why the dire predictions of physician shortages have not become apparent. As noted earlier, the Association of Medical Colleges has been predicting a physician shortage since at least 2006. Yet the average time to schedule an appointment with a family physician was 20 days in 2009 and and 20.6 days in 2022 (AMN Healthcare 2022). Waiting times do appear to have increased recently, reaching 31 days in a 2025 survey, though the longer wait times are driven by specialists; the same survey reports 23.5 days for family medicine (AMN Healthcare 2025). This difference is consistent with mid-level practitioners effectively substituting for physicians in the specialties where they can do so best—and with entry barriers remaining high for specialists.

The rapid growth in physician substitutes, whose training is closest to that of primary care physicians, limits the ability of primary care physicians to earn economic rents. Gottlieb et al. (2025a) compare lifetime earnings of primary care physicians, all physicians, and lawyers. Gottlieb et al. (2025a, App. D) find that, after accounting for length of training and work hours, primary care physicians earn slightly less than lawyers. Other physicians earn substantially more. This finding suggests that primary care physicians may not earn economic rents, at least relative to law—though law has its own barriers to entry.

These results feed back to our discussion in the previous section, as they imply a substantial wage premium that motivates high-achieving physicians-in-training to specialize. Taking a longer historical perspective, Nicholson (2008) summarizes eight studies that estimate the rate of return to a medical school education in the United States between 1929 and 1990. Averaged across all specialties, the general conclusion is similar: there were financial returns from entering medicine through 1966 relative to alternative professions like dentistry, presumably due to barriers to entering the profession. But between 1966 and 1990, the estimated overall rates of return for physicians compared with plausible alternatives have been modest, although they were large for non-primary-care specialties (Nicholson 2003).

Besides competing directly with primary care practitioners, mid-level providers may buttress the value of specialization in a second way. Within a practice, mid-level practitioners may be complements to specialists, for example, by handling the low-fee patient office visits and freeing the specialist up to perform more high-fee procedures. If they enable surgeons and medical specialists to increase productivity, the existence of mid-levels may contribute to the specialization premium.

Conclusion

Despite high per capita levels of US health spending compared to other countries, the United States does not always appear to have an abundance of providers available when needed. This may reflect limits to competition in the health care market, which traditional antitrust remedies do not appear to address. We argue that analysis of competition among physician, provider group, and health insurer should also consider the fundamental questions about entry barriers: who becomes a physician, and which types of workers are allowed to compete in health care provision?

The traditional justification for regulation is to address asymmetric information: consumers may not be able to determine the quality of services provided by the medical workforce (Arrow 1963). From this perspective, the regulator seeks to assure consumers that inputs into the health production function exceed a minimum acceptable quality level.

Regulating the physician market can hamper market functioning by placing constraints on the health production function and raising costs, reducing product variety, increasing wages due to restricted entry, and restricting access to physicians. Diminished product variety—specifically through the absence of lower-quality

and lower-price medical services in the market—is likely to be most detrimental to low-income consumers. Furthermore, other mechanisms can help to address asymmetric information; in the health care context, Nicholson and Propper (2011) discuss providing consumers with information about provider-specific expected health outcomes and making health professionals legally liable for poor quality, and thus willing to purchase malpractice insurance.

But this quality regulation is in tension with competition policy. Quality is generally regulated by the profession in question—and not very effectively, according to Allensworth (2025). The regulation may be more effective at maintaining entry barriers. This shows up in at least three distinct levels: medical school capacity, specialty regulation, and scope of practice.

The difficulty of entering medical school, and the various types of workarounds to these entry barriers, suggests that slot limitations may be binding and creating rents for those who succeed. Specialties represent more concentrated interests, who may have even stronger incentives to limit training and competition. This occurs through Residency Review Committees, board certification, and maintenance of certification. Finally, state scope of practice laws regulate the training required to provide particular types of care. When evaluating physician competition, all of these policies should be considered rather than focusing exclusively on downstream organization of firms.

This essay has not sought to address competition between physician practices once those practices are assembled with the available labor inputs, bargaining between the resulting practices and private health insurers, or the monopsony power that Medicare and other insurers may have with respect to physicians and related professions. These questions are all downstream of our focus—but the upstream competition to enter these professions and downstream competition in care provision are connected. In one direction, changes in the output market for physician services will affect prospective physicians' occupational and specialty choices. In the other direction, changes in entry barriers affect downstream competition, the availability of medical care, and prices. Further, competition within the physician profession—and between physicians and related professions—affects the labor that is available in the output market for medical care, and which can in turn be assembled to form medical practices.

■ *We are grateful to Aidan Buehler, Cedric Elkouh, Katherine Papen, Mingzhe Zhou, Scott Blatte, Sophia Liu, and David Quiroz Barrios for excellent research assistance, and to the Becker-Friedman Institute and the Center for Incentive Design for supporting this research. Much of the prior research discussed in this paper is joint work with other coauthors, to whom we are grateful for fruitful collaboration and discussions; for these collaborations and helpful comments we thank Neale Mahoney, Maria Polyakova, Kevin Rinz, and especially Victoria Udalova, Program Director of the Enhancing Health Data (EHealth) program at the US Census Bureau, whose data products enabled this prior work. In addition to the foregoing, Pauline Mourot and JEP editors Heidi Williams, Jeffrey Kling, and Timothy Taylor have offered valuable comments for which we are grateful.*

References

Accreditation Council for Graduate Medical Education. 1996. *Policies and Procedures*. Accreditation Council for Graduate Medical Education.

Accreditation Council for Graduate Medical Education. 2024. *Data Resource Book*. Accreditation Council for Graduate Medical Education.

Adler, Loren, Samantha Crow, Matthew Fiedler, Richard Frank, Rahul Fernandez, Derek Lake, and Robert Tyler Braun. 2025. "The Changing Landscape of Primary Care: An Analysis of Payer-Primary Care Integration." *Health Affairs Scholar* 3 (7): qxaf120.

Alexander, Diane, and Molly Schnell. 2026. "Selection into Medicine." Unpublished.

Allensworth, Rebecca Haw. 2025. *The Licensing Racket: How We Decide Who Is Allowed to Work, and Why It Goes Wrong*. Harvard University Press.

Almarzooq, Zaid I., Heather A. Lillemoe, Ebony White-Manigault, Thomas Wickham, and Laurie S. Curtin. 2021. "The Single Match: Reflections on the National Resident Matching Program's Sustained Partnership with Learners." *Academic Medicine* 96 (8): 1116–19.

American Academy of Pediatrics. 1981. "GMENAC Staff Papers: Overview." *Pediatrics* 67 (5). https://publications.aap.org/pediatrics/issue/67/5.

American Association of Colleges of Osteopathic Medicine. 2024. *Trends in Osteopathic Medical School Applicants, Enrollment, and Graduates*. American Association of Colleges of Osteopathic Medicine. https://www.aacom.org/searches/reports/report/applicants-enrollment-and-graduates-1977-78—2023-24.

American Association of Nurse Practitioners. 2025. "State Practice Environment." https://www.aanp.org/advocacy/state/state-practice-environment.

American Medical Association. 2025. "Physician Practice Benchmarking Survey." November 10. https://www.ama-assn.org/about/ama-research/physician-practice-benchmark-survey.

AMN Healthcare. 2022. "2022 Survey of Physician Appointment Wait Times and Medicare and Medicaid Acceptance Rates." AMN Healthcare.

AMN Healthcare. 2025. *2025 Survey of Physician Appointment Wait Times and Medicare and Medicaid Acceptance Rates*. AMN Healthcare.

Arrow, Kenneth J. 1963. "Uncertainty and the Welfare Economics of Medical Care." *American Economic Review* 53 (5): 941–73.

Association of American Medical Colleges. 2021. *The Complexities of Physician Supply and Demand: Projections From 2019 to 2034*.Association of American Medical Colleges.

Baker, Laurence C., M. Kate Bundorf, Anne B. Royalty, and Zachary Levin. 2014. "Physician Practice Competition and Prices Paid by Private Insurers for Office Visits." *Journal of the American Medical Association* 312 (16): 1653–62.

Becker, Gary S. 1964. *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*. NBER.

Buehler, Aidan, Joshua D. Gottlieb, Jeffrey Hicks, Lisa Laun, Mårten Palme, Maria Polyakova, Victoria Udalova, and Maria Ventura. 2026. "International Comparison of Physician Incomes." *American Economic Association: Papers and Proceedings*.

Bureau of Labor Statistics. 2024. *Occupational Employment and Wage Statistics*. US Department of Labor. https://www.bls.gov/oes/ (accessed February 19, 2026).

Carter, Reginald. 2001. "Physician Assistant History." *Journal of Physician Assistant Education* 12 (2): 130–32.

Cawley, James F., Elisabeth Cawthon, and Roderick Hooker. 2012. "Origins of the Physician Assistant Movement in the United States." *JAAPA* 25 (12): 36–40.

Clemens, Jeffrey, and Joshua D. Gottlieb. 2017. "In the Shadow of a Giant: Medicare's Influence on Private Physician Payments." *Journal of Political Economy* 125 (1): 1–39.

Congressional Research Service. 2025. *Federal Support for Graduate Medical Education: An Overview*. Congressional Research Service R44376.

Cummings, Mark. 1990. "The Pull Toward the Vacuum: Osteopathic Medical Education in the 1980s." *Journal of the American Osteopathic Association* 90 (4): 353–62.

Dahlstrand, Amanda. 2025. "Defying Distance? The Provision of Medical Services in the Digital Age." URPP Equality of Opportunity Discussion Paper 70.

Dingel, Jonathan I., Joshua D. Gottlieb, Maya Lozinski, and Pauline Mourot. 2023. "Market Size and Trade in Medical Services." NBER Working Paper 31030.

Doximity. 2025. "Physician Compensation Report 2025." Doximity. https://www.doximity.com/reports/physician-compensation-report/2025.

Dunn, Abe, and Adam Hale Shapiro. 2014. "Do Physicians Possess Market Power?" *Journal of Law and Economics* 57 (1): 159–93.

Ellickson, Paul B. 2015. "Market Structure and Performance." In *International Encyclopedia of the Social and Behavioral Sciences*, Vol. 14. 2nd ed., 9211–16. Elsevier.

Emanuel, Ezekiel J., and Matthew Guido. 2024. "Free Med School Tuition Won't Solve the Shortage of Primary Care Physicians." *STAT News*, April 22. https://www.statnews.com/2024/04/22/free-medical-school-tuition-primary-care-doctor-shortage/.

Friedman, Milton, and Simon Kuznets. 1945. *Income from Independent Professional Practice.* NBER.

Fulton, Brent D. 2017. "Health Care Market Concentration Trends in the United States: Evidence and Policy Responses." *Health Affairs* 36 (9): 1530–38.

Fusco, Joseph F., and Janice Wachtler. 1992. "Osteopathic Graduate Medical Education." *Journal of the Osteopathic Medicine* 92 (11): 1391–410.

Gaynor, Martin. 2018. "Examining the Impact of Health Care Consolidation." Speech, Committee on Energy and Commerce Oversight and Investigations Subcommittee, US House of Representatives, Washington, DC, February 14, 2018. https://docs.house.gov/meetings/IF/IF02/20180214/106855/HHRG-115-IF02-Wstate-GaynorM-20180214.pdf.

Gaynor, Martin, Kate Ho, and Robert J. Town. 2015. "The Industrial Organization of Health-Care Markets." *Journal of Economic Literature* 53 (2): 235–84.

Gottlieb, Joshua D., David Hémous, Jeffrey Hicks, and Morten G. Olsen. 2023a. "The Spillover Effects of Top Income Inequality." NBER Working Paper 31366.

Gottlieb, Joshua D., and Sean Nicholson. 2026. *Data and Code for: "Physician Competition: Entry and Substitution."* Nashville, TN: American Economic Association; distributed by Inter-university Consortium for Political and Social Research, Ann Arbor, MI. https://doi.org/10.3886/E246929V1.

Gottlieb, Joshua D., Maria Polyakova, Kevin Rinz, Hugh Shiplett, and Victoria Udalova. 2023b. "Who Values Human Capitalists' Human Capital? The Earnings and Labor Supply of U.S. Physicians." NBER Working Paper 31469.

Gottlieb, Joshua D., Maria Polyakova, Kevin Rinz, Hugh Shiplett, and Victoria Udalova. 2025a. "The Earnings and Labor Supply of U.S. Physicians." *Quarterly Journal of Economics* 140 (2): 1243–98.

Gottlieb, Joshua D., Neale Mahoney, Kevin Rinz, and Victoria Udalova. 2025b. "The Rise of Healthcare Jobs." NBER Working Paper 33583.

Green, Barbara A., and Tim Johnson. 1995. "Replacing Residents with Midlevel Practitioners: A New York City-Area Analysis." *Health Affairs* 14 (2): 192–98.

Hailat, Raed, Mohamed Ridha, Chun Chieh Lin, Kevin Kerber, and James F. Burke. 2025. "Trends in the Proportion of US and International Medical Graduates at the County Level." *JAMA* 334 (14): 1294–95.

Hausman, Naomi, and Kurt Lavetti. 2021. "Physician Practice Organization and Negotiated Prices: Evidence from State Law Changes." *American Economic Journal: Applied Economics* 13 (2): 258–96.

Horowitch, Rose. 2024. "The Perverse Consequences of Tuition-Free Medical School." *The Atlantic*, October 21. https://www.theatlantic.com/ideas/archive/2024/10/perverse-consequences-tuition-free-medical-school/680321/.

Iglehart, John K. 2008. "Grassroots Activism and the Pursuit of an Expanded Physician Supply." *New England Journal of Medicine* 358 (16): 1741–49.

Koch, Thomas, and Shawn W. Ulrick. 2021. "Price Effects of a Merger: Evidence from a Physicians' Market." *Economic Inquiry* 59 (2): 790–802.

Lee, Stephanie. 2010. "Preferences and Practice Location." Unpublished.

Lipman, Jeremy M., Colleen Y. Colbert, Rendell Ashton, Judith French, Christine Warren, Monica Yepes-Rios, Rachel S. King, Beth Bierer, Theresa Kline, and James K. Stoller. 2023. "A Systematic Review of Metrics Utilized in the Selection and Prediction of Future Performance of Residents in the United States." *Journal of Graduate Medical Education* 15 (6): 652–68.

Lo Sasso, Anthony T. 2021. "Regulating High-Skilled Immigration: The Market for Medical Residents." *Journal of Health Economics* 76: 102436.

Mattingly, T. Joseph, and Rena M. Conti. 2025. "Marketing and Safety Concerns for Compounded GLP-1 Receptor Agonists." *JAMA Health Forum* 6 (1): e245015.

McNutt, David R. 1981. "GMENAC: Its Manpower Forecasting Framework." *American Journal of Public Health* 71 (10): 1116–24.

Merative. 2021. "MarketScan Commercial Claims and Encounters (CCAE) Database, 2007–2021." Merative [publisher]. Center for Health and the Social Sciences, University of Chicago [distributor], accessed via the Randi server (accessed November 28, 2023).

National Association of State Budget Officers. 2025. *2025 State Expenditure Report*. National Association of State Budget Officers.

National Matching Services Inc. 2010. "AOA Intern/Resident Registration Program: 2010 Summary by Program Type." (Accessed September 24, 2025).

National Resident Matching Program. 2025a. "Results and Data Reports, 2002–2025." National Resident Matching Program. Archive of annual PDF reports (accessed July 9, 2025).

National Resident Matching Program. 2025b. "Results by State, Specialty, and Applicant Type, 2002–2025." National Resident Matching Program. Archive of annual PDF reports (accessed July 9, 2025).

Nicholson, Sean. 2002. *Medicare Hospital Subsidies: Money in Search of a Purpose*. American Enterprise Institute Press.

Nicholson, Sean. 2003. "Barriers to Entering Medical Specialties." NBER Working Paper 9649.

Nicholson, Sean. 2008. "Medical Career Choices and Rates of Return." In *Incentives and Choice in Health and Health Care*, edited by Frank A. Sloan and Hirschel Kasper, 195–226. MIT Press.

Nicholson, Sean, and Carol Propper. 2011. "Medical Workforce." In *Handbook of Health Economics*, Vol. 2, edited by Mark V. Pauly, Thomas G. McGuire, and Pedro P. Barros, 873–925. Elsevier.

Obradovic, Joyce L., Verna M. Bronersky, and Pamela Winslow-Falbo. 2002. "Osteopathic Graduate Medical Education." *Journal of the American Osteopathic Association* 102 (11): 582–92.

Patel, Sadiq Y., David Auerbach, Haiden A. Huskamp, Austin Frakt, Hannah Neprash, Michael L. Barnett, Hannah O. James, Laura Barrie Smith, and Ateev Mehrotra. 2023. "Provision of Evaluation and Management Visits by Nurse Practitioners and Physician Assistants in the USA from 2013 to 2019: Cross-Sectional Time Series Study." *BMJ* 382: e073933.

Patel, Sadiq Y., Haiden A. Huskamp, Austin B. Frakt, David I. Auerbach, Hannah T. Neprash, Michael L. Barnett, Hannah O. James, and Ateev Mehrotra. 2022. "Frequency of Indirect Billing to Medicare for Nurse Practitioner and Physician Assistant Office Visits." *Health Affairs* 41 (6): 805–13.

Pearson, Catherine. 2022. "What Parents Need to Know about Sharing Breast Milk." *New York Times*, May 22. https://www.nytimes.com/2022/05/20/well/family/breast-milk-sharing-formula-shortage.html.

Physicians Advocacy Institute. 2024. *Updated Physician Practice Acquisition Study: National and State-Level Changes in Physician Employment*. Physicians Advocacy Institute. https://www.physiciansadvocacyinstitute.org/Portals/0/assets/docs/PAI-Research/PAI-Avalere%20Physician%20Employment%20Trends%20Study%202019-2023%20Final.pdf?ver=uGHF46u1GSeZgYXMKFyYvw%3d%3d.

Richards, Michael R., Jonathan Seward, and Christopher M. Whaley. 2025. "Returns to Trainees in a Highly Regulated Labor Market." NBER Working Paper 34365.

Roth, Alvin E., and Elliott Peranson. 1999. "The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design." *American Economic Review* 89 (4): 748–80.

Ruggles, Steven, Sarah Flood, Matthew Sobek, Daniel Backman, Grace Cooper, Julia A. Rivera Drew, Stephanie Richards, Renae Rodgers, Jonathan Schroeder, and Kari C. W. Williams. 2025. "IPUMS USA: Version 16.0." Minneapolis, MN: IPUMS.

Schmalensee, Richard. 1989. "Inter-industry Studies of Structure and Performance." In *Handbook of Industrial Organization*, Vol. 2, edited by Richard Schmalensee and Robert Willig, 951–1009. North-Holland.

Starr, Paul. 1982. *The Social Transformation of American Medicine*. Basic Books.

Stigler, George J. 1971. "The Theory of Economic Regulation." *Bell Journal of Economics and Management Science* 2 (1): 3–21.

Sun, Eric, and Laurence C. Baker. 2015. "Concentration in Orthopedic Markets Was Associated with a 7 Percent Increase in Physician Fees for Total Knee Replacements." *Health Affairs* 34 (6): 916–21.

Thorpe, Kenneth E. 1990. "House Staff Supervision and Working Hours: Implications of Regulatory Change in New York State." *JAMA* 263 (23): 3177–81.

Todd, Barbara A., Andrew Resnick, Rebecca Stuhlemmer, Jon B. Morris, and James Mullen. 2004. "Challenges of the 80-hour Resident Work Rules: Collaboration Between Surgeons and Nonphysician Practitioners." *Surgical Clinics* 84 (6): 1573–86.

US Bureau of Labor Statistics. 2025. "Consumer Price Index for All Urban Consumers: All Items in U.S. City Average [CUUR0000SA0]." Retrieved from FRED, Federal Reserve Bank of St. Louis.

US Census Bureau. 2025. "2024 American Community Survey 1-Year Public Use Microdata Samples: Top and Bottom Coded Values."

US Census Bureau, Geography Division. 2024. "2020 Census Urban and Rural Classification." Page last revised December 16, 2024 (accessed September 24, 2025).

US Department of Education, National Center for Education Statistics. 2024."Integrated Postsecondary Education Data System (IPEDS), 2000–2023, Completions and Institutional Characteristics Components."

World Bank. 2025. "Population, Total for United States [POPTOTUSA647NWDB]." Retrieved from FRED, Federal Reserve Bank of St. Louis.

Journal of Economic Perspectives—Volume 40, Number 2—Spring 2026—Pages 143–170

Substitutes for Success? Public Versus Private Competition in Medicare Advantage

Tim Layton, Luca Maini, and J. Michael McWilliams

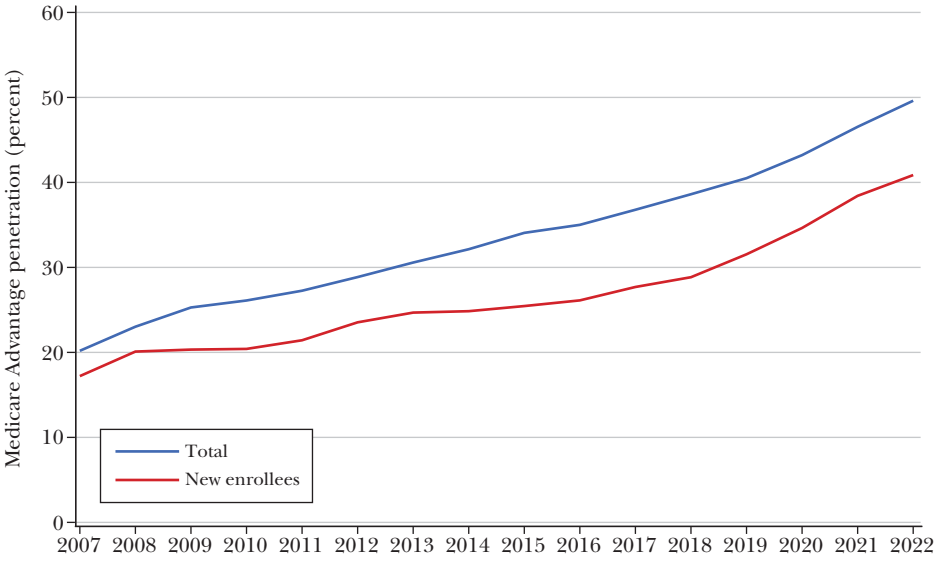
Medicare, the program that provides publicly-financed universal health insurance coverage for nearly all Americans over the age of 65, is typically characterized as a single-payer health insurance system, with government officials administratively setting fee-for-service prices and cost-sharing levels for every medical procedure and service that physicians and hospitals can provide, actuaries setting premium rates, and essentially all private healthcare providers accepting Medicare patients and payment from the program. As such, it is quite similar to single-payer health insurance systems in many European countries. It is also a simple, relatively easy-to-understand program and one of the most popular government programs in history, with over 80 percent of Americans having a favorable opinion of the program.

To most Americans, Medicare thus likely seems to be a program where we have opted for government planning and administration rather than competition among private market actors. However, over 55 percent of Medicare beneficiaries now opt out of this simple, publicly administered *traditional* Medicare program and opt instead to enroll in one of many plans offered by competing private insurers. These insurers have names that are familiar: United Healthcare, Aetna, Humana,

■ *Tim Layton is Associate Professor of Public Policy and Economics, Batten School of Leadership and Public Policy, University of Virginia, Charlottesville, Virginia. Luca Maini is Assistant Professor of Health Care Policy, Harvard Medical School, Boston, Massachusetts. Michael McWilliams is Warren Alpert Foundation Professor of Health Care Policy, Harvard Medical School, Boston, Massachusetts. Their email addresses are timothyjlayton@virginia.edu, maini@hcp.med.harvard.edu, and mcwilliams@hcp.med.harvard.edu.* *The authors have received financial support from interested parties.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20251469.

Figure 1
Medicare Advantage Growth over Time



Source: Layton, Maini, and McWilliams (2026).
Note: This figure plots the portion of Medicare beneficiaries enrolled in Medicare Advantage in each year from 2007 to 2022. New enrollees are defined as Medicare beneficiaries for whom the Medicare Beneficiary Summary File (MBSF) indicated their coverage started in the specified year. For these beneficiaries, Medicare Advantage enrollment was defined as being enrolled in Medicare Advantage any time in the first three months of Medicare enrollment. For other beneficiaries, Medicare Advantage enrollment was defined as being enrolled in Medicare Advantage in January of the specified year. The sample consists of all Medicare beneficiaries in the Medicare Beneficiary Summary File (MBSF) who were enrolled in Parts A and B of Medicare for at least one month (new enrollees) or for all twelve months of the current year (all other enrollees).

Blue Cross, and Kaiser Permanente. This "private option" is officially Medicare Part C, but more commonly known as Medicare Advantage. This program has not always enrolled such a large share of Medicare beneficiaries, but it plays an increasingly important role in modern Medicare: Figure 1 shows that the proportion of Medicare enrollees opting for Medicare Advantage increased from around 20 percent in 2007 to around 50 percent in 2022. The rapid growth of Medicare Advantage has thus made competition a key issue in Medicare.

The central goal of the Medicare Advantage program will be familiar to many economists: Leverage the power of demand-driven competition among private insurers to improve efficiency and welfare. If the traditional Medicare program mistakenly allocates too much money to surgery and not enough to primary care, then in theory, Medicare beneficiaries will demand private plans that encourage more primary care and discourage unnecessary surgeries, and private insurers will respond to this demand by offering such plans and capturing market share (and profits). Success of the program thus depends critically on competition.

In this paper, we assess the role of competition in Medicare Advantage and whether competition is playing its intended role in the Medicare program as a whole. We start by describing the choice between traditional Medicare and Medicare Advantage. We discuss the dynamics of competition between these public and private options. In particular, we discuss how the value of Medicare Advantage to Medicare beneficiaries depends both on competition from the public option of traditional Medicare and from other Medicare Advantage plans. Ultimately, we tell a story of the decline of traditional Medicare as a strong competitor to Medicare Advantage and the rise of competition among private insurers within Medicare Advantage.

In the first part of this story, we attempt to assess the reasons for the decline in the strength of the traditional Medicare program as a competitor for Medicare Advantage. We will argue and attempt to show that part (though not all) of this shift is surely related to the fact that Medicare Advantage plans are paid by the government according to a set of benchmarks and risk-adjustment rules that, in practice, tend to be quite generous relative to the cost of covering the same beneficiaries in traditional Medicare. As a result, much of what beneficiaries like about Medicare Advantage—lower out-of-pocket costs and rich supplemental benefits—is funded by higher federal payments per enrollee rather than purely by private efficiency gains driven by the invisible hand of the market. This imbalance in funding between the two programs makes it hard to evaluate whether "competition works" in Medicare, as the government has tilted the competitive playing field in favor of the private plans.

In the second part of this story, we focus on competition within Medicare Advantage and ask whether private competition has successfully replaced declining competition from the public option. Ultimately, we will show that there has been an increase in competition within Medicare Advantage. But we also argue that this increase likely only partially offsets the decline in competition from traditional Medicare, with insurers still possessing significant market power. We finish by exploring key potential barriers to entry in this market.

Choosing Between Traditional Medicare and Medicare Advantage

To understand the beneficiary's choice between traditional Medicare and Medicare Advantage, one first needs to understand that the basic traditional Medicare benefit package exposes enrollees to substantial out-of-pocket costs (known as "cost-sharing"), intended to restrain "moral hazard" or excessive utilization (Cubanski et al. 2023). This cost-sharing includes a $1,676 deductible for hospital care in 2025 and additional copayments for long stays; 20 percent coinsurance for physician services; another deductible for non-hospital services; up to $2000 in out-of-pocket costs for drugs; and significant daily coinsurance for skilled nursing faculty (nursing home) care from days 21 to 100, after which Medicare coverage ceases (meaning the beneficiary pays the full cost of care). Unlike what is mandated

for commercial insurance, there is no annual or lifetime out-of-pocket maximum in Medicare. Medicare coverage is also not free: Beneficiaries have to pay a monthly premium between $200 and $700 for "Part B" (non-hospital) coverage and also a separate premium for Part D drug coverage (totaling around $3,000 annually).

In short, a beneficiary opting for traditional Medicare can easily pay thousands of dollars in premiums and out-of-pocket expenses for prescription drugs and for services not covered by the basic benefit. Many opt to insure against this remaining risk by purchasing a supplemental "Medigap" policy to cover some or all of this cost-sharing at the nontrivial cost of around $2,000–$3,000 per year for most enrollees.

In contrast, Medicare Advantage plans typically provide enrollees with highly prized benefits not offered by the traditional program: significantly reduced cost-sharing, significant discounts on Part D and B Medicare premiums, coverage for dental, vision, and hearing care (all omitted from the traditional Medicare program), and more. Further, these insurers are required to cover all services covered by the traditional program. However, while traditional Medicare offers an essentially free choice of providers, Medicare Advantage plans generally restrict the set of providers their enrollees can see (although generally still offering relatively broad provider networks relative to employer-sponsored insurance). Also, Medicare Advantage is allowed to impose administrative hurdles such as the requirement that certain services receive prior authorization from the insurer before being provided.

In the end, the trade-off faced by beneficiaries is a traditional Medicare option with free choice of providers but high cost-sharing and no coverage for supplemental benefits like vision, dental, and hearing versus a Medicare Advantage program with much more limited cost-sharing, discounted premiums, and coverage of supplemental benefits but some restrictions on provider choice. Not surprisingly, many find the Medicare Advantage option quite attractive.

A Brief History of Medicare Advantage

The Medicare Advantage program we are familiar with today (including its name) emerged from a flurry of legislation, including the Balanced Budget Act of 1997, the Benefits Improvement and Protection Act of 2000, and the Medicare Modernization Act of 2003. However, the roots of the program go back to 1982, when Medicare received authorization to establish risk-based contracts with health plans. By the mid-1990s, the rates Medicare paid to these plans were set at 95 percent of the average traditional Medicare cost in the enrollee's county, and they were "risk adjusted" based on age, sex, and some other demographics. However, advantageous selection of healthy enrollees into the Medicare Advantage plans led to plan payments greatly exceeding actual costs. Thus, plans were easily able to provide additional benefits and earn profits, at the expense of the taxpayer. Likely at least in part due to this generous financing and in part due to increasing familiarity with managed care, by 1997 Medicare enrollment in these prepaid health plans had

climbed to 14 percent, from just 2.8 percent in 1986 (for a review of this early evolution, see Patel and Guterman 2017).

The official arrival of Medicare Advantage after 2003 brought two main changes. One was an expanded set of options for private plan enrollees, including plan options with fewer restrictions on provider choice than had previously been the norm among these plans and options that targeted specific groups (Special Needs Plans), such as Medicare beneficiaries also enrolled in Medicaid, those in institutions, and those with chronic diseases. When the Medicare Part D prescription drug benefit was rolled out in 2006, it also opened an opportunity for new combined "MA-PD" plans where a private insurer provided medical *and* drug benefits. This new, broader, and more comprehensive set of choices made Medicare Advantage a much more attractive option to a wider swath of Medicare beneficiaries.

The second set of changes sought to address the issue of advantageous selection into Medicare Advantage plans. One step was the phase-in of a new risk adjustment system, based not just on enrollee demographics but also their health status. This "diagnosis-based risk adjustment" used risk scores based on chronic conditions, inferred from diagnoses reported by physicians on claims submitted by the Medicare Advantage plans to the Risk Adjustment Processing System (RAPS) (Newhouse et al. 2012; Brown et al. 2014). The hope was that better measurement of risk would improve the alignment of payments to the (healthier) condition of those enrolling in Medicare Advantage. In addition, payments shifted to being a function of a bidding system that remains in place today, in which insurers submit bids against county-specific benchmarks that are set based on the traditional Medicare cost of an average-risk beneficiary in the county (plus any risk adjustment payments). If the bid is above the benchmark, then the plan receives a payment equal to the benchmark—and beneficiaries have to pay the difference. If the bid is below the benchmark, the plan receives the amount of its bid plus a "rebate" equal to a share of the difference between the bid and benchmark, with Medicare retaining the rest as shared savings. Initially the rebate share was set at a flat 75 percent; now it ranges from 50 to 70 percent depending on plan performance on quality measures. Plans are required to use "rebate dollars" to enhance benefits by reducing cost-sharing below the traditional Medicare level, lowering premiums for Parts B or D, or offering supplemental benefits such as vision, dental, or hearing. Importantly, plans are *not allowed* to offer these additional benefits except with rebate dollars.

The Medicare Modernization Act also increased county benchmarks to 100–115 percent of risk-standardized costs in traditional Medicare, which, in concert with imperfect risk adjustment, allowed plans to bid below benchmarks and offer attractive benefits financed by rebates. A surge in the popularity of Medicare Advantage ensued, along with major increases in Medicare costs, both of which have persisted to the present day. The Medicare Payment Advisory Commission (an independent agency that advises Congress on Medicare issues) estimated that the ratio of actual Medicare Advantage payments to the counterfactual of what program costs would have been if Medicare Advantage enrollees had been enrolled in the traditional Medicare rose from 102 percent in 2003 to 114 percent in 2009.

Before 2003 the high cost of Medicare Advantage was driven by advantageous selection, but after the 2003 legislation, generous benchmark payments and "upcoding" of beneficiaries became the more likely culprits. Because Medicare Advantage insurers were now paid more for beneficiaries with more diagnoses for chronic diseases, Medicare Advantage insurers realized that they could increase their payments substantially by ensuring that every one of their enrollees received a diagnosis for every condition with which they could credibly (and occasionally not-so-credibly) be diagnosed. This generated a situation where the risk score (the basis for risk adjustment payments) for a given individual was substantially higher when that individual enrolled in Medicare Advantage versus when the same individual enrolled in traditional Medicare, leading to higher payments to Medicare Advantage plans, above and beyond those due to high benchmarks or advantageous selection (Geruso and Layton 2020).

Importantly, these generous payments have allowed Medicare Advantage to provide more and more generous products to beneficiaries. Because of the very incomplete coverage offered by the traditional Medicare program, these more generous options, with low cost-sharing, highly subsidized premiums, supplemental benefits, and a wide choice of providers, have been welcomed by Medicare beneficiaries, making it difficult for policymakers to slow payments to Medicare Advantage. One might argue that it would be more efficient to take those excess payments and use them to improve the generosity of the traditional program, rather than indirectly providing increased generosity via a Medicare Advantage program where private insurers take a cut of each additional dollar allocated above the current cost of traditional Medicare. But doing so would require legislation with an explicitly hefty price tag, and there has not been political will to pass this type of reform. Continuing to pay Medicare Advantage plans in excess of traditional Medicare costs through unintended channels, on the other hand, requires no legislation, making this the path of least resistance to satisfying beneficiary demand for improved Medicare generosity. We are thus left in a situation where, as a society, we have determined to make Medicare more generous via these excess payments to Medicare Advantage. Importantly, in such a situation, it becomes critical that Medicare Advantage insurers face strong competition, as it is that competition that will ensure that those excess dollars go to beneficiaries instead of to insurer profits (McWilliams 2024).

Some Dynamics of Competition in Medicare Markets

We now discuss how competition leads to value for Medicare beneficiaries. We characterize the value provided by Medicare Advantage plans to beneficiaries as plan "quality." We are using "quality" in a very general way here, representing basically anything that consumers may value about the plan, such as generosity of coverage of out-of-pocket costs (including deductibles, copays, and premiums), breadth of the provider network, provision of additional benefits (like vision, hearing, and

dental), and so on. The quality of traditional Medicare is set by a political process and known. The Medicare Advantage plan will choose a level of quality to maximize its profits, which will involve comparing the marginal cost of increasing quality to the marginal benefit of attracting additional enrollees.[1]

In this framework, if we (for now) set aside issues related to selection, the decision of a Medicare Advantage plan regarding where to set its level of quality will be based on two (competing) considerations. First, the plan will consider the profit it can earn on an average enrollee at a given level of quality. Assuming that the payment level to the plan is fixed and unrelated to the quality level and that quality is costly, offering higher quality means the plan earns lower profits for each enrollee,[2] pushing the plan's desired level of quality downward (though the extent to which this is true depends on the marginal cost of quality).

Second, the plan will consider the number of Medicare beneficiaries it enrolls. As long as the plan sets quality at a level such that payments exceed plan costs per enrollee, more enrollees imply more profits. Thus, an important consideration is the extent to which setting a higher level of quality attracts more enrollees. A higher level of "quality sensitivity" among potential enrollees of the Medicare Advantage plan will thus encourage the plan to set quality higher and thereby increase its enrollment. Of course, the extent of quality sensitivity depends on quality being readily observable by beneficiaries, which is not obviously true in this market. For example, beneficiary demand for Medicare Advantage plans does not seem to follow the effect that plans have on mortality (Abaluck et al. 2021). Public disclosure of quality information may be important in Medicare Advantage, and more interestingly, the form of disclosure can have important effects on firm behavior (Vatter 2025). But, for now, let's assume some amount of quality sensitivity.

In the end, the Medicare Advantage insurer will choose a level of quality that trades off these two forces and set quality so that it is low enough to earn adequate profits on each enrollee and high enough to have a sufficiently large number of enrollees on which to earn those profits. In other words, the insurer's optimal level

[1] For more formal treatment of these issues, we present a toy model of a public-private health insurance market in the Supplemental Appendix. We consider three settings. First, we consider a setting with a monopolist Medicare Advantage insurer, where the outside option is uninsurance. Second, we consider the introduction of a public option with a monopolist Medicare Advantage insurer, where the outside option is traditional Medicare. Finally, we consider a situation in which the focal Medicare Advantage plan is now competing with both traditional Medicare and another Medicare Advantage plan. Throughout, we highlight the factors that influence quality in the focal Medicare Advantage plan to show how competition influences Medicare Advantage plan quality and, therefore, the "success" of the Medicare Advantage program and beneficiary welfare. Importantly, we show that the key determinant of quality in the focal Medicare Advantage plan is that plan's quality elasticity of demand, and the primary way that competition (with the public option or with other private plans) affects that plan's quality is via that elasticity.

[2] Essentially, we are assuming here that the Medicare Advantage plan does not charge additional premiums. This assumption is plausible to a first approximation. For example, 76 percent of Medicare Advantage enrollees in plans with drug coverage are in plans with no premium (Freed et al. 2024a). While these plans do "share" some savings with the Medicare program, assuming that they receive a fixed payment equal to the benchmark is accurate to a first approximation.

of quality will depend critically on its market share and how enrollment from Medicare beneficiaries responds to the insurer's chosen level of quality (that is, quality sensitivity).

How might competition affect these factors? Competition in Medicare comes in two flavors: competition from traditional Medicare and competition from other private Medicare Advantage plans. Let's start with competition from traditional Medicare. Clearly, in the absence of traditional Medicare, a monopolist Medicare Advantage plan will have a large market share and need not consider quality sensitivity among potential enrollees, as those enrollees have no other option other than uninsurance. In this case, the Medicare Advantage plan will set quality at a very low level in order to maximize the profits it earns on its (captive) enrollees.

If there is a public option (that is, traditional Medicare), this program will capture some share of the market and provide a viable alternative to the monopolist Medicare Advantage plan, shrinking the plan's market share and increasing the quality sensitivity of potential enrollees (as they now have another attractive option). Thus, the Medicare Advantage plan will react by setting a higher level of quality than in the absence of the traditional Medicare option. Importantly, the strength of traditional Medicare matters greatly here. A weaker traditional Medicare makes this option less attractive, leading to the Medicare Advantage plan capturing a larger market share and lowering the quality sensitivity of potential enrollees. Thus, the stronger the traditional Medicare option, the higher the level of quality the monopolist Medicare Advantage insurer will choose. The competitiveness of the public option matters.

Now consider the effect of having multiple Medicare Advantage competitors. More options likely lead the focal Medicare Advantage plan to have smaller market share and to face more quality sensitivity among potential enrollees, who now have multiple attractive alternatives. Thus, the presence of additional Medicare Advantage plans should lead the focal plan to choose an even higher level of quality, as they have fewer enrollees on which to earn more profits by lowering quality and more to gain in terms of additional enrollment by raising quality.

Combining these two types of competition—that is, between traditional Medicare and Medicare Advantage or between Medicare Advantage plans—leads to a critical insight: A strong traditional Medicare plan leads to higher quality from any Medicare Advantage plan, but it also limits the benefits of competition within Medicare Advantage. If traditional Medicare is very strong, then additional Medicare Advantage plans may have only a limited effect—if they choose to enter the market at all. Conversely, a weaker traditional Medicare option leads to lower quality from Medicare Advantage plans in general, and in this way increases the returns to competition within Medicare Advantage.

Thus, the level of quality in traditional Medicare and the extent of effective competition policy in Medicare Advantage can in some ways be thought of as substitutes: We can achieve similar benefits for Medicare beneficiaries using either lever, so in settings where it is difficult to get entry and competition, strengthening traditional Medicare may be an attractive option. Alternatively, in settings where

traditional Medicare is and always will be weak, competition policy may be more necessary.

Finally, we add a few words about selection, which is known to be important in this market and can have significant effects on the incentives of private insurers to provide higher quality products (for starting points to a vast literature, see Rothschild and Stiglitz 1976; McGuire and Glazer 2000; Layton et al. 2017). Specifically for the Medicare Advantage setting, Zahn (2025) shows that selection drives both market entry and plan design for Medicare Advantage. To understand how selection might change the effects of competition in this market, the key question still relates to quality sensitivity. However, now the question depends on *what kind* of marginal enrollee will be attracted if a Medicare Advantage plan increases its quality.

For example, consider the introduction of a strong, high-quality traditional Medicare program into a market dominated by a monopolist Medicare Advantage plan. In this case, those choosing traditional Medicare could easily be sicker than the average Medicare Advantage enrollee (as they are attracted to this strong, generous program). Thus, the potential marginal Medicare Advantage enrollees that the plan would attract by raising quality could be the sick and expensive. Thus, a Medicare Advantage plan thinking about improving quality to attract more enrollees would have less incentive to do so than it would in the absence of selection. On the other hand, if traditional Medicare is relatively weak on quality, then perhaps its beneficiaries will tend to be those who are not much concerned about quality differences between programs, perhaps because they tend to be healthier. In this case, a Medicare Advantage plan will choose an *even higher* level of quality than it would in the absence of selection, because of the possibility of attracting healthier and lower-cost enrollees. If there is competition among Medicare Advantage insurers, the problem becomes more complicated still, as no matter the strength of traditional Medicare, if a focal Medicare Advantage plan raises its level of quality, it will likely attract the sickest beneficiaries from the other Medicare Advantage insurers, again limiting its incentive to raise quality (for an example, see Kreider et al. 2024). The key insight for policymakers is that without appropriate risk adjustment to ensure that marginal consumers, whether sick or healthy, are profitable, both implementing competition policy and enhancing the strength of traditional Medicare could backfire by actually limiting the incentive of a given Medicare Advantage insurer to invest in quality.

In summary, abstracting from selection, competition from traditional Medicare unambiguously increases Medicare Advantage quality. The same goes for competition from other Medicare Advantage plans. With selection, however, all bets are off, as competition from both traditional Medicare and other Medicare Advantage plans can actually lead to worse quality in the focal plan. Based on these insights, we next proceed by assessing how competition faced by Medicare Advantage plans from traditional Medicare (plus selection between these programs) and competition within Medicare Advantage have changed over the last two decades.

The Competitiveness of Traditional Medicare Relative to Medicare Advantage

The strength of traditional Medicare as a competitor for Medicare Advantage has decreased significantly over time. As illustrated earlier in Figure 1, the market share of Medicare Advantage has shown relentless growth in the last two decades. This decline in the relative popularity of traditional Medicare occurred across essentially every subgroup of Medicare beneficiaries. For example, the popularity and growth of Medicare Advantage are fairly similar across age groups. However, the shift has happened faster in some groups than others. The share of beneficiaries enrolled in traditional Medicare tended to decline faster in zip codes with lower median income, suggesting that traditional Medicare represents a stronger competitor to Medicare Advantage among higher-income beneficiaries. Indeed, traditional Medicare today almost appears to be a kind of "luxury good," where beneficiaries with more resources are more likely to stick with it.

In addition, rural areas have traditionally had lower Medicare Advantage penetration than nonrural areas, but have seen stronger growth in more recent years. Those who are dually-eligible for Medicare and Medicaid are overrepresented among Medicare Advantage enrollees due to faster growth among this group as well. While Medicare Advantage has always been popular among Hispanic beneficiaries, its popularity has taken off among Black beneficiaries in recent years, much more so than among white beneficiaries.[3] Overall, the last two decades have thus been characterized by a universal decline in the attractiveness (and thus the competitive strength) of traditional Medicare relative to Medicare Advantage.

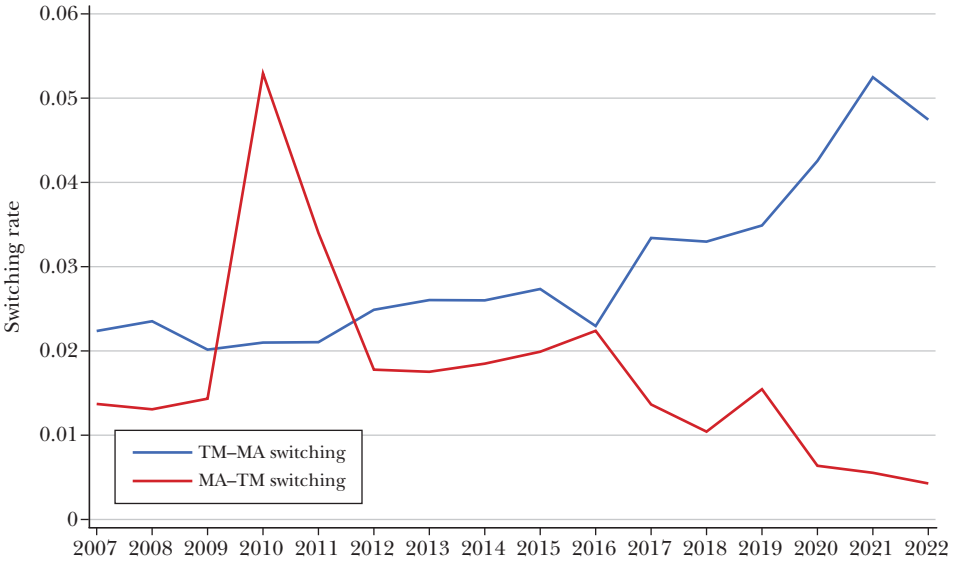
Timing and Selection

The rise of Medicare Advantage's market share has been continual. But if you look closely at Figure 1, you can see a leveling off of that growth around 2010 and then a growth acceleration starting around 2016. These changes are more visible in Figure 2, which investigates switching between traditional Medicare and Medicare Advantage. Medicare beneficiaries are highly inertial, so switching is rare. Yet, there have been some notable changes over time.

In 2010, there was a clear spike in switching from Medicare Advantage to traditional Medicare. This was a one-time shock driven by the mass exit of private fee-for-service plans that occurred in 2010 and 2011 due to a policy change that made those plans nonviable in most cases (Pelech 2018). In 2016, on the other hand, there was a clear trend break when rates of switching from traditional Medicare to Medicare Advantage increased dramatically, while rates of switching from Medicare Advantage to traditional Medicare began to decline precipitously at precisely the same time. The source of this shift is less clear.

[3] Figures showing these changes in penetration of Medicare Advantage plans by median income of zip code, rural/nonrural, dual/non-dual, ethnicity, and age are available from the authors by request.

Figure 2
Switching Between Medicare Advantage and Traditional Medicare

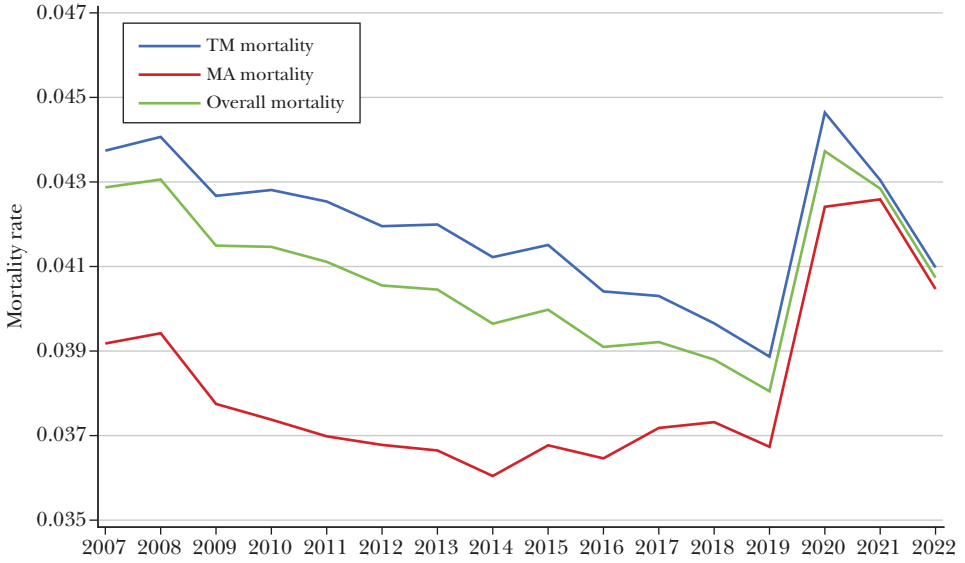


Source: Layton, Maini, and McWilliams (2026).
Note: The figure shows the portion of Medicare beneficiaries in traditional Medicare in January of year $t-1$ who are enrolled in Medicare Advantage in January of year t (blue line) and the portion of beneficiaries in Medicare Advantage in January of year $t-1$ who are enrolled in traditional Medicare in January of year t (red line) for each year from 2007 to 2022. The sample includes all Medicare beneficiaries enrolled in Parts A and B for all twelve months of the year in the Master Beneficiary Summary File (MBSF) maintained by the Centers for Medicare & Medicaid Services.

Recall that while the competitive strength of traditional Medicare matters for the level of quality chosen by Medicare Advantage plans, selection can also play an important role. A detailed investigation of this point might attempt to assess how quality sensitivity varies by health status and over time, leveraging exogenous variation in Medicare Advantage quality at various points in time over the last two decades. In Figure 3, we take a shortcut by showing the mortality rates of three groups over time: (1) all Medicare beneficiaries, (2) Medicare Advantage enrollees, and (3) traditional Medicare enrollees. If the mortality rate of Medicare Advantage enrollees is decreasing over time relative to the overall Medicare mortality rate, that suggests that the marginal Medicare Advantage enrollees are healthier than the inframarginal enrollees.

Prior to 2014, mortality was declining at a similar rate in Medicare Advantage and traditional Medicare (and overall). But after 2014, mortality rates in Medicare Advantage started to flatten. By 2019, there had been significant convergence, and the COVID years saw more convergence still. Interestingly, the timing of the shift in Medicare Advantage mortality after 2015 corresponds closely to the timing of the shift in switching patterns in Figure 2. This suggests that the marginal enrollees

Figure 3
Mortality Rates Among Traditional Medicare and Medicare Advantage Enrollees from 2007 to 2022



Source: Layton, Maini, and McWilliams (2026).
Note: The figure shows the mortality rate for Medicare beneficiaries enrolled in Medicare Advantage, traditional Medicare, or Medicare (Parts A and B) in January of the year. Mortality in a given year is defined as having a date of death during that year in the Master Beneficiary Summary File (MBSF) maintained by the Centers for Medicare & Medicaid Services.

who entered Medicare Advantage with the increased inflow of enrollees during this period were likely sicker than prior cohorts of Medicare Advantage enrollees.

While the incentives for Medicare Advantage plans to boost quality are already weakening over this period due to the decline in the competitive threat of traditional Medicare, this shift in the health status (and cost) of the marginal Medicare Advantage enrollee could weaken incentives of Medicare Advantage plans to boost quality even further. It is important to remember, however, that the extent to which these enrollees are attractive to the Medicare Advantage plans is not based on unconditional health status but instead on how Medicare Advantage plans are compensated for beneficiary risk, which is captured by risk scores and compensated by risk adjustment payments. These payments likely offset some portion of changes in risk, making the overall effects of these shifts in selection on plan incentives unclear.

So overall, traditional Medicare has weakened substantially as a competitor for Medicare Advantage. Why? The decline in the attractiveness of traditional Medicare cannot be traced to changes in that program, because traditional Medicare has not, in fact, changed much in the last two decades. Instead, we suggest four potential

hypotheses to explain the shift: (1) more money has flowed to Medicare Advantage plans, making the program attractive to marginal enrollees; (2) employers have decreased offerings of supplemental coverage for retirees, making traditional Medicare less attractive; (3) Medicare Advantage has been more attractive for some time, but because of inertia, the shift in enrollment is occurring only gradually over time; and (4) an increase in competition within Medicare Advantage over time has led these plans to compete more vigorously for enrollees by boosting quality. We discuss the first three hypotheses in turn here in this section. In the next main section, we separately investigate the question of competition within Medicare Advantage.

Higher Payments to Medicare Advantage

One potential explanation for the decline of traditional Medicare and the rise of Medicare Advantage could be that payments to Medicare Advantage have grown over time. As discussed earlier, several factors determine payments actually received by Medicare Advantage plans: the county-level benchmarks set by Congress, the bids that plans make relative to those benchmarks, the risk-adjustments applied to payments, and whether "quality bonuses" are applied to the benchmark itself and to the size of the rebate payment.

The Centers for Medicare & Medicaid Services collects data on the level of benchmarks based on county-level spending in traditional Medicare, adjusted for risk, geography, and quality.[4] From the early 1990s to around 2010, the monthly benchmarks were gradually increasing (in 1990 inflation-adjusted dollars) from about $260 to $380. However, the Patient Protection and Affordable Care Act of 2010 stopped those increases and, indeed, led to substantial *cuts* in benchmark payments. Benchmarks in 2015 were almost 25 percent lower in 2015 than they were at their peak in 2009. However, benchmarks then started growing steadily again, from about $280 in 2015 to $320 by 2025. (If quality bonuses for "five-star" plans are taken into account, the levels are slightly higher, but the change is much the same.) The timing of this rise in benchmark rates corresponds to the uptick in the Medicare Advantage growth rate and the shifts in selection. However, it clearly does not explain the earlier growth of Medicare Advantage between 2007 and 2015, as benchmarks were declining during that period.

Changes in risk scores for Medicare Advantage enrollees could also affect payments received. As discussed earlier, a large body of evidence documents significant "upcoding," where Medicare Advantage beneficiaries receive higher risk scores than they would have had they enrolled in traditional Medicare. Estimates of the "MA coding multiplier" range from 6 percent to closer to 20 percent (Geruso and Layton 2020; Jacobs and Layton 2025). Recent work by Jacobs and Layton (2025), confirmed by MedPAC, provides evidence that this multiplier has been increasing over time.

[4]The benchmark data from Centers for Medicare & Medicaid Services are available at https://www.cms.gov/medicare/payment/medicare-advantage-rates-statistics.

Ultimately, a key question is to calculate changes in the payment to Medicare Advantage plans *relative to the traditional Medicare cost.* The Medicare Payment Advisory Commission (MedPAC 2025, Figure 11-3) combined a variety of methods to address this question. In particular, the report performed a decomposition of the overall payment-cost gap into selection and coding components. The total "overpayments" to Medicare Advantage hit their low point around 2015, mostly due to declines in the payments *before* selection and coding. As discussed earlier, this is the time at which benchmarks started to increase. However, increases in benchmark levels since about 2015 seem to have been offset by decreases in bids by Medicare Advantage plans. But MedPAC still estimates that overpayments increased since 2015, attributing most of that increase to increases in coding intensity (as in Jacobs and Layton 2025).

In order for higher payments to Medicare Advantage plans to translate to higher enrollment, they must result in those plans offering beneficiaries something of value. Recall that the "rebate" is a payment to Medicare Advantage plans that have bid *below* the benchmark, equal to a fixed percentage of the gap between the benchmark and the bid. By law, these rebate dollars must be spent on additional benefits for enrollees, making these dollars the primary mechanism for offering many supplemental benefits to potential enrollees.

The same report from the Medicare Payment Advisory Commission (MedPAC 2025, Figure 11-2) plots average rebates over time for conventional Medicare Advantage plans. The average monthly rebate per enrollee is about $90 from 2010 to 2015, but then rises quickly, topping out at $196 in 2023 before levelling out. This suggests that the Medicare Advantage plans significantly improved their offerings to beneficiaries during this period. Interestingly, the timing of this increase in rebates again corresponds closely with the higher overpayments to Medicare Advantage plans, the faster rate of growth in Medicare Advantage, and the changes in switching between Medicare Advantage and traditional Medicare (as well as shifts in selection). MedPAC reports that in 2025 around 40 percent of rebate dollars are earmarked for reducing beneficiary cost-sharing and around 30 percent is going to supplemental benefits like vision, dental, and hearing. The remainder goes toward Part D and Part B premiums and beefing up Part D drug coverage.

Ultimately, the hypothesis that Medicare Advantage growth is coming from higher payments to Medicare Advantage plans (relative to costs), thus allowing these plans to provide more benefits to entice Medicare beneficiaries away from traditional Medicare and into Medicare Advantage, seems both compelling and incomplete. While increases in payments around 2015 correspond with an acceleration of Medicare Advantage growth, Medicare Advantage also continued to grow through the payment cuts of 2010 to 2015. Further, the timing of increased overpayments, rebates, and enrollment all match up nicely, but overpayments seem to have been just as high in earlier periods (throughout the 2000s) while rebates and enrollment were much lower. So, while evidence does point in the direction of payment to Medicare Advantage being an important factor in the decline in

traditional Medicare, more work is needed to further validate the connection between payments to Medicare Advantage plans, enrollment in these plans, and other potential factors that we discuss here and in the next main section.

The Demise of Retiree Supplemental Coverage

In the past, many beneficiaries had some form of supplemental coverage ("Medigap" policies) subsidized by former employers as part of their retirement package. However, the share of retirees receiving subsidized supplemental coverage from their former employers has been falling: In the 1990s, around 40 percent of firms were offering retiree health benefits; by the mid-2000s, around 30 percent; and by 2023, 21 percent (Claxton et al. 2023).[5] There is also some indication that the generosity of employer-provided retiree plans has declined, with caps on employer liability and more shifting of costs onto retirees (Freed et al. 2024c).

The *form* of employer-provided retirement health care coverage has also shifted. In 2024, more than half of large employers that offered retiree health benefits did so through an employer-sponsored Medicare Advantage plan, known as an Employer Group Waiver Plan (EGWP). This is up from around 25 percent of retiree-benefit-offering employers just seven years earlier in 2017. In 2024, employer or union-sponsored Medicare Advantage plans accounted for around 17 percent of total enrollment (Freed et al. 2024b). This share has remained fairly constant since 2010 (Freed et al. 2024c).

Thus, the decline in employer-subsidized supplemental coverage and the rise in employers offering Medicare Advantage plans to retired workers in recent years are probably parts of the story for why traditional Medicare enrollment has fallen so much over time. However, more research is needed to establish the magnitude of any causal connection between these trends and to make firmer conclusions about the role of retiree supplemental coverage in the decline of traditional Medicare.

Inertia

It is possible that the constant growth of Medicare Advantage is not due to persistent change in the attractiveness of Medicare Advantage versus traditional Medicare. Perhaps the generosity of Medicare Advantage has long exceeded that of traditional Medicare by a margin sufficient to attract a high share of Medicare beneficiaries, but because beneficiaries are highly inertial in their enrollment decisions, corresponding increases in Medicare Advantage market share have been unfolding more gradually over time.

[5] KFF uses a variety of different sources to calculate these trends in offers, with sources changing over time. These shifts in sources may result in measures being inconsistent over time. However, they used a single source from 1999 to 2017, and that source indicated 40 percent of employers offering benefits in 1999 and around 25 percent offering benefits in 2017, indicating that the overall downward trend is likely real.

A straightforward test for this hypothesis is to examine the choices of *new* Medicare beneficiaries. After all, new enrollees make an active choice of either traditional Medicare or Medicare Advantage and are thus not subject to inertia. If the sustained growth of Medicare Advantage over time is due to a large increase in the attractiveness of Medicare Advantage in the mid-2000s that has just taken a long time to play out, we would expect the share of new enrollees to have risen many years ago and remained high and relatively stable over the last 20 or so years.

Figure 1 at the start of the paper plots the Medicare Advantage share both among all beneficiaries and for 65-year-old new beneficiaries from 2007 to 2022. The share among new enrollees has actually been *lower* than the overall Medicare Advantage share since about 2009, with the gap increasing every year. Thus, much of the overall growth in Medicare Advantage is not coming from new beneficiaries making active choices, but instead from incumbent beneficiaries overcoming inertia and switching into Medicare Advantage.

At first glance, these patterns do not seem consistent with the hypothesis described above. However, perhaps both traditional Medicare and Medicare Advantage are "experience goods"—that is, it takes a few years of enrollment to be fully informed about the relative quality of the two options. In this setting, one could tell a story that, although Medicare Advantage got really great in the mid-2000s, most beneficiaries still treated traditional Medicare as a default choice. But after a few years of enrollment in traditional Medicare, some percentage are hit with an attention shock big enough to overcome inertia and make the switch. While such a story is plausible, it is difficult to test empirically.

Summary

In summary, traditional Medicare has clearly declined as a competitive threat to Medicare Advantage. We believe there is compelling evidence that the decline in market share and competitiveness of traditional Medicare can be in part explained by rising payments to Medicare Advantage. However, we readily admit that payments are an incomplete explanation, as Medicare Advantage also grew during periods in which payments declined. The decline in employer-provided supplemental coverage likely plays a role, but more work is needed to understand more definitively why traditional Medicare is just not an attractive option for millions of Medicare beneficiaries today. In the next main section, we explore one remaining explanation: the possibility that more competition among Medicare Advantage plans has made Medicare Advantage plans better and a more attractive option for beneficiaries.

Finally, we acknowledge that despite providing clear evidence that traditional Medicare has declined as a competitor to Medicare Advantage, we know little about how this decline in the importance of traditional Medicare has affected Medicare Advantage. The competitive dynamics of this market suggest that a weak traditional Medicare can weaken pressure for Medicare Advantage plans to offer generous, high-quality products, but we have little evidence of whether this is true in practice. Future work might seek to understand the causal linkage between the strength of traditional Medicare and the level of quality among Medicare Advantage plans

in order to provide evidence of the effectiveness of this tool for policymakers to improve the value Medicare Advantage offers to its enrollees.

Competing with Each Other? The Evolution of Market Structure Within Medicare Advantage

Motivated by our discussion above about the theoretical dynamics of competition in Medicare, we next focus on competition across Medicare Advantage insurers and how market structure within Medicare Advantage has evolved over the last two decades.

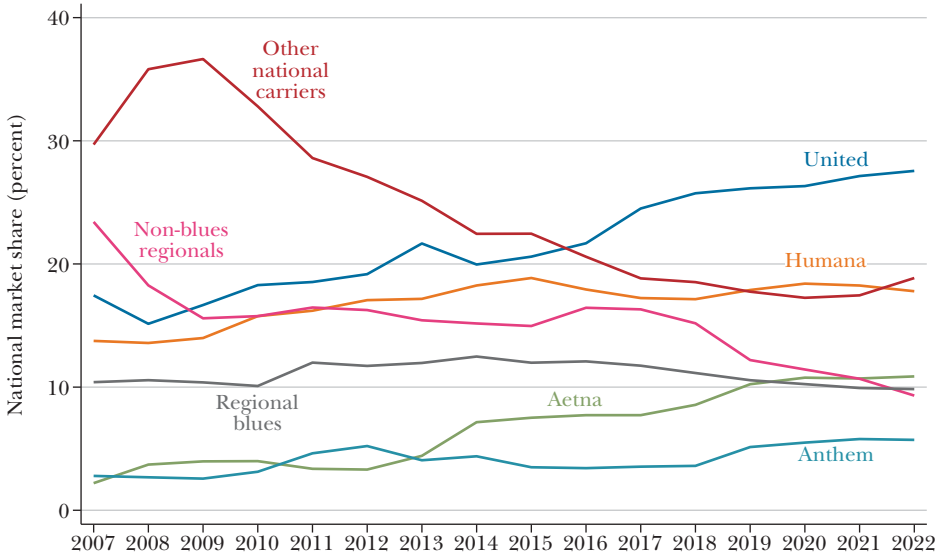
As Medicare Advantage has grown, new parent organizations have begun offering plans. However, while organizations that entered after 2007 now offer plans in over half of US counties, they only account for about 4.4 percent of 2022 Medicare Advantage enrollment. Instead of being characterized by the dynamism of expanding upstart entrants to this market, the combined market share of United, Humana, CVS/Aetna, and Blue Cross–affiliated plans (including local plans affiliated with the Blue Cross parent organization and national plans offered by Anthem/Health Care Service Corporation) has *grown* to over 70 percent in recent years. Further, other national carriers (like Kaiser and Centene) have declined in market share and sometimes exited the market altogether (for example, Cigna recently sold its Medicare Advantage business to Health Care Service Corporation). A similar pattern has played out with single-state carriers not affiliated with Blue Cross Blue Shield, whose cumulative market share has fallen from 24 percent in 2007 to less than 10 percent in 2022. Figure 4 provides a broad overview of these trends.

Although national-level statistics help visualize overall trends, Medicare Advantage markets are essentially local. Plan sponsors choose the counties in which they operate and set premiums and benefits at the county level. Most Medicare Advantage insurers only offer plans in a small fraction of counties, and even the largest national carriers tend to be dominant in some areas and relatively minor players in others. In 2022, about 300 carriers offered Medicare Advantage plans nationally, but the median Medicare Advantage beneficiary could only choose between seven different carriers in their county.

Figure 5 plots the distribution of county-level concentration in Medicare Advantage, as measured by the Herfindahl–Hirschman Index (HHI), in each year from 2007 to 2022. The HHI is defined as the sum of squared market shares (in percentage terms) of all firms in the market, and ranges from 0 to 10,000. Under current antitrust guidelines, markets with an HHI above 1,800 are considered highly concentrated, while those with an HHI between 1,000 and 1,800 are considered moderately concentrated. The horizontal lines in Figure 5 indicate these two thresholds.

Two main takeaways emerge from the figure. First, at any point in time, the vast majority of Medicare Advantage enrollees live in counties whose markets would be considered *highly* concentrated. Second, the current landscape is less concentrated

Figure 4
Market Shares of Specific Plan Sponsors in Medicare Advantage from 2007 to 2022



Source: Layton, Maini, and McWilliams (2026).
Note: This figure shows the fraction of Medicare Advantage enrollees (that is, the national market share) of select Medicare Advantage carriers from 2007 to 2022. Anthem includes all national Blue Cross Blue Shield–affiliated plans, including those sponsored by Elevance and Health Care Service Corporation. Aetna includes CVS plans. Regional carriers are defined as carriers offering plans in a single state, while national carriers are defined as those offering plans in multiple states. Only plans with at least eleven enrollees in a given county are considered part of the offerings in that county. Market shares are based on data in the Master Beneficiary Summary File (MBSF), an annual database maintained by the Centers for Medicare & Medicaid Services.

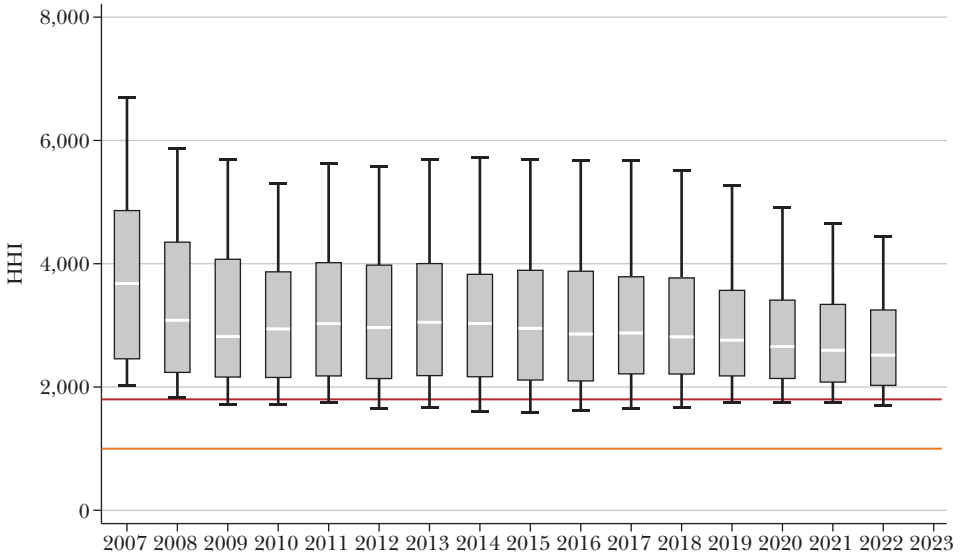
than in the late 2000s, especially in counties that had previously been highly concentrated. These results are consistent with Zahn (2025), who finds that Medicare Advantage carriers strategically avoid entering more competitive markets.

How can we reconcile this decline in local Herfindahl–Hirschman Index with the increasing national dominance of the largest insurers from Figure 4? The main force behind these county-level changes in HHI has actually been the expansion of the four largest national insurers into additional markets: United, Humana, Aetna, and Anthem. Figure 6 shows how the proportion of US counties where these national insurers operate has evolved over time. In 2007, half of all US counties had at most one of these insurers. In 2022, at least three of these insurers are present in over 80 percent of counties (and all four operate in more than half of them).[6]

[6]In more detailed analysis, we find that that counties with larger increases in Medicare Advantage enrollment also experienced larger declines in HHI concentration between 2007 and 2022, although the high-growth counties were also likely to be small in terms of population It is possible that growing

Figure 5

Boxplot of Herfindahl–Hirschman Indexes at the County-Year Level from 2007–2022

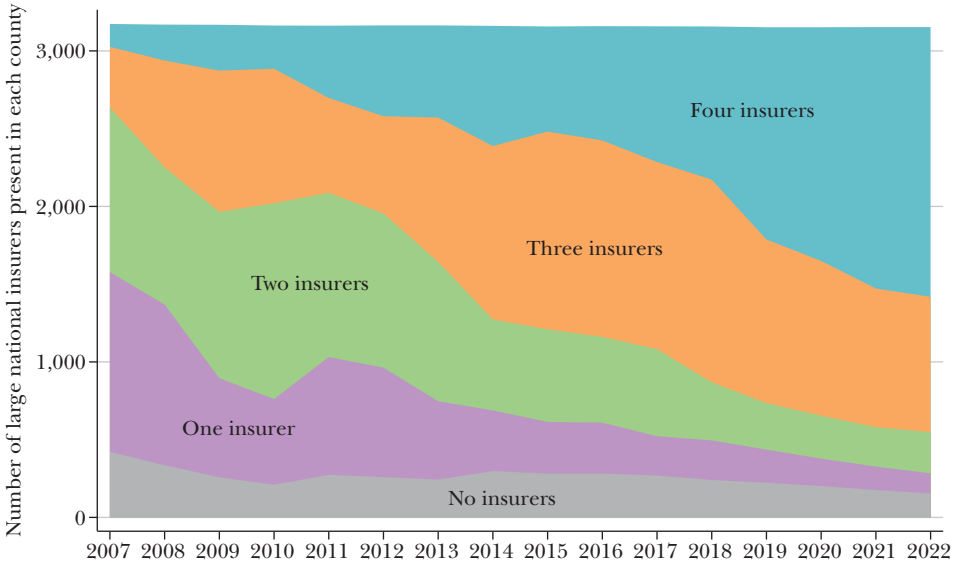


Source: Layton, Maini, and McWilliams (2026).
Note: This figure shows the distribution of county-level Herfindahl–Hirschman Index (HHI) measure of concentration in Medicare Advantage from 2007 to 2022. The white line indicates the median, and the grey boxes and confidence interval lines span the 25th–75th percentile and the 5th–95th percentile of the distribution, respectively. The orange and red lines indicate, respectively, the thresholds for moderately concentrated (1,000) and highly concentrated (1,800) markets according to the current guidelines from the Department of Justice and the Federal Trade Commission. HHIs are calculated using data from the Master Beneficiary Summary File (MBSF), an annual database maintained by the Centers for Medicare & Medicaid Services.

We conclude that county-level competition in Medicare Advantage has increased over time. Most of this competition has come in the form of competition among large, established insurers rather than new entrants to this market. Importantly, this increase in competition *within* Medicare Advantage may have offset some of the decline in competition from traditional Medicare, ensuring that Medicare Advantage plans still have incentives to deliver value to Medicare beneficiaries.

enrollment makes local markets more attractive for new or expanding sponsors; it is also possible that increasing competition leads Medicare Advantage insurers to offer better plans that are more attractive to a larger share of consumers. Also, the declines in HHI have been large and widespread in some states like Michigan, Nevada, and Kentucky, but more uneven in other states like Texas, South Dakota, and Vermont. Coastal and urban markets are relatively competitive Medicare Advantage markets, but interior and rural markets still lagging significantly behind. For details and illustrative figures, see the Supplemental Appendix.

Figure 6
Change in the Number of Largest National Insurers Across US Counties, 2007 to 2022



Source: Layton, Maini, and McWilliams (2026).
Note: This figure shows the penetration of the four largest national insurers (Aetna, Anthem, Humana, and United) at the county level over time, from 2007 to 2022. Counties are categorized by the number of insurers present. Insurers are considered present in a county if they offer at least one plan with more than ten enrollees. Market shares are calculated using data from the Master Beneficiary Summary File (MBSF), an annual database maintained by the Centers for Medicare & Medicaid Services.

How Does Increased Medicare Advantage Competition Affect Firm Behavior?

Our discussion above suggests that increased competition among Medicare Advantage insurers should put upward pressure on quality. However, this inference depends critically on the ability of consumers to observe quality, and selection could flip this result. We thus assess the evidence on whether competition within Medicare Advantage does indeed generate value for enrollees in practice.

A growing empirical literature examines how Medicare Advantage plan behavior varies with local market structure, generally finding that competition is associated with more generous coverage but that pass-through of government payments to plans has been partial, in ways consistent with the presence of insurer market power (Cabral, Geruso, and Mahoney 2018). For example, Pelech (2018) studies a reduction in Medicare Advantage competition caused by the exogenous exit of private fee-for-service insurers and finds that generosity of the remaining plans declines when competition falls, but not in the most competitive markets, suggesting that benefit generosity is particularly sensitive to changes in competition where initial competition is modest. In a study of how Medicare Advantage

plans responded to cuts in benchmark payments enacted in the Patient Protection and Affordable Care Act of 2010, Pelech and Song (2025) find that plans only partially passed payment changes through to beneficiaries—adjusting bids and less-salient benefits like cost-sharing more than premiums—indicating insurer market power. However, they also found that plans in more competitive markets changed benefits less, consistent with insurers operating closer to marginal cost when they face more competition.

Descriptive work also points to a relationship between competition and quality. Chao (2021) finds that plans in less competitive Medicare Advantage markets are less likely to offer benefits such as transportation, dental, vision, and hearing, and that they have lower per-member-per-month spending on these benefits. Ianni, McWilliams, and Curto (2025) find that plans offer new supplemental benefits to attract high-margin groups of enrollees only in less concentrated markets. Finally, Adrion (2019) finds that plans in more concentrated markets tend to have higher "star" ratings but also higher premiums.

Our own comparisons of within-geography changes in rebates (a measure of plan generosity) versus changes in market concentration indicate a positive correlation: a 1000 point decline in the Herfindahl–Hirschman Index measure of concentration is associated with a rebate increase of approximately $5.[7] The bulk of the evidence thus indicates that there is indeed a link between competition among Medicare Advantage plans and quality.

Are There Significant Barriers to Competition in Medicare Advantage?

While the recent increases in competition within Medicare Advantage are encouraging, growing evidence confirms that these insurers still possess substantial market power. Existing work in this area focuses on the overall pass-through of government subsidies to Medicare beneficiaries (rather than the correlation between pass-through and competition explored above) and falls into two main groups.

The first group uses reduced-form variation in the subsidies that insurers receive. For example, Cabral, Geruso, and Mahoney (2018) study a 2000 reform that increased Medicare Advantage subsidies and found that insurers retained about half of the additional payments, rather than fully passing them through to beneficiaries in the form of lower premiums and increased benefits. Despite the substantial changes in Medicare Advantage policy since then, follow-up studies using more recent data find strikingly similar patterns (for example, Song, Landrum, and Chernew 2013; Pelech and Song 2025). The second group of studies consists of structural models of competition tailored to the Medicare Advantage setting. For instance, Curto et al. (2021) estimate a structural model of plan bidding and conclude that Medicare Advantage insurers capture roughly two-thirds of the surplus generated by the program.

[7]For details of this analysis, see the Supplemental Appendix.

The fact that there has been an increase in competition over the last two decades is not particularly reassuring here, given that this increase in competition came mostly from expansion of large established insurers into new markets rather than entry by new competitors. Expansion by the large established insurers can definitely increase competitive pressure in a market, but this type of increase in competition is limited by the fact that there are only a few of these insurers and there remain few markets where all of them have not entered. It is thus important to ask what sorts of barriers there might be to entry from new insurers in this market. We suggest four candidates.

First, inertia in plan choice can allow incumbents to charge higher prices (or to offer less generous benefits), while also dampening incentives for entry. Although relatively few papers study inertia in Medicare Advantage directly, observed plan pricing and design are consistent with the presence of substantial inertia (Miller 2024). Inertia has also been extensively documented in similar settings, including the drug insurance in Medicare Part D (for example, Ericson 2014; Polyakova 2016; Heiss et al. 2021), and the individual market (Saltzman, Swanson, and Polsky 2025). Separately, insurers that are new to Medicare Advantage do not have a "star" rating, which may further limit take-up of new plans and discourage their entry, given that higher stars are linked with enrollment (Reid et al. 2013). Enrollment shares of new Medicare Advantage insurers remain small even several years after entry, consistent with an advantage for incumbents. On the other hand, in our data, we do not find that dominant insurers have markedly higher penetration among existing Medicare Advantage enrollees than among new enrollees: that is, the shares of Humana, United, and CVS/Aetna are comparable among new and existing enrollees. This suggests that inertia, while clearly present, may not be the only—or even the primary—force behind the dominance of these insurers.

Second, economies of scale, and particularly, costs of entry tied to building a strong provider network with a broad set of hospitals, physician groups, and other providers, could create a barrier to entry. Large national insurers, particularly those with a strong presence in the employer-sponsored insurance segment (such as United and Aetna), may have an advantage because they can leverage existing provider relationships and contracting infrastructure. Moreover, integrated firms with a large presence in commercial or employer-sponsored insurance may find it easier to transition their members into Medicare Advantage at age 65, giving them a built-in base of enrollees that new entrants lack.

Economies of scope, such as those that may stem from vertical integration, may also be important. Recent estimates suggest that United (the largest insurer in Medicare Advantage) now owns physician practices employing more than 90,000 physicians, or close to 10 percent of the entire US physician workforce (Arnold and Fulton 2025). Recent work also indicates that they pay significantly higher rates to the providers they own versus providers that they do not own. On one side, the benefits of vertical integration (elimination of double marginalization and other efficiencies) could be passed along to consumers in the form of

lower costs or improved coordination of benefits. On the other side, in the limited but suggestive evidence on this point, Frakt, Pizer, and Feldman (2013) find that Medicare Advantage plans integrated with healthcare providers tend to be more expensive than other plans of similar measured quality, which is consistent with the hypothesis that exclusive or particularly strong provider relationships confer market power and raise entry costs for potential competitors. Hnath, McWilliams, and Chernew (2024) also find that health-system-integrated plans have consistently captured between 10 and 15 percent of the Medicare Advantage market share, unlike other regional parent organizations, whose market share has declined steeply over the past two decades. When an insurer owns a hospital or provider group, that group may refuse to work with other insurers or at least charge them very high rates (Cuesta, Noton, and Vatter 2024). Such actions can make it difficult for an insurer to enter a market where many providers are affiliated with groups owned by incumbent insurers.

Third, brokers and agents may play a role in shaping competition. Virtually all Medicare Advantage plans contract with independent agents (CMS 2025). Moreover, brokers account for roughly 70 percent of new individual Medicare Advantage enrollments (McKinsey and Company 2024). While the role of middlemen and brokers has been studied extensively in other markets, there is virtually no work on the role of brokers in Medicare Advantage.[8] From a theoretical perspective, brokers in health insurance markets could lower search costs (Rubinstein and Wolinsky 1987; Gehrig 1993) by identifying the plans that are best suited for a given beneficiary's needs and act as informed monitors of plan quality (Biglaiser 1993). However, brokers and online brokerage tools generally do not represent or display all available Medicare Advantage plans: In five large counties, three national online broker tools showed on average only about 43 percent of Medicare Advantage plans available in the area, and agents, brokers, and insurers together accounted for roughly one-third of organic Medicare-related search results and 87 percent of search ads (Ali et al. 2021; Findlay, Jacobson, and Leonard 2023). Moreover, brokers are paid by insurers on a per-enrollee basis, with commission schedules that differ across coverage types and carriers and that have in recent years tended to become more favorable for Medicare Advantage referrals relative to referrals to a combination of traditional Medicare and Medigap (Ali and Hellow 2021). As such, brokers could have the effect of entrenching dominant Medicare Advantage insurers rather than steering clients to attractive new entrants. More work is needed to understand the role of brokers in Medicare Advantage.

Fourth, advertising is prominent in Medicare Advantage and related markets (Aizawa and Kim 2018; Shapiro 2020). Large national insurers with deep pockets and recognizable brands can spend heavily on television, digital, and direct-mail campaigns during the open enrollment period, which may both attract new beneficiaries and deter entry by smaller rivals that cannot match this spending. For

[8]For two recent examples of studies of intermediaries in other industries, see Salz (2022) on the trade waste industry in New York City and Biglaiser et al. (2020) on the used car market.

example, many have attributed United's large role at least in part to its advertising relationship with the American Association of Retired Persons (AARP). Moreover, insofar as advertising is a public good—benefitting all sponsors by getting seniors interested in Medicare Advantage—larger carriers are more likely to internalize its benefits, whereas smaller players who advertise may end up generating positive spillovers for everyone but themselves. However, restrictions on advertising seem unlikely to reduce barriers to entry, as new entrants would still struggle to make their products known to potential enrollees.

Conclusion

Medicare Advantage now enrolls more Medicare beneficiaries than the traditional public program. However, the ability of Medicare Advantage to deliver value—both cost savings for the government and satisfaction for enrollees—depends critically on competition.

In recent years, traditional Medicare has clearly declined as a competitor to Medicare Advantage. This decline has made competition within Medicare Advantage more important, and fortunately there is evidence that this competition has strengthened in recent years, especially among large established players. However, we are quickly reaching the limit of how much competition can improve via this route—entry from new insurers has been limited, and the large players still appear to have significant market power. Thus, either a renewal of competition from traditional Medicare or new entry by insurers into Medicare Advantage may be necessary in the future in order to ensure that the Medicare program as a whole delivers more value to beneficiaries for a given level of spending.

Looking ahead, there is substantial room for research on the sources and consequences of limited competition in Medicare Advantage. More work is needed to understand why enrollment in traditional Medicare has declined so much and how competition from traditional Medicare influences what Medicare Advantage plans offer. Within Medicare Advantage, on the supply side we know relatively little about how costly it is for new or smaller sponsors to build provider networks, how often exclusive or preferred network arrangements deter entry, and how broker compensation and marketing support shape which plans beneficiaries actually see and choose. On the demand side, more work is needed to quantify the extent to which inertia, inattention, quality ratings, advertising, and brand recognition interact to generate barriers to entry. Such work would help clarify when and where additional policy attention to competition in Medicare Advantage is most warranted.

■ *We thank the editors, Timothy Taylor, Heidi Williams, Jeff Kling, and Jonathan Parker, for helpful comments on this paper. We acknowledge financial support from the National Institute on Aging of the National Institutes of Health (P01AG032952).*

References

Abaluck, Jason, Mauricio Caceres Bravo, Peter Hull, and Amanda Starc. 2021. "Mortality Effects and Choice across Private Health Insurance Plans." *Quarterly Journal of Economics* 136 (3): 1557–610.

Adrion, Emily R. 2019. "Competition and Health Plan Quality in the Medicare Advantage Market." *Health Services Research* 54 (5): 1126–36.

Aizawa, Naoki, and You Suk Kim. 2018. "Advertising and Risk Selection in Health Insurance Markets." *American Economic Review* 108 (3): 828–67.

Ali, Riaz, Aimee Cicchiello, Morgan Hanger, Lesley Hellow, Ken Williams, and Gretchen Jacobson. 2021. "How Agents and Brokers Influence Medicare Beneficiaries' Plan Choices." Commonwealth Fund Report, April 21. https://www.commonwealthfund.org/publications/fund-reports/2021/apr/how-agents-influence-medicare-beneficiaries-plan-choices.

Ali, Riaz, and Lesley Hellow. 2021. "Agent Commissions in Medicare and the Impact on Beneficiary Choice." Commonwealth Fund Blog, October 12. https://www.commonwealthfund.org/blog/2021/agent-commissions-medicare-and-impact-beneficiary-choice.

Arnold, Daniel R., and Brent D. Fulton. 2025. "UnitedHealthcare Pays Optum Providers More Than Non-Optum Providers." *Health Affairs* 44 (11).

Biglaiser, Gary. 1993. "Middlemen as Experts." *RAND Journal of Economics* 24 (2): 212–23.

Biglaiser, Gary, Fei Li, Charles Murry, and Yiyi Zhou. 2020. "Intermediaries and Product Quality in Used Car Markets." *RAND Journal of Economics* 51 (3): 905–33.

Brown, Jason, Mark Duggan, Ilyana Kuziemko, and William Woolston. 2014. "How Does Risk Selection Respond to Risk Adjustment? Evidence from Medicare Advantage." *American Economic Review* 104 (10): 3335–64.

Cabral, Marika, Michael Geruso, and Neale Mahoney. 2018. "Do Larger Health Insurance Subsidies Benefit Patients or Producers? Evidence from Medicare Advantage." *American Economic Review* 108 (8): 2048–87.

Chao, Sandra. 2021. "Competition and Consolidation in Medicare Advantage." PhD diss., University of Maryland.

Claxton, Gary, Matthew Rae, Aubrey Winger, Emma Wager, Jason Kerns, Greg Shmavonian, and Anthony Damico. 2023. *Employer Health Benefits, 2023 Annual Survey.* KFF.

CMS (Centers for Medicare and Medicaid Services). 2025. "Agent Broker Compensation." https://www.cms.gov/medicare/health-drug-plans/managed-care-marketing/medicare-marketing-guidelines/agent-broker-compensation (accessed December 2025).

Cottrill, Alex, and Nancy Ochieng, and Tricia Neuman. 2025. "How Many Physicians Have Opted Out of the Medicare Program?" KFF Medicare Issue Brief, January 17. https://www.kff.org/medicare/how-many-physicians-have-opted-out-of-the-medicare-program/.

Cubanski, Juliette, Meredith Freed, Nancy Ochieng, Alex Cottrill, Jeannie Fuglesten Biniek, and Tricia Neuman. 2025. "Medicare 101." KFF, October 8.

Cuesta, José Ignacio, Carlos E. Noton, and Benjamin Vatter. 2024. "Vertical Integration and Plan Design in Healthcare Markets." NBER Working Paper 32833.

Curto, Vilsa, Liran Einav, Jonathan Levin, and Jay Bhattacharya. 2021. "Can Health Insurance Competition Work? Evidence from Medicare Advantage." *Journal of Political Economy* 129 (2): 359–412.

Dafny, Leemore, Mark Duggan, and Subramaniam Ramanarayanan. 2012. "Paying a Premium on Your Premium? Consolidation in the US Health Insurance Industry." *American Economic Review* 102 (2): 1161–85.

Dafny, Leemore, Jonathan Gruber, and Christopher Ody. 2015. "More Insurers Lower Premiums: Evidence from Initial Pricing in the Health Insurance Marketplaces." *American Journal of Health Economics* 1 (1): 53–81.

Emerson, Jakob. 2024. "Insurers Pull Medicare Advantage Plans as Profit Pressures Mount." Becker's Payer Issues. https://www.beckerspayer.com/payer/medicare-advantage/insurers-pull-medicare-advantage-plans-as-profit-pressures-mount/

Ericson, Keith M. Marzilli. 2014. "Consumer Inertia and Firm Pricing in the Medicare Part D Prescription Drug Insurance Market." *American Economic Journal: Economic Policy* 6 (1): 38–64.

Findlay, Steven, Gretchen Jacobson, and Faith Leonard. 2023. "The Role of Marketing in Medicare Beneficiaries' Coverage Choices." Commonwealth Fund Explainer, January 5. https://www.commonwealthfund.org/publications/explainer/2023/jan/role-marketing-medicare-beneficiaries-coverage-choices.

Frakt, Austin B., Steven D. Pizer, and Roger Feldman. 2013. "Plan-Provider Integration, Premiums, and Quality in the Medicare Advantage Market." *Health Services Research* 48 (6pt1): 1996–2013.

Freed, Meredith, Jeannie Fuglesten Biniek, Anthony Damico, and Tricia Neuman. 2024a. *Medicare Advantage 2024: Premiums, Out-of-Pocket Limits, Supplemental Benefits, and Prior Authorization*. KFF.

Freed, Meredith, Jeannie Fuglesten Biniek, Anthony Damico, and Tricia Neuman. 2024b. "Medicare Advantage in 2024: Enrollment Update and Key Trends." KFF, August 8. https://www.kff.org/medicare/medicare-advantage-in-2024-enrollment-update-and-key-trends/.

Freed, Meredith, Tricia Neuman, Matthew Rae, and Jeannie Fuglesten Biniek. 2024c. "Medicare Advantage Has Become More Popular among the Shrinking Share of Employers That Offer Retiree Health Benefits." KFF, November 18. https://www.kff.org/medicare/medicare-advantage-has-become-more-popular-among-the-shrinking-share-of-employers-that-offer-retiree-health-benefits/.

Gehrig, Thomas. 1993. "Intermediation in Search Markets." *Journal of Economics and Management Strategy* 2 (1): 97–120.

Geruso, Michael, and Timothy Layton. 2020. "Upcoding: Evidence from Medicare on Squishy Risk Adjustment." *Journal of Political Economy* 128 (3): 984–1026.

Heiss, Florian, Daniel McFadden, Joachim Winter, Amelie Wuppermann, and Bo Zhou. 2021. "Inattention and Switching Costs as Sources of Inertia in Medicare Part D." *American Economic Review* 111 (9): 2737–81.

Hnath, Joseph G. P., J. Michael McWilliams, and Michael E. Chernew. 2024. "Medicare Advantage: National Carriers Expand Market Share While Regional Carriers without Affiliation Decline, 2012–23." *Health Affairs* 43 (12): 1647–54.

Ho, Kate, and Robin S. Lee. 2017. "Insurer Competition in Health Care Markets." *Econometrica* 85 (2): 379–417.

Ianni, Katherine M., J. Michael McWilliams, and Vilsa E. Curto. 2025. "Do MA Plans Follow Incentives to Attract Enrollee Groups with New Supplemental Benefits?" *Health Affairs Scholar* 3 (9): qxaf153.

Jacobs, Paul D., and Timothy J. Layton. 2025. "Identifying Coding Intensity in Medicare Advantage through Switchers." *Health Services Research* 60 (5): e14628.

Kirzinger, Ashley, Julian Montalvo III, Audrey Kearney, Grace Sparks, Isabelle Valdes, and Liz Hamel. 2025. "KFF Health Tracking Poll: Public Weighs Health Care Spending and Other Priorities for Incoming Administration." KFF, January 17. https://www.kff.org/health-costs/kff-health-tracking-poll-public-weighs-health-care-spending-and-other-priorities-for-incoming-administration/.

Kreider, Amanda R., Timothy J. Layton, Mark Shepard, and Jacob Wallace. 2024. "Adverse Selection and Network Design under Regulated Plan Prices: Evidence from Medicaid." *Journal of Health Economics* 97: 102901.

Kronick, Richard. 2017. "Projected Coding Intensity in Medicare Advantage Could Increase Medicare Spending by $200 Billion over Ten Years." *Health Affairs* 36 (2): 320–27.

Layton, Timothy J., Randall P. Ellis, Thomas G. McGuire, and Richard van Kleef. 2017. "Measuring Efficiency of Health Plan Payment Systems in Managed Competition Health Insurance Markets." *Journal of Health Economics* 56: 237–55.

Layton, Tim, Luca Maini, and J. Michael McWilliams. 2026. Data and Code for: "Substitutes for Success? Public versus Private Competition in Medicare Advantage." Nashville, TN: American Economic Association; distributed by Inter-university Consortium for Political and Social Research, Ann Arbor, MI. https://doi.org/10.3886/E247010V1.

Lin, Haizhen, and Ian M. McCarthy. 2023. "Multimarket Contact in Health Insurance: Evidence from Medicare Advantage." *Journal of Industrial Economics* 71 (1): 212–55.

McGuire, Thomas G., and Jacob Glazer. 2000. "Optimal Risk Adjustment in Markets with Adverse Selection: An Application to Managed Care." *American Economic Review* 90 (4): 1055–71.

McKinsey and Company. 2024. "The Future of Medicare Advantage." McKinsey Healthcare Practice, December 12. https://www.mckinsey.com/industries/healthcare/our-insights/the-future-of-medicare-advantage.

McWilliams, J. Michael. 2024. "The Future of Medicare and the Role of Traditional Medicare as Competitor." *New England Journal of Medicine* 391 (8): 763–69.

McWilliams, J. Michael, Christopher C. Afendulis, Thomas G. McGuire, and Bruce E. Landon. 2011. "Complex Medicare Advantage Choices May Overwhelm Seniors—Especially Those with Impaired Decision Making." *Health Affairs* 30 (9): 1786–94.

MedPAC (Medicare Payment Advisory Commission). 2025. *March 2025 Report to the Congress: Medicare Payment Policy*. MedPAC.

Miller, Keaton. 2024. "Capturing and Harvesting in Medicare Advantage Plan Design." *Health Affairs Scholar* 2 (6): qxae077.

Newhouse, Joseph P., Mary Price, Jie Huang, J. Michael McWilliams, and John Hsu. 2012. "Steps to Reduce Favorable Risk Selection in Medicare Advantage Largely Succeeded, Boding Well for Health Insurance Exchanges." *Health Affairs* 31 (12): 2618–28.

Patel, Yash M., and Stuart Guterman. 2017. "The Evolution of Private Plans in Medicare." Commonwealth Fund Issue Brief, December 8. https://www.commonwealthfund.org/publications/issue-briefs/2017/dec/evolution-private-plans-medicare.

Pelech, Daria. 2018. "Paying More for Less? Insurer Competition and Health Plan Generosity in the Medicare Advantage Program." *Journal of Health Economics* 61: 77–92.

Pelech, Daria M. 2020. "Prices for Physicians' Services in Medicare Advantage and Commercial Plans." *Medical Care Research and Review* 77 (3): 236–48.

Pelech, Daria, and Zirui Song. 2025. "Private Medicare Plans' Responses to Benchmark Changes and Competition before and after the Affordable Care Act's Payment Cuts." *Health Services Research* 60 (2): e14392.

Polyakova, Maria. 2016. "Regulation of Insurance with Adverse Selection and Switching Costs: Evidence from Medicare Part D." *American Economic Journal: Applied Economics* 8 (3): 165–95.

Reid, Robert O., Partha Deb, Benjamin L. Howell, and William H. Shrank. 2013. "Association between Medicare Advantage Plan Star Ratings and Beneficiary Enrollment." *JAMA* 309 (3): 267–74.

Research Data Assistance Center (ResDAC). *Master Beneficiary Summary File Base A/B/C/D (MBSF).* Minneapolis, MN: ResDAC (accessed March 19, 2026). Available by DUA.

Roberts, Eric T., Brian E. McGarry, and Alexandra Glynn. 2020. "Cognition and Take-up of the Medicare Savings Programs." *JAMA Internal Medicine* 180 (11): 1529–31.

Rothschild, Michael, and Joseph Stiglitz. 1976. "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information." *Quarterly Journal of Economics* 90 (4): 629–49.

Rubinstein, Ariel, and Asher Wolinsky. 1987. "Middlemen." *Quarterly Journal of Economics* 102 (3): 581–93.

Saltzman, Evan, Ashley Swanson, and Daniel Polsky. 2025. "Inertia, Market Power, and Adverse Selection in Health Insurance: Evidence from the ACA Exchanges." *Review of Economics and Statistics.* https://doi.org/10.1162/REST.a.1615.

Salz, Tobias. 2022. "Intermediation and Competition in Search Markets: An Empirical Case Study." *Journal of Political Economy* 130 (2): 310–45.

Shapiro, Bradley T. 2020. "Advertising in Health Insurance Markets." *Marketing Science* 39 (3): 587–611.

Song, Zirui, Mary Beth Landrum, and Michael E. Chernew. 2013. "Competitive Bidding in Medicare Advantage: Effect of Benchmark Changes on Plan Bids." *Journal of Health Economics* 32 (6): 1301–12.

Trish, Erin, Paul Ginsburg, Laura Gascue, and Geoffrey Joyce. 2017. "Physician Reimbursement in Medicare Advantage Compared with Traditional Medicare and Commercial Health Insurance." *JAMA Internal Medicine* 177 (9): 1287–95.

Vatter, Benjamin. 2025. "Quality Disclosure and Regulation: Scoring Design in Medicare Advantage." *Econometrica* 93 (3): 959–1001.

Zahn, Matthew V. 2025. "Entry and Competition in Insurance Markets: Evidence from Medicare Advantage." Unpublished.

Journal of Economic Perspectives—Volume 40, Number 2—Spring 2026—Pages 171–194

Understanding Medicaid Managed Care: The Procured Competition Model

Mark Shepard and Jacob Wallace

Medicaid is the primary source of health insurance for low-income Americans and by far the largest social safety net program in the United States. As of 2025, it covers 77 million people—about one in five Americans—and accounts for about $900 billion in annual federal and state spending. By comparison, the next-largest means-tested program is the Supplemental Nutrition Assistance Program ("food stamps"), which in 2024 cost about $100 billion and served 42 million people per month.

Over the past several decades, Medicaid has undergone a major institutional transformation. Historically, states administered the program directly, paying doctors, hospitals, and other providers through a state-run fee-for-service system. Today, about 85 percent of beneficiaries are enrolled in Medicaid managed care, under which states contract with private insurers to provide coverage and manage care. This shift toward outsourcing has been driven by concerns about rising costs, program complexity, and the limits of states' capacity to administer insurance efficiently.

Yet despite its scale and growth, the case for outsourcing Medicaid to private insurers remains ambiguous. A large theoretical literature emphasizes that contracting out public services can generate both gains and losses when contracts are incomplete (Hart, Shleifer, and Vishny 1997). Private firms may have stronger

■ *Mark Shepard is Associate Professor of Public Policy, Harvard Kennedy School of Government, Cambridge, Massachusetts. His email address is mark_shepard@hks.harvard.edu. Jacob Wallace is Associate Professor of Public Health, Yale School of Public Health, New Haven, Connecticut. His email address is jacob.wallace@yale.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20251475.

incentives to reduce costs and innovate, but they may also cut quality in ways that are difficult for the government to observe or contract on. Consistent with these competing forces, the empirical evidence on Medicaid managed care shows mixed effects. There is some well-identified evidence that managed care reduces health care spending, but this has not translated into lower costs for taxpayers. Evidence on the effects of managed care on quality is inconclusive, ranging from credible evidence of harms (for example, increased mortality) to null effects or modest benefits. In short, Medicaid managed care has yet to deliver on its promise of lowering government spending or improving care.

We argue that understanding whether to privatize Medicaid and how best to do so requires a framework that captures how Medicaid managed care operates in practice. Rather than fitting neatly into standard models of either public provision or regulated competition, Medicaid managed care is best understood as a hybrid model of what we call "procured competition." States use a procurement process to select a menu of participating insurers and then rely on competition among those insurers to allocate enrollees and discipline performance.

This hybrid structure blends two approaches that have typically been studied separately. In standard public procurement—such as contracting for infrastructure projects—the government specifies desired services, solicits bids, selects winners, and manages contracts, often resulting in a single contractor. In regulated insurance markets, by contrast, firms generally enter freely and compete under rules governing prices, plan design, and subsidies—as in programs like the Affordable Care Act state-level Marketplaces for individual health insurance or Medicare Advantage. Medicaid managed care combines these models: states select participating insurers via a procurement process, but then allow those insurers to compete for enrollees within a tightly regulated program.

From a procurement perspective, introducing consumer choice and competition can help incentivize insurer effort by allowing beneficiaries to "vote with their feet," rewarding higher-performing plans. From a market perspective, the procurement stage allows the state to shape the set of competitors—excluding low-quality plans and potentially extracting better terms from insurers. But combining these approaches also creates a more complex system that is difficult to design and implement well, particularly in a setting where plan quality is hard to observe and insurers face incentives to select based on risk rather than to improve care.

A central theme of this paper is that we still know relatively little about how to make this hybrid model work effectively. The design of procurement processes, market rules, and enrollee choice mechanisms all play critical roles in shaping insurer incentives, yet there is limited theoretical analysis and empirical evidence on how these policies interact or how they affect costs, quality, and access. Thus, policymakers have little guidance on how to design rules for procurement and market competition. However, the enormous variation in how state-level Medicaid programs set these policies presents both opportunities and challenges for economic research.

In the remainder of the paper, we use the procured competition framework to explain how Medicaid managed care operates in practice and to highlight open questions for research and policy. We begin with background on the Medicaid program and the rise of managed care, including reviewing the mixed evidence on the effects of managed care on costs and quality. We then develop the conceptual framework of procured competition and use it to organize the key policy levers in the program, including procurement design, market rules, and plan choice provisions. Throughout, we emphasize where theory provides guidance, where evidence is informative, and where important gaps remain. Understanding Medicaid managed care is increasingly important as states continue to rely on private insurers to deliver public insurance—and as similar hybrid models are considered in other areas of social policy. Medicaid offers a valuable laboratory for studying the promise and perils of market forces in the provision of essential social goods.

Should States Contract Out? The Make-or-Buy Decision in Medicaid

Why do state governments contract out the administration of Medicaid benefits to private insurers? This section introduces Medicaid managed care as a policy response to rising Medicaid costs, provides background on the economics of Medicaid managed care, and discusses theory and evidence on the trade-offs associated with such "make-or-buy" decisions.

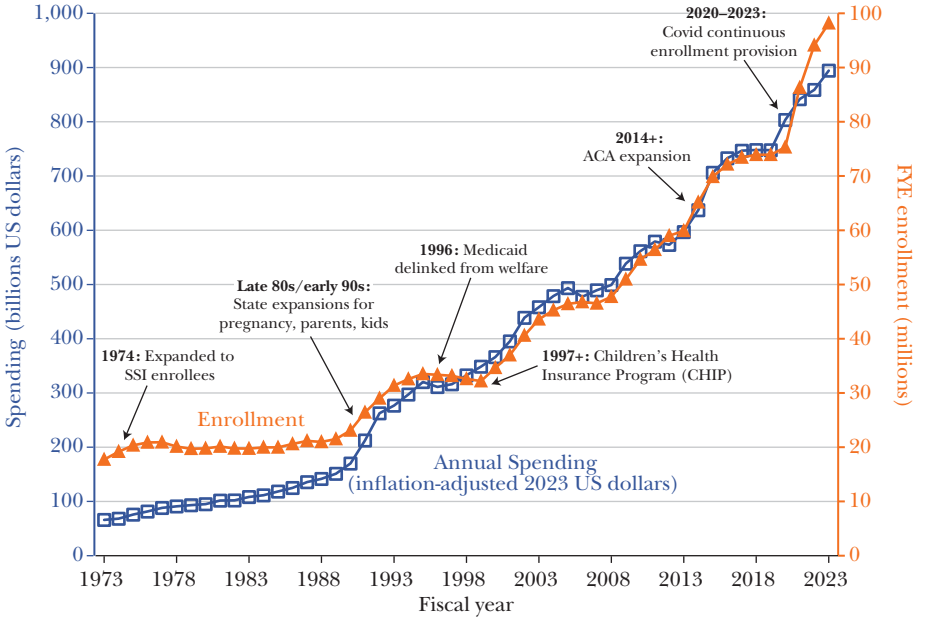
The Medicaid Program

Medicaid was created in 1965 alongside Medicare (the health program for seniors age 65 and over) as a federal-state program originally aimed at covering specific groups of low-income Americans—primarily mothers and children, the elderly, and people with disabilities—but has expanded over time toward broader coverage of low-income adults. Medicaid is both a key source of health insurance and a central piece of America's social safety net. Without Medicaid, many more low-income individuals would be uninsured, as they generally lack employer-provided health insurance that covers most higher-income Americans.

Medicaid covers a broad set of benefits, encompassing a wider range of services than most other health insurers. In addition to mandatory coverage of hospital and physician care, states may include "optional" benefits such as dental and vision care for adults (all children receive dental and vision benefits). Medicaid also finances a large share of long-term services and supports for individuals with complex medical or functional needs, including care in nursing homes and home-based care. Coverage is provided with little to no patient cost-sharing, and beneficiaries generally do not pay premiums.

Eligibility for Medicaid varies by state and population group but has broadened considerably since its inception. States must cover certain mandatory populations—primarily low-income children, pregnant individuals, seniors, and people with disabilities—and may also opt to cover additional groups. The Patient

Figure 1
Medicaid Program Growth and Key Policy Events



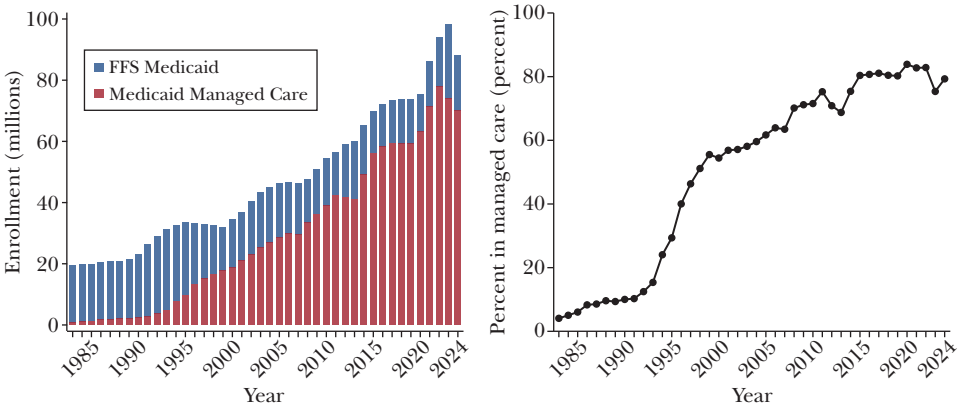
Source: Figure based on data from MACPAC (2024).
Note: Enrollment is full-year equivalent enrollment in Medicaid and the Children's Health Insurance Program. Spending is total annual Medicaid and Children's Health Insurance Program spending (federal + state), inflation-adjusted to 2023 dollars using the Consumer Price Index for All Urban Consumers (CPI-U). SSI is Supplemental Security Income, in which the Social Security system provides income to the low-income elderly and disabled. ACA stands for the Patient Protection and Affordable Care Act of 2010.

Protection and Affordable Care Act of 2010 intended to expand eligibility further by making all adults with incomes below 138 percent of the federal poverty level eligible. However, the US Supreme Court decision in *NFIB v. Sebelius* made this expansion optional.[1] As of 2025, 40 states and the District of Columbia have adopted the Medicaid eligibility expansion.

Medicaid's history is marked by two key trends: growing size and increasing privatization. Figure 1 shows Medicaid's growth in enrollment and inflation-adjusted spending over time, labeling key policy events. From about 20 million enrollees in the early 1970s, it grew to over 90 million at its Covid-era peak in 2023, after which it declined to 77 million as of late 2025 (not shown in the figure). Total federal and state spending (in 2023 dollars) rose from about $100 billion in the 1970s to over $900 billion today. Almost all recent spending growth has come from expanded

[1] *National Federation of Independent Business v. Sebelius*, 567 US 519 (2012).

Figure 2
Growth of Medicaid Managed Care over Time



Source: Managed care enrollment data come from several sources: Kaiser Family Foundation (1995) for data on 1983–1994, Kaiser Family Foundation (2001) for 1995–2000, Kaiser Family Foundation (2011) for 2002–2011, and Kaiser Family Foundation (2026) for 2013–2024. Managed care enrollment for 2001, 2012, and 2023 is linearly interpolated from adjacent years. Total Medicaid enrollment is from MACPAC (2026, exhibit 10).

Note: The figure documents the shift in Medicaid enrollment from state-run fee-for-service (FFS) plans to private managed care plans over 1983–2024. The left panel shows total Medicaid enrollment (in full-year equivalent enrollees) decomposed into fee-for-service and managed care. The right panel plots the managed care share over time (managed care enrollment divided by total enrollment), which rises from near zero in the mid-1980s to 84 percent by 2024. Managed care enrollment is defined broadly to include risk-based plans and primary care case management. About 78 percent of enrollees are in comprehensive risk-based managed care plans.

enrollment rather than rising costs per person. Real per-enrollee Medicaid spending has been constant at around $10,000 per year since the late 1990s.

Medicaid's steady growth has placed pressure on policymakers to control costs. Historically, states have primarily done this by reducing optional benefits, restricting eligibility, or cutting payments to doctors, hospitals, and other medical providers. This pressure has contributed to a long-running search for more efficient delivery models, including whether outsourcing to private managed care insurers can save money.

Figure 2 shows the trend towards outsourcing Medicaid delivery to private managed care insurers. From a tiny share of Medicaid enrollment in the 1980s, Medicaid managed care grew sharply during the 1990s to cover about 60 percent of Medicaid enrollees by 2000, following a broader shift towards managed care in US health insurance during the 1990s. Whereas managed care experienced a nationwide "backlash" in the late 1990s, Medicaid managed care enrollment continued to grow during the 2000–2020 period. By 2024, roughly 85 percent of Medicaid beneficiaries were enrolled in some form of managed care. This sharp growth raises important questions about the efficiency of outsourcing Medicaid delivery, which we discuss next.

The Rise of Medicaid Managed Care

The shift toward Medicaid managed care—under which states contract with private managed care organizations—began in the early 1970s, though Medicaid managed care did not become the dominant delivery system until decades later. Medicaid managed care represents the "buy" rather than "make" choice in the classic "make-or-buy" decision. Under traditional fee-for-service Medicaid, the state Medicaid agency pays providers directly for medical care they deliver. Under managed care, by contrast, the state makes a per-member-per-month "capitation" payment to a private managed care organization. The payment is risk-adjusted based on enrollee demographics—such as age, sex, and eligibility category—as well as health conditions to account for differences across plans in the expected cost of enrollees. For example, a plan would receive a higher capitation payment for an adult enrollee with a disability than for a healthy child. In exchange for these capitation payments, managed care organizations—generally large commercial insurers, Medicaid-focused insurers, or provider-sponsored plans affiliated with safety-net providers—assume responsibility for building a network of contracted providers, paying those providers, and managing enrollee care. In effect, the government outsources care management to private firms while continuing to provide financing and set program rules.

Outsourcing to private firms with high-powered incentive contracts involves a classic set of trade-offs studied in a large literature in contract theory (for example, Laffont and Tirole 1993; Shleifer 1998). On the one hand, under capitated insurance contracts, private managed care organizations capture the full benefits of making cost-reducing changes like eliminating unnecessary care, negotiating lower prices, and finding ways to keep people healthy. Private insurers may be better able to reduce costs and improve quality than a program operated by public sector employees. (An important assumption here is that there exist socially efficient ways to reduce costs in health insurance and that private insurers have the tools to implement them, a point to which we will return.)

On the other hand, private insurers may go too far in cutting costs because they do not internalize the negative impacts of cost-cutting on "non-contractible" quality (Shleifer 1998). For example, an insurer could cut costs by denying claims for truly necessary services or by maintaining an inadequate network of doctors or hospitals. If it is infeasible for states to specify what "necessary" or "adequate" means in every situation, then this incomplete contracting may lead to under-provision of quality by private firms.

Thus, the superiority of the "buy" option (private contracting) depends on the extent to which quality is observable and contractible (Hart, Shleifer, and Vishny 1997). If quality is easily specified—as it is likely to be in, say, cement production or garbage collection—then contracting out can leverage private efficiency without suboptimal quality. However, if quality is hard to observe, the effects of contracting on welfare become ambiguous. Because the contractor has the ability to stint on dimensions of quality not observed by policymakers, the result may be lower-quality output, even if private firms are in principle more efficient. Hart, Shleifer, and

Vishny (1997) develop this argument through a discussion of private prisons where, like health care, quality is important but non-contractible. This is because prison quality depends on factors such as guards' judgment, which are impossible to specify contractually in advance. Prison contracts can mandate training hours or restrict the use of force, but these crude proxies leave substantial discretion to contractors. Hence, prison contractors have too strong an incentive to cut costs, because they capture the savings but do not internalize the quality reductions.

The experience of California as a pioneering Medicaid managed care state illustrates the potential pitfalls of contracting out. California expanded Medicaid managed care in 1972—taking a laissez-faire approach to encourage rapid market entry—and within a year, there were 132,688 beneficiaries enrolled in 22 Medicaid managed care plans (Chavkin and Treseder 1977). However, subsequent investigations revealed that this rapid growth was fueled by predatory marketing practices. Plans engaged in aggressive door-to-door solicitation, utilizing inducements ranging from fried chicken to football tickets, as well as deceptive tactics such as having enrollers disguised in medical uniforms threaten a loss of benefits. Once enrolled, patients often faced significant barriers to access care, with restrictive policies linked to at least one patient death. There were also allegations of "cream skimming," in which plans selectively enrolled healthier, lower-cost beneficiaries. Auditors found that only 48 percent of the capitation payments to plans had been expended on health care services.

Ultimately, the California Medicaid scandals precipitated a swift policy backlash; the resulting state and federal reforms stalled the momentum of Medicaid managed care for decades. However, the reforms also shaped the regulatory framework that would eventually govern a nationwide resurgence for Medicaid managed care.

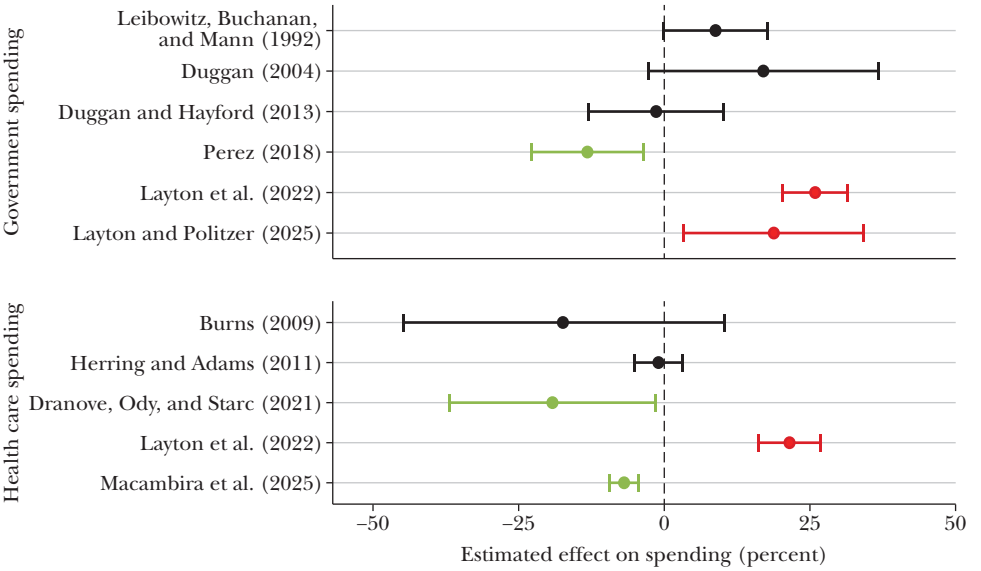
Medicaid managed care enrollment accelerated during the "managed care revolution" of the 1990s. One catalyst was rising Medicaid costs—often termed the "Pac-Man" of state budgets. Another was the removal of the "75/25 rule" that no more than 75 percent of the enrollment of a managed care organization be composed of Medicaid and Medicare enrollees. This rule was originally designed as a "private market test" for plan quality, limiting participation in the Medicaid program to plans that also operated in the commercial market. It was repealed as part of the Balanced Budget Act of 1997, with the intent of removing a barrier to entry in Medicaid.

Enrollment in Medicaid managed care surged from under 10 percent in 1990 to over 50 percent by 1998, and has now reached approximately 85 percent of all beneficiaries (as shown in Figure 2). Over the past decade, the federal government issued Medicaid managed care regulations in 2016 and updated those regulations in 2020 to tighten oversight, but there remain open questions as to whether the Medicaid managed care model is the right approach and how best to structure it.

Make or Buy? Evidence on the Effects of Medicaid Managed Care

States have articulated several rationales for the shift to Medicaid managed care. One is cost containment. Another is budget predictability: states claim that the capitation model stabilizes expenditures relative to the open-ended liability of fee-for-service

Figure 3
The Effects of Private Managed Care Versus State-Run Fee-for-Service on Spending



Note: The figure shows estimates of the causal effects of private Medicaid managed care (relative to a state-run fee-for-service plan) on Medicaid spending, drawing on key papers from the literature. The top panel shows the impact on government spending—what the state pays to insurers. The bottom panel shows impacts on health care spending—the cost of health care actually used by patients. The gap between the two includes insurer administrative costs and profits. Green indicates a statistically significant decrease in spending. Red indicates a statistically significant increase in spending. The included studies differ based on the geography, populations, and services studied. Most studies are national in scope—with the exception of Leibowitz, Buchanan, and Mann (1992) [Unidentified state], Duggan (2004) [California], Layton et al. (2022) [Texas], and Macambira et al. (2025) [Louisiana]—and focus on total spending (government or health care). The exception is Dranove, Ody, and Starc (2021), which studies the effects of Medicaid managed care on pharmacy spending.

models. Yet another is the potential for competition among managed care organizations to spur innovation and improve the efficiency and quality of care delivery.

The evidence is mixed on whether Medicaid managed care achieves cost control, with estimated effects going in both directions and differing across states, populations, and services studied. Figure 3 presents estimates from several key studies of the effects of Medicaid managed care on government spending (top panel) and health care spending (bottom panel). The distinction between government spending and health care spending turns out to be critical for interpreting the evidence. Studies that focus on government spending as the outcome generally report null or positive effects of Medicaid managed care. That is, privatization does not appear to have reduced, and may have increased, the cost to taxpayers. This is a striking result given that cost containment was the primary rationale for the shift to managed care.

By contrast, studies of the effects of Medicaid managed care on health care spending tend to find negative or null effects of Medicaid managed care. For

example, Macambira et al. (2025)—which uses variation due to the random assignment of enrollees to Medicaid managed care versus fee-for-service in Louisiana—finds that managed care reduces overall health care spending by 7 percent relative to fee-for-service, driven by a 25 percent reduction in pharmacy spending.[2] Hence, while there is limited evidence that Medicaid managed care reduces government spending, there is some well-identified evidence that it reduces health care spending.

The gap between these two sets of spending results is an important area for future research. Reduced health care spending need not translate into lower government spending for several reasons: managed care plans may capture the savings as profits; the increased administrative costs of operating a managed care system may offset lower health care spending; or fiscal dynamics—under which states base future capitation payments on plans' historical costs—may increase capitation payments to plans over time (Layton and Politzer 2025). At the same time, the evidence that managed care reduces health care spending suggests there may be efficiency gains from contracting out and raises the question of whether procurement can be redesigned so that states capture more of the savings.

Research on the impacts of privatization on quality of care and enrollee health is inconclusive (Sparer 2012; Franco Montoya, Chehal, and Adams 2020). While several studies document severe health harms, including increased mortality and worse birth outcomes (Aizer, Currie, and Moretti 2007; Kuziemko, Meckel, and Rossin-Slater 2018; Duggan, Garthwaite, and Wang 2021) and lower enrollee satisfaction (for example, Macambira et al. 2025), others find evidence of expanded access and patient benefit or no evidence of harm (for example, Chorniy, Currie, and Sonchak 2018; Dranove, Ody, and Starc 2021; Layton et al. 2022). The variation in findings is not easily explained by differences in the populations or states studied, making it difficult to draw general conclusions about when managed care helps or harms enrollees. Understanding the sources of this heterogeneity remains an important area for future research.

Collectively, the evidence on cost and quality suggests that the "make-or-buy" decision remains an open question in Medicaid. We note, however, that many states have moved on from this issue and are committed to the Medicaid managed care contracting approach. Therefore, how to design Medicaid managed care, rather than whether to do it, is the relevant policy question for most policymakers.

Medicaid's "Procured Competition" Approach

Classic debates in economics focus on whether governments *should* intervene in markets, but an equally important question is *how* governments intervene. Public

[2] This result is similar to Dranove, Ody, and Starc (2021), who find that Medicaid managed care reduces pharmacy spending by 21 percent using a different source of identifying variation—the state-level carve-ins of pharmacy services from fee-for-service to Medicaid managed care.

involvement can range from lightly regulated competition to direct public provision. Between these extremes lies a wide range of institutional arrangements whose structure and economic logic can be difficult to parse. Nowhere is this clearer than in health care, where governments rely on a diverse and often confusing mix of market- and non-market-based approaches.

We start from the premise that Medicaid's goal is to ensure low-income people can access needed basic health care at an efficient public cost. Federal and state rules specify a detailed set of covered services—hospital care, physician care, prescription drugs, and more—but determining when care is truly needed and how it can be delivered most efficiently is operationally complex. The central function of modern health insurance is to "manage" care and negotiate prices to contain costs. But how can the government ensure that the private insurers perform well, rather than stint on needed care or engage in risk selection to boost profits?

The Role of Government in Health Insurance

Figure 4 presents a framework for organizing the government's involvement in delivering social goods, from free markets at one end to direct public provision at the other. Along this spectrum, the government's role varies across three primary dimensions: (1) which firms can compete (firm entry), (2) how services are designed and priced (product design and pricing), and (3) the degree of consumer choice.

At one extreme is unregulated (or free) markets, in which firms can freely enter and set prices and product features, and consumers can freely choose among competing options. While free markets can be efficient under textbook conditions, health insurance is an industry that departs sharply from this benchmark. Insurance is a classic selection market, in which competition is distorted by firms' incentives to avoid high-risk consumers (the sick) rather than improve quality and value (Einav, Finkelstein, and Fisman 2023). These problems are compounded by high concentration and market power, for which there is increasing evidence and which interact with selection incentives (Dafny, Duggan, and Ramanarayanan 2012; Mahoney and Weyl 2017; Kong, Layton, and Shepard 2024). In addition, insurance is a complex product, and extensive evidence shows that consumers struggle to make fully informed and rational choices among plans (Abaluck and Gruber 2011; Handel and Kolstad 2015; Bhargava, Loewenstein, and Sydnor 2017; Ericson and Sydnor 2017). Finally, efficiency is not the sole objective: ensuring equitable access to health care—especially for the poor, for whom most modern medicine is unaffordable—is a priority that markets alone cannot deliver.

Some argue that health insurance departs so fundamentally from a standard market good that direct government provision is warranted. While the classic example is municipal fire departments, direct provision is also common in health care. Examples include the UK National Health Service and US Veterans Health Administration—systems that are owned, operated, and financed entirely by government. Direct provision of health insurance is used in nations like Canada and Taiwan, even if most hospitals and doctors are private. Fee-for-service Medicaid is an example of this type of direct public provision of insurance.

Figure 4
Framework for Government's Role in Social Services

	Unregulated markets	Regulated competition (*ACA markets*)	Procured competition (*Medicaid*)	Public procurement (*road building*)	Direct public provision (*fire dept.*)
Firm entry	Free entry	Free entry	Govt. selects multiple competing firms	Govt. selects single firm per job	N/A (no firms)
Product design, pricing	Not regulated	Flexible, within regulated limits	Firms follow state contract, with some flexibility	Firm provides contracted service at auctioned price	Govt. designs and delivers service
Consumer choice	Free choice	Choice among qualifying plans	Choice among procured firms' plans	No choice	No choice

Source: Author's creation.
Note: The figure shows five models for government's role in social services, ranging from minimal role (free markets) to full control (direct public provision). The models differ along three dimensions: (1) which firms can compete (firm entry), (2) how services are designed and priced (product design and pricing), and (3) the degree of consumer choice. Examples of each model are shown. ACA Markets stands for the state-run Marketplaces for individual health insurance established by the Patient Protection and Affordable Care Act of 2010.

Between these extremes lie two common middle-ground approaches for health insurance: regulated competition and public procurement. Regulated competition retains market entry and consumer choice but imposes rules on pricing and product design to address selection, quality, and market power concerns. The goal of such regulations is not to replace markets, but to "manage competition" to make insurance markets work (Enthoven 1993). The health insurance Marketplaces established under the Patient Protection and Affordable Care Act of 2010 are a familiar example.

By contrast, public procurement starts from government control but outsources implementation to private firms. In this model—familiar from road construction and public infrastructure—the government specifies the desired product and then invites firms to compete for a contract, often via bidding on a formal auction. The government selects a single "winning" firm, which then supplies the contracted services at the agreed price. The government's role then shifts to contract management and enforcement—raising familiar principal-agent issues around effort, incentives, and monitoring.[3]

[3]Although this factor is often overlooked, both traditional Medicare and fee-for-service Medicaid rely heavily on contracting for administrative functions like claims processing, with evidence that these contractors matter for outcomes like claims denials (League 2024) and fraud detection (Shi 2024).

Despite decades of experience, there is no consensus on whether regulated competition or more centralized approaches like public provision or procurement work better for health insurance. Market-based models have expanded substantially, as seen in the growth of Medicare Advantage and the Affordable Care Act Marketplaces.[4] On the other hand, dissatisfaction with cost, complexity, and access in these programs remains widespread.

Medicaid managed care offers a distinctive alternative. By combining elements of public procurement with regulated competition, it occupies a middle ground that seeks to harness the strengths—and limit the weaknesses—of both approaches. We turn next to the defining features of this model of procured competition.

Procured Competition: Medicaid's Hybrid Model

Medicaid managed care blends public procurement with market competition. Unlike standard procurement, where a single winner takes all, Medicaid selects multiple insurers that then compete for enrollees in a market. This blended model injects market discipline into a classic contracting environment, but it also creates a complex set of incentives that states must actively manage.

To understand this hybrid model, note that Medicaid managed care is, at its core, much closer to public procurement than regulated markets along three main dimensions. First, government selects insurers via procurement. To participate in Medicaid, insurers must submit bids and be chosen by the state; participation is determined by state selection rather than free entry.

Second, government specifies most plan features and prices. States specify detailed provisions via managed care contracts that are often hundreds of pages long. These contracts define covered services, patient cost-sharing (essentially zero), rules for provider networks, insurer compensation arrangements, and provisions for quality reporting and oversight. They also specify insurer prices (payments per enrollee) and other compensation provisions, such as risk adjustment. Insurer flexibility is limited to a few features like provider networks and care management rules. Additionally, Medicaid insurers each offer a single plan, limiting insurer flexibility but simplifying enrollee choice. By contrast, markets like the Affordable Care Act Marketplaces for individual health insurance and Medicare Advantage allow insurers to offer multiple plans and feature dozens or even hundreds of available plans.

Third, government oversees insurer performance. States retain ultimate responsibility for program performance, even as they delegate implementation to private firms. States actively manage contracts over time, monitor insurer behavior, and can impose sanctions or exclude underperforming plans.

At the same time, Medicaid managed care incorporates a limited role for market competition. Because states contract with multiple insurers, beneficiaries

[4] From a small program in the early 2000s, Medicare Advantage has grown to cover over half (54 percent) of Medicare enrollees in 2025. Enrollment in the Affordable Care Act state-run Marketplaces for individual health insurance more than doubled between 2020 and 2025, driven by larger subsidies and stronger insurer competition.

can choose their preferred plan and switch if they become dissatisfied. Evidence suggests enrollee choices respond to plan quality, implying that choice injects real market discipline (Geruso, Layton, and Wallace 2023).[5]

Why adopt this hybrid structure? Consider the overall goals of Medicaid managed care: (1) selecting the right health insurers to participate (procurement); (2) getting those plans to exert costly effort (contract management); and (3) sorting consumers into plans that best fit their needs (matching). Contracting with multiple insurers helps limit concerns about incumbent lock-in and market power, especially given the disruption associated with insurer exits (Politzer 2025). At the same time, consumer choice can complement traditional contract tools—such as monitoring and pay-for-performance—by rewarding plans that perform well on dimensions that consumers can see but that are difficult to specify or enforce in contracts. Finally, offering multiple plans allows the state to accommodate heterogeneous preferences, particularly for provider networks.

Conversely, procurement can be understood as a way of addressing well-known market failures in insurance (Cuesta and Tebaldi 2025). Even with strong regulations and corrective incentives, growing evidence suggests that adverse selection, market power, and consumer choice frictions remain potent forces that distort insurance competition. Medicaid addresses these problems by tightly standardizing plan features while giving states authority to select the menu of competitors. If used well, this authority could let states curate a desirable set of plan options, limit insurer profits, and allow consumers to choose within an approved menu of "sufficiently good" options. If used poorly, however, outcomes could be worse than what emerges from a less tightly controlled approach.[6]

Together, these features illustrate the central theme of this paper: Medicaid's procured-competition model blends government control with selective use of market forces in an effort to manage incentive problems inherent in outsourcing. The effectiveness of this hybrid design depends not only on whether states contract out to managed care, but on how they design the specific policies governing procurement and competition, which we examine next.

Key Policies for Medicaid's Procured Competition

Having outlined the goals of contracting out Medicaid to private managed care insurers and introduced the "procured competition" model at a high level, we

[5] Of course, if the consumers who switch plans are unhealthy and therefore unprofitable, the incentive can go in the opposite direction (Shepard 2022; Kreider et al. 2024). This is the perverse logic of adverse selection.

[6] Medicaid is not unique in using elements of procured competition. Large employers commonly select a limited set of health plans through procurement, and some Affordable Care Act Marketplaces—such as those in California and Massachusetts—use "active purchasing" models in which regulators negotiate terms and determine which plans are offered (Shepard and Forsgren 2023). These suggest additional settings in which economists can gain insights about the effectiveness of procured competition policies.

Table 1
Key Features of Medicaid's Procured Competition Model

Policy area	Medicaid's distinct features	Key questions for policy/research
Procurement process	1. Insurer costs/prices play little role in procurement (*prices are set administratively*)	Should states use *competitive bidding* to set prices subject to regulatory constraints?
	2. Medicaid selects multiple winning insurers (*versus standard auctions with a single winner*)	Should states be more selective? Or should they forgo auctions and accept "any willing plan"?
	3. Medicaid uses complex scoring auction for selecting winners, often with little weight on cost.	What are the optimal scoring auction weights to put on cost, quality, and other features?
Market design rules	4. Most benefits are standardized (*exception: provider networks, utilization management, customer service*)	Is standardization optimal? How best to ensure adequate networks and appropriate prior authorization?
	5. Enrollees pay zero premiums (→ *no incentive to choose cheaper plans*)	Should there be premium competition? Should low-cost plans be allowed to offer supplemental benefits (as in Medicare Advantage)?
	6. Cost-cutting incentive limited by *risk sharing* and *dynamic cost-plus rate setting*	Do these provisions raise costs? What is the optimal balance between risk sharing versus cost-cutting incentives?
Plan choice rules	7. Large share of enrollees are passively *auto-assigned* to plans (∼45 percent in median state)	What are implications of assigning based on quality or cost? What does an optimal auto-assignment algorithm look like?
	8. Public fee-for-service plan operates alongside private managed care organizations	Should market be public only, private only, or a choice ("*public option*")?

Note: The table shows distinct features of Medicaid's procured competition model, grouped into three policy areas. The final column lists key open questions for policymakers and economic researchers.

now turn to the key policy choices involved with implementing it. We organize our discussion around three core areas: (1) the procurement process, (2) market design rules, and (3) plan choice rules. For each, we describe current practice, how it maps to economic theory, and the existing evidence and open questions for policy and research. Table 1 summarizes these features and the main areas where additional research is needed.

Medicaid's Procurement Process

A central feature of the Medicaid managed care model is that the state selects insurers via a procurement process. The entire process can take 18 to 24 months and involves a strategic planning phase in which the state specifies the desired product characteristics, issues a request for proposals, collects proposals, announces the winning bidders, and negotiates final contracts. Economic theory points to three

key considerations for states: (1) the role of cost and quality in the auction scoring rules; (2) how many winners to select; and (3) for how long to award contracts.

A growing body of evidence suggests that the stakes of procurement auctions are high in Medicaid, due to the large differences in cost and quality of performance across health insurers. For example, exploiting random assignment across Medicaid plans, Geruso, Layton, and Wallace (2023) find that insurer costs varied by 25 percent—with the enrollees preferring higher spending plans. In a commercial setting, Handel and Kolstad (2015) found similar-sized differences in health care spending across insurers. Insurers also seem to vary in quality. Studying plan exits in Medicare Advantage, Abaluck et al. (2021) find meaningful differences in mortality and show that consumers placed little weight on this in their choices. These results also raise questions about the apparently limited ability of consumers to observe quality.

States must decide how to incorporate cost and quality into the procurement process. Medicaid contracts involve multiple dimensions—such as price, benefits, provider networks, and quality—so states typically rely on scoring auctions that evaluate bids along several attributes rather than price alone. This approach is well suited to settings where key aspects of performance vary in their levels of contractibility (Asker and Cantillon 2010). Evidence from other public procurement settings suggests that these design choices can matter. For example, when the California Department of Transportation awarded some highway contracts using a scoring system that rewarded both lower prices and faster completion times, projects were completed 30–40 percent faster, and the resulting benefits to commuters exceeded the increase in procurement costs (Lewis and Bajari 2011).

Despite the widespread use of scoring auctions in Medicaid, states vary substantially in how they structure these auctions—for example, in the relative weight placed on price versus quality or provider network adequacy (Baumgarten 2020). This variation suggests a promising opportunity for empirical research. Several states have revised their procurement methods over time or adopted more formal scoring systems, and procurement records and contract outcomes are often publicly available. Yet there is little systematic evidence on how these auction designs affect which insurers are selected, the prices states pay, or the quality of care delivered.

One defining feature of Medicaid's procurement model is a near absence of cost/price competition across most states. States generally set capitation rates administratively using actuarial analyses and select the winners based on non-price dimensions (that is, with essentially zero weight on price/cost in scoring auctions).[7] Even states that collect competitive cost bids are constrained in rate-setting by federal rules that specify "actuarially sound" rate ranges.

[7] Based on data for 2014, Layton, Ndikumana, and Shepard (2018) report that 32 states use administrative rate setting, six states use negotiated rates, and six use competitive bidding. A more recent report by MACPAC (Forbes and Dunbar 2022) does not report data but gives the following summary: "Most states have MCOs [managed care organizations] compete on program elements, not price."

At first blush, the limited role of cost in Medicaid procurement auctions appears to be a puzzle. However, cost-based auctions may lead to a "winner's curse" where the winning firms are those most likely to have underestimated their costs (Decarolis 2014). Minnesota's 2015 procurement offers one such example. To reduce cost, Minnesota allowed cost to account for 45 percent of the auction score, and two new entrant insurers with low bids won most of the statewide contracts (Baumgarten 2020). Within months, these insurers reported large losses, and one eventually exited. Minnesota subsequently removed cost from its Medicaid procurement scoring. While Minnesota's experience illustrates the risks of weighting cost heavily in procurement, it is only a single case; more systematic evidence is needed on whether and how states should incorporate cost into scoring auctions.

That Medicaid nonetheless remains among the lowest-cost forms of US health insurance suggests that aggressive price bidding may not be essential for cost control. But it also leaves open whether factoring price into procurement could generate even lower costs.

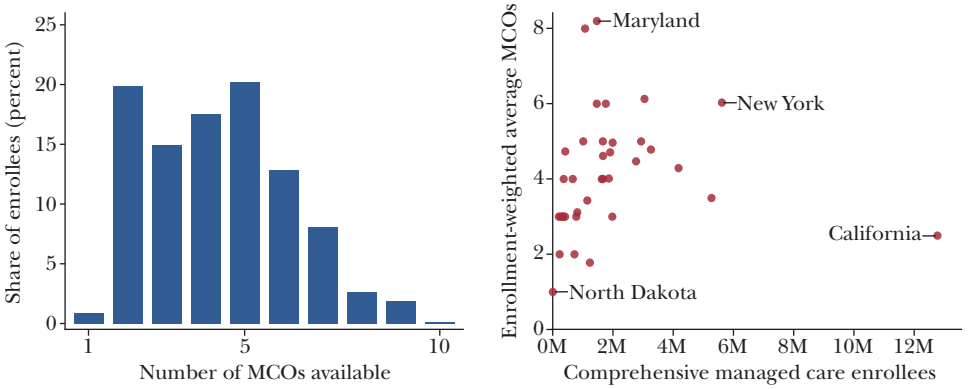
A second defining feature of Medicaid procurement is how many winners are selected. States vary widely in the number of competing managed care organizations per county, even after accounting for market size, as shown in Figure 5. Commonly, about half or more of bidding insurers win a contract.[8] However, a few states allow "any willing plan" to participate if they satisfy contract requirements—a policy closer to free entry than competitive procurement. States may have several reasons to favor broader participation.

Contracting with multiple insurers can preserve beneficiary choice, improve plan-consumer matching (for example, access to preferred providers), and create market discipline through consumers' ability to switch plans. A larger set of participating plans may also limit incumbency advantage and strengthen states' ability to discipline underperformers (Anton and Yao 1987). Finally, Medicaid procurements are often contentious, and greater selectivity would increase the risk of costly and time-consuming legal challenges from losing bidders. These legal challenges are an interesting area for research. Anecdotally, they play a large role in the formal structure of procurement, because agencies know they must justify decisions or face the risk of lawsuits, but whether this threat improves procurement or makes it overly "rigid" is an open question.

These benefits must be weighed against arguments for greater selectivity, including the potential to secure better terms and reduce cross-plan adverse

[8]While comprehensive data on bidding and plan selection do not exist, a few examples are illustrative. Oregon's 2019 procurement awarded contracts to 15 of 19 bidders (Baumgarten 2020). California's 2022 procurement ultimately awarded contracts to five of eight bidders (three bidders initially, and another two after appeals, as reported in Kelly 2023). Florida's 2024 procurement awarded contracts to five of eleven bidders: see bidders at ahca.myflorida.com/download/23423 and winners at ahca.myflorida.com/download/24478. Nebraska's 2022 procurement selected three of five bidders: bidders are listed at das.nebraska.gov/materiel/purchasing/112209-O3/Respondents.pdf and winners at das.nebraska.gov/materiel/purchasing/112209-O3/Intent-To-Award.pdf.

Figure 5
Variation in Number of Medicaid Managed Care Organizations per County



Source: Data are from T-MSIS Analytic Files (TAF), December 2022 (the most recent month available), and the sample is restricted to 36 states with comprehensive Medicaid managed care programs for which data quality is sufficiently high. Comprehensive managed care enrollment in Panel B is from the CMS Medicaid Managed Care Enrollment Report, 2022 (Centers for Medicare & Medicaid Services 2024).
Note: The figure shows the distribution of the number of Medicaid Managed Care Organizations (MCOs) available per county. Panel A shows the enrollment-weighted distribution of MCO availability; bars reflect the share of enrollees residing in counties with a given number of MCOs. Panel B plots each state's total managed care enrollment against its enrollment-weighted average number of MCOs per county.

selection—in this case, any factors or strategies that may tend to concentrate high-cost patients in a few plans (Ryan 2023; Cuesta and Tebaldi 2025).

Finally, states must decide for how long to award Medicaid managed care contracts. Anecdotal evidence suggests that these contracts are generally more than three years long (Baumgarten 2020). On the one hand, longer contracts avoid the administrative costs of reprocurement and strengthen incentives for long-term investments (Laffont and Tirole 1993). On the other hand, shorter contracts increase performance incentives due to a more credible threat of near-term replacement (Albano and Cesi 2008) but raise reprocurement costs and may lead to more frequent disruptions of members. Research is needed to document the variation in contract lengths and terms, and then to tie these differences to strategic insurer behavior and program outcomes.

Market Design Rules

After states select participating insurers for their Medicaid program through procurement, how should they structure market design rules to create incentives for insurers to deliver high-quality care while controlling costs for taxpayers? Again, the goal of market design is to induce insurers to assemble provider networks, manage care, and deliver services in ways that optimally balance access to care versus costs. In practice, this effort involves trade-offs.

The classic concern is that plans will stint on quality—limiting access "too much" to cut costs and earn profits. The incentive to stint, and to hope that stinting might not be observed by the state, arises from two types of asymmetric information. First, insurance "quality" is complex: it depends on detailed features of networks and care-management practices that are hard for both the state and enrollees to observe. In other words, there is meaningful non-contractible quality. Second, the enrollees best positioned to detect quality are sicker individuals who attempt to use complex care. These sicker individuals are often unprofitable to insurers. An insurer who improves quality may be penalized by attracting many sick enrollees (that is, by adverse selection).

As a concrete example illustrating these forces, Kreider et al. (2024) study access to a world-class cancer hospital in New York's Medicaid program. Although access to the hospital was highly valued by cancer patients, it was also substantially more expensive. Prior to 2005, no Medicaid managed care plan covered the hospital. When one plan briefly added the hospital to its network, it experienced a large influx of cancer patients, an increase of about 50 percent. Because of relatively weak risk adjustment at the time, the change proved unprofitable, and the managed care organization dropped the hospital the next year. The result was an equilibrium in which no plan covered the top cancer hospital—a pattern echoed for other "star" hospitals in other insurance settings (Shepard 2022; Serna 2025).

Medicaid programs address these concerns using a mix of two broad approaches. First, states impose regulatory limits on plan design. For example, states use a standardized benefit design for covered services and cost-sharing rules across insurers. As a result, competition is limited to a few flexible (and harder to standardize) attributes: provider networks, care-management strategies, and customer service.[9] Wallace (2023) demonstrates that the network differences across plans have consequences. Exploiting random plan assignment in New York's Medicaid program, he finds that narrower networks decrease spending by generating hassle costs that constrain the utilization of needed and unneeded services and reduce enrollee satisfaction, suggesting networks are an important but blunt instrument for reducing spending. In addition, states set premiums equal to zero for all Medicaid managed care plans, eliminating premiums as a possible competitive margin. This prevents low-cost plans from growing by offering discounts. It also limits the ability of low-quality plans to risk-select healthy enrollees by undercutting higher-quality plans.[10]

[9]While states set network adequacy requirements—rules specifying minimum numbers of provider by type and maximum time and distance to reach them—these are crude measures of access. This is a concrete illustration of the non-contractibility of quality problem discussed by Hart, Shleifer, and Vishny (1997): what matters for enrollees is whether they can see a qualified provider in a timely manner, but what states can measure is only a rough approximation. States also struggle to enforce compliance with network adequacy standards. A growing body of evidence documents that many of the providers listed in provider network directories do not actually see Medicaid patients (Ludomirsky et al. 2022; Zhu et al. 2022).

[10]For health care systems as a whole, the question of whether it is useful to have competition in premiums paid remains open. For instance, national insurance systems in Switzerland and the Netherlands include premium competition, whereas the main systems in Germany and Colombia do not. Medicare Advantage

In the second broad approach, states use compensation incentives that encourage insurers to provide quality and/or to soften cost-cutting incentives. For example, states use pay-for-performance incentives that often give bonuses or tie a portion of compensation to meeting quality metrics, such as ensuring enrollees get preventive screenings or can access care in a timely manner.[11] While useful, these metrics tend to capture only a small dimension of quality. Another approach involves risk sharing in compensation. Rather than having insurers bear full risk if enrollees have unexpectedly high costs—and therefore, giving insurers a strong incentive to cut costs—states share risk through risk adjustment (which pays more for sicker enrollees), reinsurance (which covers a portion of enrollees' very high costs), and risk corridors (in which the state shares a portion of insurers' overall profits/losses). A final approach to compensation incentives involves dynamic "cost-plus" contracting, in which states base payments on plans' historical costs. While this approach reduces the risk of current underpayment if past costs were already high, it also creates disincentives for cost-cutting because successful cost containment today leads to lower payments tomorrow (Layton and Politzer 2024).

Together, these regulatory limits and compensation incentives make Medicaid managed care quite different from a standard regulated insurance market. Managed care organizations operate less like risk-bearing firms competing on price and product design, and more like state contractors whose incentives are shaped to limit the profitability of selection and protect access.

The key question is whether Medicaid managed care is striking the right balance between cost-cutting versus quality-encouraging incentives. This question has not yet been answered. At a concrete level, if an insurer cuts costs by $1 today (whether that cut happens by stinting on care or scrupulous cost containment), how much of that dollar does it retain versus share with the state via the many layers of risk sharing, dynamic cost-plus contracting, and other provisions? It may be worthwhile to consider potential gains from alternative payment mechanisms—such as yardstick competition (Shleifer 1985) or greater reliance on external benchmarks—but more empirical work is needed to understand the trade-offs.

Plan Choice Rules

Medicaid managed care relies on rules for plan choice that shape how competition operates. In a standard market, consumer choices (demand) are the main driver of competitive incentives. But in Medicaid managed care, two additional provisions play a large role and offer opportunities for future research.

---

allows premium competition, but in practice a large number of plans "bunch" premiums at exactly $0 (Stockley et al. 2014). For Medicaid, zero premiums reflect equity goals and limit adverse selection, but they also prevent plans from funding more generous (and more expensive) offerings through higher premiums. Outcomes from Medicaid may offer an opportunity to quantify the trade-offs of a zero-premium design.

[11] These metrics are often based on the widely used "HEDIS measures" (which stands for Healthcare Effectiveness Data and Information Set) developed by the National Committee for Quality Assurance (NCQA).

First, many Medicaid beneficiaries—45 percent in the median state, as of the latest available data (Smith et al. 2015)—do not actively choose a plan; instead, they are auto-assigned by the state. This makes the auto-assignment algorithm a key driver of market shares and incentives for managed care organizations. In principle, auto-assignment rules could reward high-quality or low-cost plans by directing passive enrollees toward them, thereby spurring competition (Buitrago et al. 2025). In practice, however, relatively few states take this approach, perhaps because traditional measures of plan performance are mired in selection concerns (Wallace et al. 2022, 2025). As of 2022, only ten states reported incorporating quality-related performance measures into auto-assignment, and none explicitly assign enrollees based on plan costs (Hinton et al. 2022). States vary widely in how they approach auto-assignment: while some states assign enrollees in equal shares across plans, others weight assignment toward larger plans (to reward sought-after plans) or smaller plans (to balance enrollment and ensure sufficient scale). This variation likely reflects, in part, competing policy priorities, but the absence of a clear best practice suggests there is significant scope to improve program efficiency through smarter defaults. For interested researchers, the fundamental starting points are to collect updated data on both how many enrollees are auto-assigned and how states set their auto-assignment rules—neither of which has been measured systematically since 2015—and to seek out administrative data that contain indicators for which enrollees are auto-assigned.

Second, states vary in whether they rely entirely on private managed care organizations for Medicaid (34 states), entirely on a public fee-for-service plan (five states), or on a mixed model. A few states offer "primary care case management," which is a state-run plan that pays most claims as with standard fee-for-service, but also pays a primary care physician a small fee to manage an enrollee's care. Although primary care case management is not full risk-managed care—the state, not the primary care physician, retains the insurance risk—it introduces gatekeeping and coordinating functions that distinguish it from pure fee-for-service. Eight states offer enrollees a choice between private managed care organizations and primary care case management, effectively creating a public option within Medicaid markets. Some form of "public option" has been a major proposal for the state-run Marketplaces for individual health insurance, and Medicaid provides an interesting opportunity to gain evidence on its impact.

These plan choice rules remain largely unexplored in the Medicaid context and have relevance to broader debates about health insurance reform. They would seem to be a particularly promising area for future work.

Conclusion

Medicaid managed care sits at the center of some of the most important questions in health economics and public policy. Medicaid is a massive public program with the dual goals of ensuring access to essential health care for low-income

populations and doing so at an acceptable public cost. Over time, states have increasingly turned to private insurers to help meet these goals, fundamentally reshaping how Medicaid operates. Yet Medicaid managed care does not resemble either a standard insurance market or a conventional public procurement setting.

We have argued that Medicaid managed care is best understood as a hybrid model of procured competition. States use procurement to tightly specify benefits, prices, and oversight, while preserving a limited role for market competition through plan choice and enrollment. Viewing Medicaid through this lens helps clarify both its distinctive institutional features and the policy trade-offs embedded in its design. Compared with regulated insurance markets, Medicaid relies far more heavily on standardization and administrative control; compared with traditional procurement, it incorporates consumer choice and multiple contractors.

This framework highlights a set of unresolved policy questions. How selective should procurement be? How much risk should insurers bear? How should plan choice, auto-assignment, and public options be designed? Do current payment and incentive structures strike the right balance between cost control and quality? As Medicaid continues to evolve—and as other insurance settings consider whether to draw lessons from Medicaid—understanding how procured competition works in practice remains a central task for researchers and policymakers alike.

■ *We thank Eliana Buckner and Grant Kovel for excellent research assistance. We thank Tim Layton, Chima Ndumele, Ruochen Sun, and the JEP editors (Timothy Taylor, Heidi Williams, and Jonathan Parker) for feedback on this draft. We acknowledge use of AI tools in researching and drafting this paper.*

References

Abaluck, Jason, Mauricio Caceres Bravo, Peter Hull, and Amanda Starc. 2021. "Mortality Effects and Choice across Private Health Insurance Plans." *Quarterly Journal of Economics* 136 (3): 1557–610.

Abaluck, Jason, and Jonathan Gruber. 2011. "Choice Inconsistencies among the Elderly: Evidence from Plan Choice in the Medicare Part D Program." *American Economic Review* 101 (4): 1180–210.

Aizer, Anna, Janet Currie, and Enrico Moretti. 2007. "Does Managed Care Hurt Health? Evidence from Medicaid Mothers." *Review of Economics and Statistics* 89 (3): 385–99.

Albano, Gian Luigi, and Berardino Cesi. 2008. "Past Performance Evaluation in Repeated Procurement: A Simple Model of Handicapping." Unpublished.

Anton, James J., and Dennis A. Yao. 1987. "Second Sourcing and the Experience Curve: Price Competition in Defense Procurement." *RAND Journal of Economics* 18 (1): 57–76.

Asker, John, and Estelle Cantillon. 2010. "Procurement When Price and Quality Matter." *RAND Journal of Economics* 41 (1): 1–34.

Baumgarten, Allan. 2020. "Analyzing Medicaid Managed Care Organizations: State Practices for Contracting with Managed Care Organizations and Oversight of Contractors." Unpublished.

Bhargava, Saurabh, George Loewenstein, and Justin Sydnor. 2017. "Choose to Lose: Health Plan Choices from a Menu with Dominated Option." *Quarterly Journal of Economics* 132 (3): 1319–72.

Buitrago, Giancarlo, Paul Rodríguez-Lesmes, Natalia Serna, and Marcos Vera-Hernández. 2025. "How to Promote Health Insurer Competition? Evidence from Automatic Enrollment Rules." Unpublished.

Burns, Marguerite E. 2009. "Medicaid Managed Care and Cost Containment in the Adult Disabled Population." *Medical Care* 47 (10): 1069–76.

Centers for Medicare & Medicaid Services. 2024. "Medicaid Managed Care Enrollment and Program Characteristics, 2022." https://www.medicaid.gov/medicaid/managed-care/downloads/2022-medicaid-managed-care-enrollment-report.pdf (accessed March 23, 2026).

Chavkin, David F., and Anne Treseder. 1977. "California's Prepaid Health Plan Program: Can the Patient Be Saved." *Hastings Law Journal* 28: 685–760.

Chorniy, Anna, Janet Currie, and Lyudmyla Sonchak. 2018. "Exploding Asthma and ADHD Caseloads: The Role of Medicaid Managed Care." *Journal of Health Economics* 60: 1–15.

Cuesta, José Ignacio, and Pietro Tebaldi. 2025. "Public Procurement vs. Regulated Competition in Selection Markets." NBER Working Paper 34141.

Dafny, Leemore, Mark Duggan, and Subramaniam Ramanarayanan. 2012. "Paying a Premium on Your Premium? Consolidation in the US Health Insurance Industry." *American Economic Review* 102 (2): 1161–85.

Decarolis, Francesco. 2014. "Awarding Price, Contract Performance, and Bids Screening: Evidence from Procurement Auctions." *American Economic Journal: Applied Economics* 6 (1): 108–32.

Dranove, David, Christopher Ody, and Amanda Starc. 2021. "A Dose of Managed Care: Controlling Drug Spending in Medicaid." *American Economic Journal: Applied Economics* 13 (1): 170–97.

Duggan, Mark. 2004. "Does Contracting Out Increase the Efficiency of Government Programs? Evidence from Medicaid HMOs." *Journal of Public Economics* 88 (12): 2549–572.

Duggan, Mark, Craig Garthwaite, and Adelina Yanyue Wang. 2021. "Heterogeneity in the Impact of Privatizing Social Health Insurance: Evidence from California's Medicaid Program." NBER Working Paper 28944.

Duggan, Mark, and Tamara Hayford. 2013. "Has the Shift to Managed Care Reduced Medicaid Expenditures? Evidence from State and Local-Level Mandates." *Journal of Policy Analysis and Management* 32 (3): 505–35.

Einav, Liran, Amy Finkelstein, and Ray Fisman. 2023. *Risky Business: Why Insurance Markets Fail and What to Do about It*. Yale University Press.

Enthoven, Alain C. 1993. "The History and Principles of Managed Competition." *Health Affairs* 12 (S1): 24–48.

Ericson, Keith Marzilli, and Justin Sydnor. 2017. "The Questionable Value of Having a Choice of Levels of Health Insurance Coverage." *Journal of Economic Perspectives* 31 (4): 51–72.

Forbes, Moira, and Sean Dunbar. 2022. "Understanding Medicaid Managed Care Procurement Practices across States." Unpublished.

Franco Montoya, Daniela, Puneet Kaur Chehal, and E. Kathleen Adams. 2020. "Medicaid Managed Care's Effects on Costs, Access, and Quality: An Update." *Annual Review of Public Health* 41: 537–49.

Geruso, Michael, Timothy J. Layton, and Jacob Wallace. 2023. "What Difference Does a Health Plan Make? Evidence from Random Plan Assignment in Medicaid." *American Economic Journal: Applied Economics* 15 (3): 341–79.

Handel, Benjamin R., and Jonathan T. Kolstad. 2015. "Health Insurance for 'Humans': Information Frictions, Plan Choice, and Consumer Welfare." *American Economic Review* 105 (8): 2449–500.

Hart, Oliver, Andrei Shleifer, and Robert W. Vishny. 1997. "The Proper Scope of Government: Theory and an Application to Prisons." *Quarterly Journal of Economics* 112 (4): 1127–61.

Herring, Bradley, and E. Kathleen Adams. 2011. "Using HMOs to Serve the Medicaid Population: What Are the Effects on Utilization and Does the Type of HMO Matter?" *Health Economics* 20 (4): 446–60.

Hinton, Elizabeth, Lina Stolyar, Madeline Guth, and Mike Nardone. 2022. "State Delivery System and Payment Strategies Aimed at Improving Outcomes and Lowering Costs in Medicaid." Kaiser Family Foundation Issue Brief, January 12. https://www.kff.org/medicaid/state-delivery-system-and-payment-strategies-aimed-at-improving-outcomes-and-lowering-costs-in-medicaid/.

Kaiser Family Foundation. 1995. "Medicaid and Managed Care: Policy Brief." KFF, May 30. https://www.kff.org/medicaid/medicaid-and-managed-care-policy-brief/.

Kaiser Family Foundation. 2001. "Medicaid and Managed Care." KFF, December. https://www.kff.org/wp-content/uploads/2013/01/medicaid-and-managed-care-fact-sheet.pdf.

Kaiser Family Foundation. 2011. *Medicaid Managed Care Enrollment Report*. KFF.

Kaiser Family Foundation. 2026. *Medicaid Managed Care Enrollment by Plan Type*. https://www.kff.org/medicaid/state-indicator/enrollment-by-medicaid-mc-plan-type/ (accessed March 23, 2026).

Kelly, Susan. 2023. "California Revises Medicaid Contract Awards, Adding 2 Insurers." *Healthcare Dive*, January 4. https://www.healthcaredive.com/news/California-MediCal-Medicaid-contracts/639580/.

Kong, Edward, Timothy Layton, and Mark Shepard. 2024. "Adverse Selection and (Un)natural Monopoly in Insurance Markets." NBER Working Paper 33187.

Kreider, Amanda R., Timothy J. Layton, Mark Shepard, and Jacob Wallace. 2024. "Adverse Selection and Network Design under Regulated Plan Prices: Evidence from Medicaid." *Journal of Health Economics* 97: 102901.

Kuziemko, Ilyana, Katherine Meckel, and Maya Rossin-Slater. 2018. "Does Managed Care Widen Infant Health Disparities? Evidence from Texas Medicaid." *American Economic Journal: Economic Policy* 10 (3): 255–83.

Laffont, Jean-Jacques, and Jean Tirole. 1993. *A Theory of Incentives in Procurement and Regulation*. MIT Press.

Layton, Timothy J., Nicole Maestas, Daniel Prinz, and Boris Vabson. 2022. "Health Care Rationing in Public Insurance Programs: Evidence from Medicaid." *American Economic Journal: Economic Policy* 14 (4): 397–431.

Layton, Timothy J., Alice Ndikumana, and Mark Shepard. 2018. "Health Plan Payment in Medicaid Managed Care: A Hybrid Model of Regulated Competition." In *Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets: Theory and Practice*, edited by Thomas G. McGuire and Richard C. van Kleef, 523–61. Academic Press.

Layton, Timothy, and Eran Politzer. 2024. "The Dynamic Fiscal Costs of Outsourcing Health Insurance—Evidence from Medicaid." NBER Working Paper 33302.

Layton, Timothy J., and Eran Politzer. 2025. "The Dynamic Fiscal Costs of Outsourcing Health Insurance—Evidence from Medicaid." *Journal of Public Economics* 247: 105417.

League, Riley. 2024. "Administrative Burden and Consolidation in Health Care: Evidence from Medicare Contractor Transitions." Unpublished.

Leibowitz, Arleen, Joan L. Buchanan, and Joyce Mann. 1992. "A Randomized Trial to Evaluate the Effectiveness of a Medicaid HMO." *Journal of Health Economics* 11 (3): 235–57.

Lewis, Gregory, and Patrick Bajari. 2011. "Procurement Contracting with Time Incentives: Theory and Evidence." *Quarterly Journal of Economics* 126 (3): 1173–211.

Ludomirsky, Avital B., William L. Schpero, Jacob Wallace, Anthony Lollo, Susannah Bernheim, Joseph S. Ross, and Chima D. Ndumele. 2022. "In Medicaid Managed Care Networks, Care Is Highly Concentrated among a Small Percentage of Physicians." *Health Affairs* 41 (5): 760–68.

Macambira, Danil Agafiev, Michael Geruso, Anthony Lollo, Chima D. Ndumele, and Jacob Wallace. 2025. "The Private Provision of Public Services: Evidence from Random Assignment in Medicaid." NBER Working Paper 30390.

MACPAC. 2026. *MACStats: Medicaid and CHIP Data Book*. Medicaid and CHIP Payment and Access Commission.

MACPAC. 2024. *MACStats: Medicaid and CHIP Data Book 2024*. Medicaid and CHIP Payment and Access Commission.

Mahoney, Neale, and E. Glen Weyl. 2017. "Imperfect Competition in Selection Markets." *Review of Economics and Statistics* 99 (4): 637–51.

Perez, Victoria. 2018. "Does capitated managed care affect budget predictability? Evidence from Medicaid programs." *International Journal of Health Economics and Management* 18 (2): 123–52.

Politzer, Eran. 2025. "A Change of Plans: Switching Costs in the Procurement of Health Insurance." *Journal of Health Economics* 102: 103021.

Ryan, Conor. 2023. "Mergers in the Presence of Adverse Selection." Unpublished.

Serna, Natalia. 2025. "Exogenous Exits, Market Structure, and Equilibrium Contracts in Health Care." *American Economic Review: Insights* 7 (3): 325–39.

Shepard, Mark. 2022. "Hospital Network Competition and Adverse Selection: Evidence from the Massachusetts Health Insurance Exchange." *American Economic Review* 112 (2): 578–615.

Shepard, Mark, and Ethan Forsgren. 2023. "Do Insurers Respond to Active Purchasing? Evidence from the Massachusetts Health Insurance Exchange." *Journal of Risk and Insurance* 90 (1): 9–31.

Shi, Maggie. 2024. "Monitoring for Waste: Evidence from Medicare Audits." *Quarterly Journal of Economics* 139 (2): 993–1049.

Shleifer, Andrei. 1985. "A Theory of Yardstick Competition." *RAND Journal of Economics* 16 (3): 319–27.
Shleifer, Andrei. 1998. "State versus Private Ownership." *Journal of Economic Perspectives* 12 (4): 133–50.
Smith, Vernon K., Kathleen Gifford, Eileen Ellis, Robin Rudowitz, Laura Snyder, and Elizabeth Hinton. 2015. *Medicaid Reforms to Expand Coverage, Control Costs and Improve Care: Results from a 50-State Medicaid Budget Survey for State Fiscal Years 2015 and 2016*. Kaiser Family Foundation.
Sparer, Michael. 2012. "Medicaid Managed Care: Costs, Access, and Quality of Care." Robert Wood Johnson Foundation Synthesis Project Research Synthesis Report 23.
Stockley, Karen, Thomas McGuire, Christopher Afendulis, and Michael E. Chernew. 2014. "Premium Transparency in the Medicare Advantage Market: Implications for Premiums, Benefits, and Efficiency." NBER Working Paper 20208.
Wallace, Jacob. 2023. "What Does a Provider Network Do? Evidence from Random Assignment in Medicaid Managed Care." *American Economic Journal: Economic Policy* 15 (1): 473–509.
Wallace, Jacob, J. Michael McWilliams, Anthony Lollo, Janet Eaton, and Chima D. Ndumele. 2022. "Residual Confounding in Health Plan Performance Assessments: Evidence from Randomization in Medicaid." *Annals of Internal Medicine* 175 (3): 314–24.
Wallace, Jacob, Chima D. Ndumele, Anthony Lollo, Danil Agafiev Macambira, Matthew Lavallee, Beniamino Green, Kate A. Duchowny, and J. Michael McWilliams. 2025. "Attributing Racial Differences in Care to Health Plan Performance or Selection." *JAMA Internal Medicine* 185 (1): 61–72.
Zhu, Jane M., Christina J. Charlesworth, Daniel Polsky, and K. John McConnell. 2022. "Phantom Networks: Discrepancies between Reported and Realized Mental Health Care Access in Oregon Medicaid." *Health Affairs* 41 (7): 1013–22.

Journal of Economic Perspectives—Volume 40, Number 2—Spring 2026—Pages 195–214

A Users' Guide to Uncovering Worker and Firm Effects: The ABC of AKM

Stéphane Bonhomme, Thibaut Lamadon, and Elena Manresa

Linked employer-employee data are increasingly used to study key questions about labor markets related to worker and firm heterogeneity, sorting, and the sources of wage dispersion. The main feature of linked data is that they keep track of the identity of workers and firms while following workers across employers. While linked datasets were initially rare and can be difficult to access for confidentiality reasons, they have become increasingly widely available and have been used to study a host of economic questions.

The workhorse method in the literature is the so-called AKM method introduced in Abowd, Kramarz, and Margolis (1999). The insight of AKM is to use linked data in order to tell apart the roles of worker and firm heterogeneity in wage determination. By estimating firm and worker components, researchers can then answer a variety of questions related to the wage potentials of workers, the pay policies of firms, and the sorting patterns between workers and firms.

A key motivation for the method is to identify the sources of the wage differences across firms that characterize modern labor markets: inter-industry differentials are sizable, larger firms tend to pay better than smaller ones, multinationals pay higher wages than national firms, and there are also substantial within-group wage differentials (say, between large firms in a similar industry). However, these differences may come from different sources: they may arise because firms pay similar workers differently, or because they employ different

■ *Stéphane Bonhomme is Professor of Economics and Thibaut Lamadon is Associate Professor of Economics, both at the University of Chicago, Chicago, Illinois. Elena Manresa is Professor of Economics, Princeton University, Princeton, New Jersey. Their email addresses are sbonhomme@uchicago.edu, lamadon@uchicago.edu, and em7118@princeton.edu.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20231372.

types of workers. The first explanation points to the existence of high- and low-paying firms, while the second one points to high- and low-wage workers being employed by different firms.

The AKM methodology provides a way to quantify these two mechanisms, and thus to document how heterogeneity in permanent differences among workers and heterogeneity in permanent differences in firms' pay policies shape individual wages. The AKM model postulates that, in addition to worker and firm characteristics that are observed in the data (such as the worker's experience or the firm's size), the wage is determined by two key factors. The first component, denoted as ψ_j, is specific to the firm. Hence, two firms may pay the same worker differently, implying the existence of firm-specific wage premia. The second component, denoted as α_i, is specific to the worker. Hence, two workers in the same firm may earn different wages. The AKM model postulates that, net of covariates and some idiosyncratic shocks, wages (expressed in logarithms) are additive in the worker-specific and firm-specific components.

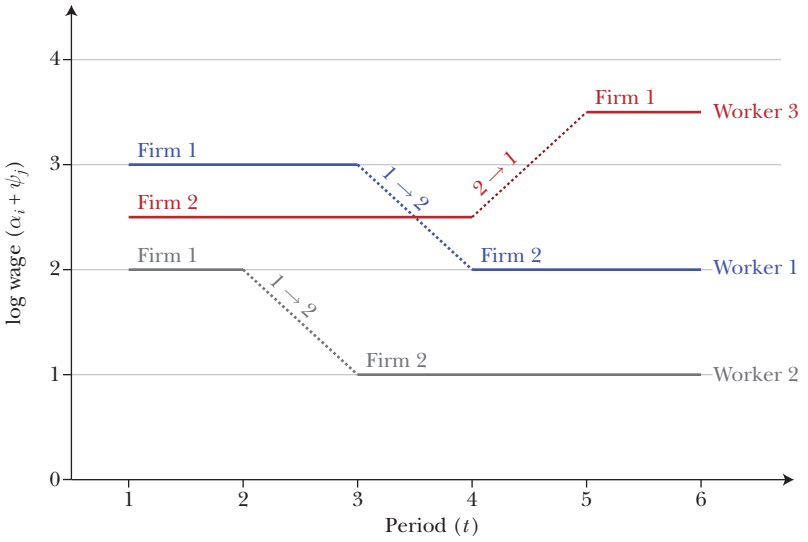
Figure 1 illustrates the main implications of the model by plotting the log wage over time for three hypothetical workers. Initially, in period 1, workers 1 and 2 are employed in firm 1, and earn different wages due to α_1 being greater than α_2. When worker 2 moves to firm 2, she experiences a wage reduction because ψ_2 is lower than ψ_1—that is, firm 2 pays similar workers less than firm 1. When, in a later period, worker 1 also moves to firm 2, she experiences exactly the same wage drop in percentage terms as worker 2. Moreover, the wage gap between workers 1 and 2 in firm 2 is the same as in firm 1. Next, worker 3, who was initially employed in the lower-paying firm 2, moves to the better-paying firm 1. The wage gain she experiences exactly mirrors the losses experienced by workers 1 and 2 in their respective job moves.

At its core, the AKM methodology leverages information from workers moving between firms, extracted from linked employer-employee data, to separately identify the firm components ψ_j and the worker components α_i. Given estimates of these parameters, which are commonly referred to as worker and firm "effects," researchers can assess how much of the cross-sectional wage variation and its evolution over time can be explained by worker heterogeneity, firm heterogeneity, and the sorting between workers and firms.

The AKM approach has been influential in labor economics. In this article, we review empirical findings in the literature while exploring the subtleties of the model and its estimator. We will highlight best practices for overcoming established estimation hurdles and discuss the methodological challenges that still remain, particularly regarding the static and additive nature of the model.

Throughout the article, we will emphasize that the apparent simplicity of the AKM approach is deceptive. The model does not restrict how job mobility relates to permanent worker and firm heterogeneity, and it can speak to economic models of job choice and wage determination. This explains the growing popularity of AKM in applied work. Naturally, this flexibility introduces complexity; the large number of parameters and the network structure of the worker-firm data create important

Figure 1
Wages over Time in the AKM Model



Note: The figure plots the wage trajectories of three workers with various α_1 values ($\alpha_1 > \alpha_2$ and $\alpha_1 < \alpha_3$) moving through two firms with different ψ_j values ($\psi_2 < \psi_1$). Time is on the x-axis, and log wages are on the y-axis. Colors indicate workers.

challenges for the estimation of the model, and for the development of extensions that relax some of its most restrictive assumptions.

The AKM methodology was originally developed in the context of wages, workers, and firms, and we will use this setting to illustrate the approach throughout our presentation. Because many linked datasets share a similar structure, researchers have successfully adapted the model to a wide diversity of questions. We will mention applications to health economics and corporate finance, but note that the AKM approach has been applied to diverse questions in a variety of fields.

Our discussion is intended to provide a concise introduction to the AKM methodology and a users' guide for implementation. For more details, we refer readers to the specialized surveys by Abowd, Kramarz, and Woodcock (2008), Card et al. (2018), Bonhomme (2020), and Kline (2024).[1]

A Conceptual Framework for AKM

A useful way to think about the AKM framework introduced by Abowd, Kramarz, and Margolis (1999) is as a simple model of wage determination. Workers

[1] To facilitate the understanding and adoption of the methods, a notebook is available at https://github.com/tlamadon/abc-of-akm.

are characterized by a fixed level of productivity, α_i, and firms are characterized by a pay policy, ψ_j, representing how much the firm pays per unit of effective labor. Wages reflect the combination of these two components. The structure is intentionally parsimonious: one part of pay is tied to who the worker is, the other to where the worker is employed.

The worker component, α_i, can be interpreted broadly. It may represent something concrete, such as task completion speed, or something more abstract, such as overall labor market ability. If two workers have productivities α_1 and α_2, the difference $\alpha_1 - \alpha_2$ reflects the percentage gap in the value they would generate over a fixed period. In this sense, α_i summarizes all persistent, worker-specific determinants of earnings.

The firm component, ψ_j, captures systematic differences in pay across employers. A high-ψ_j firm may share rents with workers through bargaining or profit-sharing, or compensate workers for less desirable working conditions. In turn, a low-ψ_j firm may be exercising monopsony power. The framework does not commit to a particular mechanism; instead, it treats ψ_j as a reduced-form measure of firm-level wage-setting. Differences in ψ_j therefore correspond to wage premia that apply to any worker employed at that firm.

In logarithms, wages are additively separable as $\alpha_i + \psi_j$. This additive structure yields transparent comparative statics. Holding the firm fixed, a worker with higher α_i earns proportionally more than a worker with lower α_i. Holding the worker fixed, differences in pay across firms equal the difference in their firm effects ψ_j. The model thus provides a simple accounting framework for wage dispersion across both workers and firms.

To complete the picture, one must describe how workers and firms match together. Workers may survey multiple firms and choose the one offering the highest wage. Alternatively, they may encounter firms sequentially and decide whether to move based on wages and nonpecuniary considerations. The AKM framework remains largely agnostic about these mechanisms. It allows for sorting—for example, high-productivity workers may be disproportionately employed at high-paying firms—and for mobility over time, provided that worker and firm components are stable and enter wages additively.

Observed wages, of course, do not perfectly equal $\alpha_i + \psi_j$. An empirical specification therefore includes a residual term. In its simplest form, this residual, a luck of the draw, is unrelated to the underlying matching process. More elaborate interpretations are possible, but the baseline specification abstracts from them.

The model is estimated using linked employer-employee data. Each observation corresponds to a worker i in period t. We will denote as j_{it} the firm that employs worker i in period t, and the observed compensation as Y_{it} (in logs). Because workers move across firms over time, these data allow researchers to disentangle persistent worker effects from firm-specific wage premia, and thereby to decompose overall wage dispersion into its constituent components.

The setting we will use as a running illustrative example throughout the paper concerns the role of firms in shaping wage inequality, which is a central question

addressed in the AKM literature. A key tool is a decomposition of wage dispersion into components that reflect dispersion in workers' α_i (*worker heterogeneity*), firms' ψ_j (*firm heterogeneity*), and whether high-α_i workers tend to work in high-ψ_j firms (*sorting*). In the cross-section, these decompositions are informative about the share of variance explained by firms and workers in the overall wage distribution. When applied to various countries, or to various sub-periods in the same country, they shed light on how the role of firm heterogeneity, worker heterogeneity, and sorting, vary across contexts and over time.

The Econometric Model

In the AKM model introduced by Abowd, Kramarz, and Margolis (1999), workers i are characterized by their components, or "effects," α_i, and firms j are characterized by their components, or "effects," ψ_j. The model postulates the following expression for the wage (in logs) that worker i would earn at time t if she was employed in firm j:

$$Y_{it}(j) = X_{it}\beta + \alpha_i + \psi_j + U_{it}.$$

The wage depends on the sum of the worker and firm components, $\alpha_i + \psi_j$. As a result, the difference in the wages that a worker would earn in firms 1 and 2 is precisely equal to the difference in their components $\psi_1 - \psi_2$. This wage differential between the two firms is the same for all workers and is constant over time. Hence, in Figure 1, workers 1 and 2 experienced exactly the same wage drop when moving, despite the fact that they moved in different periods.

The actual wage in AKM is not strictly equal to the sum of worker and firm components, for two reasons. First, wages also depend on some idiosyncratic shocks U_{it}. These shocks account for variation over time in a worker's wage even if she remains in the same firm. Second, wages depend in addition on some covariates X_{it} that are observed in the data—such as the labor market experience of the worker—with their associated coefficient β.

The quantity $Y_{it}(j)$ is the wage that would result from the worker being exogenously assigned to firm j. Hence, for any given worker in the economy, in any given period, the wage equation specifies *all* the wages that the worker could have earned in *all* possible firms. However, the data are not directly informative about those potential wages. The observed wage in the sample, say Y_{it}, corresponds to the firm that worker i is employed at in period t. Because we have denoted this firm as j_{it}, the wage that is recorded in the data is thus $Y_{it} = Y_{it}(j_{it})$.

It is important to note that there are typically many thousands of firms in the sample, and the data only provide information about the wage at one firm in a given period. Hence, recovering potential wages in all firms requires solving a formidable extrapolation problem. Remarkably, the structure of the AKM model gives researchers the ability to predict the wages of a worker in all (connected) firms in

the economy. In practice, this is achieved through the use of the AKM estimator that we will describe in the next section.

Under the AKM model, the average wage in the firm reflects a combination of firm and worker effects. Abstracting from covariates and idiosyncratic shocks for simplicity, the average wage in firm j is the sum of the firm component ψ_j and the average of the worker components α_i for the workers employed in the firm. A potential explanation for a high mean wage in the firm is that it has a high ψ_j, through which it consistently pays its workers a high wage. However, an alternative explanation for a high wage is that workers in the firm have high α_i, meaning that the firm employs workers who would earn good wages irrespective of where they work. Hence, average wages in the firm are not sufficient to disentangle firm and worker components.

The key feature of the model that allows researchers to separately recover worker and firm effects is job mobility, under the assumption that the latter is *exogenous*. As we indicated in the conceptual framework, the AKM model does not restrict how worker-firm matches, here denoted as j_{it}, are related to the worker and firm components, α_i and ψ_j. At the same time, the model does impose that the shock U_{it} be uncorrelated with the set of all worker-firm matches in the data. This assumption, which is referred to as *exogenous mobility* in the literature, is central to the methodology. Under exogenous mobility, the difference in average wages of workers who switch between firms is directly informative about the firms' components. While some evidence has been offered in support of exogenous mobility, the assumption has been debated and remains controversial. We will mention some limitations of the model's assumptions, as well as extensions of the model, in the last section of the article.

Taking stock, in the AKM setup, the worker's wage Y_{it} at match j_{it} is a linear function of the worker component, the firm component, and the covariates. Moreover, under the exogenous mobility assumption, the error term U_{it} is unrelated to the worker-firm matches. The ordinary least squares estimator is the most popular estimator in such settings, and it is the estimator on which the AKM approach relies. We present the estimator in the next section.

For example, Card, Heining, and Kline (2013) study how firm and worker heterogeneity shape the structure and evolution of wages in Germany. In their 2002–2009 sample, they follow 16 million workers. The AKM model provides predictions of the wages those workers would earn in all 1,500,000 firms (establishments) in the sample. For this reason, the authors argue that it is important to offer evidence in support of the model's assumptions, and they report a number of diagnostic checks to this end. In the last section of the article, we will mention several diagnostics and extensions of the AKM model.

The AKM Estimator

In this section, we describe the AKM estimator introduced by Abowd, Kramarz, and Margolis (1999). The starting point is that the researcher has access to linked

employer-employee data. These data contain a panel component, because they follow workers over time, as well as a firm component, because they keep track of the firm identifiers. Linked employer-employee datasets are now available in many countries.

Estimating a linear regression in typical applications of AKM requires handling the presence of a large number of parameters. In many linked datasets, there are hundreds of thousands or millions of workers and firms, hence the same number of worker- and firm-specific parameters to estimate. While initially challenging, estimation algorithms are now well understood, and efficient computational routines have been developed. A state-of-the-art computer package, which also includes the improvements to the original AKM approach that we will discuss in the last two sections of this article, is *pytwoway*.[2]

A specific feature of the AKM setting is that the variables on the right-hand side of the regression are not linearly independent. Dependence between right-hand-side variables arises mechanically because, in the wage equation, only the sum of worker and firm components can be identified but their separate levels are not. Hence, one cannot hope to recover the average α_i separately from the average ψ_j. This issue is straightforward to address by normalizing one of the coefficients, such as the average firm component in the sample, to zero.

Another reason why right-hand-side variables are linearly dependent is related to the *lack of connectivity* of the worker-firm network. Consider the case where the workers in a firm never leave the firm, and the firm does not hire new workers during the entire observation period. In this case, there is no way to know whether the workers have high α_i or the firm has a high ψ_j, and consequently the firm's ψ_j is not identified. More generally, identification is only secured within *connected sets* of the worker-firm graph. In such a set, it is possible to link any two firms by tracing out workers' movements.

For example, to identify the ψ_j components of firms 1 and 2 one can compare the wages of workers who are employed at both firms at some point in the sample (that is, $1 \leftrightarrow 2$ movers). Alternatively, if no worker moves between 1 and 2, one can consider another firm, say firm 3, compute the wage difference for $1 \leftrightarrow 3$ movers, and subtract of the wage difference for $2 \leftrightarrow 3$ movers.

The leading approach to deal with lack of connectivity is to focus on a single (typically, the largest) connected set of the worker-firm graph, often called "the" connected set. Abowd, Creecy, and Kramarz (2002) propose a simple algorithm to compute connected sets, which works as follows. Start by including a worker in the set, and include all the firms where the worker has been employed at some point during the sample period. Then, include in the set all the workers who have been employed at some point in those firms. This process is continued until the size of the set ceases to increase. In many applications, the largest connected set is large and contains most workers and most medium and large firms, yet it typically leaves out some small firms and their workers.

[2] The *pytwoway* computer package can be found at https://tlamadon.github.io/pytwoway/.

When focusing on the connected set, computing the AKM estimator requires solving a linear system. This is a large, yet sparse, linear system, as workers only visit a handful of firms during the period. Reliable computational routines with minimal memory requirements have been developed, notably the iterative "zigzag" algorithm of Guimarães and Portugal (2010).[3]

The zigzag algorithm works as follows. Instead of trying to jointly recover the parameters for the worker component α_i, the firm component ψ_j, and the coefficient β on other covariates, the algorithm recovers them one at a time in an iterative fashion. Indeed, given ψ_j and β, α_i can be estimated as a simple worker-specific average. In turn, given α_i and β, ψ_j can be estimated as a firm-specific average. Lastly, given α_i and ψ_j, β can be efficiently estimated by standard linear regression (because β is a "usual," low-dimensional parameter). Starting from some initial values, iterating between these three sets of parameters until convergence provides a fast and reliable estimation routine.

The most popular use of the AKM methodology is as a way to quantify variance components, using the following decomposition of the total variance of log wage residuals:

$$\underset{\text{total log wage variance}}{\text{var}(Y_{it} - X_{it}\beta)} = \underset{\text{worker variance}}{\text{var}(\alpha_i)} + \underset{\text{firm variance}}{\text{var}(\psi_{jit})} + \underset{\text{covariance}}{2\text{cov}(\alpha_i, \psi_{jit})} + \underset{\text{idiosyncratic variance}}{\text{var}(U_{it})}.$$

In this decomposition, the worker and firm components quantify how much of the dispersion in log wages can be attributed to dispersion in worker and firm effects, respectively. For example, in an economy where firms pay similar workers identically, ψ_j is constant among all firms and its variance is equal to zero. In such an economy, wage differences between firms solely reflect differences in the types of workers they employ.

An important term in the decomposition is the covariance. Note that the firm component ψ_j is evaluated at the actual worker-firm match j_{it}. A positive covariance thus indicates that high-α_i workers tend to work in high-ψ_j firms, while a negative covariance indicates the opposite. This covariance, and the associated correlation coefficient, are commonly interpreted as measuring the contribution of worker-firm *sorting* to the overall dispersion in log wages. Consistently with this interpretation, if workers moved randomly between firms and employment patterns were independent of worker and firm heterogeneity, then the covariance component would be equal to zero.

Decompositions of this form are commonly reported in empirical work. A direct extension is to decompose the total variance of log wages that includes covariate effects, by adding variance and covariance terms quantifying the contribution of covariates. Another common approach is to decompose the between-firm variance in addition to the total variance.

[3] An alternative estimation approach is to estimate the AKM model in first differences, exploiting that $Y_{it} - Y_{i,t-1}$ no longer depends on α_i. One can then estimate the ψ_j parameters using least squares, and recover the α_i parameters after the fact.

Returning to our example concerning the role of firms in shaping wage inequality, an important goal in the literature has been to assess the quantitative magnitude of worker variance, firm variance, and covariance in explaining wage dispersion. For example, Card, Heining, and Kline (2013) and Song et al. (2019) report estimates of variance components in the cross-section, and study how they evolve over time and help explain the evolution of wage inequality in Germany and the United States, respectively. We will mention some of their main empirical findings in the next section.

A common approach to estimation is to compute the variances and covariances of the estimated worker and firm effects α_i and ψ_j. For example, the estimator of the worker variance proposed by Abowd, Kramarz, and Margolis (1999) is simply the variance of the estimated worker parameters α_i. However, in many empirical applications of AKM, this approach leads to an estimator with a potentially large *bias*. In the second-to-last section of the article, we will mention alternative estimators of variances and covariances that correct for the bias and lead to more reliable estimates.

In addition to variance decompositions, the AKM method is commonly used for a variety of other purposes in the literature. In some settings, researchers are interested in the coefficient β of a covariate. In this case, the presence of worker and firm components on the right-hand-side of the regression captures potential confounding factors. As an example, Lavetti and Schmutte (2016) control for worker and firm effects when estimating compensating wage differentials for occupational fatality risk, thus accounting for the possibility that fatality rates may differ across workers and firms.

In other settings, researchers are interested in averages of firm components ψ_j for various groups of firms. As an example, Setzler and Tintelnot (2021) compare averages of ψ_j parameters for multinational and national firms. This methodology accounts for the fact that firms are heterogeneous within groups, while controlling for worker heterogeneity as well. Another example is the analysis and decomposition of inter-industry wage differentials pioneered by Abowd, Kramarz, and Margolis (1999) and studied recently in Card, Rothstein, and Yi (2024).

More broadly, AKM estimates of worker and firm components are also commonly used on the left-hand-side, or on the right-hand-side, of regressions. In addition, researchers often correlate worker and firm effects with other characteristics of workers and firms in order to relate them to interpretable dimensions that are measured in the data. Recovering AKM estimates of worker and firm components to later use them in a subsequent part of the analysis has become ubiquitous in applied labor economics.

Some Applications of AKM

Since its introduction in Abowd et al. (1999), the AKM methodology has been widely used to produce estimates of worker and firm effects and to shed light on

features of labor markets. Here we review several findings reported in the literature and then mention some applications of AKM to other fields within economics.

The starting point, and the motivation, for the analysis is the large share of wage dispersion that occurs *between firms*. In the United States and continental Europe, for example, the between-firm share of variance in log wages typically accounts for between 30 percent and 50 percent of the total variance. This indicates that wages differ sharply between firms, not only within. However, through the lens of the AKM model, these wage differences may reflect several distinct mechanisms.

Consider an extreme case where firms pay workers equally, that is, the ψ_j components are identical in all firms in the economy. It is possible to reconcile this hypothetical scenario with large wage differences between firms if they employ different types of workers. Although firms pay *similar workers* the same because ψ_j is constant, they pay *the workers they employ* differently because some firms employ high-α_i workers while other firms employ low-α_i workers. At the other extreme, substantial wage differences between firms are also consistent with workers being identical in the sense that their α_i components are the same, yet high-ψ_j and low-ψ_j firms pay workers differently.

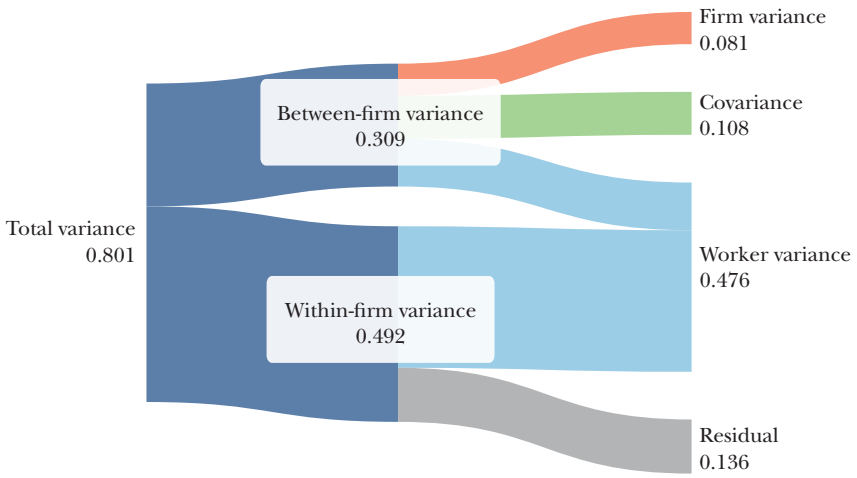
The AKM model and the use of linked worker-firm data allow researchers to identify these distinct mechanisms and shed light on the sources of the wage differences between and within firms. The literature has studied both the sources of cross-sectional wage dispersion and the factors explaining the evolution of wage inequality over time.

In the cross-section, all three shares of variance—that is, worker variance, firm variance, and covariance—have been found to explain sizable shares of the overall wage dispersion, although there is substantial variation in estimated relative shares across settings and periods. Across a set of empirical studies, Bonhomme et al. (2023) report that the interquartile range of the percentage of variance explained by firm components is 15 percent to 25 percent, while the corresponding range for the percentage of variance explained by the covariance term includes estimates larger than 15 percent as well as some negative estimates.[4]

As an example, in Figure 2 we report estimates from Song et al. (2019), based on the United States for the 2007–2013 period. The figure shows that the between-firm variance amounts to 39 percent of the total variance of log annual earnings. The AKM decomposition gives a sharper conclusion: 10 percent of the variation is explained by firms, 59 percent by workers, and 13 percent by the covariance, the remainder being explained by idiosyncratic shocks. Note that, as Figure 2 illustrates, while the firm variance and covariance arise solely from between-firm dispersion, the worker variance reflects a combination of within- and between-firm dispersion.

[4] These numbers are based on the original AKM method and are not corrected for bias.

Figure 2
Decomposition of the Variance of log Annual Earnings in the United States



Source: The figure is constructed from Tables III and IV in Song et al. (2019).
Note: The numbers correspond to the 2007–2013 period and are net of observed covariates. Worker variance is indicated in blue, firm variance in red, covariance in green, and residual variance in gray.

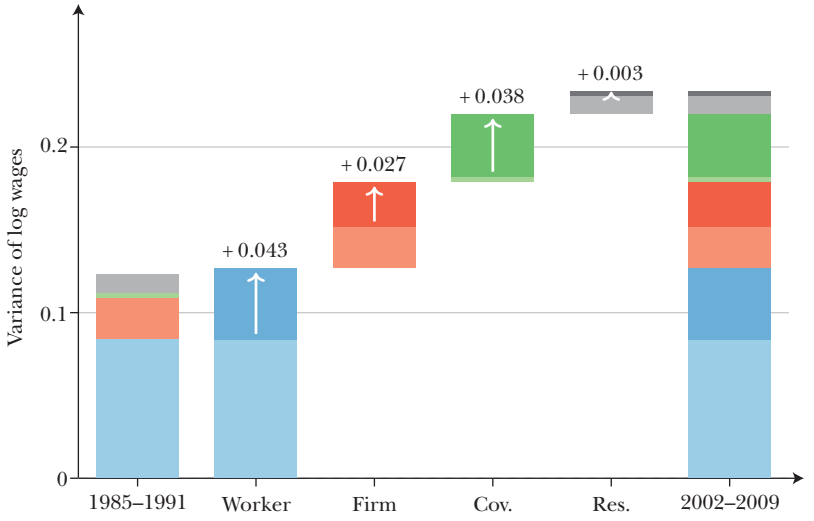
AKM decompositions are also commonly used to interpret changes in wage inequality over time. As an example, in Figure 3 we report changes between 1985–1991 and 2002–2009 in German daily wages, as estimated by Card, Heining, and Kline (2013). Out of a substantial 0.111 variance increase, the authors estimate that the change in worker variance accounted for 39 percent, the change in firm variance accounted for 24 percent, and the change in covariance accounted for 34 percent of the increase, with a small residual part explained by idiosyncratic shocks. Song et al. (2019) produced a similar analysis of the changes in earnings inequality in the United States between 1978 and 2013, also finding a substantial increase in the covariance term.

However, some of the findings based on the AKM methodology have been found to be fragile. An issue that has been extensively studied is the *bias* of variance shares, due to the imprecise estimation of the worker and firm parameters α_i and ψ_j. In the next section, we will explain the source of the bias and review some available approaches to alleviate the issue. Importantly, using available correction methods can make material differences to the estimates.

While Abowd, Kramarz, and Margolis (1999) proposed their method to study workers and firms, many other empirical settings have a similar structure, and AKM has been used in a variety of fields. Domains of application include the economics of education, health economics, international trade, and corporate finance, among others.

An example in health economics is Finkelstein, Gentzkow, and Williams (2016), who study the sources of geographic differences in health care utilization.

Figure 3
Change in the Decomposition of the Variance of log Daily Wages over Time in Germany



Source: The figure is constructed from Table IV in Card, Heining, and Kline (2013).
Note: The numbers are net of observed covariates. Worker variance is indicated in blue, firm variance in red, covariance in green, and residual variance in gray.

The existing disparity could reflect place-specific factors, such as doctors' incentives or the quality of hospitals. Alternatively, it could be due to differences in patients' levels of sickness or preferences. While the former explanation could be used to justify policies changing doctors' incentives, the latter cannot. To identify these two possible mechanisms, the authors exploit patients' migration across regions using the AKM methodology. They find that both patient-specific components and place-specific components contribute substantially to variation in utilization across areas, and that the covariance between these two components is positive.

An example in the context of corporate finance is Amiti and Weinstein (2018), who develop a methodology to separately recover bank-specific supply shocks and firm-specific demand shocks using lending data. Similarly to the AKM approach, their methodology accounts for the presence of firm and bank components in a linear regression. However, unlike in labor market settings where workers are not employed in multiple firms in a given period, at each point in time a firm may borrow from various banks—and, of course, banks lend to multiple firms as well. This permits researchers to estimate time-varying firm and bank parameters, interpreted as shocks. Using those in an analysis of the determinants of firms' investment decisions, the authors find that bank-specific shocks contribute significantly to aggregate investment fluctuations.

The Bias of AKM

Although the AKM methodology pioneered by Abowd, Kramarz, and Margolis (1999) has been highly influential, it suffers from some important limitations. Those can be separated into two categories. The first one concerns the issues with the AKM estimator, and chiefly the problem of bias. The second category concerns the model's assumptions, and whether the AKM model provides a good description of actual wages. We will now review these two types of limitations in this section and the next, while mentioning some extensions of the original approach that aim at providing improvements.

To understand the bias issue, it is important to note that estimates of worker and firm components α_i and ψ_j (which themselves are unbiased) are commonly contaminated by a substantial amount of *noise*. To shed light on the source of the noise, recall that differences between the ψ_j components of firms are identified by the wages of the workers moving between these firms. Hence, in order to reliably estimate firm components ψ_j, the number of job movers is key.

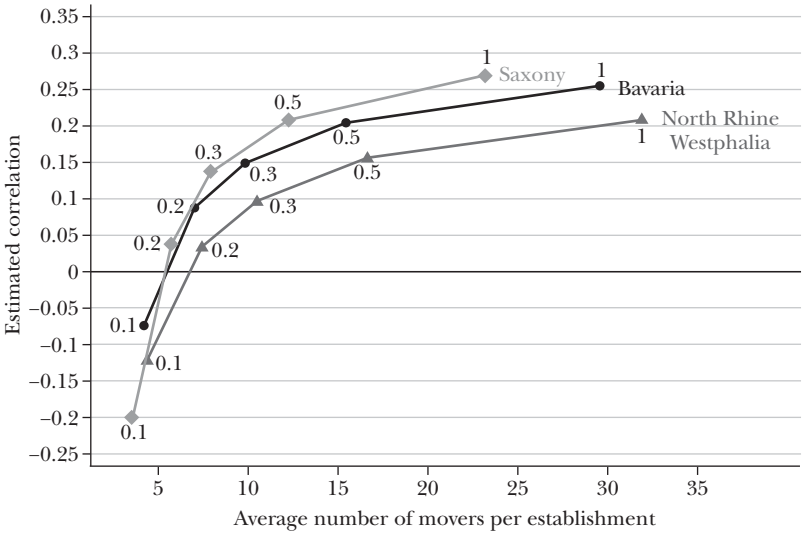
Moreover, as we indicated when we described the AKM estimator, the ability to recover firm and worker components hinges on a *connectivity* condition, which requires that any two firms be connected through job movers. When two firms are not connected—either directly or indirectly—it is not possible to recover the relative components ψ_j of these two firms. Likewise, when two firms are weakly connected through only a handful of job movers, the estimates of the firms' components tend to be noisy.

Due to insufficient job movers in some firms and an insufficient degree of connectivity, AKM estimates of firm and worker components are often noisy. For many quantities of interest such as variance shares, the noise in worker and firm effects creates a *limited mobility bias* on estimates of those quantities. Bias arises because estimation noise does not average out for quantities such as variances and covariances. Importantly, bias may still be substantial despite the presence of very large numbers of workers and firms. Qualitatively, AKM estimates tend to overestimate the contribution of firms to wage dispersion. In turn, the covariance between worker and firm components is also biased, but the bias tends to be downward so AKM tends to underestimate the contribution of the covariance term.

Nevertheless, concerns with bias do not apply equally to all quantities of interest. For example, if one is interested in the coefficient β of a covariate, or in comparing the average firm components in two groups of firms, then the estimation noise is likely to average out. However, it is difficult to know how particular dimensions of the data, such as the number of movers per firm or the degree of connectivity of the network, affect the noise in the AKM estimates and the resulting bias on the quantity of interest (Jochmans and Weidner 2019).

As a way to assess the importance of bias for the question at hand, we recommend implementing the *sub-sampling plot* proposed by Andrews et al. (2012). Start with a sub-sample of large and well-connected firms. Then, remove a random subset of job movers within each firm. As the share of movers removed increases, the

Figure 4
Sub-Sampling Experiment (German Data)



Source: The figure is reproduced from Andrews et al. (2012).
Note: The average number of job movers per establishment is shown in the x-axis, while the estimate of the correlation between worker and firm effects is shown on the *y*-axis. The three curves correspond to three German regions. The numbers 0.1, 0.2, 0.3, . . . indicate the share of the sample used in each estimation.

extent of the bias on the quantity of interest is likely to increase because firms in the sub-sample have fewer movers and tend to be less well-connected. The graph simply plots estimates of the quantity of interest on the y-axis against the share of job movers removed on the x-axis.

We show an example of a sub-sampling plot in Figure 4, taken from Andrews et al. (2012), using German data for 1998–2007. The authors first select a random sub-sample of 10 percent of workers, indicated as 0.1 in the figure. They then add back the workers to achieve 20 percent, 50 percent, and 100 percent of the original sample, respectively (in the largest connected set). Given a share of workers, they report the AKM estimate of the correlation between worker and firm effects on the y-axis and the average number of movers per establishment on the x-axis, separately for three German regions.

The figure shows that in smaller, less connected samples, the correlation is low or negative, while in larger, more connected samples the correlation is positive and quite large. The correlation coefficient is commonly interpreted as reflecting the sign and strength of sorting in the economy, and the figure provides clear evidence of bias that is consistent with AKM-based measures of covariances and correlations being biased downward. Similar findings have been reported for other countries and other quantities of interest, such as the variance of firm effects and the covariance component, as shown in Bonhomme et al. (2023).

In addition to reporting a sub-sampling plot, we recommend implementing *bias correction methods* and applying them to the quantity of interest. Available methods to correct the bias can be divided into two categories: fixed-effects methods and correlated random-effects methods. Reliable implementations of both approaches are available in the *pytwoway* package that we have already mentioned, and we now review them in turn.

Fixed-effects methods exploit linear regression algebra to construct an estimate of the bias on the quantity of interest, such as the variance of firm effects ψ_j. This construction requires specific assumptions on the idiosyncratic shocks. Under the assumption that shocks are independent and have a common variance, Andrews et al. (2008) propose an unbiased estimator of the variance of ψ_j obtained by subtracting an estimate of the bias. Their approach equally applies to other variance and covariance terms in standard decompositions. Kline, Saggio, and Sølvsten (2020) propose an extension of the correction that allows for unrestricted variance heterogeneity in idiosyncratic shocks.

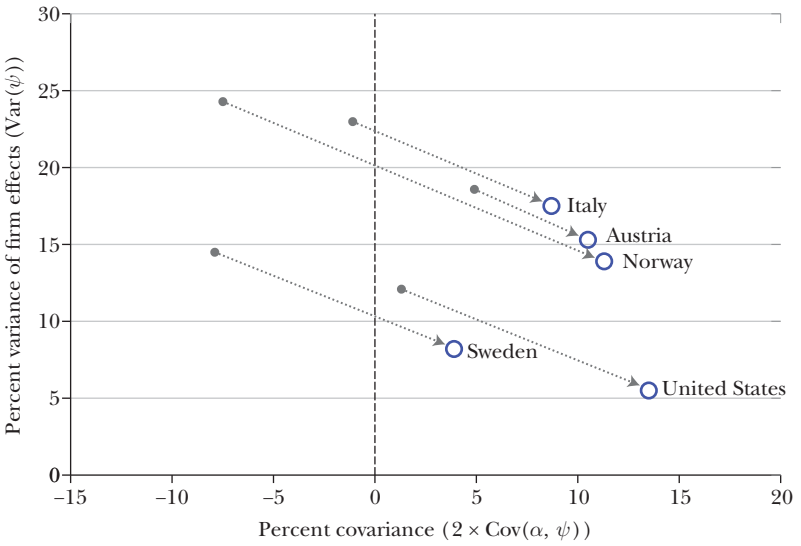
Correlated random-effects methods rely on a model for the distribution of worker and firm components α_i and ψ_j in the economy. Estimates of the parameters of this distribution can be used to produce estimates of the quantities of interest, including (but not limited to) variance shares. Moreover, from a Bayesian perspective, one can interpret the distribution as a prior and report posterior average quantities in the spirit of Empirical Bayes. Woodcock (2015) pioneered this approach in the AKM setting, while Bonhomme et al. (2023) and recently Cheng, Ho, and Schorfheide (2025) proposed and estimated more flexible models.

To illustrate the magnitude of the bias in practice, in Figure 5 we plot original estimates based on AKM as well as their bias-corrected counterparts. We report results for five countries, using data from Bonhomme et al. (2023). We show the bias correction proposed by Andrews et al. (2008), but other methods (Kline, Saggio, and Sølvsten 2020; Bonhomme et al. 2023) tend to give similar or larger differences on these data. The figure shows that the original estimates of the variance share of firm effects (on the y-axis) range between 12 and 24 percent of the total variance, whereas their bias-corrected counterparts range between 5 and 17 percent. In turn, the AKM estimates of the covariance share (on the x-axis) are negative for three of the five countries, whereas the bias-corrected counterparts are all positive, ranging between 4 and 13 percent of the total variance. These findings confirm that, without correction, AKM estimates of variances tend to be overstated while covariances tend to be biased downwards, and they demonstrate that the correction can have a material impact.

The Frontier of AKM

The AKM approach introduced by Abowd, Kramarz, and Margolis (1999) has been highly influential in labor economics and other fields. However, the model's assumptions are not uncontroversial, and they are the subject of a growing literature.

Figure 5
Impact of Bias Correction on Variance Components



Source: The figure is constructed from Table F.2 in the online appendix to Bonhomme et al. (2023).
Note: We show the share of variance explained by the variance of firm effects on the y-axis and the share explained by the covariance between worker and firm effects on the x-axis. The numbers correspond to six-year panels, in the largest connected set. We show uncorrected AKM estimates as dots, and bias-corrected estimates from Andrews et al. (2008) in empty circles.

AKM fundamentally relies on two key assumptions: that the log wage depends on the sum of the worker and firm components $\alpha_i + \psi_j$ (*additivity*), and that the idiosyncratic shocks are unrelated to the worker-firm matching process (*exogenous mobility*). When interpreted through the lens of economic models of the labor market, both assumptions are restrictive.

Additivity rules out the presence of interactions between the worker and the firm components $\alpha_i \times \psi_j$. However, in many models following the classical theory of sorting proposed by Becker (1973), production complementarities between worker and firm inputs are key drivers of sorting patterns. Hence, assuming a particular additive functional form for how worker and firm components affect wages may be restrictive, as emphasized by Eeckhout and Kircher (2011).

In turn, exogenous mobility implies that the AKM model cannot account for a worker leaving the firm because of a negative wage shock. Moreover, the model does not allow for the history in past firms to affect a worker's wage. For example, only the new firm affects the wage after a job move, but the model rules out an effect of the previous firm. These features are at odds with search models with wage posting or offer/counteroffer mechanisms (Postel-Vinay and Robin 2002). They also preclude the presence of time-varying, firm-specific human capital.

In reaction to these concerns, researchers have developed a number of diagnostics aimed at assessing the plausibility of the AKM model's assumptions. Notably, Card, Heining, and Kline (2013) present an event study graph where they plot wage changes around a job move event. They interpret the graph as suggesting that additivity between worker and firm effects may be a good approximation, and that the absence of a pre-mobility "dip" in earnings alleviates concerns about the exogenous mobility assumption.

However, such diagnostics may not fully resolve concerns about the AKM assumptions. Researchers have documented empirical violations of additivity, finding that different groups, such as men and women, or racial groups, are affected by different firm effects (Card, Cardoso, and Kline 2016; Gerard et al. 2021). Such findings suggest that AKM's key assumption that the firm effect ψ_j is the same for all workers does not hold. Moreover, the static nature of AKM is difficult to reconcile with the influential body of work on dynamic economic models of job mobility and wage determination.

The literature is only starting to explore model specifications that account for mechanisms absent in AKM. Bonhomme, Lamadon, and Manresa (2019) propose and estimate a wage model that allows for interaction effects between worker and firm heterogeneity. In addition, the framework they introduce allows for dynamic effects of past firms on future wages and relaxes the exogenous mobility assumption. Their findings based on Swedish data suggest that a log-additive model provides a reasonable approximation to the variance of log wages. At the same time, they find evidence of interactions between worker and firm effects, dynamic effects of past firms, and failure of exogenous mobility, all of which are assumed away in AKM. Abowd, McKinney, and Schmutte (2019) propose a model of wages and job flows that explicitly allows for endogenous mobility by letting the decision to leave the firm be influenced by match quality.

A parallel development is the growing literature on structural models of the labor market with two-sided—worker and firm—heterogeneity. As recent examples, Sorkin (2018) estimates workers' preferences for firms and relate them to AKM firm effects; Card, Cardoso, and Kline (2016) and Lamadon, Mogstad, and Setzler (2022) bring in information about the value added of the firm and use structural models of job choice and wage determination to study the pass-through of productivity shocks; Hagedorn, Law, and Manovskii (2017), Lentz, Piyapromdee, and Robin (2023), and Lamadon et al. (2024) estimate dynamic structural models of wages and mobility with two-sided heterogeneity; and Borovičková and Shimer (2024) study the selection issue caused by failure of exogenous mobility within a search model of the labor market.

Methodological developments are also needed for other settings, beyond the traditional labor market applications of the AKM methodology. For example, interaction effects between the α_i of different workers are likely to be important in settings with economic spillovers or team production (Arcidiacono et al. 2012; Cornelissen, Dustmann, and Schönberg 2017; Ahmadpoor and Jones 2019; Bonhomme 2021). Likewise, accounting for complementarity between managers' talent and

the tasks they perform requires extending the AKM model by allowing for interactions between α_i and ψ_j (Crippa 2025).[5] Methods to group similar individuals and firms, say, together (Bonhomme, Lamadon, and Manresa 2019) are useful to reduce the dimensionality of the model and relax various of the key assumptions in AKM. However, there is a need for more theoretical work and applications.

■ *The authors acknowledge helpful comments by Martyn Andrews, Erik Hurst, Jeffrey Kling, Timothy Taylor, and Heidi Williams.*

References

Abowd, John M., Robert H. Creecy, and Francis Kramarz. 2002. "Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data." Unpublished.

Abowd, John M., Francis Kramarz, and David N. Margolis. 1999. "High Wage Workers and High Wage Firms." *Econometrica* 67 (2): 251–333.

Abowd, John M., Francis Kramarz, and Simon Woodcock. 2008. "Econometric Analyses of Linked Employer–Employee Data." In *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice*, edited by László Mátyás and Patrick Sevestre, 727–60. Springer.

Abowd, John M., Kevin L. McKinney, and Ian M. Schmutte. 2019. "Modeling Endogenous Mobility in Earnings Determination." *Journal of Business and Economic Statistics* 37 (3): 405–18.

Ahmadpoor, Mohammad, and Benjamin F. Jones. 2019. "Decoding Team and Individual Impact in Science and Invention." *Proceedings of the National Academy of Sciences* 116 (28): 13885–90.

Amiti, Mary, and David E. Weinstein. 2018. "How Much Do Idiosyncratic Bank Shocks Affect Investment? Evidence from Matched Bank-Firm Loan Data." *Journal of Political Economy* 126 (2): 525–87.

Andrews, Martyn J., Len Gill, Thorsten Schank, and Richard Upward. 2008. "High Wage Workers and Low Wage Firms: Negative Assortative Matching or Limited Mobility Bias?" *Journal of the Royal Statistical Society Series A: Statistics in Society* 171 (3): 673–97.

Andrews, Martyn J., Leonard Gill, Thorsten Schank, and Richard Upward. 2012. "High Wage Workers Match with High Wage Firms: Clear Evidence of the Effects of Limited Mobility Bias." *Economics Letters* 117 (3): 824–27.

Arcidiacono, Peter, Gigi Foster, Natalie Goodpaster, and Josh Kinsler. 2012. "Estimating Spillovers Using Panel Data, with an Application to the Classroom." *Quantitative Economics* 3 (3): 421–70.

Becker, Gary S. 1973. "A Theory of Marriage: Part I." *Journal of Political Economy* 81 (4): 813–46.

Bonhomme, Stéphane. 2020. "Econometric Analysis of Bipartite Networks." In *The Econometric Analysis of Network Data*, edited by Bryan Graham and Áureo de Paula, 83–121. Academic Press.

Bonhomme, Stéphane. 2021. "Teams: Heterogeneity, Sorting, and Complementarity." Preprint, arXiv. https://doi.org/10.48550/arXiv.2102.01802.

Bonhomme, Stéphane, Kerstin Holzheu, Thibaut Lamadon, Elena Manresa, Magne Mogstad, and Bradley Setzler. 2023. "How Much Should We Trust Estimates of Firm Effects and Worker Sorting?" *Journal of Labor Economics* 41 (2): 291–322.

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa. 2019. "A Distributional Framework for Matched Employer Employee Data." *Econometrica* 87 (3): 699–739.

[5] Two recent studies by Mourot (2025) and Villacorta, Villacorta, and Gutierrez (2023) emphasize the importance of surgeon—hospital and bank—firm interactions, respectively.

Borovičková, Katarína, and Robert Shimer. 2024. "Assortative Matching and Wages: The Role of Selection." NBER Working Paper 33184.

Card, David, Ana Rute Cardoso, Joerg Heining, and Patrick Kline. 2018. "Firms and Labor Market Inequality: Evidence and Some Theory." *Journal of Labor Economics* 36 (S1): S13–70.

Card, David, Ana Rute Cardoso, and Patrick Kline. 2016. "Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women." *Quarterly Journal of Economics* 131 (2): 633–86.

Card, David, Jörg Heining, and Patrick Kline. 2013. "Workplace Heterogeneity and the Rise of West German Wage Inequality." *Quarterly Journal of Economics* 128 (3): 967–1015.

Card, David, Jesse Rothstein, and Moises Yi. 2024. "Industry Wage Differentials: A Firm-Based Approach." *Journal of Labor Economics* 42 (S1): S11–59.

Cheng, Xu, Sheng Chao Ho, and Frank Schorfheide. 2025. "Optimal Estimation of Two-Way Effects under Limited Mobility." NBER Working Paper 34014.

Cornelissen, Thomas, Christian Dustmann, and Uta Schönberg. 2017. "Peer Effects in the Workplace." *American Economic Review* 107 (2): 425–56.

Crippa, Federico. 2025. "Identification, Estimation, and Inference in Two-Sided Interaction Models." Preprint, arXiv. https://doi.org/10.48550/arXiv.2510.22884.

Eeckhout, Jan, and Philipp Kircher. 2011. "Identifying Sorting—In Theory." *Review of Economic Studies* 78 (3): 872–906.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. 2016. "Sources of Geographic Variation in Health Care: Evidence from Patient Migration." *Quarterly Journal of Economics* 131 (4): 1681–726.

Gerard, François, Lorenzo Lagos, Edson Severnini, and David Card. 2021. "Assortative Matching or Exclusionary Hiring? The Impact of Employment and Pay Policies on Racial Wage Differences in Brazil." *American Economic Review* 111 (10): 3418–57.

Guimarães, Paulo, and Pedro Portugal. 2010. "A Simple Feasible Procedure to Fit Models with High-Dimensional Fixed Effects." *Stata Journal* 10 (4): 628–49.

Hagedorn, Marcus, Tzuo Hann Law, and Iourii Manovskii. 2017. "Identifying Equilibrium Models of Labor Market Sorting." *Econometrica* 85 (1): 29–65.

Jochmans, Koen, and Martin Weidner. 2019. "Fixed-Effect Regressions on Network Data." *Econometrica* 87 (5): 1543–60.

Kline, Patrick M. 2024. "Firm Wage Effects." NBER Working Paper 33084.

Kline, Patrick, Raffaele Saggio, and Mikkel Sølvsten. 2020. "Leave-Out Estimation of Variance Components." *Econometrica* 88 (5): 1859–98.

Lamadon, Thibaut, Jeremy Lise, Costas Meghir, and Jean-Marc Robin. 2024. "Labor Market Matching, Wages, and Amenities." NBER Working Paper 32687.

Lamadon, Thibaut, Magne Mogstad, and Bradley Setzler. 2022. "Imperfect competition, Compensating Differentials, and Rent Sharing in the US Labor Market." *American Economic Review* 112 (1): 169–212.

Lavetti, Kurt, and Ian M. Schmutte. 2016. "Estimating Compensating Wage Differentials with Endogenous Job Mobility." Unpublished.

Lentz, Rasmus, Suphanit Piyapromdee, and Jean-Marc Robin. 2023. "The Anatomy of Sorting—Evidence from Danish Data." *Econometrica* 91 (6): 2409–55.

Mourot, Pauline. 2025. "Should Top Surgeons Practice at Top Hospitals? Sorting and Complementarities in Healthcare." Unpublished.

Postel-Vinay, Fabien, and Jean-Marc Robin. 2002. "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity." *Econometrica* 70 (6): 2295–350.

Setzler, Bradley, and Felix Tintelnot. 2021. "The Effects of Foreign Multinationals on Workers and Firms in the United States." *Quarterly Journal of Economics* 136 (3): 1943–91.

Song, Jae, David J. Price, Fatih Guvenen, Nicholas Bloom, and Till von Wachter. 2019. "Firming Up Inequality." *Quarterly Journal of Economics* 134 (1): 1–50.

Sorkin, Isaac. 2018. "Ranking Firms Using Revealed Preference." *Quarterly Journal of Economics* 133 (3): 1331–93.

Villacorta, Alonso, Lucciano Villacorta, and Bryan Gutierrez. 2023. "(Unobserved) Heterogeneity in the Bank Lending Channel: Accounting for Bank-Firm Interactions and Specialization." Preprint, SSRN. http://dx.doi.org/10.2139/ssrn.4608656.

Woodcock, Simon D. 2015. "Match Effects." *Research in Economics* 69 (1): 100–121.

Journal of Economic Perspectives—Volume 40, Number 2—Spring 2026—Pages 215–242

The Economics of Paid Sick Leave

Stefan Pichler, Christopher Prinz,
Stefan Thewissen, and Nicolas R. Ziebarth

Paid sick leave was integral to the world's first public health insurance system. In 1884, Germany introduced mandatory public health insurance for blue-collar workers. Among the core benefits was paid sick leave, paid at 50 percent of the wage, starting on the third day of a sickness spell and lasting up to 13 weeks (Deutsches Reichsgesetzblatt 1883). At the time, paid sick leave spending made up half of overall health insurance expenditures (Finkenstädt 2010). Today, most high-income countries provide "universal" or "statutory" access to paid sick leave, although "universal" often excludes the self-employed and, in some countries, also leaves out part-time employees or those who work for small firms.

Taking sick leave is common. In Germany in 2022, 66 percent of all employees called in sick at least once. Over five years, however, this share goes up to more than 90 percent. Moreover, in a given year, 5 percent experience a long sickness spell of at least six weeks, and such long-term sickness accounts for 40 percent of all paid sick leave days and associated costs (Ziebarth 2013; Gürtzgen and Hiesinger 2025; AOK Bundesverband 2025). In other words, as with healthcare use and spending, the

■ *Stefan Pichler is Associate Professor of Public Health, University of Groningen, Groningen, Netherlands. Christopher Prinz is a Senior Labour Market Analyst, OECD (Organisation for Economic Cooperation and Development), Paris, France. Stefan Thewissen is Professor of Inclusive and Sustainable Wellbeing and Pensions, Nyenrode Business University, Breukelen, Netherlands. Nicolas R. Ziebarth is Professor of Economics, University of Mannheim, and Head of Research Unit "Labour Markets and Social Insurance," ZEW Mannheim Leibniz Centre for European Economic Research, both in Mannheim, Germany. Ziebarth is the corresponding author at nicolas.ziebarth@zew.de.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20251479.

distribution of sick-leave days is heavily skewed with mass at zero and a long right tail (Ziebarth and Karlsson 2010, 2014; Atal et al. 2025).

In contrast, the United States does not guarantee universal paid sick leave. Since 2004, a Healthy Families Act has been repeatedly introduced in Congress, but has never passed. The recent version of the Healthy Families Act (2023)[1] envisions granting private-sector employees the right to earn and use up to eight sick days annually at 100 percent of their wages. While legislation is gridlocked at the federal level, 18 states, along with dozens of cities and counties, have enacted paid sick leave since 2007, using the Healthy Families Act as a model. These regional laws are all employer mandates to require individual sick time savings accounts, akin to health savings accounts (Goodman and Musgrave 1992). By working, employees accumulate sick time credit that does not expire. Workers then take their paid sick leave when it yields the greatest individual benefit (Pauly et al. 1995; Feldstein and Altman 2007).[2]

As a result, the share of US private sector jobs with paid sick leave has risen from 63 to 77 percent between 2010 and 2023. Moreover, firms have provided paid sick leave voluntarily at increasing rates. The substantial rise in jobs with paid sick leave is perhaps one of the most significant yet overlooked developments in the US labor market over the past decade. This 14-percentage-point increase is particularly notable given that the share of US employees with employer-provided health insurance (about 72 percent) or paid vacation (about 78 percent) barely budged over the same period (Pichler and Ziebarth 2024; BLS 2025).

Figure 1 plots average sick leave rates for 2023 by country. It uses consistently elicited labor force survey data, which ask workers how many days of work they missed due to sickness in the previous week, which are then divided by the total number of contracted workdays. As seen, sick leave rates vary substantially across OECD countries—from around 1 percent of contracted workdays in Korea, the United Kingdom, and the United States to more than 4 percent in Chile, Germany, and France. It seems plausible that these differences reflect differences in sick leave design, rather than differences in demographics or workplace conditions.

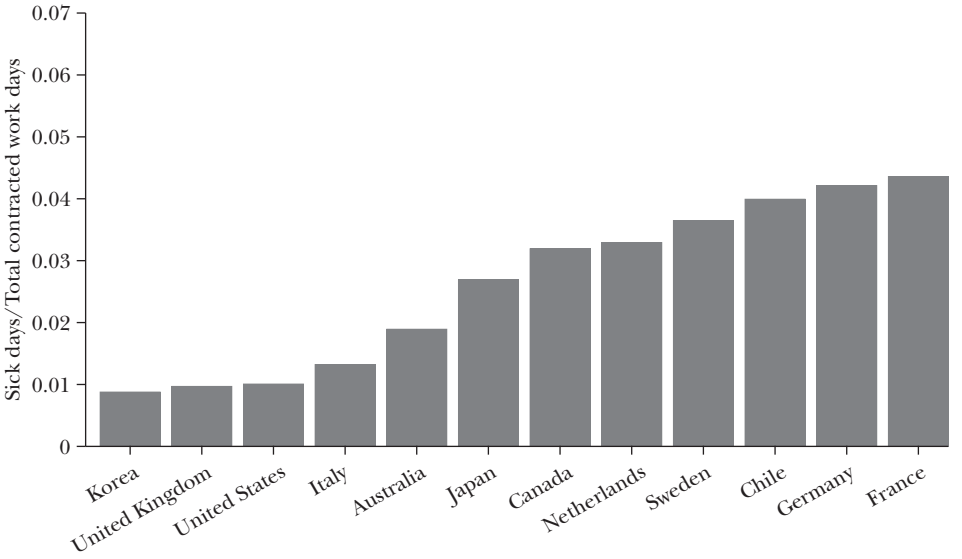
We begin by reviewing the theory and evidence on the economics of paid sick leave, focusing on potential economic justifications for government-mandated sick leave. In this context, the strongest rationale likely concerns the negative externalities arising from the spread of contagious diseases when employees work while sick. COVID-19 has made this case particularly clear: Access to paid sick leave provides economic incentives for workers to self-isolate, thereby reducing transmission risks (Pichler and Ziebarth 2017).

We then turn to the design of paid sick leave schemes. We first delve into the policy parameters of paid sick leave: what level of income is replaced, for how long,

[1] Healthy Families Act, S.1664, 118th Congress (2023).

[2] For public state employees such as public school teachers, sick leave policies usually stipulate that unused credit can be cashed out at retirement and/or in the form of increases in public pensions (Cronin, Harris, and Ziebarth forthcoming). The US state-level mandates largely lack such provisions.

Figure 1
Selected International Sick Leave Rates



Source: Authors' illustration for a selection of countries based on data from the European Union Labour Force Survey (EU-LFS), the Current Population Survey (CPS) from the United States, the Canadian Labour Force Survey (CA-LFS), the UK Labour Force Survey (UK-LFS), the Australian Bureau of Statistics (ABS) Labour Force Survey for Australia, Encuesta Nacional de Empleo (ENE/LFS) conducted by the Instituto Nacional de Estadísticas de Chile, the Korea National Health and Nutrition Examination Survey (KNHANES), and Nagata et al. (2023) for Japan.
Note: These representative labor force surveys all use identical questions on sickness-related absence from work. Data are from 2023 and show sick leave rates by country, based on survey data asking how many days of work employees missed in the previous week due to sickness, and then dividing the total number of sick days by the total number of contracted workdays. Data are for 2023, except for Australia and Chile (2024). Data for Australia use actual hours and workdays rather than contracted hours and days. Data for Korea and Japan have a recall period of one month. Data for Korea are based on all employees, while data for the remaining countries are based on full-time employees. Self-reports may contain recall biases, even more so the longer the recall period. Note that reported sick leave is a function of sick pay access, generosity, labor force composition, and types of jobs in the industry.

and so on. Even within Europe, countries vary considerably in the duration of paid sick leave they offer and in the replacement rate. For example, Germany's sick leave mandate requires employers to pay 100 percent of wages for up to six weeks. In contrast, the employer mandate in the Netherlands covers spells up to 104 weeks, with a statutory minimum replacement rate of 70 percent of the regular gross wage. However, the Netherlands is clearly an outlier when it comes to the extensiveness of their employer mandate. The other end of the spectrum includes Italy, which has no employer mandate. Instead, Italy's paid sick leave is paid out by its social security system from day four of a sickness spell, up to 20 days, at 50 percent of wages.

Most OECD countries, including the United States, provide access to short-term sick leave—that is, for high-frequency but relatively mild conditions like the common

cold—through employer mandates. By contrast, most of these countries organize their longer-term sick leave through separate state institutions, such as a part of social security systems, sometimes called "statutory sickness insurance."

Furthermore, most OECD countries design their paid sick leave schemes in a manner analogous to unemployment insurance, which requires traditional insurance tools to limit moral hazard. These include formal monitoring, waiting periods, and replacement rates as a fraction of regular wages. When employees are on sick leave, they generally remain employed and are expected to return to their previous jobs. Only when the work-disabling condition is expected to last permanently do employees typically apply for disability insurance, which exists in all OECD countries via separate social insurance institutions (Banks, Blundell, and Emmerson 2015; Koning and Lindeboom 2015; MacVicar et al. 2022). Workers generally exhaust their sick leave before transitioning to disability benefits, though some countries either make no formal distinction between sick leave and disability benefits or allow paid sick leave of unlimited duration.

Ultimately, the provision of paid sick leave through government regulation faces a central trade-off. On one side, ensuring universal and generous access to paid sick leave may encourage excessive absenteeism from work ("moral hazard"). On the other side, lacking universal access to paid sick leave may produce excessive "presenteeism" where workers come to work sick, with lower productivity, a higher risk of workplace injury, and a greater chance of infecting coworkers and customers. The differences in sick leave rates in Figure 1 suggest that some countries may have either too much inefficient absenteeism or too much inefficient presenteeism in the provision of paid sick leave.

Sick Pay Mandates: Theory and Rationale

The United States, in the period before the implementation of the 18 state-level mandates, gives a sense of how sick leave access emerges in an unregulated labor market. In 2011, based on calculations from the National Compensation Survey by the Bureau of Labor Statistics, 63 percent of all US jobs came with (voluntary) paid sick leave provision for short-term sickness. However, heterogeneity by job type was substantial: 76 percent of full-time employees but only 28 percent of part-time employees had paid sick leave coverage. Further, private-sector employees in large firms with more than 100 employees were more likely to be covered by paid sick leave than those in firms with fewer than 100 employees (74 percent versus 54 percent). Unionized jobs had a higher probability of coming with paid sick leave than nonunionized jobs (72 percent versus 63 percent). Notably, the income gradient was steep: only 33 percent of jobs in the bottom quartile of the wage distribution offered paid sick leave, compared with 86 percent in the top quartile.

So why did a substantial number of employers in 2011 voluntarily offer paid sick leave when not required to do so? Indeed, why do many US employers do so today, even when not required by state or local law?

First, offering paid sick leave may attract and retain qualified or scarce labor (Dube, Lester, and Reich 2016). Second, offering paid sick leave incentivizes employees to recover and self-isolate when their productivity is low due to illness, and when they may be contagious. Third, these first two reasons suggest that offering paid sick leave (at least partly) pays for itself. Moreover, to the extent employees value paid sick leave, they would accept a lower wage in return, potentially even lower than the cost to employers of offering it (Summers 1989). Finally, social norms or employer altruism may also play a role in the decision to provide paid sick leave voluntarily.

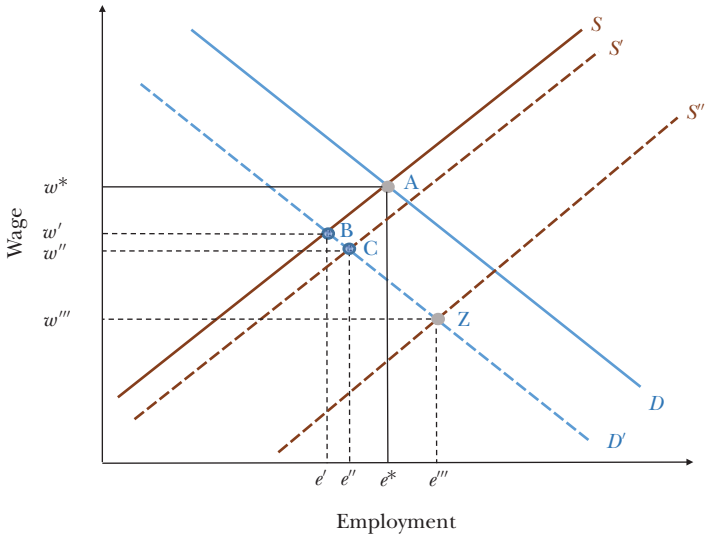
Critics of government regulation argue that employers—not the state—are best positioned to decide whether to offer paid sick leave. Also, they argue that employees—not the state—know their preferences best and are thus best positioned to decide whether to work in jobs with paid sick leave. Because employees have heterogeneous preferences, shaped, for example, by their risk tolerance (Adams-Prassl et al. 2023), some maximize their utility when choosing a job with paid sick leave but a lower wage, while others may prefer to self-insure by accepting a higher wage in a job without paid sick leave (Maestas et al. 2023). Thus, under the assumptions of a frictionless labor market, rational agents, and no externalities, economic theory implies that it is optimal for employers and employees to negotiate paid sick leave as a fringe benefit in exchange for lower wages. Workers will then sort into jobs with and without paid sick leave, depending on their preferences. Under these assumptions, the outcome of such an unregulated market for paid sick leave—say, something comparable to the US economy circa 2011—would be efficient.

The Basic Theory of Mandating Paid Sick Leave

If the government mandates paid sick leave, firms that already provide it voluntarily are unaffected by the regulation. By contrast, firms that did not voluntarily offer paid sick leave must now provide it. Labor costs increase. As the textbook example in Figure 2 illustrates, the firm's labor demand curve thus shifts downwards by the amount it costs the firm to provide paid sick leave, from D to D'.

In addition, the employees' labor supply curve shifts outwards by the amount that employees value the benefit. Here, we can distinguish between three main cases. First, if employees do not value the benefit at all, their labor supply curve will not shift and stay as S. In this case, wages and employment both fall from equilibrium A (with e^* and w^*) to B (with $e' < e^*$ and $w' < w^*$). Second, if employees value the benefit a little bit, but not as much as it costs employers to provide, employment falls by less than under the first case and wages fall by more: the labor supply curve shifts outwards by the amount that employees value the benefit, here to S', to reach the new equilibrium C (with $e'' > e'$ and $w'' < w'$). In the third case, if employees value the mandated benefit by *more* than it costs employers to provide, the labor supply curve shifts substantially to S'' and the new equilibrium Z has the lowest wages of all scenarios ($w''' \ll w^*$), but the highest employment level ($e''' > e^*$), even higher than before the mandate's introduction.

Figure 2
Wage and Labor Supply Effects under Sick Pay Mandates



Source: Authors' illustration based on Summers (1989).
Note: The graph shows potential labor supply and labor demand shifts following paid sick leave mandates. "A" is the initial equilibrium. See main text for details.

Empirical evidence suggests that the United States lies somewhere around point Z in Figure 2. For example, Maestas et al. (2023) find that US employees highly value paid time off and would be willing to forgo 16 percent of their wages in exchange for 10 days of paid leave—although 10 days of leave in a work-year of 250 days would cost only 4 percent of their wages. In other words, employees value paid time off at a multiple of its cost. In addition, there is little evidence that employer mandates for paid sick leave have significantly reduced overall employment (or wage) growth in US labor markets (Pichler and Ziebarth 2020; Woods, Schneider, and Harknett 2025). Slopen (2024) even finds a 1 percentage point increase in women's employment due to mandated sick pay, with the effect particularly pronounced among women with children and without college degrees.

There are several explanations for why wages in the United States do not appear to decrease substantially due to paid sick leave mandates. First, implied relative wage effects are small and difficult to identify empirically. For example, most sick pay mandates allow employees to earn 1 hour of paid sick time for every 40 hours worked, which is 2.5 percent of wages in a static calculation. In practice, Maclean, Pichler, and Ziebarth (2025) estimate an actual labor cost increase of $0.06 per hour worked due to mandates, on average across all firms. For context, this is 0.3 percent of employees' hourly gross wage of $22.63, while health insurance costs average $2.42 per hour (BLS 2025).

Second, being able to take sick leave when being sick may enhance workers' own and their coworkers' labor productivity, as we discuss in more detail in the next section (Pichler, Wen, and Ziebarth 2020; Chunyu, Volpin, and Zhu 2024). Taking reduced contagion effects into account would thus reduce the cost pressure for employers; the labor demand curve would then shift by less than D' in Figure 2.

Contagion Externalities

Perhaps the most compelling economic rationale for mandating paid sick leave is contagion externalities, that is, the externalities arising from the spread of contagious diseases. Since COVID-19 shocked the world, the case for mandated paid sick leave to reduce infection externalities has gained substantial traction (OECD 2020). The argument is straightforward: When employees lack paid sick leave, they "work sick"—a phenomenon known as "presenteeism"—and spread contagious diseases. One driver of presenteeism is that contagiousness is at least partially unobservable, so firms, workers, coworkers, and customers cannot internalize the risks and costs of contagion (Pichler and Ziebarth 2017). In this context, paid sick leave serves an essential economic function by incentivizing workers to self-isolate.

The contagion externality argument weakens as the share of employees who work remotely in the economy increases. For example, in the United States, full days worked at home have risen from 7 percent pre-COVID in 2019 to 28 percent in 2023 (Barrero, Bloom, and Davis 2023). However, the numbers also imply that 60 percent of all US jobs are still fully onsite. Furthermore, occupations with particularly low sick leave coverage rates, absent any mandates, were "Food preparation & serving" (around 20 percent) and "Retail and related" (around 50 percent), as documented by Maclean, Pichler, and Ziebarth (2025). In these service industry occupations, remote work is difficult, if not impossible.

Several researchers have studied contagion effects by focusing on the expansion of sick leave access through employer mandates in the United States. Using data from before the main expansion of paid sick leave mandates from the 2011 American Time Use Survey, Susser and Ziebarth (2016) estimate that about three million US employees worked while sick every week. Several empirical reduced-form articles provide evidence of "contagious presenteeism"—working sick with contagious diseases. Pichler and Ziebarth (2017) offer a theoretical model in which employers cannot perfectly observe contagiousness: in this model, contagious presenteeism behavior decreases when employees gain access to paid sick leave. Empirically, Pichler and Ziebarth (2017) exploit the staggered enactment of paid sick leave mandates in 8 ("treated") US cities, linking these policy changes to weekly Google Flu (2015) data for a total of 81 cities (8 treated and 73 control) from 2003 to 2015 (N=48,333). Their difference-in-differences analysis shows that population-level influenza-like illness rates (defined as having fever of 100°F (37.8°C) or greater, accompanied by a cough and/or sore throat) declined by roughly 6 percent in the first two years after mandate adoption. Because Google Flu trends correlate strongly, but not perfectly, with actual test-based rates of influenza-like illness collected by

the Centers for Disease Control and Prevention (CDC), Pichler, Wen, and Ziebarth (2020) replicate and reinforce these findings using state-level variation in paid sick leave mandates, difference-in-differences models, and CDC's Weekly US Influenza Surveillance Report from 2010 to 2018 ($N = 20{,}319$). In this setting, they estimate a decline in population-level rates of influenza-like illness of about 11 percent in the first post-mandate year.

There is even some evidence, albeit mixed, that access to paid sick leave may *reduce* overall sick leave use—taken literally, this finding would imply that the reduction in contagious presenteeism outweighs the increase in use due to more absenteeism. Specifically, Stearns and White (2018) find that Connecticut's (very limited) paid sick leave mandate reduced overall sick leave rates, employing a measure for sick leave use in the Current Population Survey's basic monthly files.[3] However, Maclean, Pichler, and Ziebarth (2025) cannot confirm this finding. Using firm-job level data from the National Compensation Survey conducted by the US Bureau of Labor Statistics, Maclean, Pichler, and Ziebarth (2025) study the effects of enacting 13 state-level mandates from 2010 to 2022 ($N = 443{,}740$). As expected, when expanding paid sick leave access, sick leave use increases by about two sick days in the first year among those newly covered by paid sick leave. Consistent with that finding, using the 2011 and 2017/2018 waves of the American Time Use Survey Leave Supplement, Callison and Pesko (2022) provide evidence for a reduction in "unmet sick leave needs," that is, fewer respondents who say they needed to take time off from work due to their own illness in the past week, but could not.

In April 2020, when COVID-19 hit the United States with full force, one of the legislative responses was the first federal paid sick leave provision ever—the Families First Coronavirus Response Act (2020).[4] This law temporarily granted up to two weeks of emergency sick leave for COVID-19-related symptoms for firms with fewer than 500 employees, funded via an advanced tax credit. At least 8 million employees used COVID-19 emergency sick leave during 2020, while an estimated 15 million per week still reported unmet sick leave needs (Jelliffe et al. 2021). Nevertheless, the spread of the disease was less dynamic due to the law (Pichler, Wen, and Ziebarth 2020; Andersen et al. 2023).

Taken together, high-quality, mostly causal evidence shows that contagious presenteeism and infection activity significantly decrease when paid sick leave is expanded through government regulation. Note that this evidence is primarily based on the United States and its approach of implementing short-term sick leave via individualized sick time credit accounts.

[3] The US Current Population Survey, like the EU-Labour Force Survey, asks whether employees worked fewer than their standard hours in the past week and then elicits reasons, among them "own illness" (see Figure 1). To our knowledge, this is the only consistently collected sick leave measure across the countries.

[4] Families First Coronavirus Response Act, H.R.6201, 116th Congress (2019–2020).

Adverse Selection and Moral Hazard

Paid sick time is a form of insurance, and thus potentially susceptible to the problems that all insurance markets face: adverse selection and moral hazard. These arguments are also behind "Some Simple Economics of Mandated Benefits," in which Summers (1989) lists reasons that could justify an employer mandate for certain benefits.

The standard insurance market phenomenon of adverse selection arises due to information asymmetries (Pauly 1974; Rothschild and Stiglitz 1976). In the case of sick leave, some employees may have a high latent demand for paid sick leave that employers cannot observe. Workers with existing health issues could then sort into jobs that offer paid sick leave and use it extensively. As employers would be aware of this "adverse selection" sorting dynamic, they may deliberately underprovide sick leave.

Thus, adverse selection offers a plausible explanation for the low job coverage rates of paid sick leave in unregulated labor markets, despite strong latent worker demand and willingness to trade off wages (as mentioned above). However, we are unaware of empirical evidence on adverse selection in sick leave markets. The lack of evidence is both a data problem (Chiappori and Salanie 2000) and due to a lack of natural variation: institutional settings often do not provide enough observable between- and within-firm variation to study whether more generous sick leave attracts a sicker workforce, while most available data even entirely lack the information needed to study adverse selection empirically.

Moral hazard arises when insurance coverage reduces the incentives to avoid a loss, thereby increasing the probability that the insured take risks and use the benefit when covered. Thus, moral hazard can take two conceptual forms. Economists distinguish between *ex ante moral hazard* before becoming sick ("acting riskier because of paid sick leave") and *ex post moral hazard* after becoming sick ("given sickness, more use of paid sick leave"). In reality, it is challenging to distinguish between these two phenomena empirically. Instead, economists have estimated the increase in utilization resulting from more generous insurance coverage.

For the United States, Maclean, Pichler, and Ziebarth (2025), as noted earlier, find a causal increase of about two paid sick days taken in the first year after private-sector employees gain access to paid sick leave under a mandate. The implied elasticity is around 0.9 on the extensive margin—for a 10 percent increase in worker access to paid sick leave, use increases by 9 percent. Studying paid sick leave generosity on the intensive margin among US public school teachers, Cronin, Harris, and Ziebarth (forthcoming) show that a 10 percent higher balance of sick leave credit increases sick leave use by 4.5 percent on a given school day.

Other Externalities

Additional potential externalities might also justify mandating paid sick leave. In the broader discussion of employer mandates, Summers (1989) argues that the high share of the uninsured in the United States imposes significant externalities on hospitals, physicians, and consumers due to uncompensated care, which must

be provided to the uninsured for ethical and legal reasons. A similar argument applies to paid sick leave if sick employees, who cannot take sick leave to recover, face an increased risk of long-term sickness or permanent work incapacity and produce higher public disability or healthcare spending. While empirical evidence documenting such long-term externalities for health insurance coverage has grown strong, particularly for children (Goodman-Bacon 2018, 2021; East et al. 2023), there is a lack of data and variation for high-quality causal effect studies that would measure how paid sick leave affects health and labor market outcomes in the long run.[5] In a study of the Spanish public sector, Marie and Vall Castelló (2023) find an increased relapse rate and more work accidents after the wage replacement rates of paid sick leave were cut. The findings illustrate how externalities can arise with low paid sick leave generosity on the intensive margin.

Basic Worker Rights

Societies may want to ensure that all workers have access to what is perceived as a basic worker right or "merit good." The representative National Paid Sick Days Study (Smith and Kim 2010), conducted before the state-level mandates and COVID-19, found that 69 percent of Americans view paid sick leave as a "very important" workers' right that should be guaranteed by the government. The poll also indicates strong bipartisan support, with even 64 percent of "strong Republican" supporters viewing paid sick leave as "a very important" workers' right. A follow-up poll during the COVID-19 pandemic uncovered similar bipartisan support for paid sick days without specifically inquiring about worker rights or government regulation (National Partnership for Women and Families 2020).

Also, policymakers may view an employer mandate for paid sick leave as a means to reduce job inequality. As witnessed in the US labor market, without government regulation, low-wage and service sector workers, minorities, and women are less likely to have access to paid sick leave on their job (Susser and Ziebarth 2016; Johnson et al. 2022; Harknett and Schneider 2022; Maclean, Pichler, and Ziebarth 2025). Note that, in many OECD countries, paid sick leave also entails the right to call in sick to care for sick children, and sometimes also for sick parents or other relatives.

[5]There is, however, evidence that access to paid sick leave causally leads to more preventive care visits (Callison et al. 2023) and a stronger involvement with the healthcare system (Gupta et al. 2025). Regarding disability insurance, Stepner (2019) finds that short- and long-term disability insurance are intertemporal complements rather than substitutes in Canada. In other words, he does not find that short-term disability insurance ("long-term sick leave") reduces the probability of long-term disability insurance, but increases it. However, the proposed mechanism is moral hazard that arises because Canada's short-term disability insurance covers waiting periods for long-term disability insurance. Similar findings with US data are inconclusive (Autor et al. 2013). Still, novel research suggests a positive causal relationship driven by disability claims for children due to more informal caregiving and more healthcare utilization (Assamidanov et al. 2025).

Key Design Choices for Paid Sick Leave

The theoretical arguments for government regulation of paid sick leave do not specify a program design. Table 1 provides a concise overview of sick leave systems across a dozen OECD countries from five continents, including the United States, Canada, the United Kingdom, France, Germany, Italy, the Netherlands, and Sweden, as well as Korea, Japan, Chile, and Australia. It focuses on the key design choices for paid sick leave: waiting periods, monitoring requirements, wage replacement, and maximum payment duration.

For policymakers outside the United States and Australia—the countries that operate short-term sick leave through individualized credit accounts—these policy parameters involve trade-offs that determine a system's generosity and its incentives. As such, these parameters also affect the share of employees who engage in inefficient absenteeism versus inefficient presenteeism.

Waiting Periods and Monitoring

Two approaches that countries use to reduce the risk of moral hazard and excessive absenteeism are waiting periods and monitoring requirements. Several countries have waiting periods of one or several days before sick leave can be claimed (OECD 2020). Waiting periods imply that sick pay on these days, by federal minimum standards, is zero—although employers are always free to provide more generous sick pay than stipulated by the statutory minimum conditions. Waiting periods reduce the rate of shirkers, but increase the rate of those who work while sick. Moreover, waiting periods lead to sharp benefit discontinuities that can result in unintended consequences.

A randomized field experiment conducted in Sweden in 1988 abolished the waiting day and increased replacement rates for longer durations (Pettersson-Lidbom and Thoursie 2013). The combined effect was a *reduction* in sick leave rates, although the system became more generous. Employees on sick leave went back to work more quickly after the waiting day was abolished. Apparently, Sweden's one-day waiting period incentivized sick employees to prolong existing sickness spells to avoid the risk of another uncompensated day in the event of a relapse. In line with this reasoning, Barone (2023) finds that optimal paid sick leave schemes should avoid sharp benefit discontinuities and zero-sick-pay waiting periods.

Monitoring requirements are a second crucial element in containing moral hazard, especially in European-style "traditional" sick leave schemes (as opposed to US- or Australian-style sick time credit schemes) that operate similarly to unemployment insurance. Initially, monitoring is typically carried out by (primary care) physicians who assess applicants' work capacity and issue sick notes. The requirement for such sick notes varies by country, as shown in Table 1. For long-term spells, sick pay schemes often require recertification by a doctor from the social institution.

More restrictive monitoring requirements reduce absence rates and the share of shirkers, but also increase the share of employees who come to work sick and spread diseases, work at low productivity, or experience work-related accidents

Table 1

Paid Sick Leave Policy Parameters in Selected OECD Countries

Country	Waiting period	Monitoring	Short-term sick leave replacement rate	Short-term sick leave maximum duration	Long-term sick leave replacement rate	Long-term sick leave maximum duration
Europe						
France	3 days	Doctor's note from day 3	90% paid by employer[a]	60 days	50% paid by health insurance	12 months in 3 years[b]
Germany	No	Doctor's note from day 4	100% paid by employer	42 days per illness	70% paid by health insurance	78 weeks in 3 years
Italy	3 days	Doctor's note from day 4	50% paid by social insurance	20 days	67% paid by social insurance	180 days
Netherlands	2 days	Meeting with company doc after 6 weeks[c]	70% paid by employer	-	No separate system	2 years (no separate system)
Sweden	No	Doctor's note within 8 days	80% paid by employer	14 days	80% paid by social insurance	No time limit
United Kingdom	3 days	Doctor's note from day 8	GBP 119 (USD 160) per week paid by employer	-	No separate system	196 days (no separate system)
Asia						
Japan	3 days[d]	Doctor's note from day 4	66.67% paid by health insurance	-	No separate system	18 months
South Korea	N/A[d]	N/A	N/A	N/A	N/A[e]	N/A
Americas						
United States	N/A[f]	N/A; some states require doctor's notes after 3–4 days	100% of available credit	Up to available credit	N/A	N/A
Canada	1 week (no STSL, N/A)	Doctor's note when applying for LTSL	No federal STSL for first week of sickness spell	N/A	55% paid by employment insurance	26 weeks
Chile	3 days[d,g]	Doctor's note from day 1	100%	10 days	100% paid by health insurance	As long as certified as temporary
Oceania						
Australia	N/A	Only if employer demands	100% of available credit	10 days, unused leave carries over	N/A	N/A

Note: Own collection and categorization. Data for 2025. The overview excludes top-up payments as specified in collective agreements which exist in basically all countries. Only federal minimum regulations are displayed. STLT=short-term paid sick leave, LTSL=long-term paid sick leave ("medical leave", "TDI"), N/A=not applicable.

[a] Level and duration is tenure specific. Typically, 90 percent for 30 days followed by 66.66 percent for another 30 days.

[b] Payable for up to three years (without waiting period) for long-term sickness.

[c] The meeting with the company doctor can be earlier if demanded by the employer.

[d] No employer mandate for short-term paid sick leave.

[e] No institutional system for long-term paid sick leave, except for government officials who receive 100 percent for up to 60 sick days per year without a waiting period (short-term leave) and 50–70 percent for 24 months (long-term leave).

[f] Mandated credit accounts in 18 states and Washington, DC, see Maclean, Pichler, and Ziebarth (2025) for more details.

[g] Retroactive 100 percent wage replacement for the first three days if the sick leave spell exceeds ten days.

(*continued*)

Table 1
Paid Sick Leave Policy Parameters in Selected OECD Countries (*continued*)

Source:

France: Assurance Maladie (2025): Arrêt de travail pour maladie : les indemnités journalières du salarié, April 9, 2025, https://www.ameli.fr/assure/remboursements/indemnites-journalieres-maladie-maternite-paternite/arret-maladie-salarie. Retrieved Sept. 17, 2025. Ministère du Travail (2025): L'indemnisation légale des absences pour maladie ou accident, July 7, 2025, https://travail-emploi.gouv.fr/lindemnisation-legale-des-absences-pour-maladie-ou-accident. Retrieved Sept. 17, 2025. Legifrance (2025): Code du Travail, https://www.legifrance.gouv.fr/codes/id/LEGISCTA000018537772. Retrieved Sept. 17, 2025.

Germany: Bundesministerium für Gesundheit (2025): Krankengeld, Bundesministerium für Gesundheit, April 24, 2025, https://www.bundesgesundheitsministerium.de/krankengeld.html. Retrieved Sept .17, 2025.

Italy: INPS (2025): Indennità di malattia e visite mediche di controllo, Istituto Nazionale della Previdenza Sociale, July 15, 2025, https://www.inps.it/it/it/dettaglio-approfondimento.schede-informative.indennit–di-malattia-e-visite-mediche-di-controllo.html. Retrieved Sept. 17, 2025.

Netherlands: UWV (n.d.): Loondoorbetaling bij ziekte, https://www.uwv.nl/nl/ziek/loondoorbetaling. Retrieved Sept. 17 2025; Business Gov NL (n.d.): Reintegration obligations for employers, https://business.gov.nl/regulation/reintegration-obligations. Retrieved Sept. 17, 2025.

Sweden: Försäkringskassan (2025): Sick pay, Feb. 28, 2025, https://www.forsakringskassan.se/english/for-employers/illness-and-injury/sick-employee-days-1-90/sick-pay. Retrieved Sept. 17, 2025. Norden (n.d.): Sickness benefit in Sweden, https://www.norden.org/en/info-norden/sickness-benefit-sweden. Retrieved Sept. 17, 2025.

United Kingdom: Gov.uk (n.d.): Statutory Sick Pay. https://www.gov.uk/statutory-sick-pay. Retrieved Sept. 17, 2025.

Japan: IBM Japan Health Insurance Society (n.d.): If you take time off from work due to sickness. https://www.ibmjapankenpo.jp/eng/member/benefit/rest_a.html. Retrieved Sept. 17, 2025. IBM Japan Health Insurance Society (n.d.): FAQ—Injury and Sickness Allowance. https://www.ibmjapankenpo.jp/asp/faq/faq.asp?articleid=6174. Retrieved Sept. 17, 2025.

South Korea: MOEL (n. d.): Labor Standards, Ministry of Employment and Labor. https://www.moel.go.kr/english/policy/laborStandards.do. Retrieved Sept. 17, 2025. MOHW (2022): Pilot project on paid sickness benefits (press release), Ministry of Health and Welfare, July 4, 2022. https://www.mohw.go.kr/board.es?act=view&bid=0032&list_no=372137&mid=a20401000000&nPage=14&tag=. Retrieved Sept. 17, 2025. OECD (2023): Disability, Work and Inclusion in Korea, OECD Publishing, Paris. https://doi.org/10.1787/bf947f82-en

United States: Currently no federal law, FFCRA has expired, see Department of Labor (2020) https://www.dol.gov/sites/dolgov/files/WHD/Pandemic/FFCRA-Employer_Paid_Leave_Requirements.pdf. Retrieved Sept. 20, 2025. Overview of State-Level Mandates including links to laws are in A Better Balance (2025): https://www.abetterbalance.org/our-issues/paid-sick-time/

Canada: Government of Canada (2025): Employment Insurance Sickness Benefits, Jan. 10, 2025. https://www.canada.ca/en/services/benefits/ei/ei-sickness.html. Retrieved Sept. 17, 2025.

Chile: Superintendencia de Salud (n.d.): Licencia Médica. https://www.superdesalud.gob.cl/tax-temas-de-orientacion/licencia-medica-2403. Retrieved Sept. 17, 2025. Barone, V. (2023). On the design of paid sick leave: A structural approach. Job Market Paper, https://viquibarone.github.io/baronevictoria/Barone_paidsickleave.pdf, retrieved Jan. 23, 2025.

Australia: Fair Work Ombudsman (n. d.): Sick and carer's leave. https://www.fairwork.gov.au/leave/sick-and-carers-leave. Retrieved Sept. 17, 2025.

(Haller, Staubli, and Zweimüller 2023; Marie and Vall Castelló 2023). Monitoring also imposes costs on employees and the healthcare system, which is why most OECD countries do not require doctors' notes from day one of a disease. However, all OECD countries require some form of medical documentation after a certain duration of illness. Public health insurers, unemployment agencies, and/or employers often have additional options to assess employees' work capacity while on sick leave.

For example, unemployed individuals may report sick for strategic reasons to avoid sanctions by the unemployment agency and to extend the benefit duration (van den Berg, Foerster, and Uhlendorff forthcoming). In Italy, doctors from the social security agency carry out checks and visit employees on sick leave at home. Shirkers are then liable to disciplinary action, including getting fired (D'Amuri 2017).

A monitoring policy can also include mandatory return-to-work activities. For example, in the Netherlands and according to the Gatekeeper Improvement Act (Ministry of Social Affairs and Employment 2002), during the first eight weeks on sick leave, employers and employees must develop a workplace reintegration plan; occupational or firm physicians oversee and assess employees' work capacity every six weeks and recommend (graded) return-to-work (Kools and Koning 2019; Godard, Koning, and Lindeboom 2024).

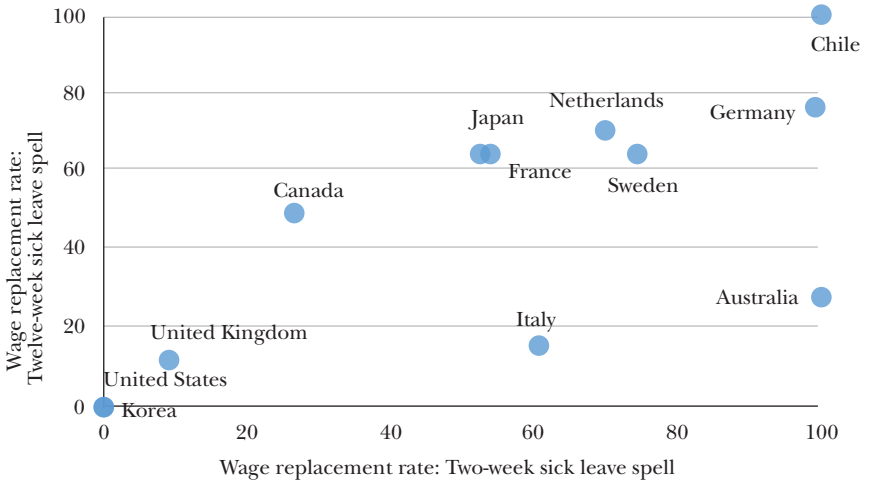
Replacement Rates and Benefit Durations

In traditional "pooled" systems for paid sick leave, laws stipulate minimum statutory wage replacement rates of typically less than 100 percent. The reasoning follows basic economic logic, which suggests cost-sharing to avoid overutilization (Manning et al. 1987). The trade-offs for higher versus lower replacement rates are again intuitively clear: Lower replacement rates lead to less shirking, but more presenteeism. In addition, lower-income populations and those with higher care needs are more likely to be sick and require more healthcare and sick leave. As a result, cost-sharing and less generous benefits tend to be regressive. As shown in Table 1, wage replacement rates vary from 0 percent to 100 percent for short-term sick leave in our set of twelve OECD countries.

Figure 3 summarizes selected countries' paid sick leave generosity by plotting federal, statutory wage replacement rates for a two-week (x-axis) and a twelve-week (y-axis) sickness episode. Given that the United States has no federal sick leave provision, it ranks at the bottom. Especially in conjunction with the rather restrictive credit accounts that provide up to six annual *days* of paid sick leave for full-time employees, many employees are likely engage in inefficient presenteeism in the United States, while inefficient absenteeism is likely less of an issue. On the contrary, countries that rank at the top of the paid sick leave generosity distribution—Germany or Chile—likely exhibit high rates of "inefficient absenteeism" and shirking. For example, a recent German poll suggests that 10 percent of Germans admit to calling in sick "often" (and another 23 percent "from time to time"), despite being able to work (Pronovo BKK 2024). Comparing these replacement rates to the sick leave rates in Figure 1, we obtain a correlation of 0.66 between the sick leave rate and the two-week wage replacement rate and of 0.93 for the four-week wage replacement rate.

As seen in Table 1, the statutory benefit duration also shows substantial heterogeneity across OECD countries, ranging from 0 percent due to waiting days to six weeks of employer-provided sick pay at 100 percent of the gross wage in Germany and up to 104 weeks at 70 percent in the Netherlands. Usually, when employer-provided (short-term) sick leave ends, public social insurance institutions—may it be health insurance, unemployment insurance, or other social security institutions—approve

Figure 3
Generosity of International Sick Leave Systems



Source: Author's visualization based on OECD (2023).
Note: Table 1 provides details on specific paid sick leave policy parameters such as wait times, benefit durations, and replacement rates. The categorization here shows average wage replacement rates for a two- and twelve-week sickness spell, respectively, solely considering federal minimum standards. Employers can always top up these federally regulated minimum benefits.

and pay out sickness benefits; often, a new maximum benefit duration and replacement rate are determined when the responsibility is transferred from the employer to the public system. It is worth noting that there is no definitive, internationally recognized definition of what constitutes short- and long-term sick leave. In the United States, a week of sick leave may be considered "long-term," whereas in Germany, it has become common to refer to the lower, health-insurer-provided sick pay after week six as "long-term sick leave."

In most countries, replacement rates decline with the benefit duration, a pattern also common in unemployment insurance. What are the trade-offs between rising or falling replacement rates?

First, note that labor supply elasticities, and thus distortions as we will see below, tend to fall the longer the duration of the illness (Ziebarth 2013; Böckerman, Kanninen, and Suoniemi 2018). This supports Barone's (2023) argument that replacement rates should rise with the benefit duration, as the distortions from inefficient absenteeism are smaller when the labor supply elasticity is smaller. Rising replacement rates also lead to better risk protection and less inequality, as they redistribute toward the long-term sick, who have permanently lower incomes (Meyer and Mok 2019). On the other side, rising replacement rates also imply less generous benefits for short-term sick leave—and thus more contagious presenteeism. In addition, as long as labor supply elasticities are nonzero, rising replacement rates weaken return-to-work incentives and may foster long-term work incapacity.

Finally, the longer the illness and benefit duration, the lower the risk of contagiousness, which usually lasts only a few days. However, many European countries with relatively long durations offer long-term sick leave in combination with work reactivation measures such as compulsory dialogue meetings (Markussen, Røed, and Schreiner 2018), workplace accommodation (Maestas, Mullen, and Rennane 2019), or graded return-to-work requirements (Kools and Koning 2019). Fundamentally, the trade-offs are a faster return-to-work with the risk of relapses and lower productivity (Marie and Vall Castelló 2023) versus costly and unnecessarily long sick leave episodes that may result in permanent labor market withdrawal (Hernæs 2018).

The Structure and Financing of Paid Sick Leave

High-income countries also differ in how they organize their sick leave systems. Most of these countries have employer mandates for short-term sick leave, but separate social insurance institutions covering long-term sick leave. However, as mentioned, there is no definitive "official" distinction between short-term and long-term sick leave, and there is no general consensus that short-term sick leave must be provided through an employer mandate and long-term sick leave through what we henceforth call "risk-pooled statutory paid sick leave."

Moreover, the underlying structure and financing of sick leave systems pose policy choices of their own. The main US approach, as noted earlier, involves an employer mandate under which employees accumulate sick leave credits, which they can use when they choose. The government could also reimburse firms for sick leave with revenue coming from general taxes. Or it could levy firm-specific, possibly experience-rated, premiums or wage-proportional contribution rates shared between firms and workers. In what follows, we begin with the choice between "risk-pooled" versus "credit account" approaches and then discuss the choice between funding options.

Program Design: Pooled Versus Credit Accounts

When analyzing the design of paid sick leave systems, the United States remains an outlier. There is no permanent US federal regulation to ensure access to either short-term or long-term sick leave. As of writing, the 18 US states (along with Washington, DC, and dozens of US cities) that have implemented employer mandates for short-term sick leave have implemented individualized sick pay credit accounts, akin to health savings accounts for routine healthcare visits (Pauly et al. 1995). In theory, the government could directly top up sick time credit accounts, similar to health savings accounts and common practice in, for example, Singapore (Sun 2025).[6]

[6]When COVID-19 hit and a bipartisan bill established emergency sick leave for COVID reasons within weeks, immediate funding was provided through advanced tax credits for firms, as part of the 2020 Families First Coronavirus Response Act (Pichler, Wen, and Ziebarth 2020; Jelliffe et al. 2021).

These US employer mandates commonly earmark 2.5 percent of work time as sick time credit—one hour per 40 hours worked. This sick time credit can be taken as needed at a 100 percent wage rate, often on an hourly basis, with minimal monitoring. It can also be used to care for sick relatives, such as sick children. Unused credit carries over to the following year. Critics of individualized credit accounts point toward inequality concerns and the low generosity for people with high needs, for example, the chronically sick or those who have many children. Australia also operates such a system at the national level.

In addition, several US states have implemented long-term sick leave, called "medical leave" (and "temporary disability insurance") for private sector employees. It is established via separate institutions and funded through payroll taxation of between 0.2 and 1.2 percent of gross wages; that is, through risk-pooled paid sick leave. In most states, these long-term benefits are provided for up to twelve weeks, ranging from 50 percent (New York) to 100 percent (Oregon) of the qualifying gross wage.

The US approach of sick time credit accounts for short-term sickness and risk-pooled statutory sick leave for long-term sickness resembles what has been called Consumer-Driven Health Insurance, in which medical savings accounts cover routine outpatient visits, while high-deductible traditional health insurance covers major health shocks and diseases (Pauly et al. 1995).

In contrast, most OECD countries have opted for a traditional risk-pooled social insurance design along the lines shown in Table 1 and discussed in the previous section. Such a risk-pooled statutory design entails paid sick leave with (1) replacement rates as a fraction of the wage, sometimes combined with (2) one or several waiting days before benefits kick in, (3) a maximum duration up to which benefits can be received, and (4) monitoring through medical professionals who issue doctors' notes.

Risk-pooled statutory paid sick leave systems tend to have much more generous coverage per sickness episode as can be seen in Table 1.[7] In such systems, countries set a minimum replacement rate that employers can top up, for example, through individual or collective bargaining. One main disadvantage of these risk-pooled statutory sick leave models is that moral hazard is primarily contained through cost-sharing and monitoring via doctors' notes, both of which are imperfect (Markussen, Røed, and Røgeberg 2013).

Table 2 summarizes differences between (1) a risk-pooled statutory paid sick leave system and (2) an individualized sick pay credit account system. The primary advantage of an individual credit account system is that individual intertemporal incentives replace formal monitoring by doctors. After all, physicians usually do not have strong incentives to scrutinize their patients' sick leave requests. Under credit accounts, formal monitoring is less needed, as sick time credit is individualized and carries future value (Cronin, Harris, and Ziebarth forthcoming).

[7] In principle, of course, individualized credit accounts could be designed to be more generous than "pooled" paid sick leave.

Table 2
Trade-Offs in Pooled Versus Credit Paid Sick Leave Systems

	Risk-pooled statutory paid sick leave	Individualized sick time credit accounts
Formal monitoring ("doctor's note")	Inherent feature, typically required after several days, see Table 1	Not needed due to intertemporal incentives; doctors' notes are often prohibited in the United States
Shirking ("moral hazard")	Inherent concern; the more so, the more generous the benefits	Controlled by individualized design, carry over, and cash out rules
Contagious presenteeism ("working sick with infectious disease")	Exists even in most generous schemes, but open-endedness (no cap on the number of spells within a year) lowers the risk	Individualized credit without explicit mechanism to internalize externalities
Benefit generosity (waiting days, replacement rate, duration)	For long-term sickness and the unhealthy, usually more generous than saving accounts, as open-ended and pooled	For short-term sickness and the healthy, 100 percent replacement rates; unused credit tends to carry a monetary value
Inequality	Those on sick leave incur costs as replacement rates often are $<$ 100 percent; otherwise, low inequality due to open-endedness and insurance pooling with maximum benefit durations up to 52 weeks at 100 percent (Norway), see Table 1.	No wage losses for accumulated credit; otherwise, high inequality as credit is usually capped at six to eight sick days per year in the United States (see previous row)

Source: Author's summary.

While doctors' notes are not instrumental in an individual credit account system due to the inherent intertemporal incentives that sick time credit accounts provide, they are also explicitly prohibited under most state-level sick pay mandates. One reason is that requiring doctors' notes in the United States would impose a very high cost on those without health insurance or high cost-sharing. Even in states without sick pay mandates, employers may voluntarily waive doctors' note requirements, as requiring them also implies higher healthcare use and thus higher spending and higher future insurance premiums if they provide employer-sponsored health insurance.

Not having to rely on doctor's notes as the key monitoring instrument is also a big advantage as, unsurprisingly, physician leniency in issuing doctors' notes varies tremendously (Riise, Willage, and Willén 2025). For example, Germany has free choice of providers and little to no cost-sharing in its public health insurance system. Hence, if one doctor refuses to approve a patient's sick leave request, the patient can go doctor-shopping.

Further, assessing work capacity is costly on the patient and the provider side. It includes type I ("doctor's note despite work capacity") and type II ("no doctor's note despite work incapacity") errors (Low and Pistaferri 2025). A range of medical conditions are inherently difficult for medical practitioners to verify, similar to assessing permanent work disability in disability insurance (Autor and Duggan 2006; Cao et al. forthcoming). Sick leave due to such hard-to-verify conditions, like mental health, has more than tripled in Germany since 2000 (Ziebarth and Pichler 2024; Best et al. 2025).

On the other hand, contagious presenteeism is likely better contained in classic risk-pooled social insurance systems due to their open-endedness and the absence

of a limit on the number of days covered. However, no data exist for researchers to benchmark shirking and contagious presenteeism rates by country. Conceptually, a key trade-off would be to weigh the share of employees who engage in inefficient absenteeism ("shirking") versus inefficient presenteeism ("working sick with low productivity"). An optimal system would seek to minimize both phenomena.

Funding Paid Sick Leave through Employer Mandates Versus Taxes

Paid sick leave can be financed directly by employers, with some portion of the cost likely passed on to employees as lower wages. Or it could be financed by direct government provision of benefits, funded by general or firm-specific taxes. The usual pattern across OECD countries is that employers are mandated to provide (and fund) short-term sick leave, whereas long-term sick leave is provided by tax-funded social insurance.

Under employer mandates, firms bear the full risk and costs of their workforce's paid sick leave use. Economists tend to agree that employer mandates carry a smaller deadweight loss than direct government provision of benefits, funded by general or firm-specific taxes. Why might an employer mandate be more efficient than a general tax: say, a payroll tax to fund government-provided paid sick leave?

Mandated benefits largely avoid the deadweight loss associated with taxation, as Summers (1989) points out. In the graphical example of Figure 2 above, in equilibrium Z, where employees value the benefit more than it costs, employment is even larger than in equilibrium A, as employees expand supply and pay for the benefit through lower wages. While a general tax would equal the costs to the government of providing the benefit, mandated benefits imply a lower (and thus less distortionary) tax: it equals only the difference between employer costs and employee benefit valuation.[8]

Moreover, when employers face the burden of sick leave costs, they have incentives to prevent sickness and accommodate frail workers. Making employers financially responsible has brought down disability case rolls in the Netherlands (Koning and Lindeboom 2015). The same mechanism likely operates for paid sick leave, although we are not aware of plausibly causal evidence on this outcome.

On the other hand, a broad-based pooling of sick leave costs has certain advantages as well. Levying taxes and spreading the costs across all firms and workers reduces firms' financial risk. It also eliminates negative unintended consequences, such as discrimination against frail employees. Note that, in general, long-term sick leave carries high(er) employer costs and uncertainty about expected employee returns (similar to parental leave; see Ruhm 1998; Bailey et al. 2019).[9]

[8]We refer to the optimal taxation literature for a discussion on the pros and cons of using general tax revenue versus earmarked social insurance taxes to fund public benefits.

[9]As an alternative, one can imagine an individual insurance mandate for sick leave, which might be practical in countries with well-functioning insurance markets for longer-term sick pay, as they exist in many European countries like Germany or Switzerland, but no country seems to have pursued such a policy.

Employer mandates may entail high costs for small firms that cannot sufficiently pool risks—and thus may suffer substantial financial losses unrelated to their underlying business model when several employees fall sick at the same time. In the Netherlands, a private insurance market offers voluntary reinsurance options for firms. Germany, by contrast, has a mandatory reinsurance mechanism administered by public health insurers. Specifically, a law requires firms with fewer than 30 employees to select a reinsurance cost share of between 20 percent and 60 percent and pay premiums accordingly (*Aufwendungsausgleichsgesetz—AAG*).

Finally, providing paid sick leave through a public institution also requires employees or their employers to apply for the benefit, provide medical documentation, and, upon approval, receive a portion of their salary over several weeks. The process is bureaucratic and costly. It is, thus, less suited for short-term sickness episodes of one or a few days.

In conclusion, the dominant form of designing sick leave systems—employer mandates for the short-run and social insurance for the long run—can be viewed as a middle ground to balance pros and cons of employer mandates.

Modelling Optimal Paid Sick Leave

Here, we build on the trade-offs and policy design issues discussed above to describe two methods for modeling optimal paid sick leave: an optimal social insurance framework model and a structural modeling approach.

An Optimal Social Insurance Framework

A common approach to studying optimal social insurance is known as the Baily-Chetty framework (Baily 1978; Chetty 2006). It was originally used to study the optimal level of unemployment insurance. The core of this approach is to identify the key trade-offs and a few key elasticities: in the unemployment insurance version of the model, these include how the generosity of benefits affects the duration of unemployment and how much consumption drops when receiving unemployment benefits (depending in part on what percentage of wages are replaced and for how long).

In the context of paid sick leave, Maclean, Pichler, and Ziebarth (2020) develop such a sufficient statistics approach following the Baily-Chetty framework and building on Barmby, Sessions, and Treble (1994), who model employees' health as private information.[10] The model features a representative employee and

[10] This assumption is a standard one in the theoretical literature on this topic: for example, see Hirsch, Lechmann, and Schnabel (2017), who build on the efficiency wage literature (Shapiro and Stiglitz 1984) and theoretically produce presenteeism behavior via firms' setting of average wages to workers with heterogeneity in disutility from work. Unlike Brown and Sessions (1996), their model produces presenteeism behavior precisely because statutory sick pay is fixed at 100 percent of the wage as in Germany, and as the dismissal risk increases with absenteeism (by assumption). Brown and Sessions (1996) provide an early literature review that includes a static labor supply model.

employer. The model does not focus on long-term equilibria but instead assumes short-run wage rigidities due to social norms, collective bargaining, or legal barriers. Further, the model specifies employee utility as a function of wage, leisure time, and health. Employees either call in sick or work. Firms maximize profits, which depend on wages and work productivity. A key innovation is to endogenize work productivity, which is lower when employees work sick. Furthermore, health is treated as a continuous parameter (as in Barmby, Sessions, and Treble 1994), and for any given generosity level of paid sick leave, there is a health reservation level at which employees are indifferent between attending work and calling in sick.

In such a context, when is it welfare-improving for a social planner to mandate paid sick leave access for all employees? The marginal welfare benefit of paid sick leave arises when employees value it at least as much as it costs employers to provide. This monetized "benefit wedge" is simply wages multiplied by the difference in how employees value increases in sick pay and the cost of employers in providing it. Notice that in the analysis of Figure 2, the condition that employees value the sick leave benefit by at least as much as it costs employers to provide it, is also the condition in which we find no employment losses, or even employment gains.

This marginal benefit wedge must then be compared with the marginal "cost wedge" of increasing the generosity of paid sick leave, which specifies the effect of more sick pay on firm production and employee compensation. In particular, it includes the difference between "presenteeism productivity" (which is wages multiplied by labor productivity when working sick) and the "sick pay price" (which is the percentage of the wage not replaced by paid sick leave).[11] Note that this framework also considers moral hazard: more workers stay home and receive sick pay when paid sick leave becomes more generous. Also note that this model does not incorporate explicit consumption smoothing or a utility functional form that imposes risk aversion. The latter is implicitly reflected in how much employees value sick pay.

In summary, under a Baily-Chetty-type sufficient statistics condition, the social planner would increase sick pay until the higher employee utility from receipt of sick leave equals the sum of higher firm labor costs plus productivity losses due to presenteeism. Adding externalities due to workplace infections to this framework is straightforward.

[11] The text discussion focuses on the intuition behind the main trade-off. The formal optimality condition, assuming that the social planner assigns equal weight to employee utility and employer profits, is as follows, see Maclean, Pichler, and Ziebarth (2020): $\lambda w = \varepsilon\big(\delta w - (1-\alpha)w\big)/\alpha$. The left-hand side of this optimality condition specifies the benefits of marginally increasing the generosity of sick pay. It is the difference between higher worker utility versus higher firm costs (λ) due to paid sick leave, multiplied by the hourly wage (w). The right-hand side of the optimality condition represents the costs of marginally increasing the generosity of sick pay. "Presenteeism productivity" is wages (w) multiplied by labor productivity when working sick (δ). The "sick pay price" is $(1-\alpha)w$, where α is the replacement rate for sick pay. The right-hand side is then weighted by the labor supply elasticity (ε), which has little effect when the elasticity is close to zero but matters for short-term sickness with relatively large elasticities (Johansson and Palme 2002; Ziebarth and Karlsson 2010, 2014; Böckerman, Kanninen, and Suoniemi 2018; Maclean, Pichler, and Ziebarth 2025).

When plugging empirical estimates into this optimality condition for the US economy, leaving out externalities from infections and with an average presenteeism productivity of 77 percent from Maestas et al. (2023), employees must value paid sick leave at least at 142 percent of employers' costs to make mandates welfare-improving. As mentioned, Maestas et al. (2023) find that employees value paid leave at four times its costs. Thus, even without externalities, expanding sick leave coverage is very likely welfare-improving in the United States. When contagion externalities are taken into account, this model suggests that sick pay mandates are essentially always welfare-improving.

Kanninen, Böckerman, and Suoniemi (2025) offer another example of tailoring the standard Baily-Chetty approach to derive optimal sick pay. They focus on the employee side, and also model employee health as private information and incorporate a fixed cost of sickness. Using data on sick leave and well-being from the German Socio-Economic Panel (SOEP), they find a higher marginal utility of income when individuals are on sick leave.

Optimal Paid Sick Leave in a Structural Model

Optimal paid sick leave can also be modeled and estimated within a structural economic model, as in Barone (2023).[12] In this model, risk-averse employees incur uncertain health shocks and either call in sick or work while sick. Presenteeism generates higher income, but entails disutility. Firms can only partially observe illness, and illness can either be contagious or not. Presenteeism leads to production losses relative to working healthily. In this setting, a social planner chooses a paid sick leave benefit schedule to maximize social welfare, balancing workers' utility against production losses and moral hazard. This structural model enables the analysis to model different combinations of sick leave policy tools.

Barone (2023) estimates the model using administrative data from Chile, which operates a sick leave system similar to those in many European countries, see Table 1. The data consist of sick leave claims from 2015 to 2019, which are then linked to medical estimates of expected recovery times by diagnoses. Sick leave behavior is calibrated based on empirically observed sick leave use by day of the week and health diagnoses. From these data, employee preference parameters like the value of leisure and the utility cost of working sick are retrieved. Importantly, deviations between actual sick leave use and externally validated expected recovery times identify moral hazard, which is defined as being on sick leave longer than medically necessary.

[12] To the best of our knowledge, Gilleskie (1998) was the first economist to specify and estimate a structural model of US work absences. She models the decision to seek healthcare use and call in sick as sequential choices and uses 1987 Medical Expenditure Panel Survey data to estimate the worker's optimization problem with maximum likelihood. Policy simulations that eliminate health insurance cost-sharing and expand paid sick leave access to all employees suggest an increase in sick leave use by 11 percent. Grossmann, Schünemann, and Strulik (2025) estimate a dynamic general equilibrium model with overlapping generations, endogenous labor supply, and health and aging processes. The findings show that cutting the replacement rate increases distributional disparities in income and health.

To derive the optimal sick pay design, the social planner predicts the distribution of sick leave spells, given estimated health shocks, under alternative sick pay designs. Chile's sick pay schedule offers a key source of variation: a three-day waiting period, 100 percent wage replacement from days four to ten, and retroactive 100 percent replacement also for the first three days if the sick leave spell exceeds ten days.

The optimal paid sick leave design then minimizes "inefficient absenteeism" and "inefficient presenteeism." Inefficient absenteeism occurs when, induced by sick pay, employees incur marginal utility from work below their labor productivity and then call in sick. A higher wage replacement rate increases inefficient absenteeism. By contrast, inefficient presenteeism occurs when employees work sick as their marginal utility from sick leave is below their absence costs, but higher than their work productivity when working sick. A higher replacement rate reduces inefficient presenteeism.

This structural model suggests that optimal sick pay schemes should have no "cliff effects" due to waiting periods with a 0 percent replacement rate, nor should they have retrospective 100 percent after-the-fact compensation for waiting periods when crossing particular sick leave durations. Further, optimally set replacement rates should increase smoothly with the spell duration and be below 100 percent. This combination of smooth but increasing replacement rates aims to contain inefficient absenteeism by shifting costs to workers for short spells and avoid inefficient presenteeism by providing coverage for time off. The results from this model apply most directly to sick leave systems outside the United States, featuring a standard social insurance replacement rate design.

Conclusion

The United States has traditionally not regulated employee access to paid sick leave. However, as state- and local-level governments have rolled out sick pay mandates across the US economy over the past 15 years, most sick leave policy action and a rising share of economic articles have come out of the United States.[13] Despite this wave of research, many questions remain underexplored. For example, we know little about how peers—at the workplace or in private settings—shape sick leave behavior (Hesselius, Nilsson, and Johansson 2009). Also, little is known about optimal paid sick leave systems over the working lifecycle of an employee. Grossmann's (1972) human capital model may suggest that early paid sick leave interventions may result in a lower risk of longer-term work incapacity at older ages.

[13] Although growing, the economics literature on paid sick leave has arguably not kept pace with the amount of change, especially when compared to the extensive literature on the causes and consequences of health insurance coverage. Supplemental Appendix Figure A1 illustrates the number of peer-reviewed, Web of Science articles on paid sick leave in economics, which has about tripled between 2000 and 2024. In contrast, articles on health insurance have increased more than sixfold from 76 to 507 during the same period.

The existing research has clarified the fundamental trade-offs involved in paid sick leave—that is, an optimal paid sick leave system minimizes both absenteeism (the costs of calling in sick) and presenteeism (the costs of working while sick). In general, when a paid sick leave system becomes more generous, the share of shirkers will increase and the share of employees working sick will decrease.

Furthermore, labor supply elasticities for short-term sickness are relatively large (around one) but fall with the benefit duration. Thus, when employees have no access to paid sick leave, as continues to hold true in many states and professions in the United States, many employees will work sick and the costs of presenteeism will be high as workers spread contagious diseases and have low work productivity. On the other hand, a generous sick leave system like Germany's, where employer mandates require a 100 percent replacement rate for up to six weeks per sickness episode, leads not only to high employer costs, but also to an inefficiently high share of workers who call in sick, thereby raising the costs of absenteeism.

However, the generosity of paid sick leave does not vary along a single dimension: instead, it can include a variety of policy design tools, including waiting periods, monitoring mechanisms, duration periods, and the level of sick pay as well as the fundamental question of whether to organize paid sick leave as an individual credit account system or to follow traditional pooled social insurance models. We still know relatively little about the combinations of tools that would be most effective for firms to reduce both absenteeism and presenteeism (Hassink and Koning 2009; Alfitian, Sliwka, and Vogelsang 2024).

■ *We thank JEP's editorial board, in particular Timothy Taylor, as well as Victoria Barone, Pierre Koning, Catherine Maclean, Mark Pauly, and Danny Schneider for excellent comments and suggestions, and Daseul Moon, Tomohisa Nagata, and Pascal Marianna for data.*

References

Adams-Prassl, Abi, Teodora Boneva, Marta Golin, and Christopher Rauh. 2023. "The Value of Sick Pay." *European Economic Review* 151: 104314.

Alfitian, Jakob, Dirk Sliwka, and Timo Vogelsang. 2024. "When Bonuses Backfire: Evidence from the Workplace." *Management Science* 70 (9): 6395–414.

Andersen, Martin, Johanna Catherine Maclean, Michael F. Pesko, and Kosali Simon. 2023. "Does Paid Sick Leave Encourage Staying at Home? Evidence from the United States during a Pandemic." *Health Economics* 32 (6): 1256–83

AOK Bundesverband. 2025. "40 Prozent aller Fehltage entfielen 2024 auf Langzeit-Erkrankungen von mehr als sechs Wochen." March 11. https://www.aok.de/pp/bv/pm/fehlzeiten-2024.

Assamidanov, Anwar, Dhaval M. Dave, Jooyoung Kim, Brandy Lipton, Xuechao Qian, and Joseph J. Sabia. 2025. "The Effect of Paid Sick Leave Mandates on Individuals with Disabilities: Evidence from Social Security Disability Claims." NBER Working Paper 34542.

Atal, Juan Pablo, Hanming Fang, Martin Karlsson, and Nicolas R. Ziebarth. 2025. "German Long-Term Health Insurance: Theory Meets Evidence." *Journal of Political Economy* 133 (6): 1840–85.

Autor, David H., and Mark G. Duggan. 2006. "The Growth in the Social Security Disability Rolls: A Fiscal Crisis Unfolding." *Journal of Economic Perspectives* 20 (3): 71–96.

Autor, David H., and Mark Duggan. 2010. "Supporting Work: A Proposal for Modernizing the US Disability Insurance System." Unpublished.

Autor, David, Mark Duggan, Jonathan Gruber, and Catherine Maclean. 2013. "How Does Access to Short Term Disability Insurance Impact SSDI Claiming?" NBER Retirement and Disability Research Center Working Paper NB13-09.

Bailey, Martha J., Tanya S. Byker, Elena Patel, and Shanthi Ramnath. 2019. "The Long-Term Effects of California's 2004 Paid Family Leave Act on Women's Careers: Evidence from US Tax Data." NBER Working Paper 26416.

Baily, Martin Neil. 1978. "Some Aspects of Optimal Unemployment Insurance." *Journal of Public Economics* 10 (3): 379–402.

Banks, James, Richard Blundell, and Carl Emmerson. 2015. "Disability Benefit Receipt and Reform: Reconciling Trends in the United Kingdom." *Journal of Economic Perspectives* 29 (2): 173–90.

Barmby, Tim, John Sessions, and John Treble. 1994. "Absenteeism, Efficiency Wages and Shirking." *Scandinavian Journal of Economics* 96 (4): 561–66.

Barone, Victoria. 2023. "On the Design of Paid Sick Leave: A Structural Approach." Unpublished.

Barrero, José María, Nicholas Bloom, and Steven J. Davis. 2023. "The Evolution of Work from Home." *Journal of Economic Perspectives* 37 (4): 23–50.

Best, Rachel Kahn, Yan Fang, Catherine Fisk, Linda Hamilton Krieger, and Diana Reddy. 2025. "Disputed and Disfavored: Pain, Mental Illness, and Invisible Conditions in Disability Discrimination Cases." *Social Science and Medicine* 371: 117885.

Böckerman, Petri, Ohto Kanninen, and Ilpo Suoniemi. 2018. "A Kink That Makes You Sick: The Effect of Sick Pay on Absence." *Journal of Applied Econometrics* 33 (4): 568–79.

Brown, Sarah, and John G. Sessions. 1996. "The Economics of Absence: Theory and Evidence." *Journal of Economic Surveys* 10 (1): 23–53.

BLS (Bureau of Labor Statistics). 2025. *Quarterly Census of Employment and Wages (QCEW).* https://www.bls.gov/cew (accessed July 8, 2025).

Callison, Kevin, and Michael F. Pesko. 2022. "The Effect of Paid Sick Leave Mandates on Coverage, Work Absences, and Presenteeism." *Journal of Human Resources* 57 (4): 1178–1208.

Callison, Kevin, Michael F. Pesko, Serena Phillips, and Julie A. Sosa. 2023. "Cancer Screening after the Adoption of Paid-Sick-Leave Mandates." *New England Journal of Medicine* 388 (9): 824–32.

Cao, Yaming, Nicolas R. Ziebarth, Björn Fischer-Weckemann, Johannes Geyer. Forthcoming. "Fundamentally Reforming the DI System: Evidence from Germany." *Review of Economics and Statistics.*

Chatterji, Monojit, and Colin J. Tilley. 2002. "Sickness, Absenteeism, Presenteeism, and Sick Pay." *Oxford Economic Papers* 54 (4): 669–87.

Chetty, Raj. 2006. "A General Formula for the Optimal Level of Social Insurance." *Journal of Public Economics* 90 (10–11): 1879–901.

Chiappori, Pierre-Andre, and Bernard Salanie. 2000. "Testing for Asymmetric Information in Insurance Markets." *Journal of Political Economy* 108 (1): 56–78.

Chunyu, Liangrong, Paolo F. Volpin, and Xingchen Zhu. 2024. "Do Paid Sick Leave Mandates Increase Productivity?" Preprint, SSRN. http://dx.doi.org/10.2139/ssrn.4096707.

Cronin, Christopher J., Matthew C. Harris, and Nicolas R. Ziebarth. Forthcoming. "The Anatomy of US Sick Leave Schemes: Evidence from Public School Teachers." *Review of Economics and Statistics.*

D'Amuri, Francesco. 2017. "Monitoring and Disincentives in Containing Paid Sick Leave." *Labour Economics* 49: 74–83.

Deutsches Reichsgesetzblatt.1883. Gesetz, betreffend die Krankenversicherung der Arbeiter, Fassung vom 15. Juni 1883, Nr 9: 73–104. https://de.wikisource.org/wiki/Gesetz,_betreffend_die_Krankenversicherung_der_Arbeiter (accessed July 1, 2025).

Dube, Arindraji, T. William Lester, and Michael Reich. 2016. "Minimum Wage Shocks, Employment Flows, and Labor Market Frictions." *Journal of Labor Economics* 34 (3): 663–704.

East, Chloe N., Sarah Miller, Marianne Page, and Laura R. Wherry. 2023. "Multigenerational Impacts of Childhood Access to the Safety Net: Early Life Exposure to Medicaid and the Next Generation's Health." *American Economic Review* 113 (1): 98–135.

Einav, Liran, Amy Finkelstein, Stephen P. Ryan, Paul Schrimpf, and Mark R. Cullen. 2013. "Selection on

Moral Hazard in Health Insurance." *American Economic Review* 103 (1): 178–219.
Feldstein, Martin, and Daniel Altman. 2007. "Unemployment Insurance Savings Accounts." *Tax Policy and the Economy* 21: 35–63.
Finkenstädt, Verena. 2010. *Das Spannungsverhältnis von Äquivalenz und Einkommensumverteilung in der GKV—Eine Analyse der historischen Entstehungszusammenhänge.* Wissenschaftliches Institut der PKV.
Gilleskie, Donna B. 1998. "A Dynamic Stochastic Model of Medical Care Use and Work Absence." *Econometrica* 66 (1): 1–45.
Godard, Mathilde, Pierre Koning, and Maarten Lindeboom. 2024. "Application and Award Responses to Stricter Screening in Disability Insurance." *Journal of Human Resources* 59 (5): 1353–86.
Goodman, John C., and Gerald L. Musgrave. 1992. *Patient Power: Solving America's Health Care Crisis.* Cato Institute.
Goodman-Bacon, Andrew. 2018. "Public Insurance and Mortality: Evidence from Medicaid Implementation." *Journal of Political Economy* 126 (1): 216–62.
Goodman-Bacon, Andrew. 2021. "The Long-Run Effects of Childhood Insurance Coverage: Medicaid Implementation, Adult Health, and Labor Market Outcomes." *American Economic Review* 111 (8): 2550–93.
Google Flu. 2015. *Google Flu Trends.* https://www.google.org/flutrends/about/data/flu/us/data.txt (accessed September 1, 2024).
Grossman, Michael. 1972. "On the Concept of Health Capital and the Demand for Health." *Journal of Political Economy* 80 (2): 223–55.
Grossman, Volker, Johannes Schünemann, Holger Strulik. 2025. "Sick Pay Policies and the Socioeconomic Gradient in Health and Welfare." CESifo Working Paper Series 12265.
Gruber, Jonathan. 1994. "The Incidence of Mandated Maternity Benefits." *American Economic Review* 84 (3): 622–41.
Gupta, Sumedha, Johanna Catherine Maclean, Christopher J. Ruhm, and Kosali I. Simon. 2025. "The Impact of State Paid Sick Leave Mandates on Medicaid-Financed Prescription Medications." NBER Working Paper 34485.
Gürtzgen, Nicole, and Karolin Hiesinger. 2025. "Dismissal Protection and Long-Term Sickness Absence: Evidence from a Policy Change." *Industrial Relations: A Journal of Economy and Society* 64 (3): 318–42.
Haller, Andreas, Stefan Staubli, and Josef Zweimüller. 2023. "Designing Disability Insurance Reforms: Tightening Eligibility Rules or Reducing Benefits?" *Econometrica* 92 (1): 79–110.
Harknett, Kristen, and Daniel Schneider. 2022. "Mandates Narrow Gender Gaps in Paid Sick Leave Coverage for Low-Wage Workers in the US." *Health Affairs* 41 (11): 1575–82.
Hassink, Wolter H. J., und Pierre Koning. 2009. "Do Financial Bonuses Reduce Employee Absenteeism? Evidence from a Lottery." *ILR Review* 62 (3): 327–42.
Hernæs, Øystein. 2018. "Activation against Absenteeism—Evidence from a Sickness Insurance Reform in Norway." *Journal of Health Economics* 62: 60–68.
Hesselius, Patrik, J. Peter Nilsson, and Per Johansson. 2009. "Sick of Your Colleagues' Absence?" *Journal of the European Economic Association* 7 (2-3): 583–94.
Hirsch, Boris, Daniel S. J. Lechmann, and Claus Schnabel. 2017. "Coming to Work While Sick: An Economic Theory of Presenteeism with an Application to German Data." *Oxford Economic Papers* 69 (4): 1010–31.
Jelliffe, Emma, Paul Pangburn, Stefan Pichler, and Nicolas R. Ziebarth. 2021. "Awareness and Use of (Emergency) Sick Leave: US Employees' Unaddressed Sick Leave Needs in a Global Pandemic." *Proceedings of the National Academy of Sciences* 118 (29): e2107670118.
Johansson, Per, and Mårten Palme. 2002. "Assessing the Effect of Public Policy on Worker Absenteeism." *Journal of Human Resources* 37 (2): 381–409.
Johansson, Per, and Mårten Palme. 2005. "Moral Hazard and Sickness Insurance." *Journal of Public Economics* 89 (9–10): 1879–90.
Johnson, Candice Y., Kristen Said, Ashley E. Price, Dennis Darcey, and Truls Østbye. 2022. "Paid Sick Leave among US Private Sector Employees." *American Journal of Industrial Medicine* 65 (9): 743–48.
Kanninen, Ohto, Petri Böckerman, and Ilpo Suoniemi. 2025. "Income–Well-Being Gradient in Sickness and Health." *Health Economics.* https://doi.org/10.1002/hec.70063.
Koning, Pierre, and Maarten Lindeboom. 2015. "The Rise and Fall of Disability Insurance Enrollment in the Netherlands." *Journal of Economic Perspectives* 29 (2): 151–72.
Kools, Lieke, and Pierre Koning. 2019. "Graded Return-to-Work as a Stepping Stone to Full Work Resumption." *Journal of Health Economics* 65: 189–209.

Low, Hamish, and Luigi Pistaferri. 2025. "Disability Insurance: Error Rates and Gender Differences." *Journal of Political Economy* 133 (9): 2962–3018.

Maclean, Johanna Catherine, Stefan Pichler, and Nicolas R. Ziebarth. 2020. "Mandated Sick Pay: Coverage, Utilization, and Welfare Effects." IZA Discussion Papers 13132.

Maclean, Catherine, Stefan Pichler, and Nicolas R. Ziebarth. 2022. "Mandated Sick Pay: Coverage, Utilization, and Welfare Effects." Washington Center for Equitable Growth Working Paper.

Maclean, Johanna Catherine, Stefan Pichler, and Nicolas R. Ziebarth. 2025. "Mandated Sick Pay: Coverage, Utilization, and Crowding-In." *Journal of the European Economic Association* 23 (5): 1868–907.

Maestas, Nicole, Kathleen J. Mullen, David Powell, Till von Wachter, and Jeffrey B. Wenger. 2023. "The Value of Working Conditions in the United States and the Implications for the Structure of Wages." *American Economic Review* 113 (7): 2007–47.

Maestas, Nicole, Kathleen J. Mullen, and Stephanie Rennane. 2019. "Unmet Need for Workplace Accommodation." *Journal of Policy Analysis and Management* 38 (4): 1004–27.

Manning, Willard G., Joseph P. Newhouse, Naihua Duan, Emmett B. Keeler, Arleen Leibowitz, and M. Susan Marquis. 1987. "Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment." *American Economic Review* 77 (3): 251–77.

Marie, Olivier, and Judit Vall Castelló. 2023. "Sick Leave Cuts and (Unhealthy) Returns to Work." *Journal of Labor Economics* 41 (4): 923–56.

Markussen, Simen, Knut Røed, and Ole Røgeberg. 2013. "The Changing of the Guards: Can Family Doctors Contain Worker Absenteeism?" *Journal of Health Economics* 32 (6): 1230–39.

Markussen, Simen, Knut Røed, and Ragnhild C. Schreiner. 2018. "Can Compulsory Dialogues Nudge Listed Sick-Listed Workers Back to Work?" *Economic Journal* 128 (610): 1276–303.

McVicar, Duncan, Roger Wilkins, and Nicolas R. Ziebarth. 2022. "Five Decades of Disability Benefit Policies in Five OECD Countries." In *Work and the Social Safety Net: Labor Activation in Europe and the United States*, edited by Douglas J. Besharov and Douglas M. Call, 150–82. Oxford University Press.

Ministry of Social Affairs and Employment. 2002. *Wet verbetering poortwachter.* Government of the Netherlands. https://wetten.overheid.nl/BWBR0013063.

Meyer, Bruce D., and Wallace K. C. Mok. 2019. "Disability, Earnings, Income and Consumption." *Journal of Public Economics* 171: 51–69.

Nagata Tomohisa, Kiminori Odagami, Masako Nagata, Nuri Purwito Adi, and Koji Mori. 2023. "Protocol of a Study to Benchmark Occupational Health and Safety in Japan: W2S-Ohpm Study." *Front Public Health* 11: 1191882.

National Partnership for Women and Families. 2020. "Voters Show Bipartisan Support for Permanent Paid Sick Days and Paid Family and Medical Leave." https://nationalpartnership.org/wp-content/uploads/2023/02/voters-show-bipartisan-support-for-permanent-paid-sick-days-and-paid-family-and-medical-leave.pdf, retrieved November 2, 2024.

OECD. 2010. *Sickness, Disability and Work: Breaking the Barriers: A Synthesis of Findings across OECD Countries.* OECD Publishing.

OECD. 2020. *Paid Sick Leave to Protect Income, Health and Jobs through the COVID-19 Crisis.* OECD Publishing.

OECD. 2023. *Disability, Work and Inclusion in Korea: Towards Equitable and Adequate Social Protection for Sick Workers.* OECD Publishing.

Pauly, Mark V. 1974. "Overinsurance and Public Provision of Insurance: The Roles of Moral Hazard and Adverse Selection." *Quarterly Journal of Economics* 88 (1): 44–62.

Pauly, Mark V., John C. Goodman, Judith Feder, Larry Levitt, Stuart M. Butler, David M. Cutler, and Gail R. Wildensky. 1995. "Tax Credits for Health Insurance and Medical Savings Accounts." *Health Affairs* 14 (1): 125–39.

Pettersson-Lidbom, Per, und Peter Skogman Thoursie. 2013. "Temporary Disability Insurance and Labor Supply: Evidence from a Natural Experiment." *Scandinavian Journal of Economics* 115 (2): 485–507.

Pichler, Stefan, Katherine Wen, and Nicolas R. Ziebarth. 2020. "COVID-19 Emergency Sick Leave Has Helped Flatten the Curve in the United States." *Health Affairs* 39 (12): 2197–204.

Pichler, Stefan, and Nicolas R. Ziebarth. 2017. "The Pros and Cons of Sick Pay Schemes: Testing for Contagious Presenteeism and Noncontagious Absenteeism Behavior." *Journal of Public Economics* 156: 14–33.

Pichler, Stefan, and Nicolas R. Ziebarth. 2020. "Labor Market Effects of US Sick Pay Mandates." *Journal of Human Resources* 55 (2): 611–59.

Pichler, Stefan, and Nicolas R. Ziebarth. 2024. "Sick and Medical Leave in the United States: A Categorization and Recent Trends." IZA Policy Paper 206.

Pronovo BKK. 2024. "Arbeiten 2023." January 22. https://www.pronovabkk.de/unternehmen/presse/pressemitteilungen/pressemitteilungen-2024/arbeitnehmer-innen-melden-sich-trotz-arbeitsfaehigkeit-krank.html#section-id-pr-senz-trotz-krankheit.

Riise, Julie, Barton Willage, and Alexander Willén. 2025. "Intergenerational Effects of Sick Leave on Child Human Capital." *Journal of Health Economics* 103: 103047.

Rothschild, Michael, and Joseph Stiglitz. 1976. "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information." *Quarterly Journal of Economics* 90 (4): 629–49.

Ruhm, Christopher J. 1998. "The Economic Consequences of Parental Leave Mandates: Lessons from Europe." *Quarterly Journal of Economics* 113 (1): 285–317.

Schmutte, Ian M., and Meghan M. Skira. 2025. "The Response of Firms to Maternity Leave and Sickness Absence." *Journal of Human Resources* 60 (6): 2125–67.

Shapiro, Carl, and Joseph E. Stiglitz. 1984. Equilibrium Unemployment as a Worker Discipline Device. *American Economic Review* 74 (3): 433–44.

Slopen, Meredith. 2024. "The Impact of Paid Sick Leave Mandates on Women's Employment and Economic Security." *Journal of Policy Analysis and Management* 43 (4): 1129–51.

Smith, Tom W., and Jibum Kim. 2010. *Paid Sick Days: Attitudes and Experiences*. Public Welfare Foundation.

Stearns, Jenna, and Corey White. 2018. "Can Paid Sick Leave Mandates Reduce Leave-Taking?" *Labour Economics* 51: 227–46.

Stepner, Michael. 2019. "The Long-Term Externalities of Short-Term Disability Insurance." Unpublished.

Summers, Lawrence H. 1989. "Some Simple Economics of Mandated Benefits." *American Economic Review* 79 (2): 177–83.

Sun, Jessica Ya. 2025. "Is There Moral Hazard in Medical Savings Accounts? Evidence from Singapore." *Journal of Economic Behavior and Organization* 231: 106938.

Susser, Philip, and Nicolas R. Ziebarth. 2016. "Profiling the US Sick Leave Landscape: Presenteeism among Females." *Health Services Research* 51 (6): 2305–17.

van den Berg, Gerard J., Hanno Foerster, and Arne Uhlendorff. Forthcoming. "Structural Empirical Analysis of Vacancy Referrals with Imperfect Monitoring and the Strategic Use of Sickness Absence." *Journal of Labor Economics*.

Woods, Tyler, Daniel Schneider, and Kristen Harknett. 2025. "Estimating the Impact of State Paid Sick Leave Laws on Worker Outcomes in the US Service Sector, 2017–2023." *SSM—Population Health* 31: 101830.

Ziebarth, Nicolas R. 2013. "Long-term absenteeism and moral hazard—Evidence from a natural experiment." *Labour Economics* 24: 277–92.

Ziebarth, Nicolas R., and Martin Karlsson. 2010. "A Natural Experiment on Sick Pay Cuts, Sickness Absence, and Labor Costs." *Journal of Public Economics* 94 (11–12): 1108–22.

Ziebarth, Nicolas R. and Martin Karlsson. 2014. "The Effects of Expanding the Generosity of the Statutory Sickness Absence Insurance System." *Journal of Applied Econometrics* 29 (2): 208–30.

Ziebarth, Nicolas R. und Stefan Pichler. 2024. "Einordnung des deutlichen Anstiegs der krankheitsbedingten Fehlzeiten seit 2022." ZEW policy briefs 18/2024, ZEW—Leibniz Centre for European Economic Research. https://www.zew.de/publikationen/einordnung-des-deutlichen-anstiegs-der-krankheitsbedingten-fehlzeiten-seit-2022 (accessed May 12, 2025).

Journal of Economic Perspectives—Volume 40, Number 2—Spring 2026—Pages 243–258

Retrospectives: The Great Dollar-Shortage Debate

Harris Dellas and George S. Tavlas

This feature addresses the history of economic terms and ideas. The hope is to deepen the workaday dialogue of economists, while perhaps also casting new light on ongoing questions. If you have suggestions for future topics or authors, please contact Joseph Persky, University of Illinois at Chicago (jpersky@uic.edu).

The controversy over the existence of a global "dollar shortage" dominated discussions on international economics in the 15 years following the end of World War II. During the 1940s, the rest of the world—primarily Western Europe—had capital account surpluses vis-à-vis the United States (dollars borrowed from the United States) that fell short of what was needed to finance their corresponding trade and current account deficits versus the United States (the amount needed to pay for their net imports and the interest owed to their lenders). This shortfall, defined as the "dollar shortage" by Charles Kindleberger, was financed by using official dollar and gold reserves—that is, by running a balance-of-payments deficit. The scarcity of reserves available to countries other than the United States limited the size of feasible current account deficits that those countries could

■ *Harris Dellas is President of the Karl Brunner Institute, Bern, Switzerland, and Professor of Economics, University of Surrey, Surrey, United Kingdom. He is also a Research Fellow, Federal Reserve Bank of St. Louis, St. Louis, Missouri, and at the Centre for Economic Policy Research, London, United Kingdom. George S. Tavlas is Alternate to the Governor of the Bank of Greece for the European Central Bank Governing Council, Athens, Greece, and Distinguished Visiting Fellow, Hoover Institution, Stanford University, Stanford, California. Tavlas is the corresponding author at gtavlas@bankofgreece.gr.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20251468.

finance and endangered reconstruction efforts, hindering the improvement of livings standards. The IMF Articles of Agreement even included a "scarce currency" clause that permitted and expected countries to discriminate against US goods in their trade policies if they faced a dollar shortage; although the clause was never implemented (for discussion, see Horsefield 1969, p. 193).

The dollar-shortage debate started in earnest in the early 1940s and then ran through much of the 1950s, generating scores of articles and books, with many prominent economists and journalists weighing in. As early as 1943, *The Economist* published two articles titled "The Dollar Problem—I" and "The Dollar Problem—II" (*Economist* 1943a, b). The editors of that periodical called the dollar shortage "the world's financial problem number one" (1943b, p. 750). Fifteen years later in his textbook *Economics: An Introductory Analysis*, Paul Samuelson called the dollar shortage "the big open question of our time" (Samuelson 1958, p. 707).

The debate centered on the key questions of what caused the dollar shortage and how large and persistent it was going to be. There were two distinct views concerning the causes, represented by Kindleberger and Milton Friedman. Kindleberger attributed the existence of the shortage to *real* factors, namely, to frictions in international financial markets that limited the flow of capital into Europe. He attributed its size to differences in cross-country productivity—the income convergence process and the associated current account deficit. In his view, the dollar shortage varied with the current account deficit and the latter varied mostly as a function of the productivity gap between Europe and the United States. Friedman ignored financial frictions and the role of income convergence, choosing instead to focus exclusively on nominal rigidities. He believed that the dollar shortage was simply the current account deficit, whose cause was purely *nominal*—namely, overvalued currencies against the US dollar under the Bretton Woods system of pegged exchange rates that were leading to loss of trade competitiveness.[1]

Kindleberger attributed the large size of Western Europe's current account deficits to Europe's large productivity (technological) gap with the United States. He expected the deficit and the associated dollar shortage to be chronic because he predicted a slow pace of technological catch-up that would keep Europe long dependent on United States goods for consumption and investment. Other participants were less pessimistic about the chronicity of the shortage. Most prominently, Arthur Bloomfield expected Europe's current account deficits to self-correct quicker than Kindleberger's prediction through the process of *income* growth in Europe.

[1] The global monetary setting in the post–World War II period was the Bretton Woods system. It involved a system of fixed-but-adjustable exchange rates versus the US dollar and a peg of the US dollar to gold; a multinational institution, the IMF, whose main mission was to oversee the system of fixed exchange rates and provide *short-term* loans to countries experiencing nonstructural balance-of-payments problems; and the use of capital controls to manage the pegged exchange rates. A notable exception to the last element was that of the United States, which, during the 1940s and 1950s, maintained free capital mobility with the rest of the world. This feature implies that the dollar shortage is not the standard case of disequilibrium resulting from fixing both the price (the exchange rate) and the quantity (the volume of capital flows).

Friedman held a completely different perspective on the dollar shortage—one that emphasized nominal rigidities. As fixed-exchange-rate misalignments were endemic in the Bretton Woods system, he suggested that simply letting the exchange rate float freely would quickly eliminate the trade and balance-of-payments imbalances. But this at best may have eliminated the part of the imbalance that was driven by nominal considerations (nominal price, wage, and exchange-rate stickiness), leaving the chronic, structural component emphasized by Kindleberger intact. Kindleberger was doubtful that nominal considerations played any role in the dollar shortage, especially its chronic part: low trade elasticities (as there were no good substitutes for vital US goods) made the exchange-rate adjustment ineffective in moving trade quantities. Moreover, short-lived price stickiness made the effect of such adjustment short lived. In his view, nominal rigidities could not cause—or resolve—chronic imbalances, because these were due to real factors.

In 1950, the US balance of payments swung into deficit, following a succession of annual surpluses during the previous decade.[2] During the 1950s, countries outside the United States experienced a large increase in their foreign exchange reserves, with their gold and dollar reserves rising from $15.1 billion in 1950 to $22.8 billion in 1956 (Klopstock 1957, p. 10, Table 2). By the end of the 1950s, the US total external dollar liabilities equaled the US monetary gold stock, and the US share of the world's monetary gold had fallen from 70 percent of the global stock at the end of World War II to less than 50 percent, a situation that was viewed with alarm (Bordo 1993, p. 56; Eichengreen 2007, pp. 15, 42).

Part of the improvement came from a large and sustained increase in private capital flows from the United States to Western Europe that began in the early-1950s (IMF 1955, p. 35, Table 9). Another part was attributed to the widespread devaluation of foreign currencies versus the US dollar in September 1949 and the Marshall plan, in which the US government provided dollar loans and grants to Western European countries. As these two latter factors represented one-time events, Kindleberger viewed them as having only a temporary, short-term effect on the dollar shortage. As we will discuss, it was not until the late 1950s, after Kindleberger observed the closing of the productivity gap between the US and Western European economies, that he became less pessimistic about the chronicity of the dollar shortage. This signaled the end of the dollar-shortage debate.

In the conclusion, we compare the post–World War II dollar shortage to those that have occurred in the last 15 years. We argue that some of the recent shortages—namely, the ones involving pegged exchange rates in certain lower-income countries, and which are manifested in persistent balance-of-payments deficits—are very much

[2]There is some confusion regarding usage of the term "balance of payments," because it is often meant to refer to all the foreign accounts. The "balance of payments"—also called the reserve or official settlements account—records central banks' reserve asset transactions with each other, namely transactions involving gold, foreign exchange reserves, bank deposits, and "special drawing rights" of the IMF. Nowadays, this account tends to be negligible, but under the Bretton Woods system of pegged exchange rates, it was a critically important account, because official reserves were used to manage the exchange-rate peg.

like the earlier dollar shortage. As in the earlier episode, the recent balance-of-payments deficits arose mostly from the reluctance of countries with a current account surplus to finance the trade deficits of their trading partners because of repayment risk and other financial frictions. But some other recent shortages that involve flexible-exchange-rate regimes and are manifested in deviations from "covered interest rate parity" are different from the earlier shortage. They relate to the function of the US dollar as the primary international reserve currency/safe asset and reflect a liquidity problem. Moreover, they can be addressed through central bank currency swap agreements and are subject to the Triffin dilemma.

The Main Views in the Dollar-Shortage Debate

The US Productivity Advantage and a Lack of Lending: Kindleberger

Kindleberger began his engagement in the debate about a dollar shortage with two 1943 papers written from distinct vantage points: one focused on the US balance-of-payments surplus and the other focused on the corresponding balance-of-payments deficits of US trading partners.[3] In "Planning for Foreign Investment," published in the March 1943 issue of the *American Economic Review*, Kindleberger (1943a, pp. 350–51) argued that the basic cause of the dollar shortage was the perpetual proclivity of the United States to run current account surpluses and hoard the acquired gold rather than lend it long term to foreign countries. Kindleberger wrote: "The world 'chronic shortage of dollars' . . . is basically ascribable to the United States' failure to lend abundantly, or rather more continuously." Kindleberger contrasted the United States' situation in the 1930s and early-1940s with that of Britain in the nineteenth century. During the earlier period, Britain ran current account surpluses, but unlike the United States in the latter period, Britain kept "reinvesting her current account surplus." Consequently, a sterling shortage did not emerge. The solution to the dollar shortage was for the United States to "lend abroad continuously until productivity in other countries has increased to the point where the demand for [US] . . . goods is reduced in intensity" (Kindleberger 1943a, p. 351). Not coincidentally, Kindleberger worked at the US Department of State in 1947 and 1948 and directed the governmental committee that put the Marshall Plan together and steered it through the US Congress (Mehrling 2022, p. 83).

In "International Monetary Stabilization," published in an edited volume (by Seymour Harris), *Postwar Economic Problems*, Kindleberger (1943b, p. 379) argued that the primary driver of the shortage was that US trading partners were running their large current account deficits as a result of "the technical superiority of the United States in the production of many goods necessary to a high modern standard

[3] After receiving his PhD from Columbia University in 1936, Kindleberger worked for several US government agencies, as well as the Federal Reserve Bank of New York, the Federal Reserve's Board of Governors, and the Bank of International Settlements, before joining MIT's department of economics in 1948.

of living and to the natural desire in other countries to raise real incomes faster than the basic conditions of their economic productivity justify."

Could those current account imbalances correct themselves naturally through the process of growth in the rest of the world (and increased exports to the United States), or through exchange-rate (or terms-of-trade) adjustments? Kindleberger, who maintained that a "chronic world dollar shortage" had existed since 1919, was skeptical of achieving an equilibrium with a balanced current account. In his book *The Dollar Shortage* (Kindleberger 1950), a comprehensive study of the issue, he defined a dollar shortage as "a condition of persistent departure or of persistent tendency to depart from equilibrium in the balance of payments of the United States in the direction of a [current account] surplus in excess of net long-term capital outflows. Viewed from abroad, the dollar shortage is the tendency of the current accounts of foreign balances of payments to show larger deficits than are covered by long-term borrowing" (p. 170).[4]

Kindleberger (1950, p. 175) considered several possible methods for addressing the dollar shortage—but argued that none of them would fully address the issue. First, he was dismissive of the view that "the whole question of the dollar shortage comes down to the overevaluation [against the dollar] of foreign currencies and the undervaluation of the dollar, which could be rapidly set to rights by changing the exchange rate . . . [so that] there is really no problem." In this regard, he believed that trade volumes were inelastic with respect to exchange-rate changes. Earlier, Kindleberger (1943b, p. 381) had written:

> It may be suggested that the United States has a comparatively low propensity to import and a low ratio of exports to national income, whereas the rest of the world has a relatively high elasticity of demand for United States exports of manufactured goods and a relatively high ratio of exports to income. If this be true . . . additional dollars made available to foreigners by increased United States imports may lead to a greater increase in foreign expenditures for American products, leaving the world still short of dollars.

With these assumptions, neither growth in US imports nor exchange-rate adjustment (in this case, depreciation of foreign currencies against the dollar) would succeed in ending the shortage. According to Kindleberger (1950, p. 387), "the chronic shortage of dollars would remain [in the future], albeit at higher levels of real income throughout the world, and the United States would continue to pile up [current account] surpluses."

Second, Kindleberger was not optimistic that monetary and fiscal policies could alleviate the dollar shortage. For this method to work, the United States would have to implement expansionary monetary and fiscal policies (to increase US imports), and the rest of the world would have to implement contractionary policies (to reduce

[4] In a biography of Kindleberger, Mehrling (2022, p. 94) reported that *The Dollar Shortage* was "the book that got him [Kindleberger] tenure" at MIT.

imports from the United States). However, Kindleberger (1950, p. 195) argued that "the United States is under no compunction to inflate so that other countries may meet a random disturbance of equilibrium requiring a structural adjustment, such as postwar reconstruction or a technological improvement resulting in the loss of an export market." As for the use of contractionary monetary and fiscal policies in foreign countries, Kindleberger (p. 199) stated that, because these policies would raise unemployment, they were "unlikely to meet with the approval of the people most directly concerned."

Third, Kindleberger (1950) was also pessimistic about the possibility of using US tariff reduction to expand its imports. He argued that "the removal of the United States tariff [is] analogous to appreciation of the dollar," a policy which he had already dismissed as ineffective (p. 225). He further argued that a US tariff reduction would reduce US income so that imports would fall on balance (p. 253).

In sum, Kindleberger believed that Europe would suffer a chronic dollar shortage, the roots of which were reflected in financial frictions, productivity gaps, and Europe's impatience to improve its living standards. The shortage would hinder its ability to purchase investment goods from the United States and, thus, hamper European growth.

Other Proponents of a Dollar Shortage

During the late-1940s and the 1950s, explanations of the dollar shortage were legion, many with similarities to Kindleberger's but with their own spin.[5] To give a sense of the breadth of views, we briefly describe those of three other prominent proponents of the dollar shortage: Sirs Roy Harrod, John Hicks, and Dennis Robertson.

Harrod (1947) had argued that the "allegation of a 'world dollar shortage' is surely one of the most brazen pieces of collective effrontery that has ever been uttered." Six years later, Harrod had become an advisor to the IMF and became convinced that the world economy did indeed suffer from a dollar shortage. Harrod (1953) agreed with Kindleberger on the importance of low elasticities of demand and supply in international trade (p. 18) and more rapid productivity growth in the United States than in the rest of the world (p. 30) as causes of the dollar shortage. He added two other factors: higher inflation in Europe than in the United States which diminished the competitiveness of European economies (p. 35) and Europe's "worsened" terms of trade vis-à-vis the United States (p. 21). Harrod was not optimistic about a solution to "the dollar crisis" (p. 33).

John Hicks (1953), in "An Inaugural Lecture" at Oxford's All Souls College, referred to the "dollar problem" as "perhaps the fundamental economic problem confronting this [the UK] country" (p. 121). Like, Kindleberger, Hicks emphasized that US productivity was growing rapidly, but he also emphasized the difference between productivity growth in import-competing industries and export industries.

[5] Some of the literature is reviewed in Yeager (1965, pp. 458–63). Yeager (p. 459) called the dollar-shortage discussion a "fad" under which the "billowing academic smoke strengthened belief in a real-world fire."

To hold their own in US markets, Hicks argued, other countries would have to lower the prices of their exports while enjoying only slightly reduced prices on their imports from the United States (Hicks 1953, pp. 132–33). Hicks did not believe that a devaluation of foreign currencies against the dollar would help. He considered the 1949 devaluation of sterling (discussed below) to have been a failure, because domestic nominal wages rose following the devaluation, negating its effects on competitiveness: "wages chased after the rising prices [on UK imports], and the rise in wages made the devaluation largely ineffective" (pp. 133–34).

Dennis Robertson, in his 1954 book *Britain in the World Economy*, agreed with Kindleberger that dollar shortage was due to the combination of the destruction of productive capacities in Europe due to World War II, the tendency of Western European countries to overspend, and the high rate of technological progress in the United States.

A Self-Correcting View of the Dollar Shortage: Bloomfield Versus Kindleberger

Arthur Bloomfield's engagement in the dollar-shortage debate began in 1947 through the publication of a textbook, *International Economics*, by Stephen Enke and Virgil Salera.[6] In their book, Enke and Salera pointed out an implication of Kindleberger's earlier argument that the US dollar tends to be chronically scarce because increased US imports will raise nominal incomes abroad, producing increased demands for US products in excess of the initial increase in US imports. In criticizing Kindleberger's thesis, Enke and Salera expressed their gratitude to Bloomfield "for the basic ideas involved in the refutation of this [Kindleberger's] thesis" (Enke and Salera 1947, p. 600, fn. 2). Bloomfield (1949) makes this argument in his own words.

The criticism ran as follows. Recall that, ignoring the government budget deficit (or surplus), the current account balance is the difference between domestic saving and investment. Enke and Salera focused on Kindleberger's assumption—that an increase in a foreign country's exports could *increase* its trade deficit if all of the additional income from exports went into buying imports—required the country's marginal propensity to save to be zero or negative. However, so long as some fraction of income is leaked into domestic savings, a new dollar of exports would not be able to lift income by enough to call forth a full dollar of new imports. Enke and Salera (1947, p. 601) observed: "Unfortunately for the [Kindleberger] thesis, there is scarcely any case where the marginal propensity to save is not a positive quantity."

Kindleberger (1949) responded to what he called the "Bloomfield–Enke and Salera" criticism in his paper "The Foreign-Trade Multiplier: The Propensity to Import and Balance-of-Payments Equilibrium." He wrote that the 1943 formulation of his thesis was incomplete and needed two additional assumptions to complete it. First, foreign countries experiencing an increase in exports to the United States would need to have a high rate of induced domestic investment (more than offsetting the rise in saving associated with the rise in income). Second, the foreign

[6]Bloomfield received his PhD at the University of Chicago in 1942. He worked at the Federal Reserve Bank of New York from 1941 to 1958, before joining the faculty at the University of Pennsylvania in 1958.

countries would need to have a negative marginal propensity to save (p. 492). On this second point, Kindleberger (p. 494) argued: "I fail to find it strange that the savings function of a country may be negative under certain dynamic conditions . . . . it may be that countries . . . live above their means for a time in response to an increase in income. This type of response, when capital is not available for borrowing from abroad, may lead to a severe loss of reserves of gold and foreign exchange."

The Fixed-Exchange-Rate Interpretation: Friedman

A different analysis of the dollar-shortage thesis in the late-1940s and the first half of the 1950s blamed it primarily on government economic controls, especially the fixed-exchange-rate regime. The claim was that the fixed exchange rate had set the US dollar weaker (undervalued) than it would have been in a floating market, and thus led to an ongoing US trade surplus.

In a 1953 paper, "Why the Dollar Shortage?" Milton Friedman wrote that "'dollar shortage' is a strictly meaningless phrase" about which "floods of nonsense" had been written (Friedman 1953, p. 202). Why had a global shortage of dollars arisen? Friedman (p. 201) argued: "[I]ts fundamental cause and cure are alike simple: the dollar shortage is a result of governmentally controlled and rigid exchange rates." If the dollar shortage was a manifestation of fixed exchange rates, why, Friedman (p. 202) asked, "has it appeared in such virulent form only in recent years?" The answer to this apparent paradox, Friedman argued, was that, under earlier fixed-exchange-rate systems, including the classical gold standard and the interwar gold exchange standard, domestic monetary policies were determined by external forces: monetary policies would adjust so that the real price of foreign exchange maintained equilibrium in a country's balance of payments. Such adjustments became impossible once exchange-rate policy became driven by domestic considerations (mostly unemployment).

Friedman was not the first or only one to make this argument. For example, the prominent economic journalist Henry Hazlitt (1947, p. 21), in his monograph *Will Dollars Save the World?*, argued: "[T]he chief responsibility [for the dollar shortage] must be placed upon government controls. Most of the governments of the world today, by forcing commodity prices below the levels that supply and demand would bring about, are creating artificial bottlenecks and shortages." Hazlett continued: "But the gravest case of arbitrary fixing is the overvaluation that nearly all countries place on their own currencies. They will not accept the verdict of the open market as to what those currencies are really worth." Thus, Hazlitt argued in favor of floating exchange rates (p. 22) and more broadly: "There must be an end of price control, either for home-produced goods or imported goods, and an end of other regulations that prevent or unbalance trade and production" (p. 32). Similarly, Princeton University professor Frank Graham (1949, p. 9) argued in *The Cause and Cure of "Dollar Shortage"*: "The only real solution [to the dollar shortage] is the operation of the price mechanism in a free exchange market to equate national supply and demand in international trade."

An Appraisal on the Basis of Modern Macroeconomic Theory

How do the views exposited above look through the lenses of modern macroeconomic theory? The workhorse of modern international macroeconomics is the intertemporal approach to current account determination under financial frictions (Obstfeld and Rogoff 1996). Consider a real, small, open economy that faces imperfect international financial markets (say, an income-based borrowing constraint). Let technology be embodied in the capital stock, so that total factor productivity goes hand in hand with a high capital stock. Furthermore, assume that this economy's initial ratio of capital stock to total factor productivity is low relative to that of the rest of the world. The theory implies that the country will borrow in the international markets to finance investment and also to carry out consumption smoothing. There will be a *chronic* current account deficit during the convergence process of income and total factor productivity. The size and dynamics of this current account deficit will depend on the severity of the financial friction, the initial productivity gap with the rest of the world, and the country's "impatience." The current account deficit will be more persistent if the financial friction is more severe, because the friction acts as a borrowing constraint, limiting the ability to import investment goods and slow down the rate of productivity growth and catching up.

Large impatience will also increase the persistence of the current account, as income gains during the transition will be disproportionately directed towards consumption, making the marginal propensity to save (MPS) low and, perhaps, negative. Note that an income-based borrowing constraint will make the marginal propensity to invest (MPI) dependent on income as income gains relax the borrowing constraint on investment. Both of these features, namely a low MPS and a high MPI, contribute to a low-*income* sensitivity of the current account to income. Kindleberger's views on the cause of the dollar shortage (financial frictions) and on the determinants of the size and persistence of the current account dynamics (productivity gap, impatience, a low MPS and a high MPI) fit remarkably well with the framework of the intertemporal approach.

Inclusion of price stickiness and an arbitrary exchange rate adds a nominal driver to current account movements. An *overvalued* currency allegedly exacerbates current account deficits. The contribution of the nominal factor depends on two elements: the degree of nominal price/wage stickiness and the size of the trade elasticities. A small degree of nominal price stickiness makes the effects of the peg on the real exchange rate and the current account short-lived. Similarly, low trade elasticity values make the effects of the peg through the real exchange rate small. The ability of the nominal factor to account for a significant fraction of the current account size and its persistence then rests on these two elements. Kindleberger plausibly argued that the exchange rate could not do so: trade elasticities were low because of the lack of good substitutes for US goods while the technological gap was large. Moreover, he considered that the effects of exchange-rate changes on the current account would be transient.

The above competing points of view all have solid theoretical foundations. While we favor Kindleberger's view, only empirical evidence or a macro model that would be used as a laboratory for carrying out an informed variance decomposition of the historical current account can fully evaluate the relative empirical merit of the alternative theories.

Before concluding this section, we would like to make two related points. The first concerns the theoretically correct definition of the dollar shortage. It is the difference (gap) between the amount of borrowing in the first best—in our context when financial markets operate smoothly—and in a distorted equilibrium (with borrowing constraints). Such gaps or wedges are used in the literature to capture the effects of distortions and are ubiquitous in macroeconomics (for example, Chari, Kehoe, and McGratten 2007). Kindleberger's definition (the balance-of-payments deficit) is meaningful as it relates to, and varies with, the borrowing constraint of the country. However, the balance-of-payments deficit tends to underestimate the true dollar shortage, because it does not take into account the fact that a borrowing constraint also depresses the value of the current account deficit (in the extreme case where borrowing is not possible at all, the deficit is zero). But from a practical point of view, Kindleberger's definition provides a reasonable and useful measure, because it is not model dependent and is also readily measurable.

The second point is that because financial frictions make the actual current account deficit fall short of its optimal size in a frictionless economy, exchange-rate overvaluation actually mitigates the effect of the financial friction by supporting a larger deficit. This consideration is important in the context of the post–World War II reconstruction of the European economies, because imports of capital goods from the United States played a critical role in the reconstruction and the closing of the technological gap. This is standard second-best result and implies that the overvalued exchange rate may have contributed to a faster pace of Europe's income convergence to the US level.

The Resolution of the Dollar Shortage

Starting in the early 1950s, the US economy began to exhibit balance-of-payments deficits, rather than surpluses. Conversely, by the mid-1950s the Western European countries as a group moved to balance-of-payments surplus. What changed?

Several developments in the late-1940s and into the 1950s may have played a role. First, beginning in the early-1950s there was a large and sustained increase in private capital flows from the United States to Western Europe (IMF 1955, p. 35, Table 9). Second, the Marshall Plan (1948–1952) transferred $13.3 billion in economic recovery programs (grants and loans) to Western European economies—about 3 percent of the national incomes of the recipient countries. The recipient countries were permitted to export to the United States, while restricting their imports from the United States (Meltzer 1991, p. 58). The Marshall Plan allowed

European counties to purchase capital goods and raw materials needed to start up their industries, thus helping to relieve the dollar shortage (Kenen 1994, p. 492). Third, in September and October 1949, the British pound sterling was devalued against the dollar by 30.5 percent, and 30 other countries followed the British move within two weeks, with devaluations ranging from 8 percent (Italy) to 53 percent (Austria) (Eichengreen 2007, p. 77). In Hazlitt's terms, more countries became willing to accept that their currency should have a lower value. Unlike the first two direct developments, though, it is not clear that the devaluations played a significant role in the alleviation of the dollar-shortage problem (for instance, see Hicks's view above).

As these changes in the foreign accounts were taking place, the dollar-shortage debate continued through the 1950s on more or less the same terms. Proponents of the dollar-shortage thesis considered that the improvement in the balance-of-payments positions of Western European economies would be short-lived. They argued that the chronic dollar shortage would reemerge once the temporary measures, including the Marshall Plan and discriminatory trade measures against the United States, were removed, and once the effects of the 1949 exchange-rate devaluations wore off. In 1957 alone, at least six books were published on the dollar-shortage issue. Kindleberger (1958) reviewed four of those books—each of which supported the idea that a dollar shortage continued to exist.[7] The general theme running through the books was that the dollar shortage was due to high US productivity growth compared with productivity growth in other countries. After providing generally positive reviews of the books, Kindleberger (p. 394) asked: "How far have we come? Certainly a long way on the details: on the measurement of elasticities, on productivity in general (though not on the introduction of new goods), and on the general theory of balance-of-payments disequilibrium."

Others like Raymond Mikesell (1958, p. 460) took the alternative view, stating: "Dollar shortage should be regarded as a myth; it is only a cloak used to cover up the fallacies of governmental policymakers." Mikesell argued that any shortage of dollars was a consequence of overvalued exchange rates against dollar and was used by foreign governments to justify discriminatory trade practices against US exports (for a fuller exposition, see Mikesell 1959). In response to Mikesell, Kindleberger (1958, pp. 394–95) wrote: "I am still disposed to argue against Mikesell . . . . I imagine, too, that he . . . would be prepared to bet . . . that there was greater likelihood that a given future disequilibrium in the balance of payments of the United States would involve dollar shortage than dollar surfeit."

Kindleberger's view of a chronic dollar shortage derived mostly from his pessimism about the speed of elimination of this technological gap and the willingness

[7] The books were *The World Dollar Problem,* by Donald MacDougall (1957); *Balances and Imbalances of Payments,* by Geoffrey Crowther (1957); *International Monetary Policy,* by William McConnell Scammell (1957); and *Britain's Postwar Dollar Shortage,* by Elliott Zupnick (1957). Apart from the four books reviewed by Kindleberger, other books were *International and Interregional Economics,* by Seymour Harris (1957), and *Europe and the Money Muddle,* by Robert Triffin (1957).

of the United States to lend to Europe. But the productivity gap between the US and Western European economies significantly diminished in the 1950s (for example, Solomon 1977, p. 21). Kindleberger noticed this development and revised his position on the future of dollar shortages. On June 30, 1959, he appeared before the US Congress's Joint Economic Committee, and wrote in the statement accompanying his testimony: "A considerable part of the so-called dollar shortage seems to have been due to continuous innovation on the part of American industry. As foreign countries learned to imitate one product, a new technological gap was opened up in another. This technological gap is no longer so one-sidedly in favor of the United States" (Kindleberger 1959, p. 956). This signaled the end of the dollar-shortage debate.

It is beyond the scope of the present paper to determine the relative contribution of the nominal factors emphasized by Friedman and the real factors emphasized by Kindleberger to the creation and resolution of the dollar shortage. But based on estimates of trade elasticities, the degree of nominal stickiness, the size of the productivity gap after World War II, and the rapid increase in productivity in European countries—that allowed them to produce locally many of the manufacturing goods that they used to import from the United States—we are inclined to side with Kindleberger.[8]

By the late 1950s, the "dollar shortage" was taking on an altered meaning: instead of a problem caused by US current account surpluses of the 1930s and 1940s, it had become a problem of US current account deficits in the 1950s. The famous "Triffin dilemma," spelled out by Robert Triffin (1960) in his book *Gold and the Dollar Crisis*, argued that because the US dollar was the international reserve currency, it was important for the US economy to run ongoing trade deficits—to make US dollars available to the rest of the world and prevent a "dollar shortage." However, Triffin argued that the increase in the global supply of US dollars may exceed the US holdings of gold that back these dollars, creating a confidence problem that undermines the credibility of the system of fixed exchange rates and threatens the US dollar's role as a reserve currency. In addition to such sustainability concerns, other domestic considerations such as the desire to keep inflation low, may interfere with the US provision of a liquid reserve currency for the world. Indeed, excessive growth of the supply of dollar reserves by the early 1970s becomes the reason for the collapse of the Bretton Woods system of fixed exchange rates. For present purposes, suffice it to say that a Triffin-style "dollar shortage" that could arise from insufficiently large and persistent US trade deficits is not the same problem as the post–World War II "dollar shortage" that arises from large and persistent US trade surpluses.

[8]On numerous issues having to do with the 1950s and 1960s academic debates on flexible versus fixed exchange rates, including the view that the Bretton Woods fixed-rate system was prone to recurrent crises, Friedman turned out to be on the right side of the debate and Kindleberger was on the wrong side (for discussion, see Dellas and Tavlas 2018; Tavlas 2023, 2025; Nelson 2026, chapters 3, 7, and 9).

Two Recent Episodes of Dollar Shortage

The recent episodes of what have been called "dollar shortages" have been of two distinct types. One type has afflicted advanced economies that operate floating exchange rates and possess sophisticated financial markets. It has been manifested in violations of what is called the "covered interest parity" condition. The basis for covered interest parity is the difference between two US dollar interest rates, the direct interest rate (say, on US Treasury bonds) and the dollar interest rate on a currency swap agreement, in which two parties swap principal and interest payments in two different currencies. Under smoothly functioning financial markets, arbitrage dictates that these two interest rates should be the same. However, beginning around 2007 and culminating during the 2010s, the covered interest parity condition for various currency pairs with the US dollar deviated significantly from zero. For example, the US dollar–Japanese yen covered interest parity basis exceeded 75 basis points in both 2011 and 2016, an extraordinary amount. This means that investors chose to obtain US dollars in the present and carry them forward, rather than enter into a swap contract with the Japanese yen that promised to deliver the same amount of dollars in the future. By doing so, investors sacrificed more than 75 basis points of return.

A plausible reason for the existence of this differential is the anticipation of a dollar shortage in the future that might obstruct the delivery of the dollars promised. The shortage could be a reflection of heightened, anticipated demand for US dollars (say, because of maturing dollar liabilities by non-US banks) combined with strains in global interbank markets (such as constrained bank access to wholesale dollar funding because of regulatory constraints on bank balance sheets) that limited arbitrage. In response to these huge violations of arbitrage, policymakers throughout the world entered into swap agreements with the US Federal Reserve in order to allow non-US banks to *indirectly* access the Fed's *liquidity* facilities. The dollar shortage was mitigated by the US central bank's willingness to provide dollar liquidity globally. We view this type of dollar shortage as related to the Triffin dilemma—that is, it arises as a result of demand for the US dollar due to its international reserve function.

The other type of modern "dollar shortage" has afflicted emerging economies in a way that shares many of the key features characterizing the European countries during the original "dollar shortage" episode after World War II: namely, low productivity and the desire to catch up, combined with exchange-rate targeting and capital controls. As Reinhart (2016) tells the story, a number of emerging markets had high export values during the bonanza in commodity prices in the early 2000s. Global commodity prices are often denominated in US dollars, so these countries built up substantial US dollar foreign exchange reserves. Some of them even pegged the foreign exchange value of their currency to the US dollar. But when the commodity boom ended after the Great Recession of 2007–2009, countries that had fixed their exchange rates began to bleed reserves, often until the pegging arrangement was ended, and countries with flexible exchange rates saw a sharp

decline in the value of their currency. These balance-of-payments deficits, like those of the early post–World War II period, had little to do with the reserve status of the currencies of the countries that had a trade surplus with these countries that were hemorrhaging reserves. They arose from credit constraints—namely, the reluctance of a creditor (a trade surplus country) to provide credit to what appears to be a risky debtor (a trade deficit country).

■ *We are grateful to Jeffrey Kling, Ed Nelson, Jonathan A. Parker, Timothy Taylor, and Heidi Williams for helpful comments. We thank Elisavet Bosdelekidou and Maria Monopoli for excellent research assistance.*

References

Bloomfield, Arthur I. 1949. "Induced Investment, Overcomplete International Adjustment, and Chronic Dollar Shortage." *American Economic Review* 39 (5): 970–74.

Bordo, Michael D. 1993. "The Bretton Woods International Monetary System: A Historical Overview." In *Retrospective on the Bretton Woods System: Lessons for International Monetary Reform*, edited by Michael D. Bordo and Barry Eichengreen, 3–108. University of Chicago Press.

Chari, V. V., Patrick J. Kehoe, and Ellen R. McGrattan. 2007. "Business Cycle Accounting." *Econometrica* 75 (3): 781–836.

Croome, Honor. 1950. "The Dollar Siege." *Lloyds Bank Review* (July): 25–46.

Crowther, Geoffrey. 1957. *Balances and Imbalances of Payments.* Harvard University Graduate School of Business Administration.

Dellas, Harris, and George S. Tavlas. 2018. "Milton Friedman and the Case for Flexible Exchange Rates and Monetary Rules." *Cato Journal* 38 (2): 361–77.

Economist. 1943a. "The Dollar Problem—I." 145 (5231): 717–18.

Economist. 1943b. "The Dollar Problem—II." 145 (5232): 750–51.

Eichengreen, Barry. 2007. *Global Imbalances and the Lessons of Bretton Woods.* MIT Press.

Enke, Stephen, and Virgil Salera. 1947. *International Economics.* Prentice-Hall.

Friedman, Milton. 1953. "Why the Dollar Shortage?" *The Freeman* 4 (6): 201–203. Reprinted in *Dollars and Deficits: Inflation, Monetary Policy and the Balance of Payments*, edited by Milton Friedman, 211–17. Prentice-Hall.

Graham, Frank D. 1949. "The Cause and Cure of 'Dollar Shortage.'" Princeton University International Finance Section, Essays in International Finance 10.

Harris, Seymour E. 1957. *International and Interregional Economics.* McGraw-Hill.

Harrod, Roy. 1947. *Are These Hardships Necessary?* 2nd ed. Rupert Hart-Davis.

Harrod, Roy F. 1953. "Imbalance of International Payments." *International Monetary Fund Staff Papers* 3 (1): 1–46.

Hazlitt, Henry. 1947. *Will Dollars Save the World?* D. Appleton-Century Company.

Hicks, J. R. 1953. "An Inaugural Lecture." *Oxford Economic Papers* 5 (2): 117–35.

Horsefield, John K. 1969. *The International Monetary Fund, 1945–1965.* Vol. 1. IMF.

IMF. 1955. *Annual Report.* IMF.

Kenen, Peter B. 1994. *The International Economy.* 3rd ed. Cambridge University Press.

Kindleberger, Charles P. 1943a. "Planning for Foreign Investment." *American Economic Review* 33 (1, Part 2): 347–54.

Kindleberger, Charles P. 1943b. "International Monetary Stabilization." In *Postwar Economic Problems*, edited by Seymour E. Harris, 374–95. McGraw-Hill Book Company.

Kindleberger, Charles P. 1949. "The Foreign-Trade Multiplier: The Propensity to Import and Balance-of-Payments Equilibrium." *American Economic Review* 39 (2): 491–94.

Kindleberger, Charles P. 1950. *The Dollar Shortage*. MIT Press.

Kindleberger, Charles P. 1958. "The Dollar Shortage Re-revisited: A Review Article." *American Economic Review* 48 (3): 388–95.

Kindleberger, Charles P. 1959. "Statement." In *Hearings before the Joint Economic Committee Congress of the United States*, 68th Congress, 1st Session, June 29, 30, July 1 and 2, Part 5: *International Influences on the American Economy*, 954–58. Government Printing Office.

Klopstock, Fred H. 1957. "The International Status of the Dollar." Princeton University International Finance Section Essay in International Finance 28.

MacDougall, Donald. 1957. *The World Dollar Problem*. St. Martin's Press.

Mehrling, Perry. 2022. *Money and Empire: Charles P. Kindleberger and the Dollar System*. Cambridge University Press.

Meltzer, Allan H. 1991. "US Policy in the Bretton Woods Era." *Federal Reserve Bank of St. Louis Review* 73 (3): 54–83.

Mikesell, Raymond F. 1958. "Quantitative Import Restrictions and United States Foreign Trade Policy." In *Foreign Trade Policy Compendium of Papers on United States Foreign Trade Policy*, collected by the staff for the Subcommittee on Foreign Trade Policy of the Committee on Ways and Means, 457–67. Government Printing Office.

Mikesell, Raymond F. 1959. "Dollar Shortage: A Modern Myth." *Journal of Political Economy* 67 (3): 307–09.

Nelson, Edward. Forthcoming. *Milton Friedman and Economic Debate in the United States, 1973–2006*. University of Chicago Press. http://cdevcom.win02.tmd.cloud/EdwardNelson/Chapters1to10-MFED-1973to2006.pdf.

Obstfeld, Maurice, and Kenneth Rogoff. 1996. *Foundations of International Macroeconomics*. MIT Press.

Reinhart, Carmen M. 2016. "A 'Dollar Shortage' Has Returned: This is Why." *World Economic Forum*, October 27. https://www.weforum.org/stories/2016/10/a-dollar-shortage-has-returned-this-is-why/.

Robertson, Dennis H. 1954. *Britain in the World Economy*. Allen and Unwin.

Samuelson, Paul A. 1958. *Economics: An Introductory Analysis*. 4th ed. McGraw-Hill Book Company.

Scammell, William M. 1957. *International Monetary Policy*. Macmillan.

Solomon, Robert. 1977. *The International Monetary System, 1945–1976*. Harper and Row.

Tavlas, George S. 2023. *The Monetarists: The Making of the Chicago Monetary Tradition, 1927–1960*. University of Chicago Press.

Tavlas, George S. 2025. "The Debate about the Resilience of the Bretton Woods System: Kindleberger, Nurkse, and Friedman." *History of Political Economy* 57 (5): 813–58.

Triffin, Robert. 1957. *Europe and the Money Muddle*. Yale University Press.

Triffin, Robert. 1960. *Gold and the Dollar Crisis: The Future of Convertibility*. Yale University Press.

Yeager, Leland B. 1965. *International Monetary Relations: Theory, History, and Policy*. Harper and Row.

Zupnick, Elliot. 1957. *Britain's Postwar Dollar Problem*. Columbia University Press.

Journal of Economic Perspectives—Volume 40, Number 2—Spring 2026—Pages 259–266

Recommendations for Further Reading

Timothy Taylor

This section will list readings that may be especially useful to teachers of undergraduate economics, as well as other articles that are of broader cultural interest. In general, with occasional exceptions, the articles chosen will be expository or integrative and not focus on original research. If you write or read an appropriate article, please send a copy of the article (and possibly a few sentences describing it) to Timothy Taylor, preferably by e-mail at <taylort@macalester.edu>, or c/o Journal of Economic Perspectives, Macalester College, 1600 Grand Ave., Saint Paul, MN 55105.

Smorgasbord

The *World Development Report 2025* is on the theme "Standards for Development" (World Bank, December 2025, https://www.worldbank.org/en/publication/wdr2025). In the press release for the report, Indermit Gill remarked: "Standards are both central and unsung today. When they're set right, they go unnoticed: the ship sails through the canal, the building withstands an earthquake, a kilogram weighs the same in Kenya as in Canada, and no one gives the gains that come a second thought. The standardized shipping container might well have catalyzed more trade in manufactured goods than all the trade deals put together. Digital standards could do the same for the services trade. When countries are active in adapting, aligning, and authoring standards, they are a powerful tool for growth and poverty reduction. This report is the first assessment of the role of standards

■ *Timothy Taylor is Managing Editor, Journal of Economic Perspectives, based at Macalester College, Saint Paul, Minnesota.*

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at https://doi.org/10.1257/jep.20261504.

in economic development—and a call to developing nations to make them a core component of their development strategies."

Christina D. Romer discusses "Rowing Together: Lessons on Policy Coordination from American History" (Hutchins Center Working Paper #105, February 2026, https://www.brookings.edu/articles/rowing-together-lessons-on-policy-coordination-from-american-history/). "It is not enough for monetary and fiscal policy to be well coordinated. They also need to be moving toward the appropriate goal. To put it another way: Rowing together is great when the boat is headed in the right direction; it can be a disaster when the boat is headed in the wrong direction. Coordinated policy was a godsend in 2009; it was a tragedy in 1931. A corollary to this fundamental point is that sometimes rowing in opposite directions can be preferable. At least then, the boat stays where it is rather than move in the wrong direction. If monetary or fiscal policy is going astray, it is vitally important that the other tool of macropolicy be uncoordinated. . . . In describing the cases I focus on two important 'whys': (1) Why was monetary and fiscal policy coordinated in some cases and not in others? (2) Why sometimes when policy was coordinated was it headed in the wrong direction?"

Pia M. Orrenius delivered the Presidential Address of the Southern Economic Association in 2025 on the subject "Temporary Fixes, Permanent Problems: Implications of the Growing Reliance on Liminal Status in U.S. Immigration Policy," now published in a paper coauthored with Madeline Zavodny (*Southern Economic Journal*, October 2025, pp. 181–93, https://onlinelibrary.wiley.com/doi/epdf/10.1002/soej.12782). From the abstract: "Record migrant inflows in the post-pandemic period have brought renewed attention to the problems associated with temporary fixes to immigration policy. In response to surges of vulnerable migrants and because of the failure to enact legislative change, multiple presidential administrations devised new ways to enable millions of migrants to remain in the United States for years and even indefinitely without lawful permanent resident status. This lack of permanent status is costly to migrants, employers, and the broader economy. It deters investments in human and physical capital, leading to a misallocation of resources and subpar productivity growth. It also leaves migrants vulnerable to policy changes triggered by a change in administration, most recently the threat of mass deportations. Solutions require Congress to address the coming demographic winter with, among other policies, higher volumes of legal immigration and more opportunity to earn permanent residence and a pathway to U.S. citizenship."

Alexander J. Field tells the story of "The World War II US Rubber Famine" (*Business History Review*, Autumn 2025, 99: 365–90, https://www.cambridge.org/core/journals/business-history-review/article/world-war-ii-us-rubber-famine/624C7C8F548C6A1F583A395A7C2BAEC0). From the abstract: "By April 1942, Japan had cut off almost all US supplies of natural rubber, a key raw material for which the country had effectively no domestic sourcing. The resulting shortage aggravated downward pressure on manufacturing productivity and seriously jeopardized military capability. The risks that this would happen, widely foreseen, could have been mitigated by more or earlier stockpiling, subsidization of domestic plant-based sources of latex, or development of a synthetic rubber industry. At the time of Pearl

Harbor, each route had been pursued in a limited fashion or not at all. This paper explores why, highlighting the outsized role played by businessman/politician Jesse Jones, as well as the multiple channels through which the rubber famine adversely affected the country's wartime economy. This history starkly illuminates a policy dilemma still with us: How much 'insurance' should a country carry when it depends heavily on interruptible foreign sourcing of a strategic input?"

Nicolas Sommet, Adrien A. Fillon, Ocyna Rudmann, Alfredo Rossi Saldanha Cunha, and Annahita Ehsan find "No meta-analytical effect of economic inequality on well-being or mental health" (*Nature*, January 22, 2026, 649: 926–37, https://www.nature.com/articles/s41586-025-09797-z). "Exposure to economic inequality is widely thought to erode subjective well-being and mental health . . . However, existing studies face reproducibility issues . . . Here we present a meta-analysis of 168 studies using multilevel data (11,389,871 participants from 38,335 geographical units) identified across 10 bibliographical databases (2000–2022). Contrary to popular narratives, random-effects models showed that individuals in more unequal areas do not report lower subjective well-being . . . Meta-regressions revealed that the adverse association between inequality and mental health was confined to low-income samples. Moreover, machine-learning analyses indicated that the association with well-being was negative in high-inflation contexts but positive in low-inflation contexts. These moderation effects were replicated using Gallup World Poll data (up to 2 million participants). These findings challenge the view that economic inequality universally harms psychological health and can inform public health policy."

Economics of Artificial Intelligence

Daron Acemoglu, David Autor, and Simon Johnson envision "Building pro-worker artificial intelligence" (Hamilton Project at the Brookings Institution, February 2026, https://www.hamiltonproject.org/publication/paper/building-pro-worker-ai/). They write: "We define pro-worker technologies—including AI—as technologies that make human skills and expertise more valuable by expanding worker capabilities. . . . AI's demonstrated prowess in narrow domains has fostered inflated expectations about what pure automation can achieve. The logic runs as follows: If AI has already surpassed human performance in specific tasks, it must be capable of replicating everything experts do—just without the experts themselves. This assumption shapes thinking across the spectrum, from enthusiasts championing AI's potential to critics warning of its dangers. . . . In this framing, AI does not enhance expertise—it eliminates the need for it entirely. . . . Though full automation remains a distant prospect for much decision-making work, the opportunities for human-AI collaboration are immediately available. AI excels as a partner precisely because its strengths—inexhaustible memory, fast pattern recognition, and continuous operation—compensate for human limitations. Where experts struggle to recall every relevant precedent, consider all possible scenarios, or synthesize insights across disparate fields, AI can fill these gaps. In doing so, it often

enhances distinctively human capacities: interpreting context, making ethical judgments, generating novel solutions, and understanding how individual tasks advance larger objectives. The choice between automation and collaboration need not be absolute, and the division of labor between human judgment and AI capabilities will shift as the technology matures."

Eric Fruits and Kristian Stout offer "AI, Productivity, and Labor Markets: A Review of the Empirical Evidence" (International Center for Law & Economics Issue Brief, February 5, 2026, https://laweconcenter.org/resources/ai-productivity-and-labor-markets-a-review-of-the-empirical-evidence/). "The empirical literature points to a clear pattern. Controlled field experiments and randomized trials document large productivity gains at the task and firm level, often alongside quality improvements. Across writing, customer support, software development, accounting, law, and translation, studies report 15% to more than 50% reductions in task-completion time, meaningful quality gains, and disproportionately large benefits for less-experienced workers, producing 'skill compression' within occupations. At the same time, aggregate labor-market indicators through 2024–2025 show limited disruption, despite rapid adoption. . . . Adjustment to date has occurred through task reallocation and within-firm productivity gains, rather than mass displacement. Macroeconomic projections remain uncertain. Credible estimates range from modest productivity gains to large output increases, depending on assumptions about task coverage, diffusion speed, organizational redesign, and complementary investment."

Alex Imas offers a complementary overview of the evidence in "What is the impact of AI on productivity? Reconciling the micro and the macro evidence" (*Ghosts of Electricity*, January 29, 2026, https://aleximas.substack.com/p/what-is-the-impact-of-ai-on-productivity). "At the micro level, the evidence is mixed but leaning heavily towards positive productivity benefits. Studies find productivity gains ranging from modest increases on some tasks to substantial returns (50%+) to AI. There is also variation in who benefits from AI. While the literature has thus far documented a prevalence for an equalizing effect—with less experienced/skilled workers benefiting most—there are notable exceptions. At the macro level, these gains have not yet convincingly shown up in aggregate productivity statistics." Imas also provides a list of AI-and-productivity studies that he will continually update. Thus, his March 2026 update reads: "In the initial version of the post, I had said that the micro productivity data was not yet showing up in the aggregate statistics, but that I expected that to change in the near-term. I believe that the newest batch of aggregate data—which shows a big upwards revision—is showing signs of AI productivity gains. Of course this is not the final word, but the trend is worth noting."

Tim Sablik interviews Anton Korinek "On how rapid advances in AI might reshape the nature of work and how economists can help society prepare" (*Econ Focus*: Federal Reserve Bank of Richmond, Fourth Quarter 2025, https://www.richmondfed.org/publications/research/econ_focus/2025/q4_interview). "Almost all economists I talk to have come to appreciate that these tools can be very helpful. Economics is a very instrumentalist discipline. When economists realize that something is economically

useful, they won't put up a lot of barriers against it. That said, it's important to acknowledge that we need to be careful with these tools because they do sometimes produce mistakes. They need to be overseen. It's kind of like working with a research assistant. We would not take everything that research assistants produce for us without checking it, and it's the same with our AI systems. . . . [I]f we get AI systems that transform society at the same scale as the Industrial Revolution, the economy is going to be a big part of that transformation. Economists are well positioned to provide insights into how our economy might be reshaped. One thing to consider is that we may have to redesign our systems of taxation. Right now, roughly two-thirds of all income derives from labor, and probably more than two-thirds of all tax revenue comes from taxing that labor income. If the value of labor suddenly falls dramatically because of transformative AI, then we're going to have to tax differently. I prepared a paper for an NBER meeting on public finance in the age of AI in September where my co-author and I argue that if labor becomes a less important part of the economy, we may want to switch to more consumption taxation. And then if human consumption becomes a less important part of the economy, we may ultimately have to switch to taxing the capital behind the AI systems themselves."

James Tebrake, Bachir Boukherouaa, Jeff Danforth, and Niva Harikrishnan describe "StatGPT: AI for Official Statistics" (International Monetary Fund, March 9, 2026, https://www.imf.org/en/publications/departmental-papers-policy-papers/issues/2026/03/10/statgpt-ai-for-official-statistics-573514). They did an experiment with ChatGPT and other AI tools: "The prompt 'Can you generate a table of economic growth rates for the G7 countries taking the data from the latest issue of the IMF's World Economic Outlook. Can you provide data for 2018 to 2025. Can you provide the output in a CSV file.' was entered 10 times in the same conversation—10 times in 10 different conversations, and 5 times in the same conversation with a copy of the latest World Economic Outlook loaded in memory (total of 25 prompts)." They find: "Overall, ChatGPT provided a correct response 34 percent of the time when the prompts were entered into the same conversation. The level of accuracy declined to 17 percent when the request was made using unique conversations. When the latest publication of the World Economic Outlook was loaded into ChatGPT, the level of accuracy fell to 14 percent." The authors describe how to use a series of prompts so that the AI tool will focus in on a specific dataset and retrieve the specific data you want. They also dream of building "a true Global Trusted Data Commons—a comprehensive, AI-ready index of official statistics data . . ."

Global Debt

The *OECD Debt Report 2026* is subtitled: "Sustaining Debt Market Resilience Under Growing Pressure" (March 2026, https://www.oecd.org/en/publications/global-debt-report-2026_e9d80efd-en.html). "Central banks remain the largest domestic holders of government debt in many OECD countries, but as many major central banks have continued to shrink their balance sheets, the market is increasingly

dependent on more price-sensitive investors, including hedge funds, households and certain foreign holders. This transition away from price-insensitive demand, in many ways a normalisation of the investor composition, could increase market volatility. Corporate bond markets have seen a parallel shift in investor composition following the post-2008 regulatory overhaul, with open-ended investment funds, exchange-traded funds (ETFs) and, more recently, principal trading firms, playing an increasingly important role. Changes in the sovereign investor base reverberate in the corporate market. Investors willing to hold riskier securities at a time of lower yields may rebalance their portfolios as central banks withdraw, leaving the growing new corporate issuance to be absorbed by a smaller investor base. . . . These risks must be carefully managed to ensure that sovereign and corporate bond markets, with a combined size of USD 109 trillion, continue to provide stable financing to governments and corporations. This is especially important as they are set to play an increasing role in funding AI investment and defence spending, at a time when decisions on monetary policy, public debt and pension fund asset allocation are coming under growing pressure."

The January 2026 *Global Economic Prospects* report includes a special topics chapter on "Rebuilding Fiscal Space: The Case for Fiscal Rules," with a focus on emerging market and developing economies (World Bank, January 2026, https://www.worldbank.org/en/publication/global-economic-prospects). "At a time when global shocks have become more frequent and government debt among emerging market and developing economies (EMDEs) has climbed to a 55 year high, fiscal rules are an important policy tool for promoting fiscal discipline. More than half of EMDEs have at least one fiscal rule, up from about 15 percent in 2000. Fiscal rules are associated with improvements in budget balances that extend to the medium and long term. Among EMDEs, improvements in the cyclically adjusted primary balance (CAPB) peak five years after fiscal rules are adopted, reaching a cumulative 1.4 percentage points of trend GDP. The gains are more pronounced when institutions are strong and economic conditions are favorable at the time of adoption, and the use of a deficit rule is central to durable improvements. Fiscal rules are also associated with a greater likelihood of fiscal adjustment episodes—multiyear periods of improvement in the CAPB as a percent of trend GDP. . . . Further, fiscal rules need not be complex to be effective. Simple rule frameworks are associated with a higher likelihood of revenue-based adjustment."

Essays on a Theme

Melissa S. Kearney and Luke Pardue have edited a collection of four essays on the subject of *Demographic Headwinds: The Economic Consequences of Lower Birth Rates and Longer Lives* (Aspen Economic Study Group, February 2026, https://www.economicstrategygroup.org/publication/demographic-headwinds/). As one example, Nicola Bianchi and Matteo Paradisi discuss "The Age Divide in the American Workplace." "In the late 1970s, workers over 50 earned roughly 35 percent

more per week than workers under 30. Over the following four decades, this gap widened steadily, reaching a peak in the early 2010s, when older workers earned about 55 percent more. . . . And because working lives are now longer, with many employees remaining in senior roles well into their late sixties, there is little reason to expect this pay gap to return to the more modest levels observed in the 1970s . . . [O]ver time, younger workers have become considerably more likely than their 1970s counterparts to be in the bottom quarter of the wage distribution and less likely to be in the top quarter. For workers over 50, the pattern is broadly reversed. Relative to 1976, their probability of being in the bottom quartile has gradually fallen, reaching around 15–17 percent below its baseline level during the 2010s, while their probability of being in the top quartile has typically been between 5 and 15 percent higher. . . . In the mid-1970s, workers over 50 were about 5 percentage points more likely than workers under 30 to be employed in management occupations in the top quarter of the wage distribution. By 2024, this gap had widened to almost 8.3 percentage points. . . . [T]hese patterns are common to most high-income economies, rather than being unique to the United States."

Dirk Niepelt has edited *Frontiers in Digital Finance*, a collection of 20 short and readable essays (Centre for Economic Policy Research, November 2025, https://cepr.org/publications/books-and-reports/frontiers-digital-finance). From his introduction: "The digitisation of payment, trading, and settlement systems is reshaping the financial architecture. New technologies are transforming how value is created, stored, transferred, and accounted for, altering the balance between public and private money, enabling the bundling of services, challenging traditional financial institutions, and prompting a wave of regulatory and institutional responses. . . . Some regions are leapfrogging others, and conflicting ideologies about the proper role of the state in money give rise to fragmentation and concerns about monetary sovereignty. This book offers an overview of major trends, as analysed by leading researchers and policymakers. It is structured in four parts. Part 1 presents regional perspectives, examining the approaches taken by India, Brazil, sub-Saharan Africa, the United States, and the euro area. . . . Part 2 delves into stablecoins—the shooting stars in the digital financial ecosystem. . . . Part 3 turns to the concept of monetary 'singleness'—the principle that all forms of money in a currency area should be fully interchangeable and trade at par . . . Part 4 brings together chapters on tokenisation, digital platforms, and decentralized finance (DeFi), and their broader impact on service bundling, credit allocation, financial inclusion, and consumer protection."

Discussion Starters

Marc-William Palen looks back at the first age of globalization in "When Free Trade First Faltered" (*Finance & Development*, September 2025, https://www.imf.org/en/publications/fandd/issues/2025/09/when-free-trade-first-faltered-marc-william-palen). "In the 60 years or so before World War I, global trade grew rapidly despite the ever-higher tariff walls built by the rising protectionist empires of the

United States, Germany, Russia, France, and Japan. Geopolitical conflicts and trade wars grew more common even as markets became more integrated. These contradictions were at the heart of heated debates over free trade and economic nationalism that dominated the industrializing world at the time. . . . The similarities between then and now are hiding in plain sight. The quarter century between the end of the Corn Laws in 1846 and the start of the global turn to protectionism in the early 1870s saw unprecedented trade liberalization, as did the 25 years after the end of the Cold War. And much as the 19th century liberal free-trade advocates underappreciated the political appeal of nationalism and economic self-sufficiency, so too were their intellectual successors in the late 20th century precipitous in predicting the end of the nation-state—and even the end of history."

Andrew Biggs points out that "It's Fine to Embrace FDR's Vision of Social Security, But You Don't Need to Embrace Nixon's and Carter's" (*Little Known Facts*, January 29, 2026, https://littleknownfacts.substack.com/p/its-fine-to-embrace-fdrs-vision-of). "I've long been an advocate of a flat benefit structure for Social Security. That is, rather than acting as an increasingly expensive pension plan—where the top half of the income distribution receives almost two-thirds of total benefits—Social Security reform should follow other Anglo countries and gradually transition the retirement program into a robust but limited safety net for poor seniors. . . . But here's another take: even if one wanted to preserve Social Security in the wage-replacement model exactly as it was envisioned by FDR, I would take the deal. Why? . . . What makes Social Security so expensive today—and increasingly expensive in the future—is not the benefit formula that first started paying benefits out in 1940. It is a series of benefit increases enacted mostly in the 1970s by politicians seeking reelection, supported by a small group of policy leaders who envisioned a far more expansive program than FDR or ordinary Americans did."

The American Economic Association

Correspondence relating to advertising, business matters, permission to quote, or change of address should be sent to the AEA business office: aeainfo@vanderbilt.edu. Street address: American Economic Association, 2014 Broadway, Suite 305, Nashville, TN 37203. For membership, subscriptions, or complimentary access to JEP articles, go to the AEA website: http://www.aeaweb.org. Change of address notice must be received at least six weeks prior to the publication month.

The Journal of
Economic Perspectives

Spring 2026, Volume 40, Number 2

Symposium

Competition in Health Care

Articles

Features

AMERICAN ECONOMIC ASSOCIATION