Designing Dynamic Reassignment Mechanisms: Evidence from GP Allocation

Ingrid Huitfeldt   Victoria Marone   Daniel Waldinger

Supplemental Appendix

December 2025

Appendix A Data Details

A.1 Data Sources

The data for this study are primarily derived from *Fastlegedatabasen*, maintained by the Norwegian Directorate of Health, as well as the administrative registries at Statistics Norway (SSB). Individuals in the data (both patients and GPs) are each assigned a unique identifier, which can be merged across datasets. Patients' municipalities of residence are identified at the monthly level using information from *Fastlegedatabasen*. Monthly municipality of residence is carefully tracked in this data because municipalities make monthly transfer payments to one another when a patient residing in one municipality enrolls with a GP located in another.

The raw GP enrollment and waitlist data are provided at the enrollment spell and waitlist spell level, respectively. Enrollment spells start and end only on the first and last days of a month, and so are in increments of full months. Waitlist spells can begin and end in continuous time. The time at which an individual joins a waitlist governs their priority in the waitlist. We convert the spells data to an individual-month panel. If an individual is on multiple waitlists in the course of a month, we take the most recent waitlist they were on. For our descriptive analyses, we make only three restrictions: (i) dropping individual-months that occur after the individual's registered date of death, (ii) dropping individuals under age 16, and (iii) dropping individual-months in which no current GP was registered. Appendix Table A.1 shows the number of individual-months dropped by these restrictions. Children under age 16 represent 17 percent of the data and population.

With respect to the demographic characteristics reported in Table 1, temporary residents are defined as individuals whose national identity number is a "D-number," rather than a "Birth number."[57] Our measure of annual income corresponds to an individual's total labor

[57] For more information, see https://www.norden.org/no/info-norden/norsk-identitetsnummer.

Table A.1. Sample Construction Statistics

Criteria	2017 Number	2017 Pct. of initial	2018 Number	2018 Pct. of initial	2019 Number	2019 Pct. of initial
Initial patient-months	63,437,631		63,813,290		64,212,740	
Registered after death	1,918	<0.01	48	<0.01	27	<0.01
Age < 16	11,475,019	0.18	10,694,796	0.17	9,918,551	0.15
No current GP	1,801	<0.01	2,066	<0.01	2,810	<0.01
Final total	51,958,893		53,116,380		54,291,352	

Notes: This table shows the number of patient-months each year dropped due to each sample selection criterion (in the order in which drops were made). The primary restriction on the data is to remove children under the age of 16.

earnings, plus capital income and government transfers. Our flag for post-secondary education corresponds to an individual having completed any educational degree beyond high school. We classify municipalities as "urban" or "rural" based on the municipality's Centrality Index, a classification maintained by Statistics Norway.[58] We classify centrality indices [1, 2, 3] as urban, and [4, 5, 6] as rural.

Information on whether a patient has a chronic health condition is drawn from healthcare utilization data in the year 2017. These data are derived from the Norwegian Patient Registry (NPR) and the Control and Payment of Health Reimbursements Database (KUHR). The data contain the universe of publicly-financed healthcare encounters that occurred in both office-based and hospital-based settings in 2017. Importantly, the data contain diagnosis codes (ICD-10), which allow us to determine whether an individual presented with a chronic condition diagnosis in that year. We classify an individual has having a chronic condition if, based on the diagnoses recorded in 2017, their Charlson Comorbidity Index score is greater than zero.[59]

A.2 Construction of Demand Estimation Sample

We estimate our model using the set of (adult) patient-months where the patient is resident in the Trondelag region and the month is between December 2018 and November 2019. Once we have limited to patients in Trondelag, for practical reasons we do not wish to allow these patients to choose from among all possible GPs in all of Norway. We therefore make further

[58] For details, see https://www.ssb.no/en/klass/klassifikasjoner/128.

[59] For a description of the Charlson Comorbidity Index and the 16 chronic conditions it covers, see for example: https://reference.medscape.com/calculator/879/charlson-comorbidity-index-score-cci-score.

restrictions relating to the definition and construction of patients' GP choice sets.

Our baseline choice set definition is a 60 minute drive time radius around each patient's municipality of residence. Any individuals that were enrolled with or joined a waitlist for a GP further than 60 minutes driving time are dropped from the analysis. Because this restriction will in particular exclude instances of patients that moved far but did not immediately switch their GP, we exclude all patient-months in which a patient recently or imminently moved over 60 minutes (i.e., within months $[-6, 1]$).[60] In addition, we drop patient-months in which the current or requested GP exited during the subsequent month. Finally, for patient-months in which no GP switch was requested, we drop observations where the patient was currently standing on a waitlist. These months reflect the outcome of a prior decision to join a waitlist, which we already capture in the set of patient-months where was a switch was requested.

Appendix Table A.2 describes the observations dropped at each step. We are left with 4,194,737 patient-months in which a switch was not requested, and 18,312 months in which a switch was requested. These observations together are informative about when patients are attentive and consider switching GPs. Conditional on making a switch request, the switcher-month observations are informative about which GP characteristics patients value.

Table A.2. Demand Estimation Sample Construction

	Non-switches		Switches	
Criteria	Number	Pct. of initial	Number	Pct. of initial
Initial patient-months in Trondelag region	4,617,667		29,937	
Moved further than 60 min. in months [t-6, t+1]	80,105	2	5,964	20
Requested a GP further than 60 minutes			2,865	10
Current GP further than 60 minutes	118,257	3	2,309	8
Current GP exiting next month	157,832	3	394	1
Requested an exiting GP			93	<1
Currently on a waitlist	66,736	1		
Final total	4,194,737		18,312	

Notes: This table describes the sample selection criteria used in constructing the demand estimation sample. The unit of observation is a patient-month, and location of residence is measured at the monthly level. "Switches" are patient months in which the patient either joined a waitlist or else switched to a GP with open slots. "Non-switches" are patient months where a patient took no action. Both switches and non-switches are used in demand estimation.

Given the large size of the data, we proceed with estimation using all switcher patient-months and a random sample of 15,000 non-switcher patient-months. In estimation, we then

[60] In our counterfactuals, we will recover the attention probabilities for patients who moved a long distance by extrapolating the relationship between attention and move distance estimated at shorter move distances.

re-weight the sampled non-switcher observations such that they represent the full set of observations. Given that patients with recent moves are disproportionately likely to be excluded from demand estimation because they hold a GP further than 60 minutes, we also adjust the sampling weights to match the original observed proportion of movers versus non-movers in both the non-switcher and switcher samples. Finally, once the random sample of non-switchers is drawn, we enforce a final restriction that all GPs remaining the analysis are chosen a sufficient number of times for a GP fixed effect to be estimated. We require each GP to be present in at least 400 individuals' choice sets. GPs that do not meet this requirement are dropped from all choice sets, and all patients that choose such a GP are also dropped. Taken together, these collected restrictions effectively drop all patients who move into or out of Trondelag during the sample period. The exception would be if the move was from/to a neighboring municipality where the prior GP was still within 60 minutes, and where the prior GP was within 60 minutes for at least 400 patients in Trondelag. But, as discussed in footnote 35, Trondelag was selected in the first place for the rarity of such occurrences. Our final estimation sample consists of 14,820 non-switcher months and 17,864 switcher months.

Appendix B Additional Analysis

B.1 Mechanical Simulations: Implementation Details

Implementation of TTC. Only patients on waitlists participate in the TTC algorithm, as any patient not standing on a waitlist retains their slot on their current GP's panel. TTC is thus run to find a matching between (a) the set of patients standing on waitlists, and (b) the set of GP panel slots currently held by those patients. Preferences are all strict and are determined as follows:

(a) Patients standing on a waitlist first prefer their waitlist GP, then their current GP.

(b) GPs first prefer all their incumbent patients who are participating in TTC. Since these patients must be waiting on a waitlist, they have a global priority order determined by the moment in time at which they joined that waitlist. GPs prefer their incumbent patients in order of this global priority. GPs then prefer all the patients on their waitlist in the order in which they joined.

We iteratively look for cycles within chains that begin with a patient, starting with the patient who is highest on the global priority list and going down the list from there. Our procedure is based on the "you request my house, I get your turn" algorithm first proposed by Abdulkadiroğlu and Sönmez (1999).

Mechanical Simulation of TTC on Historical Data. We implement the TTC algorithm monthly on the historical waitlists data between November 2016 and December 2019. The purpose of this exercise is to generate a naive estimate of how TTC would have changed the waiting lists had it been in place during this period. In that spirit, we hold all of the following objects fixed as they are observed in the data: patient entry (births or immigration), patient exit (deaths or emigration), administrative auto reassignments of patients, GP entry, GP exit, GP panel caps, and, critically, patient actions. A patient's action each month can be (a) doing nothing and remaining with their current GP, (b) switching to an open GP, or (c) joining a waitlist for a full GP. We then run the TTC algorithm on all patients standing on waitlists each month. Patients who can participate in a cycle are reassigned to their desired GP's panel and removed from waitlists.

B.2 Preferences for GP Characteristics

This section investigates patient preferences for GP characteristics using a conditional logit analysis of GP choice. Given the size of the data, we limit attention to the demand estimation subsample described in Section IV.A and Appendix A.2. Further, for the purposes of this specific analysis, we limit to the set of patient-months in which the patient requested to switch their GP, either by switching to an open GP or else by joining a waitlist. Within this subsample, we estimate the following conditional logistic regression specification:

$$u_{ijt} = -d_{ijt} + w_{ijt}\alpha + \delta_j + X_{ijt}\beta + \sigma_\epsilon \epsilon_{ijt} \,,$$

where d_{ijt} is driving time between patient i's municipality of residence and GP j's office, w_{ijt} is the number of patients on GP j's waitlist at the beginning of the month t when patient i made their switch request, δ_j is a GP fixed effect, X_{ijt} includes interactions between patient and GP age and gender as well as the patient's chronic health status, and ϵ_{ijt} is a type-1 extreme value idiosyncratic shock. When reporting results, we normalize the coefficient on driving time to -1.

Appendix Table B.1 reports results from three specifications. Column (1) excludes GP fixed effects and interactions between patient and GP characteristics. It controls only for

GP age and gender, such that all horizontal GP differentiation comes from travel time and idiosyncratic taste shocks. Younger and female GPs are chosen more often, as are GPs with shorter waitlists. Column (2) adds observable patient-GP match-specific heterogeneity based on age, gender, and patient health status.[61] Compared to male patients, female patients have a strong preference for female GPs; they would be willing to travel more than 6 minutes longer to see one, both with or without a chronic health condition. There also appears to be homophily on age, though it is weaker and not always statistically significant. All patient groups prefer younger GPs, but compared to a patient over age 45, a younger patient would travel about a minute longer to see a GP who is also under 45.

Column (3) adds GP fixed effects, absorbing any persistent GP-specific differences in desirability. The coefficients related to match specific heterogeneity—age and gender interactions and the standard deviation of the idiosyncratic shock—are very similar to column (2).[62] However, adding GP fixed effects more than triples the magnitude of the estimated coefficient on waitlist length. This is consistent with more desirable GPs having longer waitlists. Column (3) isolates responsiveness to variation in the length of *the same GP's* waitlist (relative to other waitlists) over time. This type of variation occurs naturally in queues due to statistical fluctuations in the number and types of agents and objects arriving over time (Waldinger, 2021; Leshno, 2022). The sensitivity of patients' choices to waiting time (information) will be a key determinant of the equilibrium implications of changing the design of the waitlist mechanism.

B.3 TTC is Not Pareto Improving: Styled Examples

This section constructs simple examples in which TTC is and is not Pareto improving over Norway's status quo mechanism (Waitlists). We also compare to the patient-optimal stable match produced by DA, which is always Pareto improving over Norway's status quo mechanism. As in our mechanical simulations, each example holds initial assignments and patient switch requests fixed across assignment algorithms, and maintains FCFS priority among waiting patients. Appendix Section D.2 formally describes each algorithm.

[61] Temporary residents are treated as a distinct category because their demographic information is not available. We explored other GP and patient characteristics, and found few that were statistically and economically significant in explaining patients' GP choices.

[62] The levels of the coefficients are different than in column (2) because column (3) normalizes the preferences of temporary residents for GP age and gender to zero, but the *difference* between any two coefficients is similar.

Table B.1. Preferences for GP Characteristics: Conditional Logit

Variable	(1) β	(1) SE	(2) β	(2) SE	(3) β	(3) SE
Travel time (minutes)†	−1.000		−1.000		−1.000	
Waitlist length	−0.031	0.002	−0.031	0.002	−0.113	0.006
Female GP	1.361	0.091				
× Temporary resident			2.506	0.247		
× Perm. res. female, non-chronic, age 16–45			4.006	0.202	1.432	0.304
× Perm. res. female, non-chronic, age 45+			3.457	0.254	0.626	0.342
× Perm. res. male, non-chronic, age 16–45			−2.427	0.221	−4.905	0.330
× Perm. res. male, non-chronic, age 45+			−3.219	0.316	−5.833	0.398
× Perm. res. female, chronic, age 16–45			3.873	0.294	1.326	0.370
× Perm. res. female, chronic, age 45+			2.755	0.277	−0.079	0.359
× Perm. res. male, chronic, age 16–45			−2.598	0.455	−5.097	0.504
× Perm. res. male, chronic, age 45+			−2.244	0.386	−4.871	0.451
GP age 45+	−2.081	0.092				
× Temporary resident			−1.670	0.252		
× Perm. res. female, non-chronic, age 16–45			−2.471	0.197	−0.744	0.313
× Perm. res. female, non-chronic, age 45+			−1.674	0.254	0.025	0.351
× Perm. res. male, non-chronic, age 16–45			−2.523	0.219	−0.681	0.327
× Perm. res. male, non-chronic, age 45+			−1.063	0.294	0.758	0.381
× Perm. res. female, chronic, age 16–45			−3.049	0.302	−1.251	0.384
× Perm. res. female, chronic, age 45+			−1.684	0.278	0.031	0.368
× Perm. res. male, chronic, age 16–45			−2.529	0.441	−0.647	0.500
× Perm. res. male, chronic, age 45+			−1.826	0.378	−0.018	0.444
Coeff. on epsilon shock	6.868	0.103	6.852	0.102	6.721	0.115
GP FE	N		N		Y	
Dep. variable mean	0.005		0.005		0.005	
# Observations	3,293,694		3,293,694		3,293,694	

Notes: This table reports coefficient estimates from conditional logistic regressions predicting the GP choices of patients who requested to switch GP. The unit of observation is a (patient-month, GP) pair. The sample of switcher-months corresponds to those used in the demand estimation sample, described in Section IV.A. The average patient-month has 192 GPs within choice set. Since this analysis conditions on requesting to switch, we exclude each patient's current GP from choice set. The outcome variable is an indicator for whether the patient requested to switch to a specific GP. GP FE indicates whether the specification includes a fixed effect for each GP, with one GP's fixed effect normalized to zero. In column (3), temporary residents serve as the omitted category for interactions between GP and patient characteristics. †By normalization

Example 1. We begin with an example in which TTC generates a Pareto improvement. Appendix Figure B.1 illustrates an economy with the following primitives:

- There are five patients (i_1, i_2, i_3, i_4, i_5) and three GPs (A, B, and C), each with a panel cap of two.
- At $t = 0$, i_1 and i_2 are assigned to A; i_3 and i_4 are assigned to B; and i_5 is assigned to

C. As a result, A and B have full panels, while C has an open slot.

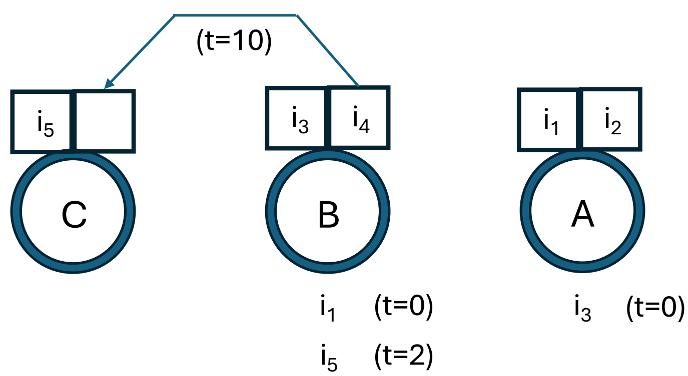
- At $t = 0$, i_1 requests to switch from A to B, while i_3 requests to switch from B to A. Since there are no other waiting patients, each patient is placed at the front of their requested GP's waitlist.

- At $t = 2$, patient i_5 requests to switch from C to B, and is placed behind patient i_1 on B's waitlist.

- At $t = 10$, patient i_4 requests to switch from B to C, which can be immediately executed because there is an open slot.

- At $t = 20$, patient i_2 dies, vacating a slot on A's panel.

Figure B.1. Example 1: TTC Generates a Pareto Improvement



We now consider when each patient is reassigned under each assignment algorithm:

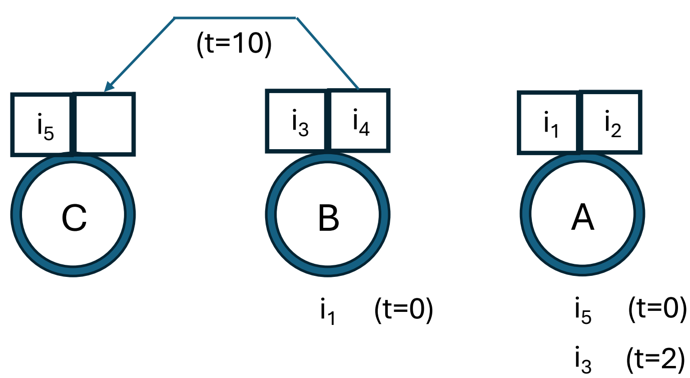
Waitlists: Under Norway's status quo mechanism, patients i_1 and i_3 cannot trade immediately, even though each is waiting for the other's GP. Instead, they must wait for a vacancy on either A's or B's panel. A vacancy arises at $t = 10$, when i_4 requests to switch from B to C. Since C has an open slot, the following steps occur: i_4 is assigned to C, creating a vacancy on B's panel; i_1 is assigned to B from the front of the waitlist, creating a vacancy on A's panel; i_3 is assigned to A from the front of the waitlist, creating another vacancy on B's panel; and i_5 is assigned to B, creating a vacancy on C's panel. At this point, no patients remain on a waitlist. Thus, all patients who have requested to switch GP are reassigned at $t = 10$.

DA: We next consider the allocation produced by running patient-proposing DA each month, after filling any vacant slots from waitlists. At $t = 0$, i_1 and i_3 are allowed to "trade" GPs since each is at the top of their respective waitlist. During the DA algorithm, i_1 first proposes to B, and i_3 first proposes to A; since neither patient is rejected, the algorithm terminates with each patient reassigned to their requested GP. Between $t = 0$ and $t = 10$, only patient i_5 is waiting to switch GP, so DA produces no additional reassignments. i_5 must still wait until $t = 10$, when i_4 open switches from B to C, to be reassigned to B. Thus, compared to Waitlists, DA allows patients i_1 and i_3 to be reassigned at $t = 0$ instead of $t = 10$, while i_5 is still reassigned at $t = 10$. This is a Pareto improvement because i_1 and i_3 wait for strictly less time, whereas i_5 waits for the same amount of time.

TTC: In this example, the TTC algorithm produces the same allocations as DA. At $t = 0$, pairs (i_1, A) and (i_3, B) immediately form a cycle, so the two patients are reassigned. TTC produces no additional reassignments, and patient i_5 is reassigned to B at $t = 10$, after patient i_4 switches to C. Like DA, TTC generates a Pareto improvement relative to Waitlists.

Example 2. Figure B.2 illustrates an example that is identical to Example 1 except that (i) patient i_5 requests GP A instead of B, and (ii) patients i_5 and i_3 request A in reverse order: i_5 requests to switch from C to A at $t = 0$, while i_3 requests to switch from B to A at $t = 2$. We again consider when each patient is reassigned under each algorithm.

Figure B.2. Example 2: TTC is not Pareto Improving



Waitlists: As in the previous example, no reassignments occur until $t = 10$. At $t = 10$, patient i_4 open switches to C's panel, leaving a vacant slot on B's panel; patient i_1 is reassigned to the vacant slot on B's panel, leaving a vacant slot on A's panel; and patient i_5 is reassigned to the vacant slot on A's panel, leaving a vacant slot on C's panel. No further reassignments occur this period, and patient i_3 remains on the waitlist for A. At $t = 20$, patient i_2 dies, vacating a slot on A's panel which is reallocated to patient i_3. No further patients are waiting to switch GP. Thus, all waiting patients are reassigned at $t = 10$ except i_3, who is reassigned at $t = 20$.

DA: In this example, patients i_1 and i_3 could execute a mutually beneficial trade beginning in $t = 2$, when i_3 joins GP A's waitlist. However, the DA algorithm will not execute this trade while patient i_5 remains on A's waitlist because it would violate waiting time priority to reassign i_3 before i_5. The specific steps would be as follows:

- In each period $t = 2, 3, ..., 9$, the DA algorithm would reassign zero patients, despite the potential gains from allowing i_1 and i_3 to trade. The sequence of proposals would be as follows. i_1 first proposes to B and i_5 and i_3 propose to A; A provisionally holds i_5's proposal and rejects i_3, who is lower on the waitlist. In the next round, i_3 proposes to B (their current GP), who then rejects i_1. In the next round, i_1 proposes to A, who then rejects i_5, who then proposes to C. At this point, all waiting patients propose to their current GPs, and since no patient is rejected, the algorithm terminates.
- At $t = 10$, DA reassigns the same patients that are reassigned under Waitlists (all except i_3). To see this, note that this is the patient-optimal *stable* match; reassigning i_3 would require not reassigning i_5, which would violate waiting time priority.
- At $t = 20$, patient i_3 is reassigned to A after patient i_2 dies.

Thus, in this example, DA produces exactly the same outcomes as Waitlists because it cannot violate waiting time priority.

TTC: In this example, TTC produces a different allocation than Waitlists and DA, and in the process, harms patient i_5 (while benefiting patients i_1 and i_3). At $t = 2$, when i_3 requests to switch from B to A, i_1 and i_3 form a cycle and trade GPs when the TTC algorithm is run. Patient i_5 remains on the waitlist for GP A. At $t = 10$, patient i_4 open switches to panel C, vacating a slot on B's panel. This slot remains vacant since there are no patients waiting for B, and i_5 remains on the waitlist for GP A. At $t = 20$, patient i_5 is reassigned to the slot vacated by i_2, who dies.

Compared to Waitlists and DA, patients i_1 and i_3 are reassigned 8 and 18 periods earlier, respectively, since they trade GPs at $t = 2$ instead of being reassigned at (respectively) $t = 10$ and $t = 20$. However, patient i_5, whose GP is not oversubscribed, waits 10 periods *longer*, being reassigned at $t = 20$ instead of $t = 10$. This example illustrates how TTC not only moves reassignments forward in time, but also *reallocates* slots towards patients that can facilitate trades. Specifically, TTC allows patient i_1's slot on A's panel to be reallocated sooner, reducing the total amount of time waited in the economy. However, it also reallocates i_1's slot from patient i_5, who cannot form a cycle, to i_1, who can.

B.4 Predictors of Switching GPs

A number of patient characteristics appear predictive of a desire to switch GPs. For example, Table 1 shows that 34 percent of patients who had switched to an open GP and never used a waitlist over the period 2017–2019 had moved at some point during that period, compared only 6 percent of patients who had neither switched GP nor used a waitlist. To investigate the relative importance of various factors that may motivate GP switching, we regress an indicator for a GP switch request on patient characteristics. A GP switch request includes either an immediate switch to an open GP or a waitlist join. Our focal patient characteristics include time-invariant patient demographics and the timing of a switch relative to a move.

Appendix Table B.2 reports these results. The outcome variable is a switching indicator. The indicator is scaled by 100 for readability, so coefficients should be interpreted as percentage points. The overall probability of observing a switch request is 0.718 percent. Specification (1) includes only a set of nine mutually exclusive and exhaustive patient demographic types. We find that temporary residents are substantially more likely to request to switch GPs than permanent residents, with a baseline probability of 1.501 percent. Among permanent residents, patients with a chronic health condition and younger patients (particularly females) are more likely to switch.

Specification (2) introduces information on timing relative to a patient move. Conditional on a move being observed in the past 6 months, the current month, or the next month, we create four categories of moves based on an intersection of timing and distance of move. We find that switches are substantially more likely to occur in the month concurrent with or directly preceding the month of a move, and also that switches are far more likely to be associated with longer moves relative to shorter moves. For example, a female moving over 30 minutes away either this month or next month (i.e., in month [t, t+1]) has an extra 12.708 percent chance of requesting to switch GPs, relative to a female who has not recently or imminently moved. Variation induced by moves therefore appears, unsurprisingly, to be an important determinant of when patients request to switch GPs.

B.5 Evidence on Endogenous Switch Requests

An important implication of our demand model is that the rate of GP switch requests does not depend on current or future waiting times for different GPs, but only on the set of GPs that can be chosen. This rules out induced demand for switching in our counterfactuals if waiting

Table B.2. Predictors of Switching GPs

Variable	(1) β	(1) SE	(2) β	(2) SE
Temporary resident	1.501	0.005	1.177	0.005
Perm. res. female, non-chronic, age 16–45	1.135	0.002	0.892	0.002
Perm. res. female, non-chronic, age 45+	0.462	0.001	0.415	0.001
Perm. res. male, non-chronic, age 16–45	0.762	0.002	0.583	0.002
Perm. res. male, non-chronic, age 45+	0.348	0.001	0.306	0.001
Perm. res. female, chronic, age 16–45	1.260	0.004	1.024	0.003
Perm. res. female, chronic, age 45+	0.557	0.002	0.512	0.002
Perm. res. male, chronic, age 16–45	0.871	0.004	0.704	0.004
Perm. res. male, chronic, age 45+	0.421	0.002	0.385	0.002
Moved ≤30 minutes, this or next month				
× Temporary resident			7.509	0.174
× Perm. res. female			3.738	0.048
× Perm. res. male			2.700	0.040
Moved ≤30 minutes, prev. 6 months				
× Temporary resident			2.277	0.075
× Perm. res. female			1.577	0.021
× Perm. res. male			1.200	0.017
Moved >30 minutes, this or next month				
× Temporary resident			21.634	0.187
× Perm. res. female			12.708	0.057
× Perm. res. male			8.924	0.047
Moved >30 minutes, prev. 6 months				
× Temporary resident			3.746	0.063
× Perm. res. female			3.561	0.021
× Perm. res. male			2.887	0.018
Dep. variable mean	0.718		0.718	
R^2	0.009		0.022	
# Observations	154,793,455		154,793,455	

Notes: This table investigates observable predictors of requesting to switch GPs. The unit of observation is the patient-month. The sample includes all adult patients in Norway over the period January 2017 to November 2019. Regressions are linear probability models where the outcome variable is an indicator for a GP switch request (either a waitlist join or a switch to a GP panel with open slots). All covariates are indicator variables. For readability, the outcome variable is scaled by 100; coefficients should therefore be interpreted as percentage points.

times change, as well as other types of forward-looking behavior. This section tests these model predictions empirically. We first use a change in the waitlist rules that reduced waitlist lengths to test whether this led to additional switch requests. We then test whether switch requests are sensitive to future waiting times. We fail to find evidence that the number of switch request responds to the returns to switching as a function or current or future waiting times.

B.5.1 Responsiveness to Aggregate Waitlist Length

An important question in our counterfactuals is whether more patients would request to switch GPs if expected waiting times fell systematically. Such a response would be predicted by a model where patients pay a "switching cost" at the time they *request* to switch GPs. In contrast, our model of exogenous inattention rules out this type of response. As there is no direct switching cost at the time of requesting to switch GP, an attentive patient will always request to switch GP if any GP is preferable to their current one.

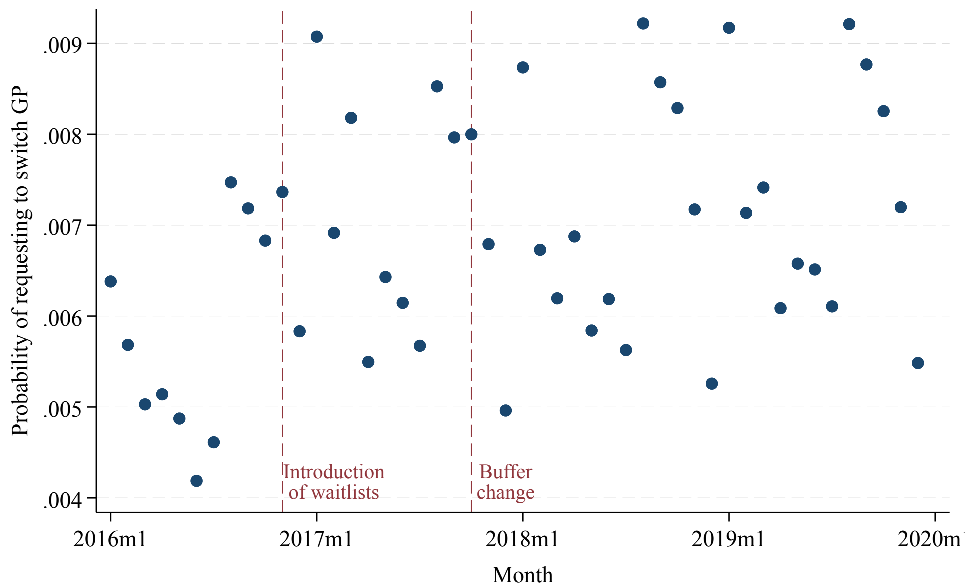
Testing for such a response is challenging because it requires variation in aggregate waitlist lengths, holding other demand and supply conditions fixed. One observation we can make is that the rates of switch requests have remained steady since waitlists were introduced in November 2016, and the number of waiting patients has grown almost linearly since then. If patients were more likely to make switch requests when waiting times were shorter, we would have expected a spike and then decline in switch requests after the introduction of waitlist. The time series evidence therefore weighs against a fully attentive model. However, it is also possible that patients gradually became aware of the new waitlist system, offsetting a deterring effect of increasing waiting times. We therefore also look for shorter-term variation in aggregate waitlist lengths.

Another source of variation that occurred during our sample period—but before the period used for structural estimation—systematically reduced both perceived and actual waiting times. Until October 2017, Norway maintained a "buffer" of 20 slots on each GP's panel, and only assigned waiting patients after 20 slots were available.[63] The buffer was intended to prevent other additions to a GP's panel, such as births and administrative reassignments, from violating the panel cap. However, recognizing that this buffer kept patients waiting for GPs with open slots, the buffer was reduced from 20 to 10 slots in October 2017. GPs with 10–19 open slots had their panels filled with patients from the waitlist at the end of the month, resulting in a one-time drop in aggregate waitlist lengths. We use this variation to test whether aggregate switching rates increased immediately after this policy change, both overall and differentially by the amount different geographic regions were impacted.

[63] The algorithm would fill all of the open slots once the buffer was exceeded, so a given GP did not always have 20 extra slots.

Appendix Figure B.3 plots the monthly share of patients requesting to switch GP in a two-year window around the month of the buffer change. Monthly switching rates range between 0.5 percent and 1 percent, but there is no visible increase in the aggregate number of switch requests beginning in October 2017. While there is not a clear aggregate response in switching requests, there was substantial variation across municipalities in the extent to which the change in the waitlist buffer affected waitlist lengths. In particular, smaller municipalities were much more affected. We can therefore explore geographic heterogeneity in the impact of the buffer change.

Figure B.3. Probability of GP Switch Request: Event Study



Notes: The figure shows the average rate of GP switch requests in Norway over time, where a switch request includes both joining a waitlist and switching to an open GP. Waiters were introduced in November 2016. The waitlist buffer was reduced from 20 to 10 in October 2017.

The GPs directly impacted by the buffer change were those with 10–19 slots available prior to the change. For these GPs, all available slots were filled once the buffer was reduced to 10, but would not have been had the buffer remained at 20. We measure exposure to the buffer change in two ways. First, we calculate the fraction of GPs in each municipality who are near the buffer ("Near-Buffer"), meaning they had 10–19 open slots on their panel as of September 2017. Second, since waitlist lengths varied among Near-Buffer GPs, we multiply each municipality's Near-Buffer share by the average change in the length of the waitlist among Near-Buffer GPs. This measure isolates the average change in the length of *all* GP waitlists induced by the buffer change, which is more than 1.5 patients in the median municipality. We interact both measures with indicators for 1–3 and 4–12 months after the buffer change.

Appendix Table B.3 reports these results. We find no differential impact of these exposure measures on switching rates, regardless of whether or not we control for a linear time trend in switching rates.

Table B.3. Switch Requests by Exposure to Buffer Change

	Near Buffer				Waitlist Change			
	(1)		(2)		(1)		(4)	
Variable	β	SE	β	SE	β	SE	β	SE
Oct.-Dec 2017 × 1 S.D. Exposure	0.120	0.153	0.121	0.153	0.000	0.117	−0.000	0.117
Jan.-Oct 2018 × 1 S.D. Exposure	0.153	0.135	0.154	0.135	−0.155	0.117	−0.155	0.117
Trend			−0.000	0.011			−0.000	0.011
Municipality FEs	Y		Y		Y		Y	
Month FEs	Y		N		Y		N	
Dep. variable mean	0.009		0.009		0.009		0.009	
# Observations	10,424		10,424		10,424		10,424	

Notes: The table reports estimates from linear regressions predicting the share of residents within a municipality-month who request to switch GP, weighted by population. The sample includes all municipality-months between October 2016 and October 2018. The outcome variable is the share of residents who requested during that month to switch to a new GP. In Columns (1) and (2), exposure is defined as the share of GPs "Near Buffer," i.e. with 10–19 open slots as of September 2017. In Columns (3) and (4), exposure is defined as "Waitlist Change," the change in waitlist length from September to October 2017 interacted with an indicator for GPs being near the buffer. Both measures are in standard-deviation units. All coefficients and standard errors are scaled by 1,000 for readability.

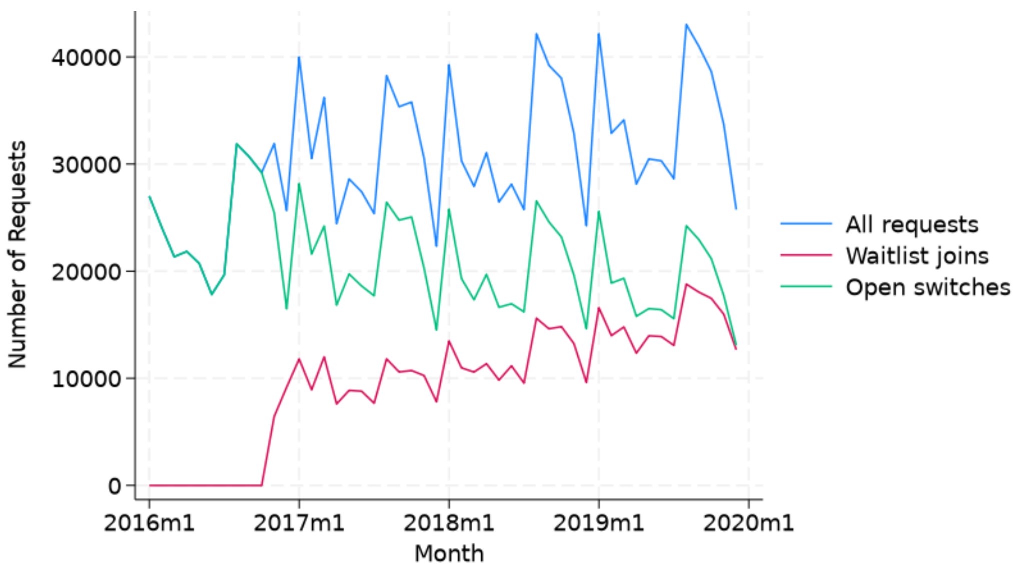
B.5.2 Responsiveness to Waitlist Growth Rates

The structural model also rules out the possibility that switch requests depend on the *relative* current and future returns to switching. If patients are forward-looking, they may adjust the timing of their switch requests to make them when waiting times are short. Since the waitlists were growing during our sample period—so waiting times are usually shorter today than in the future—our estimated attention probabilities may be biased for long-run stationary equilibria.

We test this assumption in two ways. First, we ask whether in aggregate, switch request rates are elevated in the months right after the introduction of waitlists. If patients strategically time their switch requests based on current and future waiting times, they would have an incentive to make a switch request as quickly as possible after the waitlists were introduced, when waiting times tended to be shortest as few patients had joined the newly opened waitlists. Second, to adjust for confounding time-varying factors, we test whether these rates were *relatively* more elevated in places where GP waitlists grew more quickly after the initial few waitlist months.[64] The analysis is done for all of Norway to maximize power and geographic variation.

Appendix Figure B.4 plots the raw number of switch requests in Norway each month from 2016 through 2019, also broken out by whether the request was a switch to an open GP or a waitlist join. A few things are apparent. First, there is some seasonality in the number of switch requests, which we abstract away from in the structural model but will need to account for here. Second, switch requests increased after the waitlists were introduced, and remained elevated thereafter. This is consistent with our model because the waitlists expanded patients' choice sets; GPs that would not have been available under the prior system because their panels were full could now be chosen. Third, there is a general positive trend in switch rates, including during 2017-2019. This trend combines increasing waitlist join rates and decreasing open switch rates.

Figure B.4. Number of GP switch requests over time



Notes: The figure shows the number of GP switch requests per month, overall and broken out by open switches and waitlist joins.

Appendix Figure B.5 presents plots of residualized switch requests each month during 2016–2019, adjusting for seasonal patterns, longer-term trends, and the fact that switch requests should increase due to the waitlists even without forward-looking behavior. Both panels include an indicator for post-November 2016 (i.e., the waitlists period) in the regression. Panel (a) controls for year and month-of-year fixed effects, whereas Panel (b) includes a linear time trend. In both plots, there is no jump in switch rates in the first few months of the

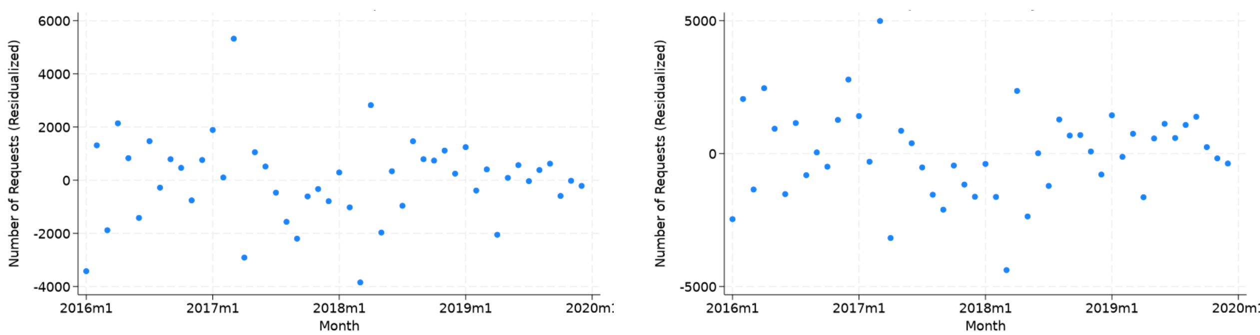
[64]We are grateful to an anonymous referee for this suggestion.

waitlists period. In fact, one would be hard-pressed to guess in which month the waitlists were introduced based on this plot. Although this is just a time series, it shows there was not a temporary surge in switch requests when the waitlists were introduced. This weighs against a model in which patients strategically time their switch requests. However, these patterns should be interpreted with caution, because other time-varying factors could have also influenced switch request rates. For example, patients may have only gradually become aware of the new waitlist system, and therefore many may not have taken advantage of the initially short waitlists, though they would have liked to.

Figure B.5. GP switch requests by month, residualized

(a) Year and Month-of-Year FEs

(b) Linear Time Trend



Notes: Switch request rates by month residualized on calendar time. Both panel include an indicator for post-November 2016, when the waitlists were introduced, and month-of-year fixed effects. Panel (a) controls for year fixed effects, while panel (b) controls for a linear time trend.

Next, we looked for evidence that switch requests surged more in places where the waitlists grew more quickly after they were introduced. A challenge here is that switch requests are mechanically related to waitlist growth. We therefore separate the first three months in the waitlist period (Nov 2016, Dec 2016, and Jan 2017) from later months, and measure the growth rate in each municipality's (or region's) waitlist growth rate beginning in February 2017. We conduct our analysis at the municipality level and cluster standard errors at the region level.

Appendix Table B.4 predicts the number of switch requests to GPs in a municipality each month. Each specification includes municipality fixed effects as well as month-of-year fixed effects and a linear time trend. Column (1) includes an indicator for the first three months (first quarter) after the waitlists were introduced. There is a statistically significant increase in switch requests, but, consistent with the earlier plots, it is estimated to be small in magnitude – 0.25 switch requests per 1,000 enrolled patients, compared to a mean rate of 5.37. Column (2) adds an interaction between the first-quarter indicator and the municipality's waitlist growth rate between February 2017 and December 2019. We estimate a small and statistically insignificant coefficient for this interaction, suggesting that first quarter switch requests are not differentially higher in municipalities where the waitlists grew more quickly afterwards. Column (3) considers an alternative proxy for waitlist growth, the share of a municipality's GPs that were near capacity in October 2016. Figure B.6 shows that this measure of capacity utilization strongly predicts waitlist growth. This interaction is also statistically insignificant and small in magnitude. Columns (4) and (5) repeat the specifications in the previous two columns, but allow for the effect of time to vary by region. This yields nearly identical results.

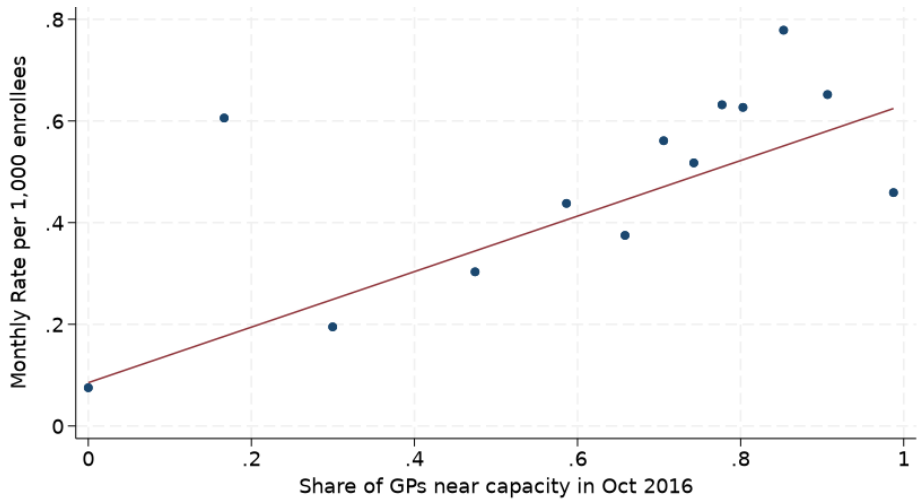
Taken together, the evidence in this section is consistent with our model's prediction that the number of switch requests does not depend on the current or future returns to switching.

Table B.4. Switch Requests After Waitlists Introduced

	Dependent Variable: Switches per 1,000 Enrollees				
	(1)	(2)	(3)	(4)	(5)
First quarter after waitlists introduced	0.246	0.249	0.116	0.240	0.0827
	(0.0817)	(0.129)	(0.329)	(0.130)	(0.331)
X growth rate after Feb 2017		-0.00844		0.0160	
		(0.306)		(0.307)	
X share GPs near cap in Oct 2016			0.245		0.307
			(0.572)		(0.575)
Observations	15170	15170	15170	15170	15170
Kommune and Month-of-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Linear Time Trend	Yes	Yes	Yes	Yes	Yes
Regional Trends	No	No	No	Yes	Yes

Notes: The table presents OLS regressions predicting the monthly rate of switch requests in a municipality. The unit of observation is a municipality month between November 2016 and December 2019. The 410 municipality's that existed throughout 2016 to 2019 are included in the sample. The growth rate after Feb 2017 is the monthly change in the number of patients on the waitlist for a GP in that municipality, per 1,000 enrollees in that municipality. The share of GPs near cap refers to the fraction of GPs in the municipality that had fewer than 20 available panel slots available in October 2016. The number of enrollees is also measured in October 2016. All specifications include fixed effects for each municipality and month-of-year, as well as a linear time trend. Columns (4) and (5) allow the fixed effects and time trend to be region-specific. Standard errors, clustered by region, are in parentheses.

Figure B.6. Subsequent Waitlist Growth by Share of GPs near Cap



Notes: The figure shows a binned scatter plot of the monthly growth rate in the number of waiting patients per 1,000 enrollees between February 2017 and December 2019, against the share of GPs near their panel cap in October 2016 (the month prior to waitlist introduction). Enrollment is measured as of October 2016. The unit of observation is a municipality. A GP is near their cap if they have fewer than 20 open slots on their panel.

Appendix C Estimation Details

This section provides details on the Gibbs' Sampler used to estimate the structural model parameters. We first describe the restrictions that the choice data place on the model primitives; then how we draw patient attention λ_{it} and flow payoffs v_{it}; and finally, how we update the discount rate ρ. The updating steps for (β, σ_ϵ) are standard. Unless otherwise specified, we condition on the month t in what follows and omit it from the notation.

Restrictions on Primitives. Consider a patient i with current GP j_0 who requests to switch in a given month. We learn two things from this patient. First, since they requested to switch GP, they were attentive: $\lambda_i = 1$. Second, their chosen GP j^* must have higher *expected net present value* than any other GP, given the patient's preferences and waiting time beliefs:

$$\underbrace{\mathbb{E}\left[e^{-\rho T_{ij^*}} \mid \mathbf{w}_i\right]}_{\omega_{j^*}(\rho)}(v_{ij^*} - v_{ij_0}) \geq \underbrace{\mathbb{E}\left[e^{-\rho T_{ij}} \mid \mathbf{w}_i\right]}_{\omega_j(\rho)}(v_{ij} - v_{ij_0}) \quad \forall j \in \mathcal{J}\,, \tag{9}$$

where we define $\omega_j(\rho)$ to be the expected discount factor to simplify notation. This set of inequalities implies a lower bound on the flow payoff v_{ij^*} from the chosen GP, and an upper bound on the flow payoff v_{ij} from each GP that was not chosen:

- **Upper Bound:** Rearranging Equation (9), for each $j \in \mathcal{J} \setminus j^*$,

$$v_{ij} \leq \underbrace{\frac{\omega_{j^*}(\rho)}{\omega_j(\rho)} v_{ij^*} - \frac{\omega_{j^*}(\rho) - \omega_j(\rho)}{\omega_j(\rho)} v_{ij_0}}_{ub_{ij}(v_{ij^*}, v_{ij_0}; \rho)},$$

where the notation $ub_{ij}(v_{ij^*}, v_{ij_0}; \rho)$ will be useful later.

- **Lower Bound** For chosen j^*,

$$v_{ij^*} \geq \underbrace{\max_{j \in \mathcal{J} \setminus j^*} \frac{\omega_j(\rho)}{\omega_{j^*}(\rho)} v_{ij} + \frac{\omega_{j^*}(\rho) - \omega_j(\rho)}{\omega_{j^*}(\rho)} v_{ij_0}}_{lb_{ij^*}(\mathbf{v}_i; \rho)}.$$

For patients who do not request to switch, it is not known whether they were paying attention. If the patient was not attentive ($\lambda_i = 0$), then not switching contains no information about their preferences. If the patient was attentive ($\lambda_i = 1$), then not switching implies that their current GP was preferable to all others: $v_{ij_0} \geq \max_{j \neq j_0} v_{ij}$.

Attention and Flow Payoffs. The Gibbs' sampler uses data augmentation to draw patients' attention and flow payoffs. The key challenge in drawing attention is calculating the likelihood that a non-switcher was attentive. By Bayes' Rule,

$$\begin{aligned} Pr(\lambda_i = 1 \mid \text{no switch}, X, j_0; \theta) &= \frac{Pr(\text{no switch} \mid \lambda_i = 1, X, j_0; \theta) Pr(\lambda_i = 1)}{Pr(\text{no switch} \mid X, j_0; \theta)} \\ &= \frac{Pr(\text{no switch} \mid \lambda_i = 1, X, j_0; \theta) Pr(\lambda_i = 1)}{Pr(\lambda_i = 0) + Pr(\text{no switch} \mid \lambda_i = 1, X, j_0; \theta) Pr(\lambda_i = 1)}, \end{aligned} \tag{10}$$

where $\theta = (\rho, \beta, \sigma_\epsilon(\cdot), p^\lambda(\cdot))$ collects the model parameters. By definition, $Pr(\lambda_i = 1) = p^\lambda(X_i)$ and $Pr(\lambda_i = 0) = 1 - p^\lambda(X_i)$, and we can express

$$\begin{aligned} Pr(\text{no switch} \mid \lambda_i = 1, X, j_0; \theta) &= Pr(v_{ij_0} \geq \max_{j \neq j_0} v_{ij} \mid X; \theta) \\ &= \int_{-\infty}^{\infty} \Phi \left[\Pi_{j \neq j_0} \left(\frac{v_{ij_0} - X_{ij}\beta}{\sigma_\epsilon} \right) \right] \phi \left(\frac{v_{ij_0} - X_{ij}\beta}{\sigma_\epsilon} \right) dv_{ij_0}, \end{aligned} \tag{11}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the standard normal CDF and PDF, respectively. For each inattentive patient, the estimator evaluates this integral using Gauss-Hermite quadrature. Each patient's λ_i is an iid Bernoulli draw with probability defined in Equation 11 for non-switchers. The attention probabilities $p^\lambda(x)$ are then drawn from their posterior Beta distributions, given attention draws across all periods t:

$$p^{\lambda}(x) \mid \{\lambda_{it}\} \sim \text{Beta}\left(\alpha + \sum_{(i,t):X_{it}=x} \lambda_{it}\ ,\ \varphi + \sum_{(i,t):X_{it}=x} (1-\lambda_{it})\right)\ , \tag{12}$$

where the prior is Beta(α, φ). In practice, we resample the individual attention draws and parameters every 10 iterations of the Gibbs' sampler.

Drawing patients' flow payoffs for each GP is more straightforward. For attentive patients, we redraw the flow payoff from each non-chosen GP from a truncated normal distribution with upper bound $ub_{ij}(v_{ij^*}, v_{ij_0}; \rho)$, and the flow payoff v_{ij^*} for the chosen GP from a truncated normal distribution with lower bound $lb_{ij^*}(\mathbf{v}_i; \rho)$.

Discount Rate. The discount rate is updated via a Metropolis-Hastings step which requires calculating the likelihood of the data given the other parameters. The algorithm begins with a previous draw ρ_{b-1} and a proposal distribution $F_\rho(. \mid \rho_{b-1}; \tau)$, which is truncated normal.[65] Each iteration, the following steps occur:

(i) Take a draw $\tilde{\rho}_b \sim F_\rho(. \mid \rho_{b-1})$

(ii) Calculate the ratio $r(\tilde{\rho}_b, \rho_{b-1}) = \frac{L(\tilde{\rho}_b)}{L(\rho_{b-1})}$, where $L(\rho)$ is the likelihood of the data given ρ and the other current draws of the model parameters.

(iii) Set $\rho_b = \left\{ \begin{array}{lll} \tilde{\rho}_b & w.p. & \min\{1, r(\tilde{\rho}_b, \rho_{b-1})\} \\ \rho_{b-1} & w.p. & 1 - \min\{1, r(\tilde{\rho}_b, \rho_{b-1})\} \end{array} \right\}$.

The main difficulty in implementing this is calculating the likelihood

$$L(\rho) = \Pi_{i:\lambda_i=1}\ Pr\left(j^* = \arg\max_{k \neq j_0}\ \mathbb{E}\left[e^{-\rho T_{ik}}\right](v_{ik} - v_{ij_0}) \mid \mathbf{X}_i, \mathbf{w}_i, \beta, \sigma_\epsilon, \rho\right) \tag{13}$$

for different values of the discount rate. This likelihood does not have a simple closed form, but it can be approximated with a high degree of accuracy using quadrature and exploiting the fact that the flow payoffs are conditionally independent given v_{ij_0}. Specifically, we can

[65] We experimented with an adaptive variance parameter τ, but settled on fixing $\tau = 0.001$.

rewrite each probability in equation 13 as

$$L_i(\rho) = \int_{-\infty}^{+\infty} \int_{v_{ij_0}}^{\infty} Pr(v_{ik} \leq ub_{ik}(v_{ij}, v_{ij_0}; \rho) \ \ \forall k \neq j, j_0 \ \mid \ \beta, \sigma_\epsilon) \ dF_{ij}(v_{ij}) \ dF_{ij_0}(v_{ij_0})$$

$$= \int_{-\infty}^{+\infty} \int_{v_{ij_0}}^{\infty} \Pi_{k \neq j^*, j_0} \Phi \left(\frac{ub_{ik}(v_{ij^*}, v_{ij_0}; \rho) - X_{ik}\beta}{\sigma_\epsilon} \right) \ dF_{ij}(v_{ij}) \ dF_{ij_0}(v_{ij_0}), \quad (14)$$

where the last equality exploits conditional independence of v_{ik} given the flow payoffs from the patient's current and chosen GPs. To reduce the dimensionality of the integral, we condition on the value of v_{ij_0} when evaluating it, and evaluate the inner integral using Gauss-Laguerre quadrature.

Appendix D Counterfactual Simulation Details

D.1 Simulated Dynamic Economy

As described in Section V.A, the simulated economy contains a finite set of patients I and GPs J. Patients have type $x \in \mathcal{X}$, where $\mathcal{X} = \{$demographic type, health type, location of residence$\}$. There are five possible demographic types: $\{$Perm. res. female ≤ 45, Perm. res. female > 45, Perm. res. male ≤ 45, Perm. res. male > 45, Temp. res.$\}$; two possible health types: $\{$chronic, non-chronic$\}$; and 45 possible locations of residence (municipalities in the Trondelag region). GPs are characterized by a demographic type (gender and over/under age 45), office coordinates, a GP fixed effect dictating their vertical quality, and a panel cap (an integer greater than zero).

GP characteristics are fixed over time. Patient characteristics evolve according to a stationary Markov process $M : \mathcal{X} \rightarrow \mathcal{X}$. Patients are assumed to indefinitely retain their gender (where the three possible genders are $\{$Perm. res. female, Perm. res. male, Temp. res.$\}$) and their health status, so aging is the only relevant demographic transition. The transition process can therefore be thought of as drawing two independent shocks: a moving shock and an aging shock, where the probability of both shocks depends on a patient's current demographic type and location of residence. We calibrate transition probabilities to match observed transitions in Trondelag over 2017–2019. We then adjust the transition process to make it stationary.[66]

[66]We begin with the probability mass distribution of patient types observed Trondelag in December 2019, where there are $|\mathcal{X}| = 5 \cdot 2 \cdot 45 = 450$ distinct patient types. We then calculate the (monthly) probability

Appendix Table D.1 reports details of the distribution of patient and GP types as well as the patient type transition process. Panel A describes patients. Seven percent of patients are temporary residents and 31 percent have a chronic health condition. Among permanent residents, half are female (with the remainder male), and 46 percent are age 16–45 (with the remainder 45 or older). Across all patients, the average probability of receiving an aging shock is 0.19 percent and the average probability of receiving a moving shock is 0.21 percent. The probability of aging is highest for temporary residents and lowest for females over age 45. The probability of moving is highest for female under age 45 and lowest for females over age 45. Conditional on moving, patients are more likely to move to a nearby location than a distant location.[67] Given the observed distribution of patient and GP locations, the average patient has 94 GPs within 15 minutes' driving time and 190 GPs within 60 minutes driving time, but there is substantial heterogeneity. In some rural areas, patients have only 5 GPs within 60 minutes, while in the central city of Trondheim, there are 189 GPs within only 15 minutes.

Panel B of Table D.1 describes the 425 GPs in the simulated economy. 48 percent of GPs are female, and 56 percent are under age 45. Among female GPs, however, 65 percent are under 45, reflecting growth of gender equity in this profession in recent decades. The average GP has a panel cap of 930 patients. The smallest GP has a cap of 80, while the largest has a cap of 1,695. For the purpose of our simulations, we do not use the panel caps observed in the data, but rather let them be determined based on the set of patients used in the simulation. For GPs that have waitlists as of December 2019, we set their panel cap equal to the observed number of patients that are enrolled with that GP and who reside in the Trondelag region. For GPs without waitlists, we set their panel cap equal to their observed enrollment among Trondelag patients times their observed ratio of enrollment to panel cap in the full data.

Finally, Panel C of Table D.1 describes the 45 possible locations where patients may live. The largest location is Trondheim municipality, with a population of 163,939 (nearly half the total population in the economy). The smallest location is Røyrvik municipality, with a

with which a "young" (≤ 45) patient becomes an "old" (> 45) patient based on our patient-month panel data from 2017–2019. In addition, we calculate the probability with which an "old" patient dies, which our transition model interprets as them transitioning to being a "young" patient again (implemented in the simulation via our "rebirth" procedure). We then also calculate the matrix of municipality-to-municipality monthly transition probabilities, separately for each of the five patient demographic types, again based on the 2017–2019 patient-month panel data. Combining these transition processes for demographics and location, we have a candidate transition matrix $\tilde{M}$, but it is not stationary. We solve for the *stationary* transition matrix M that minimizes the sum of squared differences between M and $\tilde{M}$.

[67] Appendix Table E.3 reports the corresponding rate of attention shocks received in the economy. For moves over 60 minutes, we determine the probability of an attention shock by extrapolating the relationship between attention rate and move distance derived from our demand estimates.

Table D.1. Fundamentals of Simulated Economy

		Percentile				
	Mean	Min	25th	50th	75th	Max
Panel A: Patients (N = 371,536)						
Demographic Type						
Temporary resident	0.07					
Perm. res. female, age 16–45	0.21					
Perm. res. female, age 45+	0.26					
Perm. res. male, age 16–45	0.22					
Perm. res. male, age 45+	0.25					
Other Characteristics						
Prob. aging shock	0.002	0.001	0.002	0.002	0.002	0.007
Prob. moving shock	0.002	<0.001	0.001	0.002	0.002	0.012
Num. GPs within 15 min.	95	0	15	34	189	189
Num. GPs within 60 min.	190	5	74	262	262	279
Travel time to avg. GP	94	47	47	60	118	528
Panel B: GPs (N = 425)						
Demographic Type						
Female, age≤45	0.31					
Female, age>45	0.17					
Male, age≤45	0.25					
Male, age>45	0.27					
Other Characteristics						
Panel cap	930	80	794	936	1,098	1,695
Travel time to avg. municipality	210	97	100	134	277	747
Travel time to closest municipality	6	3	5	6	7	23
Panel C: Locations (N = 45)						
Population	8,256	46	1,318	3,144	5,658	163,939
Pct. temp. res.	0.06	0.02	0.05	0.06	0.07	0.10
Pct. perm. res. female	0.47	0.41	0.46	0.46	0.47	0.54
Pct. perm. res. ≤45	0.37	0.21	0.33	0.36	0.40	0.52
Pct. perm. res. chronic	0.31	0.25	0.28	0.30	0.33	0.45

Notes: The table describes the fundamentals of our simulated economy. These fundamentals are calibrated to match the Trondelag region as of December 2019. The 45 patient locations correspond to municipalities in the Trondelag region. The reported probabilities of moving and aging shocks are monthly.

population of only 46. The average location has 6 percent temporary residents, 47 percent female permanent residents, and 37 percent young permanent residents. Urban locations have a higher fraction of temporary residents as well as young permanent residents.

D.2 Formal Description of Mechanisms

This section formally defines the allocation mechanisms studied in Section V. Let μ_0 be the allocation at the start of a given period. Each GP j has capacity N_j, a set of currently enrolled patients $\mu_0^{-1}(j)$, and a set of patients w_j on their waitlist. Each patient on a waitlist has waited a length of time t_i. Let $\succ_j$ denote GP j's (strict) preferences over patients, which encode the priority rules of the mechanism. Under a **first-come, first-served (FCFS)** priority rule, patients that have waited longer have higher priority: $\forall\, l, k \notin \mu_0^{-1}(j)$ and $\forall\, l, k \in \mu_0^{-1}(j)$, $l \succ_j k$ iff $t_l > t_k$. Under a priority rule that **respects endowments**, incumbent patients always have higher priority than non-incumbent patients at their current GP: $\forall\, l, k \in I$, $k \succ_j l$ if $k \in \mu_0^{-1}(j)$ and $l \notin \mu_0^{-1}(j)$. All priority rules we consider will respect endowments.

An attentive patient's GP request takes the form of a rank-order preference list (ROL) R_i. In our primary mechanisms of interest, patients may join at most one waitlist, so R_i has length at most two. The requested (waitlist) GP is ranked first, and the current GP is ranked second. A *matching algorithm* ϕ maps a set of patient-reported ROLs $\mathbf{R}$, GP priorities $\succ$, and panel caps $\mathbf{N}$ into an allocation μ. An *allocation mechanism* is defined as the triplet $[\mathbf{N}, \succ, \phi]$, in combination with rules regarding the maximum length of patients' ROLs and how often the matching algorithm is run. Our counterfactuals change both the priority rule $\succ$ and the matching algorithm ϕ applied each period. Patients' reported ROLs respond endogenously to these changes.

Waitlists. This is the allocation mechanism currently used in Norway. Priorities are FCFS. The matching algorithm used is the following:

> *Step 1:* For each GP j, let $O_j = N_j - |\mu_0^{-1}(j)|$ be the number of open slots on j's panel. Assign the O_j highest-priority patients on j's waitlist to j's panel; remove each of these patients from their current GP's panel and from the waitlist.
>
> *Step k:* Repeat Step 1 for the panels and waitlists resulting from Step $k-1$.

The algorithm terminates if no patients are reassigned in a step. When this occurs, there are no patients waiting for GPs with open slots.

Waitlists with Top Trading Cycles (TTC). Priorities are still FCFS. The matching algorithm first runs the Waitlists algorithm, and then runs the TTC algorithm. The TTC algorithm works as follows. Each GP begins with pseudo-capacity $\tilde{N}_j$ equal to their number of open panel slots plus the number of their *current* patients who are currently in the mechanism hoping to switch away to another GP.

> *Step 1:* Each patient "points to" their preferred GP according to R_i, and each GP points to their preferred patient according to $\succ_j$. There is at least one cycle, i.e., an ordered list $\{i_1, j_1, i_2, j_2, ..., i_k, j_k\}$ where i_1 points to j_1, j_1, points to i_2, ... , and j_k points to i_1. Further, each patient and GP can be part of at most one cycle. Each patient in a cycle is assigned to the GP they point to and is removed from their current GP's panel and the algorithm. Each GP in a cycle has their pseudo-capacity reduced by one, and is removed from the algorithm when their pseudo-capacity falls to zero.
>
> *Step k:* Repeat Step 1 with the *remaining* patients and updated GP pseudo-capacities from the end of Step $k-1$.

The algorithm terminates when no patients remain in the algorithm.

Waitlists with Top Trading Cycles and Priority (TTCP). This mechanism is identical to TTC, but modifies priorities so that patients with undersubscribed GPs are prioritized above patients with oversubscribed GPs (while still respecting endowments). Formally, $\forall\, l, k \notin \mu_0^{-1}(j), l \succ_j k$ if l currently has an undersubscribed GP and k currently has an oversubscribed GP. Among patients with the same over/undersubscribed status, there is FCFS priority. A GP is undersubscribed if it has at least one open slot on its panel. We classify a patient's current GP as over/undersubscribed at the moment they enter the mechanism.[68]

Waitlists with Deferred Acceptance (DA). Again, let $\tilde{N}_j$ denote GP j's pseudo-capacity. The DA algorithm proceeds as follows:

> *Step 1:* Each patient "proposes to" their preferred GP according to R_i. Each GP j provisionally accepts proposals from its $\tilde{N}_j$ most preferred patients according to $\succ_j$ and rejects any remaining proposals.
>
> *Step k:* Any patients who were rejected in the previous round propose to their most-preferred GP (according to R_i) who has not yet rejected them. Each GP j provisionally accepts proposals from its $\tilde{N}_j$ most preferred patients according to $\succ_j$ and rejects any remaining proposals.

The algorithm terminates when no patient is rejected.

[68] While it is possible that a GP could transition to/from being over/undersubscribed while their patient remains in the mechanism, these transitions rarely occurred in our simulations, and modeling patient beliefs about this possibility would be highly complex.

D.3 Algorithm to Compute Equilibrium

For each mechanism, we compute a stationary equilibrium in which patients' beliefs about waiting time are consistent the waiting times implied by patients' optimal decisions. We initialize the economy to have no patients standing on waitlists (i.e., no patients in the reassignment mechanism). We draw a sequence of patient types (demographic type, location of residence, identity of mother if reborn) and patient attention shocks for 600 periods (months) for all 371,536 patients in our simulation. These draws are held fixed across counterfactual mechanisms.

We search for a fixed point between patients' belief parameters $\mathbf{b} = (\eta, \kappa, \kappa_{OS}, \chi_0, \chi_1)$ and their sample analogs within the simulated economy. The algorithm works as follows. Iteration q begins with a vector of belief parameters $\mathbf{b}^q$. The following steps then occur:

(1) The simulation is run for 600 periods. Each period, the steps described in Section V.A occur, with all demographic transitions and attention shocks predetermined. Each period, attentive patients sequentially enter the reassignment mechanism and consider all GPs in the economy, observing their current waitlist lengths. Patients form beliefs about waiting time according to $\mathbf{b}^q$ and then decide which GP to choose. If they choose their current GP or an open GP, they are reassigned immediately and exit the mechanism. If they choose a GP with a waitlist, they wait there until they are successfully reassigned by the mechanism, they get another attention shock, or they die.

(2) The simulation provides data on the distribution of realized waiting times given the optimal actions implied by beliefs $\mathbf{b}^q$. We use this information to construct the sample analog of belief parameters, relying only on the last 100 periods of the simulation to allow the economy to converge to a stationary distribution.

(3) Beliefs are then updated as a convex combination of the initial and implied values: $\mathbf{b}^{q+1} = \lambda^q \mathbf{b}^q + (1 - \lambda^q)\mathbf{b}'$. The factor λ^q determines how quickly beliefs are updated.

Implied beliefs. Table D.2 reports equilibrium belief parameters. Under the status quo Waitlists mechanism, 0.53 percent of panel slots become vacant each month. Under the mechanisms with TTC, such natural vacancies arise less frequently (0.12 percent per panel slot per month), since many incumbent patients who switch away no longer leave a vacant seat in their wake.

Table D.2. Equilibrium Beliefs

	Waitlists	TTC	TTCP	DA
Panel vacancy rate (η)	0.0053	0.0012	0.0008	0.0053
Waitlist departure rate (κ)	0.0075	0.0415	0.0076	0.0075
Waitlist departure rate, curr. GP oversub. (κ_{OS})	–	–	0.0338	–
Cycle participation rate, curr. GP oversub. (χ_1)		-4.9981	-4.9156	
Cycle participation rate, curr. GP oversub. (χ_2)		-0.6331	-0.5987	

Notes: The table reports equilibrium belief parameters. In the TTCP mechanism, patients beliefs about the waitlist departure rate among waiters in front of them are κ for patients with an undersubscribed GP and $\kappa + \kappa_{OS}$ for patients with an oversubscribed GP. In all other mechanisms, patients have common beliefs about κ. For patients with an oversubscribed GP, the perceived probability of participating in a cycle in waitlist position s for a desired GP with panel cap N is given by $\chi = \exp(\chi_0 + \chi_1 \log(s/N))$.

While TTC lowers the panel vacancy rate, it raises the waitlist departure rate, as waiters may now participate in a cycle. Under TTC, all patients perceive this rate as 4.15 percent per waiter ahead of them per month. Under TTCP, patients with undersubscribed GPs perceive it to be only 0.76 percent, since the only patients ahead of them on waitlists are other patients with undersubscribed GPs, who have no chance of participating in a cycle. Under TTC, a patient with an oversubscribed GP who is 10th on the waitlist for a GP with a panel cap of 1,000 believes they will participate in a cycle with probability $0.125 = \exp(-4.9981 - 0.6331 \times \log(10/1000))$ each month. As their position-relative-to-panel-cap increases, this probability declines log-linearly. The same patient expects to participate in a cycle with monthly probability 0.029 in waitlist position 100.

Appendix Figure E.7 provides a depiction of how our beliefs model translated into expected waiting times and discount factors across mechanisms. A patient with an oversubscribed GP believes that the expected waiting time (in months) for a GP with panel size 1,000 with a waitlist length of 100 would be 18.0 under Waitlists, 16.8 under TTC, 18.0 under TTCP, and 18.0 under DA. For a patient with an undersubscribed GP, the corresponding beliefs would be 18.0 under Waitlists, 37.5 under TTC, 97.6 under TTCP, and 18.0 under DA. Note, however, that because patients with an undersubscribed GP get waitlist priority under TTCP, they would rarely find themselves so far back on a waitlist. Appendix Figure E.8 reports the density of observed chosen waitlist lengths across mechanisms. Under TTCP, almost all the mass of chosen waitlist length is below a waitlist rank of 25. Finally, Appendix Figure E.9 provides a depiction of how our beliefs model interacts with panel capacities. Under the TTC mechanism, a patient with an oversubscribed GP believes that for a GP with a waitlist length of 50, expected waiting time would be 14 months if the GP had panel capacity of 750 and 11 months if capacity were 1,250. The average panel cap among GPs in our simulation is 930 (c.f. Appendix Table D.1).

D.4 Perceived Perpetuity Equivalent Welfare Measure

The second welfare measure described in Section V.A, which we refer to in the main text as the "perceived perpetuity equivalent," is based on an attentive patient's perceived net present value (henceforth, NPV) of making their optimal GP choice. This is a natural way to compare the attractiveness of switching opportunities across mechanisms because it accounts for both the waitlist lengths patients face when making a GP choice as well as patients' beliefs about how fast waitlists will move. Such a comparison is complicated, however, by the fact that at any given moment of attention, patients' *current GP* may change across mechanisms due to prior attention shocks. Since their degree of (dis)satisfaction with their current GP will directly affect how long they are willing to wait for a new GP, a patient's current GP is an important factor driving optimal choices. It is therefore useful to decompose NPV differences across mechanisms into (i) the component driven by a change in a patient's current GP, and (ii) the component driven by everything else, namely waitlist lengths, beliefs about how waitlist lengths map to wait times, and patients' optimal GP choice *conditional on current GP*.

To perform this decomposition, we introduce some additional notation. Throughout, we consider a single attentive patient in a single period whose preferences are held fixed across mechanisms. Express an attentive patient's perceived NPV from choosing GP j while currently enrolled with GP j_0 as

$$NPV(j; j_0, \mathbf{w}, \xi) = \rho\, \mathbb{E}\left[\int_{\tau=0}^{T_j} e^{-\rho\tau} v_{j_0} d\tau + \int_{\tau=T_j}^{\infty} e^{-\rho\tau} v_{ij} d\tau \mid \mathbf{w}, \xi\right], \tag{15}$$

where $\mathbf{w}$ represents the vector of waitlist lengths for all GPs, ξ represents the patient's beliefs about the mapping from waitlist lengths to waiting time T_j, and v_j represents the per-period flow utility the patient derives from GP j. Multiplying by ρ scales the measure so it can be interpreted as a flow payoff received in perpetuity. The NPV derived from a patient's optimal choice of GP is then given by

$$NPV^*(j_0, \mathbf{w}, \xi) = \max_{j \in \mathcal{J}} \quad NPV(j; j_0, \mathbf{w}, \xi), \tag{16}$$

where $\mathcal{J}$ is the set of all GPs. We can now isolate differences in the value derived from optimal behavior under each mechanism by differences in the arguments of NPV^*. As in

Section V.C, we use the current Waitlists mechanism as our reference point. To that end, define the difference in the value derived from optimal behavior under focal mechanism M relative to the reference mechanism W as

$$\Delta NPV^{*,M-W} = NPV^{*}(j_0^M, \mathbf{w}^M, \xi^M) - NPV^{*}(j_0^W, \mathbf{w}^W, \xi^W), \tag{17}$$

where j_0^M is the patient's current GP under mechanism M, $\mathbf{w}^M$ is the vector of prevailing waitlist lengths under mechanism M, and ξ^M is the patient's beliefs about the mapping from waitlist lengths to wait times under mechanism M. Given this notation, we can then express the decomposition as

$$\begin{aligned} \Delta NPV^{*,M-W} = \quad & NPV^{*}(j_0^M, \mathbf{w}^M, \xi^M) - NPV^{*}(j_0^W, \mathbf{w}^M, \xi^M) & \text{(i)} \\ & + NPV^{*}(j_0^W, \mathbf{w}^M, \xi^M) - NPV^{*}(j_0^W, \mathbf{w}^W, \xi^W), & \text{(ii)} \end{aligned}$$

where as noted above, line (i) is the component of the welfare difference driven by a change in a patient's current GP, and line (ii) is the component driven by everything else, namely waitlist lengths, beliefs about how waitlist lengths map to wait times, and patients' optimal GP choice *conditional on current GP*. Because it more accurately represents the contemporaneous difference in the value of switching opportunities under different mechanisms, component (ii) is what is reported in the main text (Tables 5 and 6).

D.5 Benchmark Simulations

No Waitlists. An attentive patient may switch to any GP panel with open slots, but may not request a GP whose panel is currently full. The GP choice action is therefore a discrete choice among all open GPs plus the patient's current GP. We run the simulation under the assumption that patients are not strategic about when to request a GP (attention shocks are exogenously timed). Because there are no waitlists, no patients remain in the mechanism between periods.

Truthful TTC. This mechanism is identical to TTC with the exception that patients can submit ROLs of arbitrary length. We run the simulation under the assumption that when they arrive to the mechanism, attentive patients truthfully report their full ordinal preferences over GPs, truncated at their current GP.

Greedy First-Best. Reassignments maximize the sum of flow payoffs enjoyed by patients in a given month. Patients participate in the mechanism in the month when they become attentive. The set of GP slots includes all slots currently occupied by participating patients, as well as any open slots. Assignments are calculated using the Hungarian Algorithm.

No Caps. Each GP's panel cap is set to infinity, so an attentive patient may switch to any GP immediately. This simulation provides an upper bound on the welfare that can be achieved by any mechanism with capacity constraints.

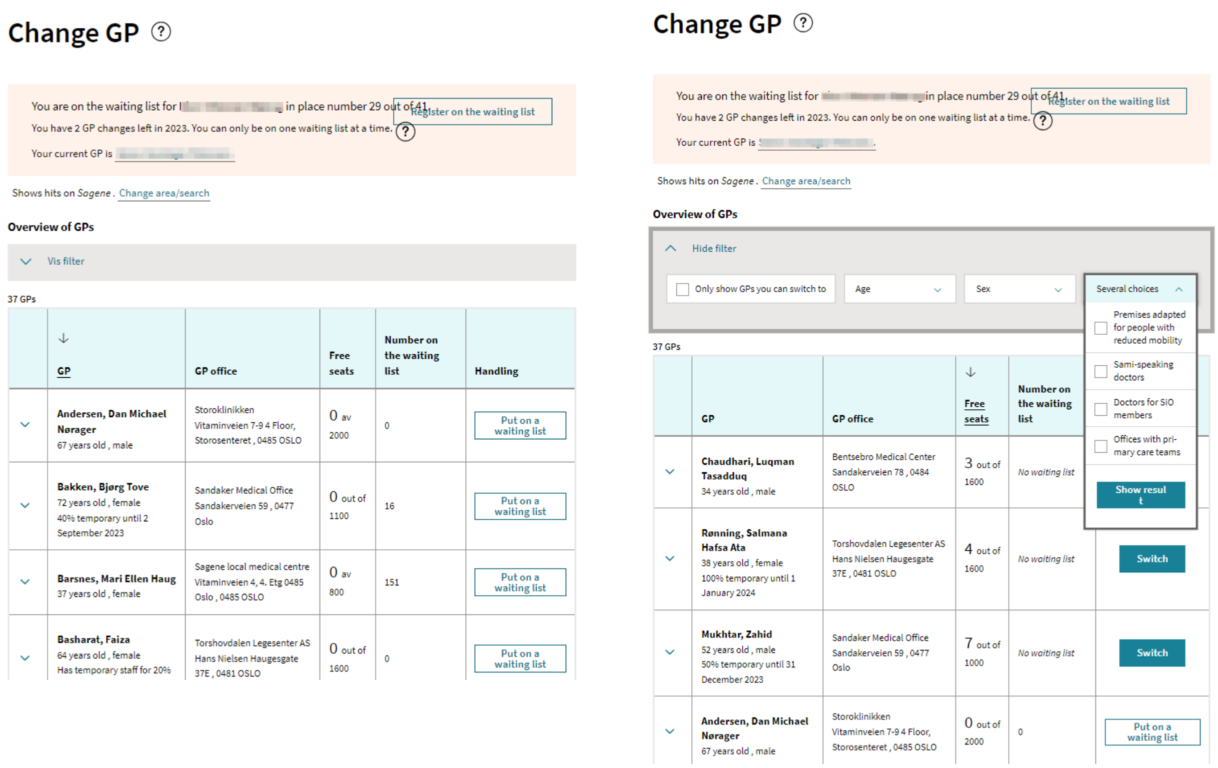
Table D.3. Results from Benchmark Simulations

	Waitlists	TTC	No Waitlists	Truthful TTC	Greedy	No Caps
GP waitlists						
Pct. of population on a waitlist	10.98	10.38	–	0.04	–	–
Pct. of GPs with a waitlist	81.3	78.9	–	32.4	–	–
Mean E(waittime) | curr. GP undersub.	18.8	23.9			–	–
| curr. GP oversub.	18.8	12.0			–	–
Attentive patient choices						
Mean E(waittime) at chosen GP	18.8	15.6				–
Pct. waitlist joins	83.1	83.9	–	93.3	–	–
| curr. GP undersub.	79.1	75.3	–	93.4	–	–
| curr. GP oversub.	83.8	85.5	–	93.3	–	–
True pref. rank of chosen GP	1.80	1.63	3.00	2.73	1.68	1.00
| curr. GP undersub.	1.9	2.2	2.9	2.8	1.7	1.0
| curr. GP oversub.	1.8	1.5	3.1	2.7	1.7	1.0
Realized assignments						
Travel time to current GP, mean (med.)	14.8 (6.4)	14.6 (6.4)	14.4 (6.5)	14.3 (6.4)	13.3 (6.4)	16.8 (6.4)
Pct. with same gender GP | young female	58.1	58.9	57.4	57.4	63.9	70.2
| young male	60.0	59.2	58.7	59.1	54.5	46.8
Welfare						
Flow payoff from current GP, mean (med.)	–†	0.72 (0.42)	0.10 (-1.43)	1.30 (-0.61)	4.80 (2.12)	6.89 (4.13)

Notes: The table reports results under the benchmark simulations described in Section V.D. Results for Waitlists and TTC are reproduced from Table 5 for comparison. Statistics are generated in months 492–551 of the simulation, out of 600 total months. They are first computed within month and then averaged across simulation months. E(waittime) is the expected waiting time implied by patients' equilibrium beliefs and current waitlist lengths. True pref. rank of requested GP is the rank of a patient's requested GP in their true flow payoff ordering. This table is referenced in Section V.D. †By normalization.

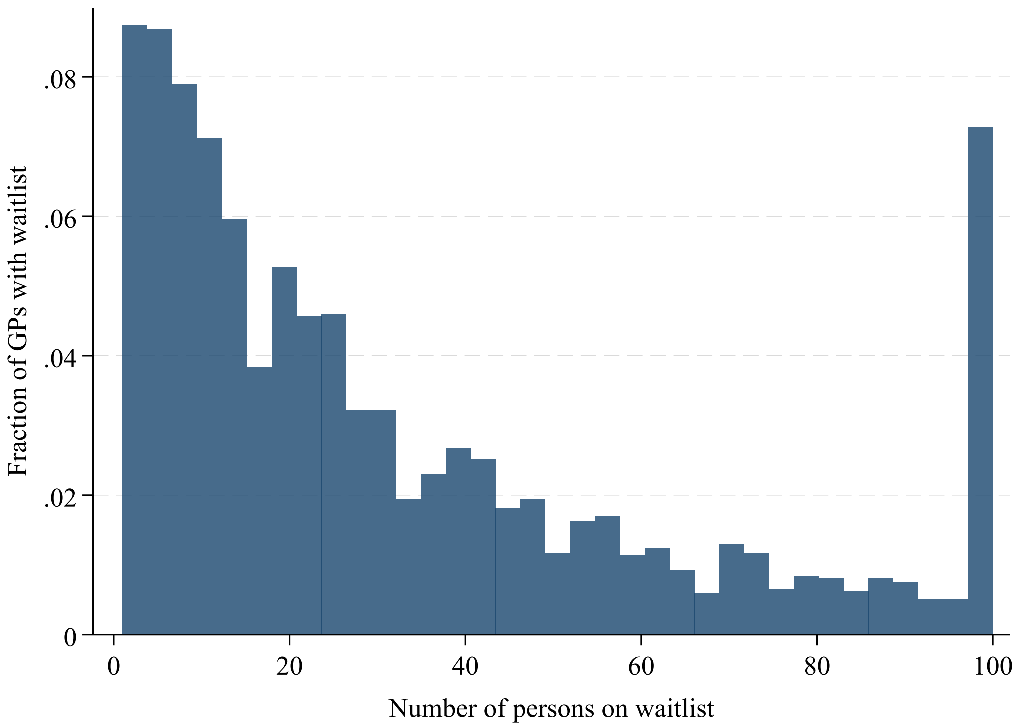
Figure E.1. HelseNorge Screenshots

(a) Sorted Alphabetically (b) Sorted by Free Seats



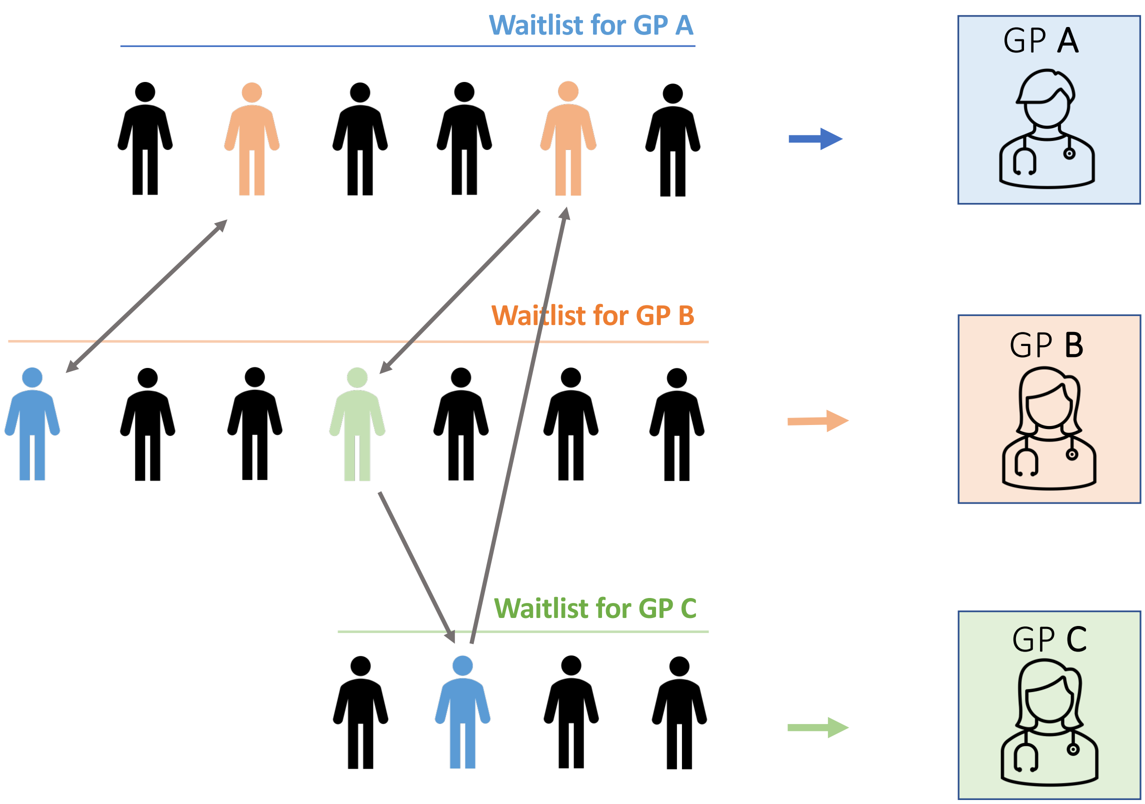
Notes: The figure shows two screenshots from the "Change GP" (*bytte fastlege*) tool on Norway's centralized online health platform, HelseNorge. The page shows the list of 37 GPs located in the Sagene neighborhood of Oslo. Panel (a) sorts this list alphabetically by GP last name (the default), and panel (b) sorts the list by the number of free slots available on each GP's panel. Only three GPs have available slots on their panel. (Webpage translated from Norwegian to English using Google Chrome, which slightly affects the rendering of graphics relative to the original.) Accessed August 18, 2023; *available at* https://tjenester.helsenorge.no/bytte-fastlege?fylke=03&kommuner=0301&bydeler=030103. This figure is referenced at footnote 16.

Figure E.2. Distribution of Waitlist lengths in December 2019



Notes: The figure shows the distribution of waiting list lengths in Norway in December 2019, among GPs that had a waiting list. Waitlist length is top-coded at 100 for readability. There were 3,695 unique GPs with waiting lists (out of 5,010 total GPs). This figure is referenced in footnote 26.

Figure E.3. Example of Top Trading Cycles



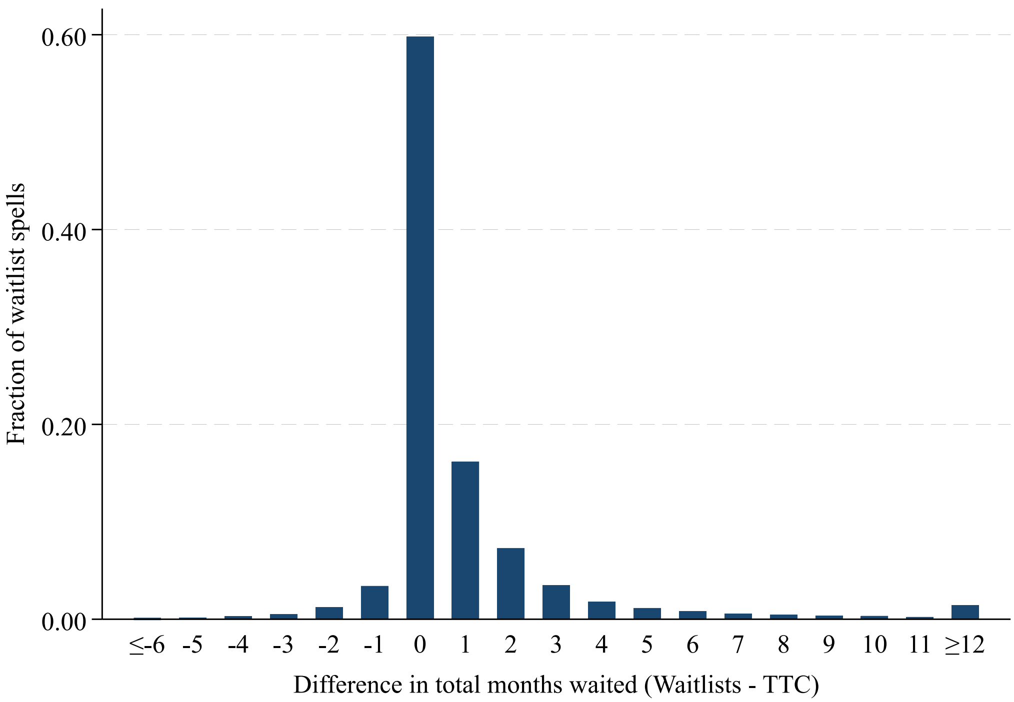
Notes: The figure shows an example of unrealized gains from trade on hypothetical GP waitlists. Each GP's waitlist is to their left, with the right-most patient at the front. Waiting patients are the same color as their currently assigned GP, or black if they are assigned to a GP other than A, B, or C. The arrows illustrate trades executed by a run of the TTC algorithm. There is a bilateral trade between the 5th patient on A's waitlist and the 7th patient on B's waitlist. In addition, there is a trilateral trade in which the 4th patient on B's waitlist takes the slot of the 2nd patient waiting for A; their slot (on GP C's panel) goes to the 3rd patient on C's waitlist, who is currently assigned to A; and their slot (on GP A's panel) goes to the 2nd patient on A's waitlist, who is currently assigned to B. This figure is referenced in Section II.C.

Table E.1. Patient Demographics by Outcome of Running TTC in December 2019

Sample demographic	Full Sample	Not on a waitlist	On a waitlist: Reassigned	On a waitlist: Not reassigned
Number of individuals	4,573,170	4,463,532	18,667	90,971
Pct. of individuals		0.98	0.00	0.02
Demographics				
Pct. female	0.50	0.49	0.66	0.66
Age	47	47	42	41
Years of education	13.1	13.1	13.4	13.3
Annual income (000 NOK)	413	414	397	373
Pct. temporary resident	0.07	0.07	0.04	0.12
Pct. ever moved	0.11	0.10	0.24	0.24
Choice of GP				
Pct. ever switched to open GP	0.13	0.13	0.17	0.30
Travel time to current GP (min.)	10.8	10.7	15.8	14.5
Pct. with GP of same gender	0.58	0.58	0.55	0.53
Use of waitlists				
Pct. ever on a waitlist	0.09	0.07	1.00	1.00
Number of months on a waitlist $\mid > 0$	6.4	4.9	7.9	10.7
Pct. waiting for GP of same gender	0.64	0.64	0.65	0.65
Travel time to wl. GP – curr. GP (min.)	-6.8	-7.2	-8.4	-5.6

Notes: The table provides summary statistics on adult patients based on the outcome of running TTC on waitlists as of December 2019. Summary statistics for each individual are calculated based on the time period 2017–2019, not just as they were observed in December 2019. "Ever" means at any point during 2017–2019. The first column reports means among all adult patients in the population. The remaining three columns are a partition of patients based on whether they were not standing on a waitlist in December 2019 and thus did not participate in TTC ("Not on a waitlist"), whether they were on a waitlist and were successfully reassigned by the TTC algorithm ("On a waitlist/Reassigned"), and finally those patients who were on a waitlist but were not successfully reassigned via TTC. This table is referenced in Section II.C.

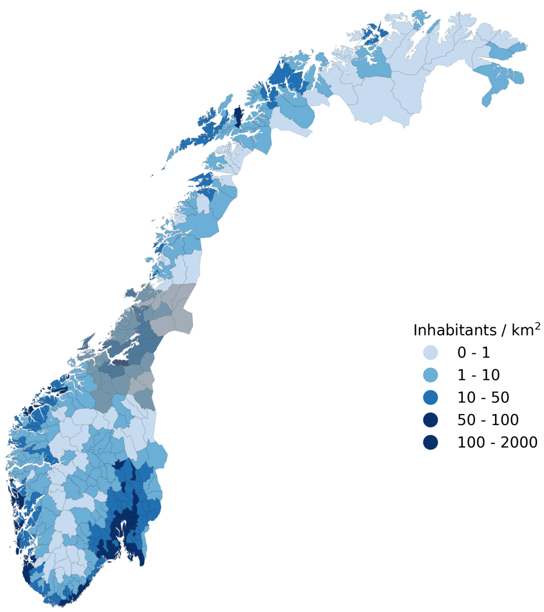
Figure E.4. Distribution of Waittime Differences Under Mechanical Simulation



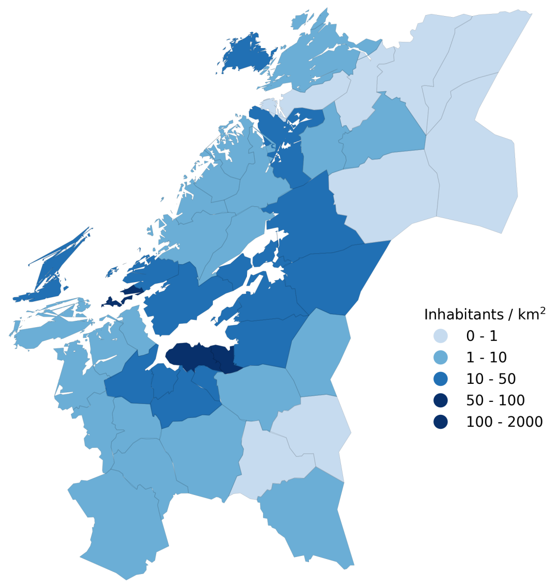
Notes: The figure shows the distribution of waiting time differences that result from the simple mechanical simulation of TTC using the historical waitlist data. An observation is a waitlist spell. The figure reports the difference between the number of months the patient waited under the status quo mechanism (Waitlists) and that under the TTC mechanism (TTC). While most patients wait for less time under TTC, 4.5 percent of patients wait for longer. This figure is referenced in Section II.C.

Figure E.5. Population Density Map

(a) All of Norway



(b) Trondelag Region



Notes: The figure shows a population density map of all of Norway as well as just the Trondelag region. In panel (a), the Trondelag region is shaded in gray. The outlined shapes within the map are municipalities. There are 421 municipalities in Norway and 58 in Trondelag (using 2019 region boundaries). The population center of Trondelag (the city Trondheim) lies at the center of the region in the darkest shaded municipality. This figure is referenced in footnote 35.

Table E.2. Comparison of Norway and Trondelag Region, 2019

Sample demographic	All of Norway	Trondelag
Panel A. Patient characteristics		
Number of individuals	4,633,395	412,774
Demographics		
Pct. female	0.50	0.49
Age	48	47
Pct. with post-secondary education	0.32	0.31
Annual income (000 NOK)	429	403
Pct. with chronic condition	0.31	0.30
Pct. temporary resident	0.07	0.07
Pct. ever moved	0.04	0.06
Choice of GP		
Pct. ever switched to open GP	0.05	0.05
Travel time to current GP (min.)	10.7	10.8
Pct. with GP of same gender	0.58	0.57
Use of waitlists		
Pct. ever on a waitlist	0.05	0.05
Number of months on a waitlist $\mid > 0$	5.0	4.9
Pct. waiting for GP of same gender	0.64	0.65
Travel time to wl. GP – curr. GP (min.)	-6.7	-7.9
Panel B. GP characteristics		
Number of GP panels	5,549	474
Panel characteristics		
Enrollment cap	1,120	1,078
Pct. months with available slots	0.32	0.27
Pct. months with temporary GP	0.12	0.12
GP demographics		
Pct. months with female GP	0.43	0.46
Pct. rural	0.37	0.52
Age	47	45
Panel enrollment stats.		
Num. waiting on waitlist	24	22
Num. enrollees / cap	0.94	0.97

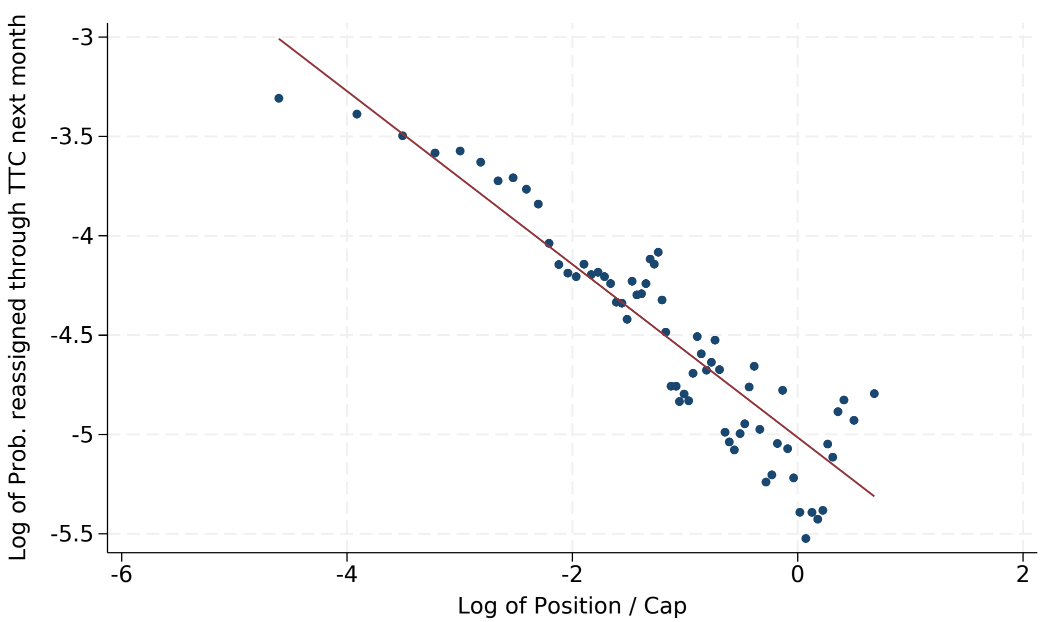
Notes: The table compares descriptive statistics between all of Norway and the Trondelag region in 2019. Panel A reports statistics on (adult) patients, and all values represent means over patient-months. "Ever" means at any point during 2019. Moves are counted only if they are across municipalities. Age, gender, education, and income data are not available for temporary residents, so those means are only among permanent residents. Panel B reports statistics on GPs, and all values in the table represent means over GP panel-months. GP enrollment and waitlist use statistics reflect the full population (including children under 16). This table is referenced in Section IV.A.

Table E.3. Summary Statistics on Realizations from Exogenous Processes

	Mean	SD
Number of patients	365,775	–
Number of attention shocks received	2,368	48
Number of moves	748	28
Number of deaths	342	19
Number of aging shocks	645	25
Number of moves in past 6 months	5,894	295
Number of attn. shocks in past 12 months	29,277	2,342

Notes: The table describes the realizations of exogenous processes in our simulated economy. These include the demographic transition processes (patients aging, dying, and moving) and the attention process (patients receiving attention shocks). Patients that die are immediately reborn, so there are a fixed number of patients in all periods. Statistics in the table are calculated across the 600 periods in the simulation. This table is referenced in Section V.A.

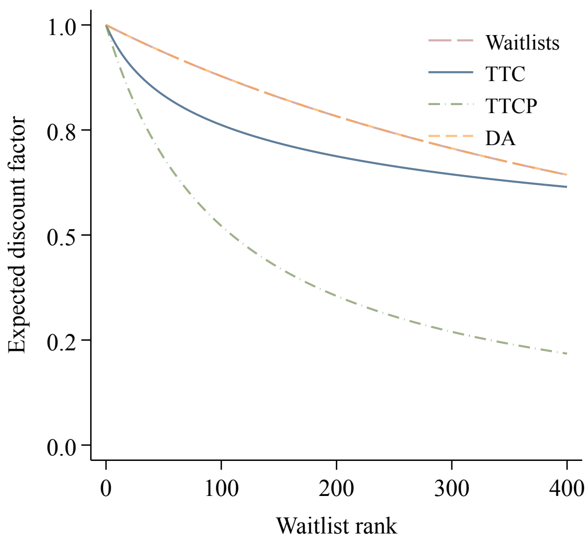
Figure E.6. Relationship Between χ and Waitlist Position Under TTC



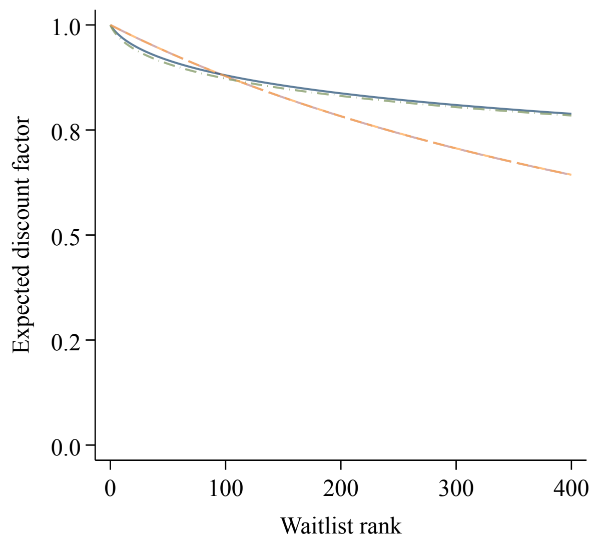
Notes: The figure shows a binned scatter plot of the log of the probability of being reassigned through TTC next month ($\log(\chi)$) versus the log of the ratio between the patient's current waitlist position and the panel size of the requested GP ($\log(s/N)$). An observation is a patient-month in the equilibrium simulation of the TTC mechanism. The sample is the set of observations in which the patient stands on a waitlist and is currently assigned to an oversubscribed GP. Reassignment probabilities are first calculated within 0.01-width bins of $\log(s/N)$ and then collapsed for the binscatter. This figure is referenced in Section V.B.

Figure E.7. Relationship Between Beliefs and Waitlist Rank by Mechanism

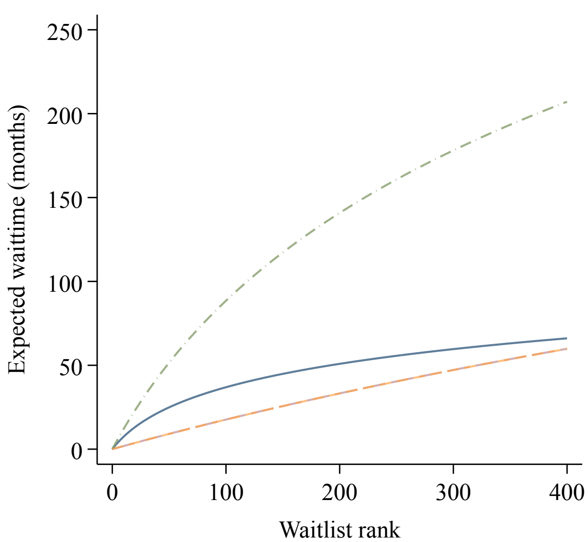
(a) EDF | curr. GP undersubscr.



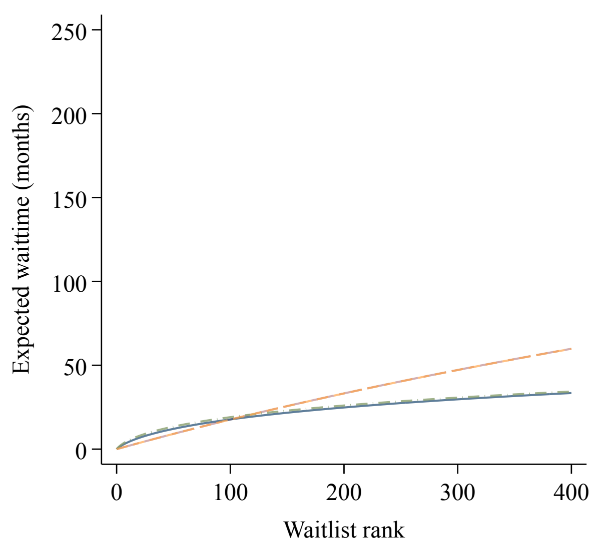
(b) EDF | curr. GP oversubscr.



(c) E(waittime) | curr. GP undersubscr.

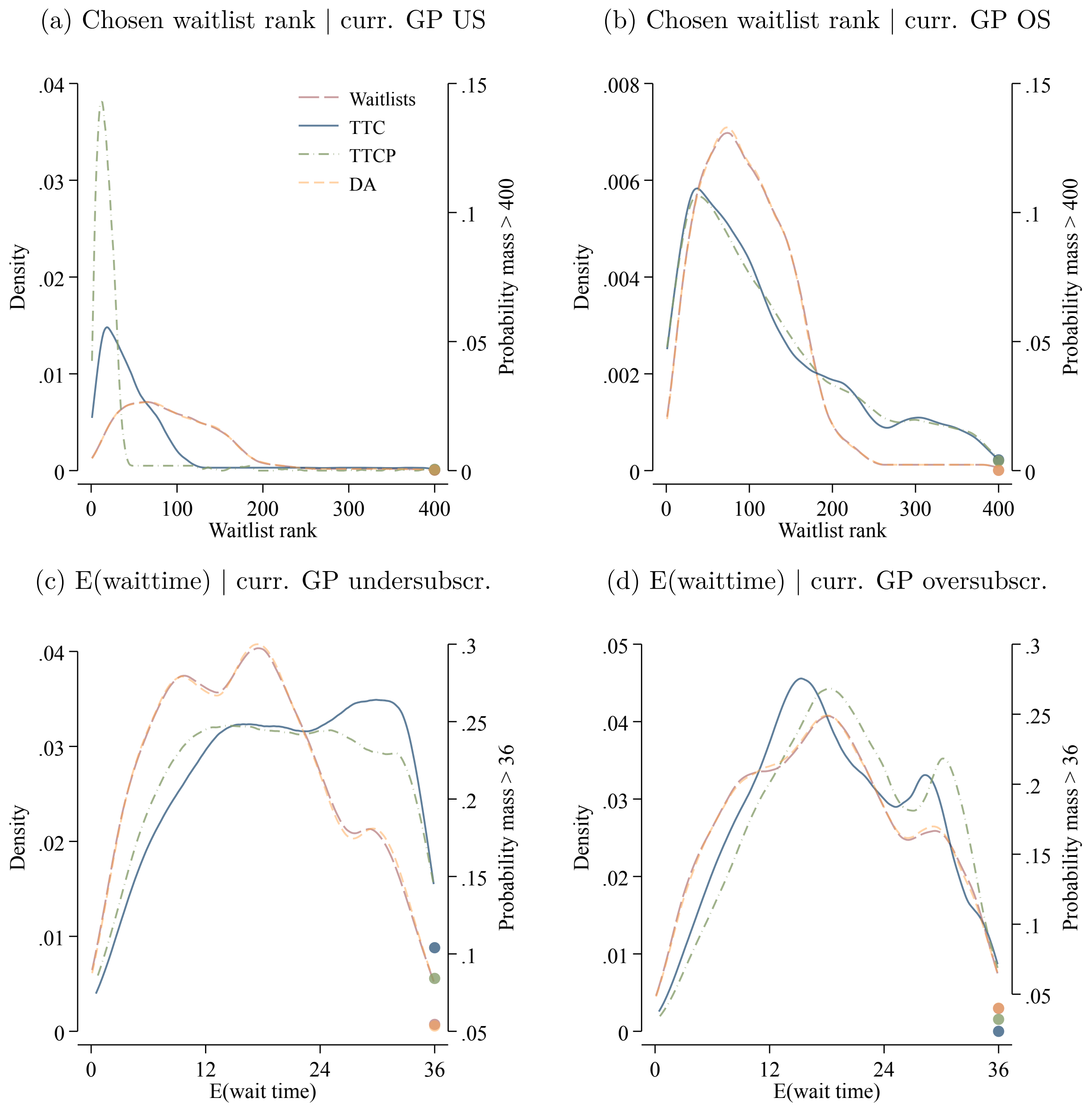


(d) E(waittime) | curr. GP oversubscr.



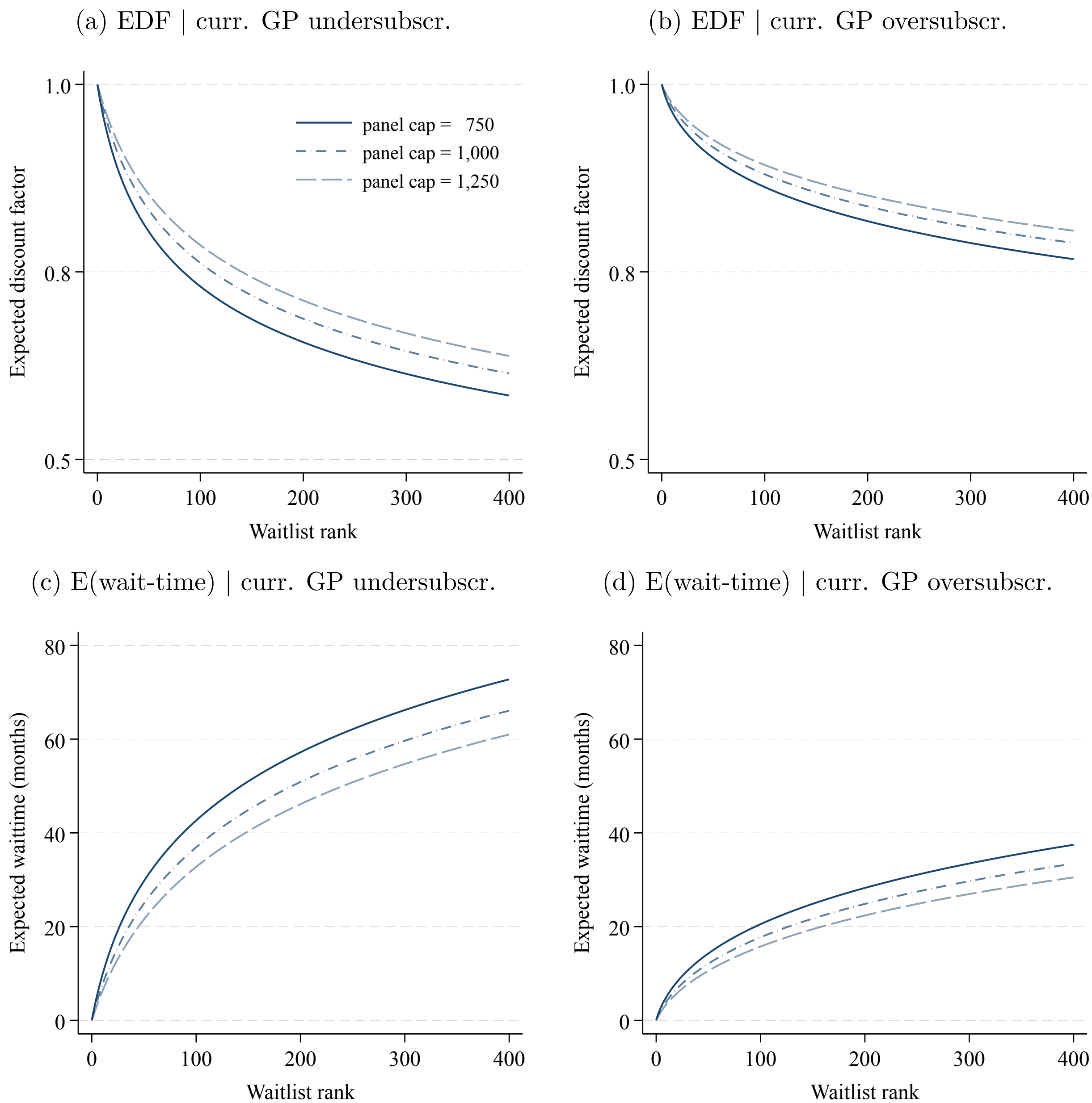
Notes: The figure shows the relationship between beliefs and waitlist rank across our four focal mechanisms, supposing all waitlists were for a GP with a panel cap of 1,000. Panels (a) and (b) report a patient's expected discount factor (EDF) as a function of waitlist rank. Panel (a) shows the EDF for a patient whose current GP is undersubscribed, while panel (b) shows the EDF for a patient whose current GP is oversubscribed. Panels (c) and (d) show the corresponding expected waiting times. This figure is referenced in Appendix D.3.

Figure E.8. Distribution of Chosen Waitlist Lengths by Mechanism



Notes: The figure shows the distribution of chosen waitlist ranks and the corresponding expected waiting times among attentive patients in each of our focal mechanisms. Panels (a) and (b) report the distribution of attentive patients' chosen waitlist rank conditional on being less than 400. The dots (scaled on the right axis) report the probability mass above this truncation point. Panels (c) and (d) report the distribution of corresponding expected waiting times, conditional on being below 36 months. Again, the dots (scaled on the right axis) report the probability mass above the truncation point. This figure is referenced in Appendix D.3.

Figure E.9. Relationship Between Beliefs and Waitlist Rank by Panel Cap (TTC Mechanism)



Notes: The figure shows the relationship between patient beliefs and waitlist rank for three different GP panel cap sizes, under the TTC mechanism. Panels (a) and (b) report a patient's expected discount factor (EDF) as a function of waitlist rank, if the waitlist considered was for a GP with a panel cap of 750, 1,000, or 1,250. Panel (a) shows the EDF for a patient whose current GP is undersubscribed, while panel (b) shows the EDF for a patient whose current GP is oversubscribed. Panels (c) and (d) show the corresponding expected waiting times. A higher panel capacity will make the waitlist move faster, and thus expected wait-time lower (and EDF higher). If a patient's current GP is oversubscribed, they will understand that they have the possibility of being reassigned via TTC, and thus have more optimistic expectations about waiting time. This figure is referenced in Appendix D.3.