

Randomizing Religion: The Impact of Protestant Evangelism on Economic Outcomes*

Gharad Bryan, James J. Choi, Dean Karlan

December 2019

Abstract

We study the causal impact of religiosity through a randomized evaluation of an evangelical Protestant Christian values and theology education program. We analyze outcomes for 6,276 ultra-poor Filipino households six months and 30 months after the program ended. At six months, we find increases in religiosity and income, no statistically significant changes in total labor supply, consumption, food security, or life satisfaction, and a decrease in perceived relative economic status. Exploratory analysis suggests that the income treatment effect may operate through increasing grit. These effects fade away at 30 months. We conclude that this church-based program may represent a method of increasing non-cognitive skills and reducing poverty in the short run among adults in developing countries, but more work is required to understand whether the effects can persist and if not, why not.

Keywords: religion, economics, poverty, non-cognitive skills

JEL: D12; I30; O12

* Bryan: g.t.bryan@lse.ac.uk London School of Economics, Innovations for Poverty Action and the M.I.T. Jameel Poverty Action Lab; Choi: james.choi@yale.edu Yale University and NBER; Karlan: karlan@northwestern.edu Innovations for Poverty Action, the M.I.T. Jameel Poverty Action Lab, NBER and CEPR. This study was registered, along with a pre-analysis plan, in the American Economic Association Registry for randomized control trials under trial number AEARCTR-0001060. Institutional Review Board approval by Innovations for Poverty Action (#1185), NBER (#17_116), and Yale University (#1410014779). For funding, we thank the Bill and Melinda Gates Foundation for funding related and overlapping data collection on access to savings, Celia and Joseph Grenny, NIH grant P01AG005842, and the Yale University Economic Growth Center. We thank Sachet Bangia, Nate Barker, Leah Bridle, Rebecca Hughes, Marius Karabaczek, Sana Khan, Megan McGuire, Neil Mirochnick, Isabel Oñate, Nassreena Sampaco-Baddiri, Cornelius Saunders, Martin Sweeney, and Sneha Stephen from Northwestern University and Innovations for Poverty Action for research assistance and management support. We thank Peter Aronow, Latika Chaudhary, Gregory Cox, Dan Hungerman, Laurence Iannaccone, and numerous seminar audiences for helpful comments. We thank Lincoln Lau, David Sutherland, Peter Nitchke, Daniel Mayhugh, Zaldy Rodriguez, the ICM Metrics team, Danilo Mijares and the ICM Bacolod staff, Lilian Barinas and the ICM Dumaguete staff, Jonathan Sanchez and the ICM Koronadal staff, and Evren Managua and the ICM General Santos staff for their collaboration, patience, flexibility, and curiosity throughout the design and implementation of this study.

A literature dating back at least to Adam Smith and Max Weber has argued that religiosity is associated with a set of characteristics that promote economic success, including diligence, thriftiness, trust, and cooperation (Iannaccone 1998; Iyer 2016). More recent research has linked religiosity to positive outcomes in domains such as physical health (Ellison 1991), crime rates (Freeman 1986), drug and alcohol use (Gruber and Hungerman 2008), income (Gruber 2005), and educational attainment (Freeman 1986; Gruber 2005). Other studies have argued for negative economic effects of some aspects of religiosity due to a focus on otherworldliness (Weber [1905] 1958 in his discussion of Catholicism) and substitution toward church attendance away from production (Barro and McCleary 2003). Despite extensive research, claims that religion *causes* outcomes remain controversial, in part because people choose their religion. Naturally occurring religious affiliation is likely to be correlated with unobserved personal characteristics, which may be the true drivers of the observed correlations. Iannaccone (1998) writes that “nothing short of a (probably unattainable) ‘genuine experiment’ will suffice to demonstrate religion’s causal impact.”

To study the causal impact of religiosity, we partnered with International Care Ministries (ICM), an evangelical Protestant anti-poverty organization that operates in the Philippines, to conduct an evaluation that randomly assigned invitations to attend Christian theology and values training. There are 285 million evangelical Christians in the world, comprising 13% of Christians and 36% of Protestants (Hackett and Grim 2011).² ICM is representative of an important sector that attempts to generate religiosity while alleviating poverty.

ICM’s program, called Transform, normally consists of three components—Protestant Christian theology, values, and character virtues (“V”), health behaviors (“H”), and livelihood (i.e., self-employment) skills (“L”)—taught over 15 weekly meetings (plus a 16th meeting for a graduation ceremony). Each meeting lasts 90 minutes, spending 30 minutes per component. ICM’s leadership believes that the values curriculum lies firmly in the mainstream of evangelical belief. Between 2009 and 2017, 194,000 people participated in Transform. The basic structure of the program, using a set series of classes outside of a Sunday worship service to evangelize, is a common model. For example, over 24 million people in 169 countries have taken the evangelistic

² The National Association of Evangelicals lists four defining characteristics of evangelical Christians that have been identified by historian David Bebbington: “the belief that lives need to be transformed through a ‘born-again’ experience and a life long process of following Jesus,” “the expression and demonstration of the gospel in missionary and social reform efforts,” “a high regard for and obedience to the Bible as the ultimate authority,” and “a stress on the sacrifice of Jesus Christ on the cross as making possible the redemption of humanity.” (<https://www.nae.net/what-is-an-evangelical/>, accessed April 20, 2018)

Alpha course since 1977 (Bell 2013), and Samaritan’s Purse has enrolled 11 million children in about 100 countries in its evangelistic Greatest Journey course since 2010 (Samaritan’s Purse 2017). Like Transform, these are courses of approximately a dozen sessions.

We randomly assigned communities to receive the full Transform curriculum (VHL), to receive only the health and livelihoods components of the curriculum (HL), to receive only the values component of the curriculum (V), or to be a no-curriculum control (C). We identify the effect of religiosity by the comparison of invited households in VHL communities to invited households in HL communities, and invited households in V communities to households in C communities that would have been invited had that community been assigned to be treated.

We measure outcomes approximately six months and 30 months after the training sessions ended and analyze them in accordance with a pre-analysis plan. At six months, we find that those who were invited to receive the values curriculum have significantly higher religiosity than those who did not receive the values curriculum, demonstrating that the treatment had its intended first-stage effect. Examining downstream economic outcomes while correcting for multiple hypothesis tests by controlling the false discovery rate (FDR), we find that the values curriculum increased household income by 9.2%, but it had no statistically significant effect on total labor supply, assets, consumption of a subset of goods, food security, or life satisfaction, and it decreased perceptions of relative economic status within one’s community by 0.11 points on a 10 point scale.⁴ Post-hoc analysis shows that the income effect is strongly concentrated on the Transform invitee and is not significant for other household members’ labor income, suggesting that the estimated income effect is not a Type I error. Additional post-hoc analysis on the health and livelihoods curriculum finds that it had no significant treatment effects on income and perceived relative economic status.⁶

Exploratory regressions suggest that the religiosity treatment effect operates by increasing grit (Duckworth et al. 2007)—specifically, the portion of grit associated with perseverance of effort (and in particular, agreement with the statements “I am a very hard worker,” “I finish whatever I begin,” and “Setbacks don’t discourage me.”). This mechanism accords with Weber’s conception of the Protestant work ethic. We find no consistent movement in the other potential mechanisms

⁴ In post-hoc analysis not contained in our pre-analysis plan, we find that the treatment effects on religiosity and income remain significant when we instead control for the family-wise error rate (FWER). We discuss in Section VII.B the conceptual differences between controlling the FDR versus the FWER.

⁶ The p -value of the null hypothesis that receiving any HL curriculum has no effect is 0.299 for income (95% confidence interval = [-2.8%, 9.0%]) and 0.395 for perceived relative economic status (95% confidence interval = [-0.13, 0.05]).

that we measured: social capital, locus of control (other than the belief that God is in control, which increases), optimism, and self-control.

By 30 months, there is no longer a statistically significant difference in religiosity between the experimental groups. Consistent with the interpretation that the six-month increase in religiosity led to an increase in grit and income, we see the treatment effects on these downstream outcomes also disappear at 30 months. A survey of the pastors who taught the values curriculum in our study indicates that the estimated treatment effects are unlikely to have disappeared because the pastors evangelized the control and HL communities.

ICM operates in a setting where most people claim to be religious. In the six-month survey, only 2.4% of those who did not receive the Values curriculum and 2.3% of those who did receive the Values curriculum indicate that they are “not religious at all” when asked, “To what extent do you consider yourself a religious person?” Our experiment should therefore be understood as measuring the effects of strengthening pre-existing religiosity or changing the emphasis of pre-existing religious beliefs, rather than the effects of causing the completely irreligious to become religious. Arguably, these intensive margin effects are the most relevant ones, since 84% of the world’s population is religious (Pew Research Center 2015). It is also important to note that ICM targets the ultra-poor within communities, and the communities in our study (including those in the no-curriculum control) are chosen by pastors who presumably believe that they would be able to run a successful program there. Most expansions by religious organizations into a community are probably based on a belief that the community would be receptive, so these are an externally relevant type of community. It is possible that the ultra-poor are more receptive to religious outreach than less impoverished individuals (Chen 2010), so ICM’s outreach may be more effective than comparable outreach to higher-income populations.

In addition, religiosity is not a singular concept, and its causal impact will likely depend on many factors. Johnson, Tompkin, and Webb (2008) differentiate between “organic” exposure to religion over a prolonged period of time (e.g., through one’s upbringing at home) and “intentional” exposure through participation in a specific program targeting a specific set of individuals. Both are important channels of religious propagation, and the type of religiosity produced may depend on the channel. Our study is about intentionally generated religiosity of a specific kind (evangelical Protestant Christian), and a significant aim of our study is to establish, in the context of a

randomized controlled trial, that intentional exposure to a religious program can generate the critical first stage: an exogenous change in religiosity.

Our paper is related to a recent literature that argues that non-cognitive skills are important drivers of economic outcomes and can be improved through specific interventions (Duckworth et al. 2007; Kautz et al. 2014; Blattman, Jamison, and Sheridan 2015). This body of work raises the possibility that programs to improve non-cognitive skills might have large positive impacts on the lives of the most disadvantaged people, but three obstacles need to be overcome to meet this goal. First, with a few exceptions (e.g., Blattman, Jamison, and Sheridan 2015), existing studies concentrate on developed countries, while most of the world's poorest people live in the developing world. Even if we can assume that non-cognitive skills are similarly malleable in the developing world, it is not clear that the environment and market structures allow for economic gains. Second, much of the literature concentrates on children, and little is known about the ability to improve the non-cognitive skills of adults, although Kautz et al. (2014) notes that non-cognitive skills are more malleable later in life than cognitive skills. Finally, it is unclear whether interventions that create large improvements can be delivered in a cost-effective, scalable manner. Our results suggest that church-based programs might be a solution. Church-based programs make use of a large existing infrastructure, teach a well-understood and developed set of values, and are often low-cost because they leverage the intrinsic motivation of church members.

Our paper also relates to a growing number of papers that use instrumental variables or natural experiments to study the causal effect of religion on economic outcomes.⁷ Clingingsmith, Khwaja, and Kremer (2009) study a randomized lottery for participation in the hajj and its impact on beliefs, values, and religious practices. Barro and McCleary (2003) conduct a cross-country analysis of economic growth using the existence of a state religion, state regulation of religion, adherence shares for the major religions, and a religious pluralism index as instruments. They find that religious beliefs increase economic growth, whereas religious service attendance decreases growth. Because our study does not induce independent exogenous variation in beliefs versus behaviors, we cannot add further evidence on this “believing versus belonging” hypothesis. Gruber (2005) uses local ancestral mix as an instrument and finds that religious participation in the U.S.

⁷ Laboratory experiments that study religious effects by exogenously varying the salience of religion include Shariff and Norenzayan (2007), Mazar, Amir, and Ariely (2008), Hilary and Hui (2009), Horton, Rand, and Zeckhauser (2011), and Benjamin, Choi, and Fisher (2016). See Shariff et al. (2016) for a review of the laboratory literature.

(which is almost entirely Christian) increases education, income, and marriage rates and decreases disability and divorce rates. Gruber and Hungerman (2008) exploit the repeal of U.S. state laws prohibiting retail activity on Sundays to find that Christian religious participation decreases drinking and drug use. Bottan and Perez-Truglia (2015) study the decline in Catholic religious participation caused by clergy scandals and find evidence that religious participation increases charitable giving.

Becker and Woessmann (2009) and Cantoni (2015) use geographic distance from Wittenberg, where Martin Luther posted his Ninety-Five Theses, as an instrument for adoption of Protestantism. Becker and Woessmann (2009) conclude that Protestantism does increase income, but this can be entirely accounted for by its effect on literacy, whereas Cantoni (2015) finds no impact on economic growth. Woodberry (2012) argues that Protestants' desire for people to read the Bible fostered mass education, mass printing, and civil society, making it more likely that a country on the receiving end of high historical Protestant missionary activity is a democracy today. Basten and Betz (2013) and Spenkuch (2017) use different peace treaties signed 500 years ago as instruments for local Protestant versus Catholic share and find support for a Protestant work ethic.

I. The ICM Transform Program

Transform's Values curriculum begins by teaching participants to recognize the goodness of the material world and their own high worth as God's creation. The theme then shifts towards humanity's rebellion against God and its negative consequences, while contrasting that with the message that "believers of Jesus will discover joy in sorrow, strength in weakness, timely provision in time of poverty, and peace in the midst of problems and pain." (Transform does not, however, teach prosperity theology—the belief that following God will guarantee economic prosperity and physical health.⁸) The Protestant doctrine of salvation by grace—people cannot earn their way into heaven by performing good works, but can only be saved by putting their faith in Jesus, upon which God forgives their sins as a free act of grace—is taught. The proper response to God's grace is to do good works out of gratitude. The final section of the curriculum covers what such good works would be. They include stopping wasting money on gambling and drinking, saving money,

⁸ The teacher's manual for the Values curriculum says that "we also see ordinary and simple people who enthrone God as their Lord and Savior discover the deep satisfaction and contentment that make them happy even in their relative poverty."

treating everyday work as “a sacred ministry,” and becoming active in a local church community. Participants are encouraged to find hope in the midst of disasters through faith and generally see that “life’s trials and troubles” are “God’s pruning knife” that will result in “more fruitfulness.” In other words, the curriculum teaches students that their suffering has meaning and purpose, and aims to build the ability to persevere through setbacks. These curricular elements dovetail with the growing literature on non-cognitive skills that emphasizes the importance of characteristics like conscientiousness, grit, resilience to adversity, self-esteem, and the ability to engage productively in society (Kautz et al. 2014).

The Health training focuses on building health knowledge and changing health and hygiene practices in the household. Additionally, ICM staff identify participants experiencing malnourishment and common health issues such as diarrhea, tuberculosis, and skin problems. They then receive nutritional supplements, deworming pills, other medical treatments, and follow-up care.

The Livelihood section of the program consists of training in small business management skills, training in one of several different livelihood options (for example, an introduction to producing compost through vermiculture), and being invited to a savings group. Minor agricultural assistance is given in the form of small seed kits. These activities are intended to provide key tools for achieving a more sustainable income and smoothing economic shocks.

The health and livelihoods components are led by two employees of ICM, while the religious training is led by a local pastor following an ICM-provided curriculum. The local pastor is not compensated by ICM but does receive training and support. Six lay volunteers from the pastor’s church serve as counselors who offer support and encouragement to the participants. For a small number of participants, ICM arranges treatment for serious medical needs.

The teacher’s manuals used by ICM are available on the authors’ websites.

II. Experimental Design

For the experiment, ICM recruited 160 pastors to each choose two communities in which (s)he did not already minister and that were at least ten kilometers away from each other. Selected communities were required to be predominantly Catholic or Protestant—which meant that Muslim-majority communities were excluded—and not to have been previously contacted by

ICM.⁹ Within each community, the pastor created a list of 40 households that (s)he considered the poorest and thus eligible for participation in Transform, and interacted with these households to assess their willingness to participate in the program should it be launched in their village. One member of the household—usually the female head of household or the female spouse of the male head of household—was identified as a potential invitee to Transform. ICM staff then administered a poverty verification questionnaire, based on indicators such as the quality of a home’s construction materials, access to electricity, clean water and sanitation, and household income—most of which do not rely upon self-reports. The previously identified individuals in the 30 households deemed poorest were invited to participate in the program if their community was selected for treatment.

The randomization was a two-stage clustered design. In the first stage, the pastors were randomly assigned to either group VHL-C or group HL-V. In the second stage, pastors in VHL-C had one of their communities randomly assigned to receive the full Transform program (VHL) and the other to be a no-treatment control group (C). Pastors in HL-V had one of their communities randomly assigned to receive only the health and livelihoods component of Transform (HL), and the other to receive only the Christian values component of Transform (V).¹⁰ We implemented this randomization scheme because each pastor had capacity to provide values training in only one community, and thus the scheme allowed every invited pastor to be involved in exactly one Transform implementation. This design also meant that the total amount of religious outreach done by ICM was not altered due to the study. Since the treatments were assigned at the community level, the estimated effect of the Values treatment on downstream economic outcomes should be interpreted as the effect of increasing religious engagement for a group of individuals in a community, rather than the effect for an isolated individual. We view this as a desirable feature, since religion is most often experienced and practiced in a communal context.

The four-month Transform program ran from February to May 2015. HL/VHL households on average attended 8.9 class sessions, and 83% attended at least one.¹¹ Participants in the VHL, HL, and V treatment arms also received food supplements, and ICM arranged treatment for serious

⁹ There is only one ICM base (located in Mindanao) that is close to any communities that are predominantly Muslim.

¹⁰ Both HL and V communities were also assisted by six counselors recruited by the pastors prior to the random assignment.

¹¹ ICM did not track attendance in the V group. If somebody was sent in the place of an invited individual, ICM recorded that individual as present. We cannot distinguish these substitute attendances from regular attendances.

medical needs (<1% of participants). We will show that the food supplements and medical treatment are unlikely to entirely explain the V curriculum treatment effects, because the HL curriculum, which is also accompanied by food supplements and medical treatment, does not have comparably significant treatment effects.

The experimental implementation was carried out by ICM independent of the researchers. ICM covered the costs of the V and VHL treatments, but the researchers raised funds to cover the costs of implementing the HL curriculum, as ICM felt uncomfortable using donor funds on a program with no religious element. The researchers provided guidance on randomization but were not paid by ICM in any way.

III. Data Collection for the Six-Month Survey

Approximately six months after Transform ended (between August 12, 2015 and January 14, 2016), we sent surveyors to the poorest 25 households selected by the pastors in each community.¹² Respondents were compensated with 100 PHP (about 2.50 USD), irrespective of whether they completed the survey.

To reduce the correlation between treatment assignment and social desirability bias in survey responses, we used surveyors from a nonprofit research organization unaffiliated with ICM, Innovations for Poverty Action (IPA). Respondents were not told of any relationship between ICM and IPA, and the informed consent script introduced the survey as follows: “Hello, my name is _____ with the research organization Innovations for Poverty Action. I am working to learn about the economic and social conditions and well-being of families in the Philippines. You are being invited to be one of the participants in this study. We expect the results from this survey will help Filipino NGOs and international organizations to develop policies and procedures that improve the lives of people.”

All of the questions about primary economic outcomes came before the main questions that asked directly about religiosity. If these direct religiosity questions caused subjects to discern a link between ICM and IPA, only some of the secondary outcome and mechanism questions would have been affected. We did mention religion at three points prior to measuring primary economic outcomes. First, the script for obtaining informed consent said, “If you agree to participate in this

¹² We sampled the 25 poorest households, rather than the full 30 identified by ICM, because of budget constraints and the programmatic importance of measuring the impact on the poorer individuals within the sample.

study, we will ask you questions about your household’s economic, health, social, and religious status.” Second, when constructing the household roster, we asked about each household member’s relation to the head of the household, permanence of his or her residence in the home, gender, age, religious denomination, marital status, schooling, literacy, and work status. Third, two of our five list-randomized questions (described in Section IV) measured religiosity in an obscured way. Given the many different characteristics mentioned in the informed consent script and measured in the household roster, the obscured nature of the list-randomized questions, and the fact that only two of the five list-randomized questions had religious content (which in turn was shown to only half the respondents; see Section IV), we think it is unlikely that respondents would have inferred a link to ICM when we were eliciting primary economic outcomes.

Surveyors attempted to interview, in descending order of preference, (a) the person previously identified as a potential Transform invitee, (b) the female head of household if the head of household was female, (c) the female spouse/partner of the male head of household, or (d) the person reporting to be responsible for health and household expense decisions. Out of 7,999 households targeted for surveying, we successfully surveyed 6,507 (81%); in 88% of these households, the respondent was the potential/actual Transform invitee. Insurgent violence and political opposition prevented the field teams from surveying in six communities (150 households), and some households either refused to be surveyed (60 households), could not be contacted (1,252 households), or suffered from survey data issues (30 households).

Management data and internal control checks identified five instances (out of the 157 pastors whose communities we surveyed) in which ICM and the pastor switched the assignments within a community pair, treating one with what the other was supposed to receive, and vice versa. Because of the paired randomization, we drop these five community pairs in our analysis without harming internal validity. There was also one community that was supposed to receive the V treatment but did not. We retain this community in our regressions, since the compliance issue was not present in both communities in the pair.¹⁴ Thus, we only use data from 6,276 households in our main

¹⁴ We show in Online Appendix Tables 3-5 the full set of analyses including the five pairs dropped in the main regressions, using the assigned treatment status for each community. Relative to Tables 1-3, the only treatment effect estimate of the V curriculum on primary outcomes that moves across the 5% or 10% significance boundaries is for perceived relative economic status, which is now significant only at the 10% level. Examining mechanisms and secondary outcomes, in the pooled specification, the negative V effect on the life orientation index loses significance even at the 10% level, while the positive V effect on grit and the negative V effect on self-control move from 10% significance to 5% significance.

analyses. Online Appendix Table 1 shows that the attrition rate and the number of days between program end and survey date do not differ significantly across the four experimental groups.

Before the intervention, we intended to conduct a baseline survey of the 7,999 households. However, we underestimated the time it would take to conduct the baseline, and we were unable to delay the start of Transform in order to complete the baseline. This means that we have baseline data on only 2,634 of the households. Online Appendix Table 1 shows that the four experimental groups are well-balanced on characteristics measured in the six-month survey that are unlikely to have changed in response to the treatment. Online Appendix Table 2 shows that in the subsample of households we were able to survey at baseline, household income and respondent age, gender, education, income, and religiosity at baseline do not predict attrition from the six-month survey.¹⁵

We filed a pre-analysis plan with the American Economic Association RCT Registry before seeing any follow-up data. In accordance with our first filing, we then examined the follow-up data blinded to treatment assignment and filed a supplement to the pre-analysis plan.¹⁶

All data supporting the findings of this study, stripped of individual-identifying information, will be posted on the IPA and JPAL Dataverse before publication.

IV. Six-Month Survey Outcome Variables

Our pre-analysis plan divided outcomes into primary religious outcomes, primary economic outcomes, mechanisms, and secondary outcomes. Many of these outcome variables are indices, which we standardize so that the control group has zero mean and unit variance. If the index is

¹⁵ In unreported regressions, we find that when baseline characteristics are interacted with treatment assignment, higher education does predict less attrition from the six-month survey among baseline respondents in the control and HL groups only. If we took this result at face value, we would expect the control and HL groups to have significantly more education than the V and VHL groups in the six-month survey. But in fact, education among *all* six-month respondents is almost perfectly balanced across experimental cells (Online Appendix Table 1). We believe the attrition regression is detecting some treatment-based differences in selection into the baseline survey, which was attempted on a rushed timeline, and not differences in selection into the six-month survey. Conditional on a community being sampled at baseline, the baseline survey successfully interviewed only 64%, 58%, 57%, and 66% of the community's targeted households in the control, V, HL, and VHL cells, respectively.

¹⁶ In accordance with the first phase of our pre-analysis plan, we analyzed the data stripped of treatment status. We randomly generated treatment assignments and checked whether including control variables from the available baseline observations reduced the standard errors of the coefficients on the randomly generated treatment dummies. We did not find any efficiency gains, so we decided not to use the baseline survey in our final regressions. We do, however, include controls for demographic variables that were collected *after* the intervention and which were unlikely to be affected by the treatment. Online Appendix Tables 6-8 show the six-month treatment effect estimates on the primary outcomes, mechanisms, and secondary outcomes if we additionally control for baseline survey measurements of household income, perceived relative economic status, and indexes of intrinsic religiosity, extrinsic religiosity, general religiosity, belief in salvation by grace, food security, life satisfaction, trust, social safety net, community activities, powerful others, locus of control, expectations, and grit.

found in previous academic literature, we use the construction method from that literature, which in our cases always involves simply summing the components (which are sometimes reverse-coded). If there is no pre-existing index, we use the index construction methodology of Kling, Liebman, and Katz (2007). We first sign all component variables such that higher is telling a consistent story for each component of the index. Then we standardize each component by subtracting its control group mean and dividing by its control group standard deviation. We compute the sum of the standardized components¹⁷ and standardize the sum once again by the control group sum's standard deviation. Appendix Table 1 shows all of the questions that comprise each of our variables. Online Appendix Tables 12-40 show the treatment effect estimates on each component of the outcome variables.¹⁸

The primary religious outcomes are the intrinsic religious orientation scale and the sum of the two extrinsic religious orientation scales of Gorsuch and McPherson (1989), a general religion index that consolidates responses to nine religious belief and practice questions, and the average of two binary indicators for whether the respondent reports that “I have made a personal commitment to Jesus Christ that is still important to me today” and “I have read or listened to the Bible in the past week.” These last two binary indicators are elicited using list randomization, a technique for eliciting responses to sensitive questions that conceals any given individual's response from the interviewer (Droitcour et al. 2011; Karlan and Zinman 2012). We do this to minimize experimenter demand and social desirability effects. In a list-randomized elicitation, participants are randomly selected to receive either a list of n non-sensitive statements or these same n statements plus a sensitive statement. They are asked to answer how many of the statements are true without specifying which ones are true. The difference in the average number of statements reported to be true between participants who received n statements and $n + 1$ statements is the estimated fraction of participants for whom the sensitive statement is true.¹⁹

¹⁷ For observations without information on one or more components of the index, we impute the missing component standardized values as the mean of the non-missing components' standardized values for that individual/household.

¹⁸ We also include Online Appendix Table 41, which shows treatment effects on consumption of “temptation goods” (cigarettes and alcoholic beverages).

¹⁹ An individual's answer about the sensitive statement can only be deduced if he or she answers 0 (implying falsity of the sensitive statement) or $n + 1$ (implying truth of the sensitive statement). An individual can answer truthfully about the longer list while being assured that her response to the sensitive statement is concealed if the number of non-sensitive questions that are true for her is not 0 or n . Among respondents who did not receive the sensitive statement, the fraction who did not give a boundary response was 73% for the list associated with the commitment to Jesus statement, and 80% for the list associated with the Bible statement. The corresponding percentages are 82%, 83%,

After data collection, we discovered an issue with our measure of intrinsic religiosity. The indexes for intrinsic and extrinsic religious orientation were measured using one 14 question block, with eight questions constituting the intrinsic index and six constituting the extrinsic index. For each question, respondents were asked to state on a Likert scale a level of agreement with a statement. In 11 out of the 14 questions, stronger agreement corresponds to stronger religiosity. In the remaining three—all of which are part of the intrinsic index—weaker agreement corresponds to stronger religiosity. We believe that respondents did not perceive the subtle changes in the direction of the questions, causing them to use stronger agreement to express stronger religiosity even for the reversed questions.²⁰ Agreement levels are positively correlated across all seven intrinsic orientation statements, regardless of whether greater agreement corresponds to greater religiosity or not. Because of this, we have chosen to exclude the three reversed questions from the intrinsic index used for the main analysis. Our broad conclusions about the six-month treatment effect on religiosity are unchanged by this choice.²¹

The primary economic outcomes are household expenditure on a sample of consumption goods, a food security index, household income, total household adult labor supply in hours, an index of life satisfaction, and perceived relative economic status.

The mechanism outcomes are three measures of social capital (a general trust index, a strength of social safety net index, and a participation in community activities index), three measures of a sense that one has control over one's life (a perceived stress index, the Levenson (1981) Powerful

and 86% for the water treatment, hand washing, and domestic abuse list-randomized questions. Therefore, the list-randomization questions concealed the truth about the sensitive statements for the majority of our respondents.

²⁰ Thirty-three percent of respondents answered “agree” or “strongly agree” to all 14 questions, regardless of whether the question was reversed, whereas only 0.02% of respondents answered “agree” or “strongly agree” to all non-reversed questions and “disagree” or “strongly disagree” to all reversed questions. The finding that many subjects indiscriminately agree with statements to express a general support for religion goes back to the earliest research on intrinsic and extrinsic religious orientation. Allport and Ross (1967) write, “In responding to the religious items these individuals seem to take a superficial or ‘hit and run’ approach. Their mental set seems to be ‘all religion is good.’ ‘My religious beliefs are what really lie behind my whole life’—Yes! ‘Although I believe in my religion, I feel there are many more important things in my life’—Yes!” They classify such types as the “indiscriminately pro-religious” and find that they are likely to be less educated.

²¹ If we instead use the eight-question intrinsic measure, as stated in our pre-analysis plan, the point estimate of the “Any V” treatment effect on intrinsic religious orientation in the pooled regression specification is 0.04 standard deviations, and its q -value rises to 0.084. In the disaggregated regression specification, the point estimate of the V versus control effect on intrinsic religious orientation is 0.01 standard deviations ($q = 0.899$), and the point estimate of the VHL versus HL effect on intrinsic religious orientation is 0.074 standard deviations ($q = 0.330$). The q -values on the other religious outcomes are qualitatively similar regardless of whether we use the eight-question or five-question intrinsic measure. Therefore, even though the estimates of the V curriculum's effect on intrinsic religious orientation weaken when we use the eight-question measure, we still find robust first-stage effects on other measures of religiosity.

Others index modified to apply to God’s control of one’s life, and a locus of control index that combines the internality and chance subscales of Levenson (1981) and the World Values Survey locus of control question), three measures of optimism (the Life Orientation Test - Revised index (Scheier, Carver, and Bridges 1994), an index of expectations about one’s life satisfaction and relative economic status five years in the future, and a general optimism index), the Short Grit Scale (Duckworth and Quinn 2009), and a subset of the Brief Self-Control Scale (Tangney, Baumeister, and Boone 2004).

The secondary outcomes are an index of belief in the Protestant doctrine of salvation by grace (an outcome of interest to ICM because the doctrine is taught in the V curriculum), an asset index, a financial inclusion index, a health index, two hygienic practice variables, a home quality index, a migration and remittance index, an absence of domestic discord index, absence of domestic violence, child labor supply, and the number of children enrolled in school.

V. Econometric Strategy

Treatment effects are estimated using ordinary least squares regressions with the following explanatory variables: treatment indicator variables, an indicator variable for the respondent’s gender, an indicator variable for the respondent being married, an indicator variable for the respondent being divorced or separated, the respondent’s years of educational attainment,²² the number of adults in the household (age ≥ 17), the number of children in the household (age < 17), and the number of days between June 1, 2015 and the interview date.²³ We also include fixed effects for each pair of communities chosen by a given pastor (“community-pair fixed effects”) where possible, as discussed in detail below. We cluster standard errors by community (the unit of randomization).

²² Pre-school only is coded as 0.5 years, some grade 12 education without high school graduation is coded as 12 years, high school graduation is coded as 13 years, partial vocational education is coded as 14 years, complete vocational education is coded as 15 years, partial college is coded as 16 years, and college graduation is coded as 17 years. In data cleaning, we discovered 27 observations in which the respondent’s name was not in the household roster, and thus respondent demographic information was missing. We code the respondent demographic variables as equaling zero for these 27 observations and control for an indicator variable equal to one if respondent demographic information is missing.

²³ These control variables were measured at the same time as the outcome variables, but are unlikely to have been affected by the treatments. Online Appendix Tables 9-11 show the treatment effect estimates on the pre-specified outcomes when the only explanatory variables are the treatment dummies and community pair or ICM base dummies. Inference about the primary outcomes is qualitatively unaffected by the absence of additional controls. Some of the mechanism and secondary outcome treatment effects cross conventional statistical significance thresholds, but because this specification was not pre-specified, we do not emphasize these results.

We estimate the treatment effect on list-randomized variables by stacking the responses of those who did and did not receive the sensitive statement in a regression that controls for treatment assignment indicator variables, an indicator variable for whether the individual received the sensitive statement, the interaction between receiving the sensitive statement and each treatment indicator variable, and all the other non-treatment variable controls from the main specification. The coefficients on the interaction variables are the treatment effects of interest. We estimate the control mean by calculating within the control group the difference (without adjusting for covariates) in the mean response between those who did get the sensitive statement and those who did not. When two list-randomized variables are combined to form an outcome variable, we stack the responses for both variables into a single regression while retaining the same control variables as above. The coefficient on the interaction variable in this case is the treatment effect on the average of the two outcomes of interest.

We test for the effect of religiosity by comparing VHL to HL respondents, and V to control respondents. We do not reject the hypothesis that the V and HL curricula have additive effects when testing jointly across all outcomes of interest; the p -values for this test are 0.344, 0.634, 0.890, and 0.234 when looking across religious primary outcomes, all primary outcomes, all primary outcomes and mechanisms, and all outcomes, respectively. Therefore, we focus—following our pre-analysis plan—on a pooled specification that estimates the effect of being invited to receive any V curriculum, while controlling for whether the household was invited to receive any HL curriculum. This pooled specification gives consistent inference on the average of the Values curriculum effect with and without a concurrent Health and Livelihoods curriculum and has greater statistical power than a specification that separately estimates the VHL-versus-HL and V-versus-control effects.²⁴

Since we conducted a matched-pair randomization, our pooled specification controls for the community-pair fixed effects previously mentioned. In our disaggregated specification, where we estimate VHL, HL, and V treatment effects separately, the estimation of the VHL treatment effect versus control also controls for community-pair fixed effects. However, the community-pair fixed effects are not possible to control for when estimating the HL and V treatment effects versus control because no pastor who selected an HL or V community also selected a control community.

²⁴ The fact that we cannot reject that the treatment effects are additive gives some confidence that this average effect is the same as the Values curriculum effect without a concurrent Health and Livelihoods curriculum.

Thus, the disaggregated specification's treatment estimates are generated from two independently estimated regressions: one to estimate the treatment effect for VHL relative to control with community-pair fixed effects, and a second to estimate the treatment effects for HL and V relative to control with fixed effects for which of the four ICM bases the community is associated with.²⁵

Because of the multiple hypotheses tested, we follow Banerjee et al. (2015): for each primary test in our pre-analysis plan, we calculate a q -value—the minimum false discovery rate (i.e., the expected proportion of rejected null hypotheses that are actually true) at which the null hypothesis would be rejected for that test (Benjamini and Hochberg 1995; Anderson 2008), given the other tests run within the family.²⁶ For the purposes of this correction, and in accordance with our pre-analysis plan, we consider the tests on primary religious outcomes to be one family (because they are a test of the study's first stage, a null result here would eliminate the justification for examining the non-religious outcomes), and the tests on primary non-religious outcomes to be another family. We implement adjustments once among the pooled specification regressions, and separately among the disaggregated specifications. In other words, the tests run within the pooled specification do not affect the q -values from the disaggregated specifications, and vice versa. Following our pre-analysis plan, we do not apply multiple hypothesis test corrections to our tests of hypothesized mechanisms and secondary outcomes because these analyses are exploratory.

VI. Results of Pre-Specified Analyses on the Six-Month Survey

The majority of our sample (69%) self-identifies as Catholic, and 21% as Protestant. Online Appendix Tables 12-15 summarize the control group's level of religiosity, and indicate that many are not maximally religiously fervent. For example, when asked, "To what extent do you consider

²⁵ Our pre-analysis plan stated that we would control for community-pair fixed effects in all regressions. We have deviated from the plan here because it is mathematically impossible to control for community-pair fixed effects in the disaggregated specification while estimating every single treatment effect. Due to the randomized design, the inability to control for community-pair fixed effects when estimating the HL and V treatment effects relative to control does not bias our estimates, but it does reduce our statistical power.

²⁶ Within each of our outcome families, let $p_1 \leq p_2 \leq \dots \leq p_m$ be the set of ordered p -values that correspond to the m hypotheses tested. For a given false discovery rate α , let k be the largest value of i such that $p_i \leq i\alpha/m$, and reject all hypotheses with rank $i \leq k$. The q -value of a hypothesis, an analog to the p -value, is the smallest α for which the hypothesis would be rejected (Anderson 2008). The Benjamini-Hochberg procedure was originally proven to work under the assumption that the test statistics were independent. Subsequent work has shown that the procedure is robust to various dependence structures (Goeman and Solari, 2014). Romano, Sheikh, and Wolf (2008) develop a testing procedure that incorporates information about the dependence structure. Benjamini-Hochberg q -values are conservative, and more powerful procedures have been more recently developed (e.g., Storey, Taylor, and Siegmund, 2004; Benjamini, Krieger, and Yekutieli, 2006). We do not follow these approaches because we wish to stay as close as possible to our pre-analysis plan, which specified the more conservative Benjamini-Hochberg procedure.

yourself a religious person?,” the average control respondent rates herself at 2.8 on a 4-point scale, where higher numbers indicate greater religiosity. Only 66% say that they have made a personal commitment to Jesus Christ that is still important to them today, and 56% have read or listened to the Bible in the past week.

Tables 1-3 contain all of our pre-specified analyses. Columns 1-4 of Table 1 show the treatment effects on the primary religious outcomes. The pooled specification (Panel A) finds that the V curriculum, offered either on its own or in conjunction with the HL curriculum, increases all four measures of religiosity, three of them at $q < 0.01$.²⁷ The effect on the three significant indices ranges from 0.08 to 0.13 standard deviations. The change in the list-randomization outcome—which we have lower statistical power to detect, both because list-randomized questions measure the outcome of interest in only half the sample and because we only have two such questions—is positive, and its 4.8 percentage point magnitude (corresponding to a 0.10 standard deviation movement given the 60.6% control group mean) is economically significant and in line with the magnitudes (in standard deviation space) we get from the three direct elicitation measures. However, the 95% confidence interval for the list-randomization index treatment effect is wide and encompasses zero. Furthermore, recent work has demonstrated a large amount of instability in estimates coming from list randomization. In a developing country context, Chuang et al. (2019) find that within a single survey of about 1,000 respondents, estimates of the prevalence of a sensitive behavior can vary by as much as 39 percentage points across two list-randomized elicitations, and there is no clear evidence that the list-randomized estimates are systematically less biased than direct responses. Thus, we believe little should be concluded from the treatment effect estimates on the list randomization outcome. The statistically significant first-stage effect of the treatment on directly elicited religiosity justifies examining differences in downstream non-religious outcomes across treatment groups to gain insight into the effects of religiosity.

We also present results for the disaggregated specification in Panel B where we estimate the impact of the V curriculum by separately comparing VHL to HL and V to control. Although the point estimates of VHL’s effect on religiosity relative to HL are always positive, they are not statistically significant. On the other hand, V significantly increases extrinsic religious orientation (0.20 *sd*, *se* = 0.06, $q = 0.013$) and marginally significantly increases intrinsic religious orientation

²⁷ Although intrinsic and extrinsic religious orientation were originally conceived of as opposing concepts on a unidimensional scale, empirical work has found the two to be orthogonal to each other (Kirkpatrick and Hood 1990).

(0.12 *sd*, *se* = 0.05, $q = 0.058$) relative to the control group. Therefore, while we report all treatment effect estimates on downstream outcomes from the disaggregated specification, we only discuss and interpret these outcomes for the V versus control comparisons, and only correct for multiple hypothesis tests within the V versus control comparisons.

The primary economic outcome effects are reported in Table 1. We find no statistically significant treatment effects on consumption, food security, total adult labor supply, or life satisfaction. We have enough statistical power to reject, at the 95% confidence level, increases in these variables of more than 0.06 standard deviations and decreases of more than 0.04 standard deviations. However, we do find a statistically significant 9.2% increase in income (386 PHP \approx 8.6 USD per month, *se* = 127 PHP \approx 2.8 USD, control group mean = 4,213 PHP \approx 94 USD, $q = 0.015$) in the pooled specification (Panel A).²⁸ In the disaggregated specification (Panel B), where we have less statistical power (the standard errors are over twice as large as in the pooled specification), the 574 PHP income effect for V compared to C is statistically significant before correcting for multiple hypothesis tests but not after ($p = 0.045$, $q = 0.271$). We also find a significant decrease in perceived relative economic status (-0.11 points on a 10-point scale, which corresponds to -0.05 *sd*, *se* = 0.05, $q = 0.050$) in the pooled specification. Perceived relative economic status is measured by one question that asks respondents to place themselves on a ladder of life where the top rung (10) represents the best-off people in their community and the bottom rung (1) the poorest people in their community. We discuss potential interpretations of these results in Section VII.

Table 2 reports tests of mechanisms that might generate the primary economic effects. The V curriculum teaches that God's love continues during adversity, which he ultimately uses for good, so participants can find hope in the midst of hardship. Correspondingly, we find in the pooled specification (Panel A) that the V curriculum leads to increases in the sense that God is in control (Powerful Others index, 0.09 *sd*, *se* = 0.03)²⁹ and a marginally significant increase in grit (0.04 *sd*, *se* = 0.02). However, there is no consistent effect on the three measures of optimism. Perceived self-control falls by a marginally significant extent (-0.03 *sd*, *se* = 0.02), which could be due to the

²⁸ The results become more statistically significant when income is winsorized at the 95th or 99th percentile, or when we use the log of income (see Online Appendix Table 42).

²⁹ Although our pre-analysis plan treats the Powerful Others index as a potential mechanism rather than a primary outcome, the increase in its value could also be seen as evidence that the V curriculum succeeded in increasing religiosity. Relative to our other primary religious outcomes, this measure may be less prone to social desirability bias.

V curriculum increasing the number of behaviors participants believe to be undesirable temptations rather than an actual reduction in self-control. There is also a marginally significant reduction in perceived locus of control (-0.04 *sd*, $se = 0.02$), although subcomponent analysis finds that V recipients report that *both* personal initiative and chance play larger roles in their life (Online Appendix Table 27). While all three of the treatment arms—VHL, HL, and V—involve group meetings that could increase social capital, we see no consistent or statistically significant effects of any of the treatments on our measures of trust, the presence of a social safety net, or participation in community activities.³⁰

Finally, we examine treatment effects on secondary outcomes (Table 3). In the pooled specification, we find that the V curriculum leads to statistically significant ($p = 0.0002$) increases in hygienic behaviors not measured by list randomization (avoiding open defecation and keeping animals in a sanitary way), but no statistically significant increase in the list-randomization response regarding washing hands after using the bathroom and treating water. We note that we find via list randomization an increase in reported domestic violence, although it is only significant at the 10% level. This finding could be interpreted either as an increase in identifying behaviors as abuse or an increase in actual abuse. Although we do not observe a statistically significant change in the non-list-randomized discord index, we do in post hoc analysis observe a significant increase in one of its components, major arguments regarding interactions with relatives (2.2 percentage points, $se = 0.8$ percentage points, Online Appendix Table 38).

The remainder of the secondary outcomes are not significant at the 5% level. We do find an unexpected, marginally significant, decrease in the index for the belief in the doctrine of salvation by grace. This may be because of the counterintuitive nature of the doctrine, which requires one to disagree with two of the three statements in our index: “I follow God’s laws so that I can go to heaven” and “If I am good enough, God will cleanse me of my sins.” In becoming more religiously fervent, subjects may have felt that they should agree more strongly with these pious-sounding statements despite the efforts of the V curriculum. The V curriculum also increases agreement with the third statement in the index, “I will go to heaven because I have accepted Jesus Christ as my personal savior,” even though that statement is *consistent* with salvation by grace. The pattern of

³⁰ Online Appendix Table 7 shows that the Any V effect on church attendance frequency is a statistically insignificant 0.9 times per year increase, versus a control mean of 39.5 times per year.

responses is consistent with the V curriculum increasing agreement with *all* pious-sounding statements.

VII. Post-Hoc Analyses on the Six-Month Survey

In this section, we discuss assorted post-hoc (non-preregistered) analyses, many of which address robustness.

VII.A. Treatment Effects on Denominational Affiliation

We find no significant evidence that any aspect of Transform increased the share of respondents identifying as Protestants, and only marginally statistically significant evidence that the V curriculum decreased identification as a Catholic by 2.7 percentage points off a control group share of 70% (Online Appendix Table 44).

VII.B. Controlling the Family-Wise Error Rate (FWER)

An alternative method of adjusting for multiple hypothesis tests is controlling the family-wise error rate (FWER) instead of the false discovery rate (FDR). The FWER is the probability of rejecting *at least one* true null hypothesis among all those tested, while the FDR is the expected proportion of rejected null hypotheses that are actually true. The following matrix, taken from Efron (2010), illustrates the difference between these two quantities.

		Decision		
		Null	Non-null	
Actual	Null	$N_0 - x$	x	N_0
	Non-null	$N_1 - y$	y	N_1
		$N - R$	R	N

There are N null hypotheses being tested, of which N_0 are actually true and N_1 are actually false. A decision rule incorrectly decides that x true null hypotheses are false, and $N_1 - y$ false null hypotheses are true. The FWER is the probability that $x > 0$, while the FDR is the expectation of x/R (defining x/R to be 0 when $R = 0$).

In post-hoc analysis, we control the FWER using the procedure of Holm (1979), which has greater power than the Bonferroni correction to detect truly false nulls while preserving the upper bound on the FWER. The FWER-adjusted p -value for a null hypothesis is the FWER tolerance level above which we would reject the null.

Table 1 shows FWER-adjusted p -values—the only non-prespecified analysis contained in this table. Although procedures that control the FDR are popular because they are more likely to reject null hypotheses, in our setting, both types of procedures lead to similar qualitative inferences, in part because of the relatively modest number of hypotheses tested. In the pooled specification, only the effect on perceived relative economic status crosses a 1% or 5% significance boundary, with an adjusted p -value of 0.083 versus a q -value of 0.050. In the disaggregated specification, the V versus C effect on intrinsic religious orientation is no longer significant even at the 10% level (adjusted $p = 0.102$ versus $q = 0.058$), but the V versus C effect on extrinsic religion orientation remains significant (adjusted $p = 0.013$).

VII.C. Naïve OLS Versus Instrumental Variable Estimates of Religiosity Effect

What would a researcher who naively runs an OLS regression of economic outcomes on religiosity in our control and HL groups find? We construct a composite religiosity index for each respondent by adding the intrinsic, extrinsic, and general religion indexes together and standardizing so that its standard deviation in the control group is 1. Online Appendix Table 45 shows that this naïve analysis leads to a negative coefficient of religiosity on monthly income of 407 PHP and a negative coefficient on adult weekly labor supply of 2.5 hours, indicating negative selection into religiosity. This is consistent with a literature that suggests that the club good provision aspects of religion are likely to generate more demand from those with low income (Chen 2010). Also interesting is that despite lower objective economic status among the more religious, the religiosity coefficients on life satisfaction and perceived relative economic status are significantly *positive*.

In contrast, an instrumental variable estimation on our full sample using receipt of the V curriculum as the instrument finds that a one standard deviation increase in composite religiosity significantly increases monthly income by 3,073 PHP and decreases perceived relative economic status by 0.9 points on a 10-point scale. These are large estimates, but they should be interpreted with caution because it seems likely that nearly all Transform participants had their religiosity increased by much less than a full standard deviation. If so, the estimated effect of a one standard deviation increase in religiosity achieved through intentional means (as opposed to organic means, as discussed in the introduction, which is probably mostly responsible for the cross-sectional variance in control group religiosity) is a linear extrapolation of an effect that is estimated over a

much smaller range. The true effect size curve may be quite concave, so the actual causal effect of increasing religiosity by a full standard deviation through intentional means may be much smaller than our estimate.

Figure 1 shows suggestive evidence on the size of the typical V curriculum effect among those whose religiosity was successfully increased. The three graphs split the sample by whether the community received the V curriculum (VHL and V groups) or not (HL and control groups), sorts each subsample by one of the directly elicited measures of religiosity, and displays the average Any V versus No V difference in this religiosity variable within each 5-percentile-wide bucket. The difference in religiosity never exceeds 0.30 standard deviations for the intrinsic index, exceeds 0.25 standard deviations only once for the extrinsic index, and never exceeds 0.16 standard deviations for the general index.³² Although the V curriculum could cause religiosity ranks to change within a population, these graphs suggest that the V curriculum seldom increases religiosity by anything close to a full standard deviation.

VII.D. Treatment Effect Heterogeneity

The fact that the differences plotted in Figure 1 tend to be smaller at higher percentiles suggests that the V curriculum increases our religiosity measures more at lower percentiles of religiosity. If that is true *and* our religiosity variables map linearly to true religiosity, then the V curriculum has a stronger impact on religiosity for the less religious.

Our ability to rigorously identify treatment effect heterogeneity is limited because we were unable to collect pre-treatment baseline data on most of our sample. What we are able to do is stratify the sample based on a small number of characteristics collected in the six-month survey that are unlikely to have been affected by the treatment (respondent age, gender, years of education, literacy, marital status, number of children in the household, and number of adults in the household). Employing the leave-one-out procedure of Abadie, Chingos, and West (2018), we use these variables to predict the composite religiosity index at six months in the HL and control

³² Because the intrinsic and extrinsic indexes take on discrete values, the difference between individuals exactly at a given integer percentile is always 0 or 0.351 for the intrinsic index, and 0 or 0.22 for the extrinsic index. Top-coding is significant for the intrinsic and extrinsic indexes; 25% of the sample has the maximum possible intrinsic index value, and 13% of the sample has the maximum possible extrinsic index value.

groups.³³ We then sort observations into terciles based on their predicted religiosity index in the absence of the V curriculum, and estimate treatment effects separately within each tercile as before.

Table 4 shows no clear pattern of treatment effect heterogeneity. While it is true that the treatment effect on the religiosity index decreases with predicted religiosity sans V curriculum (significantly only in the disaggregated specification), the treatment effect on the list-randomized religiosity measure (which was not shown in Figure 1) *increases* with predicted religiosity sans V curriculum (significantly only in the pooled specification). There is correspondingly no significant difference across terciles in the treatment effect on monthly income and perceived relative economic status—the two primary economic outcomes for which we found a significant effect over the entire sample.

VII.E. How Much of the V Treatment Effect Operates Through the V Curriculum?

Those assigned to the V treatment not only received the V curriculum, but also socialized with other classmates, spent time away from home in order to attend class, received food supplements and medical treatment, etc. How much of the V treatment effect is due to the V curriculum itself rather than the other accompanying factors?

We can gain some insight into this question by examining the effect of the HL treatment, which also brought participants together for ICM-sponsored classes. Table 1 shows that both the Any HL effects in the pooled specification and the HL effects in the disaggregated specification are not significant (even without multiple testing corrections) for the outcomes where we found significant Any V curriculum effects. This suggests that the non-curriculum aspects of the V treatment are not responsible for the entire V treatment effect.

Comparing magnitudes of the point estimates, we see that the Any V treatment effect on income is 386 PHP, whereas the Any HL treatment effect is only 131 PHP. Under the assumption that the HL curriculum did not decrease income, this gap suggests that at least 66% of the Any V treatment effect is due to the V curriculum itself. Similarly, at least 64% of the decrease in perceived relative economic status due to the Any V treatment is due to the V curriculum itself. An analogous comparison of the V treatment effect to the HL treatment effect in the disaggregated

³³ In a multivariate regression that does not leave any observations out, significant positive predictors of religiosity in the HL and control groups are being female, older, literate, less educated, and not divorced.

specification suggests a lower bound of 50% for the income effect and 45% for the perceived relative economic status effect due to the V curriculum.³⁴

VII.F. Social Desirability Bias in Survey Responses

Although it is possible that the V curriculum is causing respondents to increase the amount by which they falsely inflate reported income for social desirability reasons, this seems unlikely, since there is no positive V treatment effect on other economic outcomes—in particular, self-reported life satisfaction, a more subjective outcome than income that seems at least as susceptible to social desirability motives.

We can test for the existence of social desirability bias in some of our survey responses by using the technique of Coffman, Coffman, and Marzilli Ericson (2016). For four of the sensitive statements whose truth we elicited by list randomization, we have direct questions elsewhere in the survey that ask about the same issue. We take respondents whose list randomized question did *not* include the sensitive statement of interest, and compute how many of the list items would have been reported true if their list *had* included the sensitive statement of interest, using their response to the direct question to impute whether the sensitive statement would have been counted as true in the list randomized question.³⁵ Under the null of no social desirability bias (but keeping in mind the caveats about the instability of list randomized estimates raised by Chuang et al. (2019)), there should be no difference between (1) the number of statements that are indicated to be true by those who did receive the sensitive statement in their list randomized question, and (2) the number of

³⁴ The p -value for the difference between the Any V and Any HL treatment effects is 0.160 for income and 0.270 for perceived relative economic status. The p -value for the difference between the V and HL treatment effects is 0.257 for income and 0.628 for perceived relative economic status. We can also compare the VHL to HL treatment effects in the disaggregated specification, although this analysis is clouded by the fact that we detected no significant difference in religiosity between these two treatment cells. We find that the incremental addition of the V curriculum accounts for 45% of the VHL effect on income and 52% of the VHL effect on perceived relative economic status. The p -value for the difference between the VHL and HL treatment effects is 0.390 for income and 0.488 for perceived relative economic status.

³⁵ The directly asked questions were “How much do you agree with this statement: ‘I have made a personal commitment to Jesus Christ that is still important to me today’?”; “In the past 7 days, how many times did you read or listen to the Bible, the Koran, or other religious literature?”; “Do you wash your hands with ash or soap after using the latrine?”; and “Is the following true or false? Someone in my household is experiencing physical abuse.” We code the “personal commitment to Jesus Christ” statement as true if the respondent slightly agrees, agrees, or strongly agrees; reading or listening to the Bible as true if the respondent did so at least once; and washing hands as true if the respondent answers sometimes or always. The results are directionally identical if we count the “personal commitment to Jesus” statement as true only if the respondent agrees or strongly agrees, and if we count washing hands as true only if the respondent answers “always.” Due to a programming problem in the questionnaire, we only have 1,447 observations for the physical abuse question.

statements that we impute would have been marked as true by those who did not receive the sensitive statement in their list randomized question.

Panel A of Table 5 shows the results of a regression where the dependent variable is the number of statements in the list with the sensitive statement that are true (either actual or imputed) and the main explanatory variable is a dummy for having actually received the sensitive statement in the list. We see that the fraction who reports a personal commitment to Jesus, reading or listening to the Bible in the past week, washing their hands after going to the bathroom, or nobody in their household experiencing physical abuse is 26, 22, 23, and 9 percentage points lower, respectively, when this is elicited via list randomization instead of directly. But in Panel B, we see that the size of the reporting bias does not vary significantly with whether the respondent received the V curriculum. Although the standard errors of these interaction coefficients are relatively large, they do suggest that social desirability is not biasing our treatment effect estimates.³⁶

VII.G. Sensitivity of Estimates to Survey Attrition

As noted in Section III, the survey attrition rate did not differ across experimental cells or (for the subset of the population successfully surveyed at baseline) by characteristics measured in the baseline survey. In this subsection, we examine how our results would be affected if the outcomes of non-responders systematically differ across experimental cells.

For every missing primary outcome response except list-randomized religiosity, we impute a value x if the individual is in the VHL or V group, and a value y if the household is in the HL or control group. In the most pessimistic scenario, we set x equal to the minimum observed value and y equal to the maximum observed value of the outcome in the individual's ICM base \times treatment arm cell. We then estimate treatment effects for all the primary outcomes, setting missing explanatory variables equal to their observed base \times treatment means, and compute q -values.³⁷ In the most optimistic scenario, we set x equal to the maximum observed value and y equal to the minimum observed value of the outcome in the individual's ICM base \times treatment arm cell. We also consider the scenarios $(x, y) = (\mu - 0.25\sigma, \mu + 0.25\sigma), (\mu - 0.1\sigma, \mu + 0.1\sigma), (\mu - 0.05\sigma, \mu +$

³⁶ An alternative analysis that estimates treatment effects on the responses to the direct questions finds that none of the Any V treatment effects estimated in this way are statistically distinguishable from the Any V treatment effects estimated using list randomization, although the standard errors of the list randomization estimates are large.

³⁷ In the q -value calculation, we use the p -value from the list-randomized religiosity treatment effect without imputed observations.

0.05σ), $(\mu + 0.05\sigma, \mu - 0.05\sigma)$, $(\mu + 0.1\sigma, \mu - 0.1\sigma)$, and $(\mu + 0.25\sigma, \mu - 0.25\sigma)$, where μ and σ are the mean and standard deviation within the individual's base \times treatment.

Online Appendix Table 46 shows that the most pessimistic scenario in which the Any V treatment effect on religiosity remains significantly positive is if missing VHL and V observations have religiosity 0.1 standard deviations below their base \times treatment mean and missing HL and control observations have religiosity 0.1 standard deviations above their base \times treatment mean. The most pessimistic scenario in which the Any V treatment effect on income remains significantly positive is if missing VHL and V observations have primary economic outcomes 0.05 standard deviations below their base \times treatment mean and missing HL and control observations have primary outcomes 0.05 standard deviations above their base \times treatment mean.

VIII. Discussion of Six-Month Results

A potential puzzle regarding the treatment effect on income is that we do not observe movement in other variables that would be expected to rise with income—total labor supply, consumption, food security, and assets—while perceived relative economic status decreases.

For labor supply, although there is no change in total hours, we do see a shift from agriculture to non-agricultural self-employment, livestock tending, fishing, and other employment of unclear formality (Online Appendix Table 19), which could increase income. Furthermore, we cannot observe labor effort per hour worked, which may increase with grit and which the V curriculum encourages as “a sacred ministry” that “merits heavenly reward.” In post hoc analysis, we examine two subscales within the grit index (Duckworth et al. 2007; Duckworth and Quinn 2009) and find that all of the movement in grit is coming from the “perseverance of effort” subscale ($p = 0.00003$ for Any V, $p = 0.041$ for $V = C$)—which is the sum of agreement with the statements “I am a very hard worker,” “I finish whatever I begin,” “Setbacks don’t discourage me,” and “I am diligent”—and not the “consistency of interests” subscale ($p = 0.385$ for Any V, $p = 0.655$ for $V = C$). This is consistent with the doctrine of hard work promoted by the V curriculum.³⁹

A simple explanation could account for the lack of observed movement in consumption and assets: all of the additional income was consumed. But when we test whether the Any V income and consumption effects are equal to each other, we reject this hypothesis at $p = 0.003$. This leaves

³⁹ In Online Appendix Table 23, columns 3, 5, 8 and 9 are the subcomponents that sum up to the perseverance of effort subscale, and columns 2, 4, 6 and 7 are the subcomponents that sum up to the consistency of interests subscale.

open the possibility that there was an increase in expenditures on the goods, services, and assets that we did not measure.⁴⁰

Of course, it is possible that the income result is spurious despite the multiple-testing correction. Further evidence, however, seems inconsistent with this interpretation. Among the 88% of households where the individual identified as a potential Transform invitee was the survey respondent, the “any V” effect on labor income is 236 PHP ($p = 0.0006$) for the respondent herself and 164 PHP ($p = 0.151$) summed across all other household members. Hence, the labor income effect is strongly concentrated on the Transform beneficiary.

Another possibility is that control group respondents are understating their income to the surveyor as part of a general practice of understating their resources in order to avoid having to share them with others, and the V curriculum raises reported income because it causes respondents to be more honest about their income. But this is inconsistent with the lack of a V curriculum effect on the number of meals the household gave to others in the local community in the past 30 days (Online Appendix Table 23), although it is possible that the V treatment both increases actual meals given and reduces exaggeration in the number of meals reported given.

The negative effect on perceived relative economic status could arise from participants realizing that Transform targeted those in extreme poverty. However, the HL treatment used the same targeting process, and we do not observe a significant negative effect on perceived relative economic status for the HL curriculum. Furthermore, Banerjee et al. (2015) finds that other programs that target those in extreme poverty do not generate a negative effect on perceived relative wellbeing, although their measurements occurred two years after program completion rather than six months. The V treatment did move participants into work activities where they earned more per hour, which may have increased their contact with more economically successful individuals, thus lowering their perceived relative economic standing. Alternatively, the values program, by attempting to build hope and aspiration, may make poignant to people how others are living without as much economic hardship.

⁴⁰ For example, we did not collect data on tithing. ICM reports that its pastors collect on average 570 PHP per month from their entire congregation, and the average congregation has about 25 adults. Thus, the gap between the income and consumption treatment effects is unlikely to be entirely explained by tithing.

IX. 30-Month Survey

IX.A. Survey Administration

We sent IPA surveyors to households again from November 27, 2017 to June 6, 2018 and successfully interviewed 5,878 of them (73%). Surveyors first attempted to interview the potential/actual Transform invitee, and if he or she was not available, the potential/actual invitee's spouse or partner. In 84% of successfully interviewed households, the respondent was the potential/actual invitee. Insurgent violence prevented surveyors from entering eight communities—the six affected by violence during the six-month survey plus two others. Respondents were compensated with 100 PHP. We again drop from our analysis sample the five community pairs that were not treated in accordance with their treatment assignment. Online Appendix Table 47 shows that the attrition rate does not differ significantly across the four experimental groups and that the groups are balanced on observable characteristics. Online Appendix Table 48 shows that among those successfully surveyed at six months, attrition at 30 months is significantly higher for younger and male respondents, but is not significantly related to education, household income, or religiosity measured at six months.⁴¹

IX.B. Econometric Strategy and Outcome Variables

We did not separately pre-register the analysis for the 30-month survey, but generally follow the pre-analysis plan used for the six-month survey.

Because of the trouble respondents had in the six-month survey with the three reversed questions in the intrinsic religiosity index, we replaced those reversed questions with analogous questions for which stronger agreement indicates greater religiosity.⁴² In the analysis below, we construct the intrinsic religiosity index excluding these three revised questions, but including them does not qualitatively change our results.

Based on feedback from ICM and surveyors in the field, we modified some of the other questions that comprise our outcome variables. Although we sacrificed comparability across the

⁴¹ In untabulated results, we find that relative to the relationship in the control group, the relationship between education and attrition is significantly (at the 5% level) lower in the HL group; each year of education is associated with a 0.9 percentage point higher probability of attrition in the HL group relative to the control group. But Online Appendix Table 48 shows that when testing the equality of education levels between the control and HL groups at 30 months, the p -value is 0.972, indicating that this differential attrition created minimal imbalance in practice.

⁴² The three revised questions ask about agreement with the statements, “My religious beliefs are important as well as my behavior,” “My religion affects my daily life,” and “My religion is one of the most important things in my life.”

two surveys in order to gain precision and surveying efficiency, we do not believe that any of the changes bias the treatment effect estimates by affecting some treatment cells differently in expectation than the others.

We added questions about spending on gambling and gaming, snacks, water, and electricity, which we then include in our consumption variable. Recall periods were change from one week to 30 days for the following spending categories: phone credit, transportation, clothing and shoes, soaps, cosmetics and detergents, and gifts. As in the six-month survey, we scale all reported spending to obtain monthly spending rates.

We shifted from measuring household business and non-business income in separate sections to measuring both in the same section in a uniform manner. The recall period for non-business income changed from 30 days to seven days, and household business profit was also measured over the past seven days rather than over the most recent month with normal sales. We scale all income categories up to monthly rates for the purposes of analysis. In order to reduce the frequency of income sources falling into the “other” category, we changed the set of available categories in the survey’s income classification question, and labor supply categories changed to match the income categories.⁴³ We added a final income question asking about any other income received over the last 30 days that hadn’t been mentioned yet, such as money from friends and family, remittances, additional labor income, pensions, and government transfers.

Due to budget constraints, we dropped some questions from the 30-month survey, most of which had high overlap with other questions that we retained. We dropped three sets of questions from the life satisfaction index—whether taking all things together, the respondent would say she is happy; whether the respondent experienced enjoyment/happiness/worry/sadness during a lot of the day yesterday; and whether the respondent smiled or laughed a lot yesterday. From the three mechanism measures related to locus of control, we dropped the perceived stress scale index. From the three mechanism measures related to optimism, we dropped the life orientation index and optimism index. Among secondary outcomes, we dropped the questions about open defecation

⁴³ In the six-month survey, the income categories were agricultural labor for a non-household member, salaried/formal employment outside the household, housework in an outside household, animal tending in an outside household, operating a business that is not the household’s, daily labor, and other. In the 30-month survey, the income categories were self-employed/household business/own business, wage labor, casual labor, piece worker, and other. For those who were reported to be in wage or casual labor, we asked whether the work fell into one of 11 subcategories. For those who were self-employed or working in a household business or in their own business, we asked whether the business fell into one of eight subcategories. For those who were doing piece work, we asked whether it involved food products or non-food products.

from the non-list randomized hygiene index (leaving only a question about whether animals are kept in a stable separate from the house), the question about whether the primary latrine is in the house from the six-component house index, and the number of days migrators in the household were gone in the last six months from the five-component migration and remittance index.

IX.C. Treatment Effects on Primary Outcomes, Mechanisms, and Secondary Outcomes

Table 6 shows 30-month treatment effect estimates for the primary outcomes. (Online Appendix Tables 55-81 show the treatment effect estimates on each component of the outcome variables.) There is no evidence of a positive treatment effect for any of the religiosity measures. If anything, there is some evidence in the pooled specification that receiving the V curriculum results in lower intrinsic religious orientation, although this effect is not close to statistically significant once we control for multiple testing ($q = 0.163$). Consistent with our interpretation that the six-month income effect was driven by increased religiosity, there is no positive impact of the V curriculum on monthly income at 30 months. The coefficient on Any V is very close to zero, and the 95% confidence interval does not include a 5% increase in income.⁴⁴ There is no significant impact on any of the other primary economic outcomes either.

Turning to potential mechanisms for change in Table 7, we no longer see a positive treatment effect on grit, in accordance with what we would expect if religiosity raises income primarily through increased grit. Online Appendix Table 71 shows that unlike at six months, there is no significant movement in the components of the “perseverance of effort” subscale of grit (columns 3, 5, 8, and 9).

We do see in Table 7 positive and significant (without multiple testing corrections) Any V treatment effects on optimism and social safety net strength. Given the lack of 30-month effects in primary economic outcomes, these optimism and safety net results may be best interpreted as Type I errors. Alternatively, the V curriculum may have truly improved these outcomes, but in a way

⁴⁴ The control group’s average monthly income at 30 months is 9,707 PHP, which is much higher than the 4,213 PHP we measured at six months. The control group’s average monthly consumption level also grew from 5,001 PHP to 6,378 PHP. Although some of this growth may be due to changes in the way we measured income and consumption between surveys, at least some of it is likely to reflect real economic improvements. Food security was measured in a consistent way across surveys; the fraction of control households that reported that no household member has gone hungry in the last six months rose from 82% at six months to 94% at 30 months (Online Appendix Tables 17 and 60). This improvement is probably due in large part to regression to the mean, as households were selected for being among the poorest 30 in their community before Transform. In addition, Filipino GDP per capita grew 22% from 2015 to 2018.

that had no detectable impact on material economic circumstances. Examining the subcomponents of the social safety net index, we see that all of its movement comes from an increased ability to access money from outside the household for an urgent need (Online Appendix Table 66). If such circumstances are relatively uncommon, then household consumption may be raised in expectation but not to a statistically detectable extent. Similarly, our optimism variable measures expectations of how respondents think they will be doing five years in the future. If these expectations are rational, a measurement even further in the future might find positive V treatment effects on primary economic outcomes.

At six months, we had estimated a positive V treatment effect on the Powerful Others index—the sense that God is in control of one’s life. Table 7 shows that at 30 months, this treatment effect has reversed to be negative and significant. This is in accord with the hint in Table 6 that the V treatment effect on intrinsic religious orientation is negative at 30 months. Relatedly, among secondary outcomes in Table 8, the strongest V treatment effect is an increase in stated belief in salvation by grace. Although increasing this belief is an objective of the V curriculum, we saw that at six months, increases in religiosity due to the V curriculum are associated with *decreases* in agreement with this doctrine. Therefore, the fact that we measure increases in agreement with the doctrine at 30 months suggests that religiosity may actually be lower among V recipients.

The Any V treatment effects on other secondary outcomes in Table 8 are a mixed bag. There is a significant positive effect on the non-list-randomized hygiene index (which in this survey only measured whether animals are kept in a stable separate from the house), and a significant negative effect on the number of children enrolled in school. There are marginally significant positive effects on financial inclusion and the list-randomized hygiene outcome, which measures hand washing and treatment of drinking water.

IX.D. Did the V Curriculum Effect Disappear Because Non-V Communities Were Treated?

One potential reason why the V curriculum effects disappeared at 30 months is that the pastors evangelized in the non-V communities after the first Transform implementation.⁴⁵ To test whether

⁴⁵ Consistent with this story, the directly measured religiosity variables are higher at 30 months than at six months in *all* the treatment cells. For example, among control group respondents who were interviewed at both times, intrinsic religiosity rises by 0.18 standard deviations, extrinsic religiosity rises by 0.34 standard deviations, and general religiosity rises by 0.20 standard deviations (where standard deviation is measured at six months over control group respondents who appear in both surveys). On the other hand, religiosity measured via list randomization is much lower at 30 months than at six months in all treatment cells. The proportion of 30-month respondents for whom list

this occurred, in October 2018, we surveyed 131 of the 160 pastors involved in the study. Each pastor was presented with 45 people’s names sorted alphabetically: 15 Transform invitees from the community in which the pastor had taught the Values curriculum, 15 potential Transform invitees from the community in which the pastor had identified potential invitees but which had not been selected to receive Transform, and 15 from a placebo community that the pastor had not visited with ICM.

The survey prompt read, “We have a list of people you may have interacted with in a ministry context during and after the ICM Transform values training that you led from February to June 2015, three years ago. We believe that some of these people participated in your Transform program, and some did not.” The survey asked whether each person in the list had ever participated in a Transform Values program with the pastor, and if the person had, whether he or she participated in the Transform program in 2015 or after 2015. The survey also asked whether the pastor had interacted with the listed person in *any* ministry context (defined as “an occasion where spiritual matters were discussed, or an event sponsored by a religious ministry”), and if yes, whether that interaction happened in 2015 or after 2015.

We find that pastors report that 79% of actual Transform invitees, 58% of control group members, and 25% of placebo group members participated in the Transform Values program.⁴⁶ Similarly, pastors report having interacted in a ministry context with 65% of actual Transform participants, 46% of control group members, and 19% of placebo group members (some pastors did not classify Transform as a “ministry context”).

Importantly, there is little evidence of pastoral interactions that would have contaminated the treatment effect estimates in the 30-month survey but not the six-month survey, which completed data collection in January 2016. Conditional on believing that an individual participated in Transform, the pastor reported that that participation happened in 2015 for 99% of the individuals. Among those whom the pastor reported interacting with in a ministry context, only 3% of those

randomization was expected to successfully anonymize their response about the targeted sensitive statement (see footnote 19) is similar to the six-month survey’s proportions: 76% for commitment to Jesus, 82% for reading the Bible, 85% for water treatment, 84% for hand washing, and 90% for domestic abuse.

⁴⁶ There is a slightly negative and insignificant ($p = 0.447, 0.425, \text{ and } 0.864$ for intrinsic, extrinsic, and general religiosity, respectively) relationship between pastor-level religiosity treatment effects at six months and the difference between the fraction of Any V and No V group members the pastor reports to have participated in Transform. Therefore, we believe that the high fraction of control group members reported to have been in Transform is due to recall error rather than non-compliance with the treatment assignment.

interactions happened exclusively after 2015, 75% of them happened exclusively in 2015 or earlier, and the remaining 22% happened in both periods.

Although it is possible in principle that Transform participants evangelizing the control communities is responsible for the erosion of the estimated V curriculum effect, we believe that this is unlikely given the geographic distance between the communities and the fact that any evangelization effort in a non-V community would have been dispersed among both those who were identified as potential Transform invitees (and hence were in our survey) and those who were not.

X. Conclusion

Our work demonstrates that a randomized controlled trial is a viable tool for shifting attitudes towards and practices of religion in order to study the effect of religiosity on social and economic outcomes. As with all program evaluations, our results are, strictly speaking, specific to the program and setting we study. Having said that, Transform’s curriculum and dissemination method are similar to efforts by many religious organizations around the world, and evangelization of Catholics by evangelical Protestants is a widespread phenomenon (Pew Research Center 2014).

We find that increasing religiosity increases income while decreasing perceived relative economic status. The effects of Transform on religiosity are short-lived, however, and are insignificant 30 months after the program ends. Correspondingly, Transform’s positive effect on income and negative effect on perceived relative economic status also dissipate at 30 months. Although church-based programs like Transform may represent a method of increasing non-cognitive skills and reducing poverty in the short run among adults in developing countries, more work is required to understand whether the effects can persist and if not, why not.

References

- Abadie, Alberto, Matthew M. Chingos, and Martin R. West. 2018. “Endogenous stratification in randomized experiments.” *Review of Economics and Statistics* 100: 567-580.
- Anderson, Michael. 2008. “Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects.” *Journal of the American Statistical Association* 103 (484): 1481–95. <https://doi.org/10.1198/016214508000000841>.
- Banerjee, Abhijit, Esther Duflo, Nathanael Goldberg, Dean Karlan, Robert Osei, William Parienté, Jeremy Shapiro, Bram Thuysbaert, and Christopher Udry. 2015. “A

- Multifaceted Program Causes Lasting Progress for the Very Poor: Evidence from Six Countries.” *Science* 348 (6236): 1260799. <https://doi.org/10.1126/science.1260799>.
- Barro, Robert J., and Rachel M. McCleary. 2003. “Religion and Economic Growth across Countries.” *American Sociological Review* 68 (5): 760–81. <https://doi.org/10.2307/1519761>.
- Basten, Cristoph, and Frank Betz. 2013. “Beyond work ethic: Religion, individual, and political preferences.” *American Economic Journal: Economic Policy* 5, 67-91.
- Becker, Sascha O., and Ludger Woessmann. 2009. “Was Weber wrong? A human capital theory of Protestant economic history.” *Quarterly Journal of Economics* 124: 531-596.
- Bell, Matthew. 2013. “Alpha: The Slickest, Richest, Fastest-Growing Division of the Church of England.” *The Spectator*, November 30, 2013. <https://www.spectator.co.uk/2013/11/alpha-rising/>.
- Benjamin, Daniel J., James J. Choi, and Geoffrey Fisher. 2016. “Religious Identity and Economic Behavior.” *Review of Economics and Statistics* 98 (4): 617–37. https://doi.org/10.1162/REST_a_00586.
- Benjamini, Yoav, and Yosef Hochberg. 1995. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.” *Journal of the Royal Statistical Society. Series B (Methodological)*, 289–300.
- Benjamini, Yoav, Abba M. Krieger, and Daniel Yekutieli. 2006. “Adaptive linear step-up procedures that control the false discovery rate.” *Biometrika* 93, 491-507.
- Blattman, Christopher, Julian Jamison, and Margaret Sheridan. 2015. “Reducing Crime and Violence: Experimental Evidence on Adult Noncognitive Investments in Liberia.”
- Bottan, Nicolas L., and Ricardo Perez-Truglia. 2015. “Losing My Religion: The Effects of Religious Scandals on Religious Participation and Charitable Giving.” *Journal of Public Economics* 129 (September): 106–19. <https://doi.org/10.1016/j.jpubeco.2015.07.008>.
- Cantoni, Davide. 2015. “The economic effects of the Protestant Reformation: Testing the Weber hypothesis in German lands.” *Journal of the European Economic Association* 13: 561-598.
- Chen, Daniel L. 2010. “Club goods and group identity: Evidence from Islamic resurgence during the Indonesian financial crisis.” *Journal of Political Economy* 118(2): 300-354.
- Coffman, Katherine B., Lucas C. Coffman, and Keith M. Marzilli Ericson. 2016. “The size of the LGBT population and the magnitude of antigay sentiment are substantially underestimated.” *Management Science* 63: 3147-3529.
- Chuang, Erica, Pascaline Dupas, Elise Huillery, and Juliette Seban. 2019. “Sex, Lies, and Measurement: Do Indirect Response Survey Methods Work?” *Working Paper*, January.
- Clingingsmith, David, Asim Ijaz Khwaja, and Michael Kremer. 2009. “Estimating the Impact of the Hajj: Religion and Tolerance in Islam’s Global Gathering.” *Quarterly Journal of Economics* 124 (3): 1133–70. <https://doi.org/10.1162/qjec.2009.124.3.1133>.
- Droitcour, Judith, Rachel A. Caspar, Michael L. Hubbard, Teresa L. Parsley, Wendy Visscher, and Trena M. Ezzati. 2011. “The Item Count Technique as a Method of Indirect Questioning: A Review of Its Development and a Case Study Application.” In *Wiley Series in Probability and Statistics*, edited by Paul P. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, and Seymour Sudman, 185–210. Hoboken, NJ, USA: John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118150382.ch11>.

- Duckworth, Angela Lee, Christopher Peterson, Michael D. Matthews, and Dennis R. Kelly. 2007. "Grit: Perseverance and Passion for Long-Term Goals." *Journal of Personality and Social Psychology* 92 (6): 1087–1101. <https://doi.org/10.1037/0022-3514.92.6.1087>.
- Duckworth, Angela Lee, and Patrick D. Quinn. 2009. "Development and Validation of the Short Grit Scale (Grit–S)." *Journal of Personality Assessment* 91 (2): 166–74. <https://doi.org/10.1080/00223890802634290>.
- Efron, Bradley. 2010. *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*. Cambridge, UK: Cambridge University Press.
- Ellison, C. G. 1991. "Religious Involvement and Subjective Well-Being." *Journal of Health and Social Behavior* 32 (1): 80–99.
- Freeman, Richard B. 1986. "Who Escapes? The Relation of Churchgoing and Other Background Factors to the Socioeconomic Performance of Black Male Youth from Inner-City Tracts." In *The Black Youth Employment Crisis*, edited by Richard B. Freeman and Harry J. Holzer, 353–76. Chicago: University of Chicago Press.
- Goeman, Jelle J., and Aldo Solari, 2014. "Multiple hypothesis testing in genomics." *Statistics in Medicine* 33, 1946–1978.
- Gorsuch, Richard L., and Susan E. McPherson. 1989. "Intrinsic/Extrinsic Measurement: I/E-Revised and Single-Item Scales." *Journal for the Scientific Study of Religion* 28 (3): 348. <https://doi.org/10.2307/1386745>.
- Gruber, Jonathan. 2005. "Religious Market Structure, Religious Participation, and Outcomes: Is Religion Good for You?" *The B.E. Journal of Economic Analysis & Policy* 5 (1). <https://doi.org/10.1515/1538-0637.1454>.
- Gruber, Jonathan, and Daniel Hungerman. 2008. "The Church vs the Mall: What Happens When Religion Faces Increased Secular Competition?" *Quarterly Journal of Economics* 123 (2): 831–62. <https://doi.org/10.3386/w12410>.
- Hackett, Conrad, and Brian J. Grim. 2011. "Global Christianity – A Report on the Size and Distribution of the World’s Christian Population." *Pew Research Center*, December. <http://www.pewforum.org/2011/12/19/global-christianity-exec/>.
- Hilary, Gilles, and Kai Wai Hui. 2009. "Does Religion Matter in Corporate Decision Making in America?" *Journal of Financial Economics* 93 (3): 455–73. <https://doi.org/10.1016/j.jfineco.2008.10.001>.
- Horton, John J., David G. Rand, and Richard J. Zeckhauser. 2011. "The Online Laboratory: Conducting Experiments in a Real Labor Market." *Experimental Economics* 14 (3): 399–425. <https://doi.org/10.1007/s10683-011-9273-9>.
- Iannaccone, Laurence R. 1998. "Introduction to the Economics of Religion." *Journal of Economic Literature* 36 (3): 1465–95.
- Iyer, Sriya. 2016. "The New Economics of Religion." *Journal of Economic Literature* 54 (2): 395–441. <https://doi.org/10.1257/jel.54.2.395>.
- Johnson, Byron R., Ralph Brett Tompkins, and Derek Webb. 2008. "Objective Hope: Assessing the Effectiveness of Faith-Based Organizations: A Review of the Literature." *Baylor Institute for Studies of Religion Report*.
- Karlan, Dean, and Jonathan Zinman. 2012. "List Randomization for Sensitive Behavior: An Application for Measuring Use of Loan Proceeds." *Journal of Development Economics* 98 (1): 71–75. <https://doi.org/10.1016/j.jdeveco.2011.08.006>.
- Kautz, Tim, James Heckman, Ron Diris, Bas ter Weel, and Lex Borghans. 2014. "Fostering and Measuring Skills: Improving Cognitive and Non-Cognitive Skills to Promote Lifetime

- Success.” w20749. Cambridge, MA: National Bureau of Economic Research.
<https://doi.org/10.3386/w20749>.
- Kirkpatrick, Lee A., and Ralph W. Hood. 1990. “Intrinsic-Extrinsic Religious Orientation: The Boon or Bane of Contemporary Psychology of Religion?” *Journal for the Scientific Study of Religion* 29 (4): 442. <https://doi.org/10.2307/1387311>.
- Levenson, Hanna. 1981. “Differentiating Among Internality, Powerful Others, and Chance.” In *Research with the Locus of Control Construct*, 15–63. Elsevier.
<https://doi.org/10.1016/B978-0-12-443201-7.50006-3>.
- Mazar, Nina, On Amir, and Dan Ariely. 2008. “The Dishonesty of Honest People: A Theory of Self-Concept Maintenance.” *Journal of Marketing Research* 45 (6): 633–44.
<https://doi.org/10.1509/jmkr.45.6.633>.
- Pew Research Center. 2014. “Religion in Latin America.” *Polling and Analysis* (blog). November 13, 2014. <http://www.pewforum.org/2014/11/13/religion-in-latin-america/>.
- Pew Research Center. 2015. “The Future of World Religions: Population Growth Projections, 2010-2050.”
- Romano, Joseph P., Azeem M. Shaikh, and Michael Wolf. 2008. “Control of the false discovery rate under dependence using the bootstrap and subsampling.” *Test* 17: 417-442.
- Samaritan’s Purse. 2017. “Along the Samaritan Road: 2016 Annual Report.” Samaritan’s Purse. https://s3.amazonaws.com/static.samaritanpurse.org/pdfs/ANNUAL_REPORT_web_download.pdf.
- Scheier, Michael F., Charles S. Carver, and Michael W. Bridges. 1994. “Distinguishing Optimism from Neuroticism (and Trait Anxiety, Self-Mastery, and Self-Esteem): A Reevaluation of the Life Orientation Test.” *Journal of Personality and Social Psychology* 67 (6): 1063–78. <https://doi.org/10.1037/0022-3514.67.6.1063>.
- Shariff, Azim F., and Ara Norenzayan. 2007. “God Is Watching You: Priming God Concepts Increases Prosocial Behavior in an Anonymous Economic Game.” *Psychological Science* 18 (9): 803–9. <https://doi.org/10.1111/j.1467-9280.2007.01983.x>.
- Shariff, Azim F., Aiyana K. Willard, Teresa Andersen, and Ara Norenzayan. 2016. “Religious Priming: A Meta-Analysis With a Focus on Prosociality.” *Personality and Social Psychology Review* 20 (1): 27–48. <https://doi.org/10.1177/1088868314568811>.
- Spenkuch, Jörg L., 2017. “Religion and work: Micro evidence from contemporary Germany.” *Journal of Economic Behavior & Organization* 135, 193-214.
- Storey, John D., Jonathan E. Taylor, and David Siegmund. 2004. “Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rate: a unified approach.” *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* 66: 187-205.
- Tangney, June P., Roy F. Baumeister, and Angie Luzio Boone. 2004. “High Self-Control Predicts Good Adjustment, Less Pathology, Better Grades, and Interpersonal Success.” *Journal of Personality* 72 (2): 271–324. <https://doi.org/10.1111/j.0022-3506.2004.00263.x>.
- Weber, Max. (1905) 1958. *The Protestant Ethic and the Spirit of Capitalism*. New York: Scribner.
- Woodberry, Robert D. 2012. “The missionary roots of liberal democracy.” *American Political Science Review* 106, 244-274.

Table 1. Primary outcomes, six-month survey

Panels A and B show treatment effect estimates relative to control. In Panel A, “Any V” refers to the “Values only” and “Values, Health, and Livelihood” treatment groups, and “Any HL” refers to the “Health and Livelihood only” and “Values, Health, and Livelihood” treatment groups. Dependent variables are indicated in the column title. See Appendix for details on variable construction. Standard errors clustered by community are in parentheses. In Panel B, we do not show VHL = HL q -values and FWER-adjusted p -values for primary economic outcomes because there is no significant first-stage VHL versus HL difference in religiosity.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Primary religious outcomes				Primary economic outcomes					
	Religion intrinsic index	Religion extrinsic index	General religion index	Religion, list- randomized	Monthly consumption (PHP)	Food security index	Monthly income (PHP)	Adult weekly labor supply (hours)	Life satisfaction index	Perceived relative econ. status
Panel A: Pooled specification										
Any V	0.102 (0.024)	0.130 (0.024)	0.077 (0.023)	0.048 (0.037)	-1.1 (100.4)	0.010 (0.023)	386.1 (126.8)	0.9 (1.1)	0.019 (0.022)	-0.113 (0.047)
Any HL	0.014 (0.024)	-0.021 (0.024)	0.001 (0.023)	-0.028 (0.038)	-103.0 (93.3)	-0.044 (0.023)	131.2 (126.3)	-1.8 (1.1)	-0.010 (0.022)	-0.040 (0.047)
q -value, Any V	0.000	0.000	0.001	0.197	0.991	0.778	0.015	0.595	0.595	0.050
FWER p -value, Any V	0.000	0.000	0.002	0.197	1.000	1.000	0.015	1.000	1.000	0.083
Panel B: Disaggregated specification										
VHL	0.115 (0.034)	0.109 (0.037)	0.077 (0.031)	0.020 (0.054)	-102.2 (159.5)	-0.033 (0.037)	524.4 (175.0)	-0.9 (1.4)	0.009 (0.028)	-0.151 (0.067)
HL	0.047 (0.055)	0.073 (0.065)	-0.029 (0.054)	-0.002 (0.055)	-314.3 (203.0)	-0.050 (0.051)	287.9 (278.4)	-0.1 (2.4)	-0.031 (0.056)	-0.073 (0.112)
V	0.123 (0.050)	0.204 (0.064)	0.052 (0.051)	0.070 (0.057)	-167.4 (209.5)	-0.007 (0.050)	574.2 (285.4)	3.0 (2.3)	-0.018 (0.047)	-0.133 (0.119)
q -value, VHL = HL	0.393	0.653	0.146	0.653	--	--	--	--	--	--
q -value, V = C	0.058	0.013	0.416	0.393	0.637	0.885	0.271	0.529	0.850	0.529
FWER p -value, VHL = HL	1.000	1.000	0.330	1.000	--	--	--	--	--	--
FWER p -value, V = C	0.102	0.013	1.000	1.000	1.000	1.000	0.271	1.000	1.000	1.000
Panel C: Summary information										
Control mean	0	0	0	0.606	5,001	0	4,213	79.6	0	3.242
Control standard deviation	1	1	1	--	4,720	1	5,567	57.7	1	2.256
# observations in VHL	1,578	1,578	1,578	1,578	1,578	1,526	1,452	1,452	1,578	1,576
# observations in HL	1,549	1,549	1,549	1,549	1,549	1,521	1,440	1,439	1,549	1,548
# observations in V	1,550	1,550	1,550	1,550	1,550	1,517	1,435	1,434	1,550	1,547
# observations in C	1,599	1,599	1,599	1,599	1,599	1,567	1,490	1,490	1,599	1,596

Table 2. Mechanisms, six-month survey

Panels A and B show treatment effect estimates relative to control. In Panel A, “Any V” refers to the “Values only” and “Values, Health, and Livelihood” treatment groups, and “Any HL” refers to the “Health and Livelihood only” and “Values, Health, and Livelihood” treatment groups. Dependent variables are indicated in the column title. Indexes have been coded so that more positive numbers are better. See Appendix for details on variable construction. Standard errors clustered by community are in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Social capital			Locus of control			Optimism				
	Trust index	Social safety net index	Community activities index	Perceived stress scale index	Powerful others index	Locus of control index	Life orientation index	Expectations index	Optimism index	Grit index	Self-control index
Panel A: Pooled specification											
Any V	0.004 (0.022)	0.026 (0.024)	0.005 (0.025)	-0.011 (0.020)	0.093 (0.027)	-0.035 (0.020)	-0.050 (0.027)	-0.037 (0.025)	0.053 (0.024)	0.041 (0.022)	-0.034 (0.021)
Any HL	-0.023 (0.022)	-0.027 (0.024)	0.041 (0.025)	-0.018 (0.021)	0.044 (0.027)	-0.000 (0.020)	0.016 (0.027)	-0.016 (0.025)	-0.024 (0.024)	0.017 (0.022)	0.006 (0.020)
<i>p</i> -value, Any V	0.865	0.282	0.851	0.596	0.001	0.075	0.065	0.133	0.029	0.065	0.095
Panel B: Disaggregated specification											
VHL	-0.019 (0.032)	0.000 (0.032)	0.045 (0.034)	-0.026 (0.026)	0.135 (0.038)	-0.035 (0.029)	-0.034 (0.037)	-0.055 (0.032)	0.030 (0.032)	0.056 (0.029)	-0.027 (0.025)
HL	-0.023 (0.043)	-0.076 (0.048)	0.019 (0.058)	-0.009 (0.044)	0.031 (0.060)	-0.064 (0.057)	-0.046 (0.068)	-0.014 (0.056)	-0.007 (0.061)	0.030 (0.058)	0.039 (0.047)
V	-0.018 (0.046)	-0.023 (0.048)	-0.011 (0.059)	-0.007 (0.043)	0.073 (0.059)	-0.085 (0.050)	-0.103 (0.069)	-0.054 (0.057)	0.069 (0.066)	0.041 (0.058)	-0.001 (0.050)
<i>p</i> -value for VHL = HL	0.927	0.140	0.655	0.684	0.085	0.605	0.862	0.468	0.541	0.671	0.155
<i>p</i> -value, V = C	0.704	0.631	0.857	0.876	0.222	0.090	0.132	0.344	0.298	0.484	0.980
Panel C: Summary information											
Control mean	0	0	0	0	0	0	0	0	0	0	0
Control standard deviation	1	1	1	1	1	1	1	1	1	1	1
# observations in VHL	1,578	1,578	1,561	1,577	1,578	1,578	1,578	1,542	1,578	1,578	1,578
# observations in HL	1,549	1,549	1,542	1,549	1,549	1,549	1,549	1,508	1,549	1,549	1,549
# observations in V	1,550	1,550	1,534	1,549	1,550	1,550	1,550	1,518	1,550	1,550	1,550
# observations in C	1,599	1,599	1,592	1,599	1,599	1,599	1,599	1,567	1,599	1,599	1,599

Table 3. Secondary outcomes, six-month survey

Panels A and B show treatment effect estimates relative to control. In Panel A, “Any V” refers to the “Values only” and “Values, Health, and Livelihood” treatment groups, and “Any HL” refers to the “Health and Livelihood only” and “Values, Health, and Livelihood” treatment groups. Dependent variables are indicated in the column title. Indexes have been coded so that more positive numbers are better. See Appendix for details on variable construction. Standard errors clustered by community are in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Salvation by grace belief index	Assets index	Financial inclusion index	Health index	Hygiene index, non-list- random.	Hygiene, list random.	House index	Migration and remittance index	No discord index	No domestic violence, list-rand.	Child labor supply (hours)	# children enrolled in school
Panel A: Pooled specification												
Any V	-0.036 (0.020)	-0.027 (0.021)	0.020 (0.024)	0.000 (0.020)	0.092 (0.024)	0.043 (0.033)	0.030 (0.025)	0.027 (0.019)	-0.034 (0.024)	-0.072 (0.040)	0.2 (0.2)	-0.02 (0.02)
Any HL	-0.005 (0.020)	-0.025 (0.021)	0.157 (0.025)	0.015 (0.020)	0.030 (0.024)	0.066 (0.033)	0.007 (0.025)	-0.015 (0.019)	-0.029 (0.024)	-0.048 (0.040)	0.0 (0.2)	-0.01 (0.02)
<i>p</i> -value, Any V	0.079	0.211	0.396	0.985	0.000	0.191	0.239	0.153	0.164	0.078	0.256	0.349
Panel B: Disaggregated specification												
VHL	-0.040 (0.026)	-0.050 (0.031)	0.179 (0.038)	0.015 (0.028)	0.121 (0.034)	0.108 (0.049)	0.036 (0.036)	0.012 (0.031)	-0.063 (0.036)	-0.118 (0.055)	0.3 (0.3)	-0.03 (0.02)
HL	-0.021 (0.045)	0.014 (0.057)	0.124 (0.048)	-0.027 (0.042)	0.136 (0.070)	0.121 (0.043)	0.045 (0.059)	-0.083 (0.038)	-0.036 (0.052)	-0.081 (0.058)	-0.1 (0.4)	-0.01 (0.04)
V	-0.061 (0.041)	0.008 (0.060)	-0.010 (0.044)	-0.044 (0.041)	0.208 (0.067)	0.105 (0.045)	0.068 (0.060)	-0.039 (0.039)	-0.049 (0.049)	-0.120 (0.061)	0.1 (0.4)	-0.02 (0.04)
<i>p</i> -value for VHL = HL	0.696	0.265	0.297	0.334	0.836	0.779	0.879	0.017	0.617	0.509	0.404	0.687
<i>p</i> -value, V = C	0.143	0.899	0.811	0.285	0.002	0.020	0.258	0.317	0.326	0.050	0.775	0.618
Panel C: Summary information												
Control mean	0	0	0	0	0	0.558	0	0	0	0.903	1.6	1.67
Control std. deviation	1	1	1	1	1	--	1	1	1	0.037	12.3	1.37
# observations in VHL	1,578	1,578	1,578	1,578	1,578	1,578	1,578	1,578	1,267	1,579	1,452	1,578
# observations in HL	1,549	1,549	1,549	1,549	1,549	1,549	1,549	1,549	1,297	1,550	1,439	1,549
# observations in V	1,550	1,550	1,550	1,550	1,550	1,550	1,550	1,550	1,263	1,551	1,434	1,550
# observations in C	1,599	1,599	1,599	1,599	1,599	1,599	1,599	1,599	1,331	1,600	1,490	1,599

Table 4. Treatment effects by predicted religiosity without V curriculum, six-month survey
Panel A shows “Any V” treatment effects on the variable in the left column, separately for each tercile of predicted religiosity index value in the absence of the V curriculum. The religiosity index is the normalized sum of the intrinsic, extrinsic, and general religion indexes. Panel B shows treatment effects estimated by comparing the V group to the control group. The regression specifications are as described in Table 1. See Appendix for details on variable construction. Standard errors clustered by community are in parentheses.

	Predicted religiosity			<i>p</i> -value of joint equality across terciles
	Low	Medium	High	
Panel A: Pooled specification				
Religiosity index	0.152 (0.040)	0.146 (0.039)	0.074 (0.035)	0.101
Religion, list-randomized	-0.041 (0.065)	0.054 (0.070)	0.191 (0.068)	0.025
Monthly income	421.8 (226.3)	318.5 (188.4)	407.3 (255.9)	0.912
Perceived relative econ. status	-0.164 (0.080)	-0.092 (0.090)	0.002 (0.085)	0.291
Panel B: Disaggregated specification, V vs. control				
Religiosity index	0.252 (0.078)	0.160 (0.070)	0.027 (0.071)	0.021
Religion, list-randomized	0.008 (0.092)	0.034 (0.098)	0.212 (0.090)	0.200
Monthly income	880.6 (420.2)	471.7 (298.8)	408.3 (518.0)	0.607
Perceived relative econ. status	-0.050 (0.171)	-0.242 (0.188)	-0.054 (0.168)	0.624
Panel C: Summary information				
Mean religiosity index value in control and HL groups	-0.149	-0.049	0.221	

Table 5. Test for existence of social desirability bias in responses, six-month survey

This table shows coefficients for regressions where the dependent variable is the number of statements reported to be true in a list that includes the sensitive statement in the column label. For respondents who did not actually receive that statement in their list, the dependent variable is the number of statements they reported to be true plus an indicator for whether we impute that the sensitive statement is true for them based upon their response to a direct question about it. The key explanatory variables are a dummy for having actually received the sensitive statement in the list, treatment dummies, and interactions between sensitive statement receipt and the treatment dummies. Coefficients from additional control variables that are identical to those used for the regression in Panel A of Table 1 are not shown.

	I have made a personal commitment to Jesus Christ that is still important to me	I have read or listened to the Bible in the past week	I wash my hands after going to the bathroom	Someone in my household is experiencing physical abuse (higher = less abuse)
Panel A: Presence of social desirability bias				
Received sensitive statement	-0.262 (0.023)	-0.217 (0.024)	-0.228 (0.021)	-0.093 (0.045)
Constant	3.609 (0.103)	2.237 (0.109)	2.615 (0.179)	-1.577 (0.943)
Observations	6,276	6,276	6,262	1,447
Panel B: Interaction of social desirability bias with treatment				
Received sensitive statement	-0.286 (0.045)	-0.197 (0.041)	-0.261 (0.034)	-0.097 (0.080)
Sensitive statement × Any V	0.042 (0.046)	0.021 (0.047)	0.031 (0.041)	0.020 (0.091)
Sensitive statement × Any HL	0.007 (0.047)	-0.059 (0.048)	0.037 (0.042)	-0.012 (0.091)
Any V	0.024 (0.029)	0.074 (0.031)	0.013 (0.025)	-0.101 (0.061)
Any HL	-0.001 (0.029)	0.037 (0.030)	0.006 (0.025)	0.032 (0.060)
Constant	3.608 (0.106)	2.200 (0.110)	2.605 (0.181)	-1.584 (0.909)
Observations	6,276	6,276	6,262	1,447

Table 6. Primary outcomes, 30-month survey

Panels A and B show treatment effect estimates relative to control. In Panel A, “Any V” refers to the “Values only” and “Values, Health, and Livelihood” treatment groups, and “Any HL” refers to the “Health and Livelihood only” and “Values, Health, and Livelihood” treatment groups. Dependent variables are indicated in the column title. See Appendix and Section IX.B for details on variable construction. Standard errors clustered by community are in parentheses. In Panel B, we do not show VHL = HL q -values and FWER-adjusted p -values for primary economic outcomes because there is no significant first-stage VHL versus HL difference in religiosity in the six-month survey.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Primary religious outcomes				Primary economic outcomes					
	Religion intrinsic index	Religion extrinsic index	General religion index	Religion, list- randomized	Monthly consumption (PHP)	Food security index	Monthly income (PHP)	Adult weekly labor supply (hours)	Life satisfaction index	Perceived relative econ. status
Panel A: Pooled specification										
Any V	-0.052 (0.025)	-0.008 (0.026)	-0.023 (0.025)	0.016 (0.036)	131.9 (88.2)	-0.014 (0.024)	21.0 (207.0)	-0.8 (1.1)	-0.004 (0.022)	0.097 (0.044)
Any HL	0.035 (0.025)	0.018 (0.026)	-0.047 (0.025)	0.018 (0.036)	-77.5 (88.5)	-0.050 (0.024)	196.7 (209.9)	0.8 (1.1)	0.036 (0.022)	0.019 (0.044)
q -value, Any V	0.163	0.741	0.726	0.741	0.408	0.836	0.919	0.836	0.919	0.168
FWER p -value, Any V	0.163	1.000	1.000	1.000	0.680	1.000	1.000	1.000	1.000	0.168
Panel B: Disaggregated specification										
VHL	-0.013 (0.030)	0.012 (0.033)	-0.069 (0.031)	0.032 (0.048)	56.2 (115.5)	-0.065 (0.037)	213.4 (315.8)	0.1 (1.6)	0.032 (0.026)	0.120 (0.062)
HL	0.013 (0.067)	0.035 (0.074)	-0.027 (0.063)	0.056 (0.053)	254.6 (195.5)	-0.012 (0.052)	666.0 (433.9)	4.3 (2.1)	0.077 (0.052)	0.234 (0.109)
V	-0.070 (0.066)	0.007 (0.071)	-0.004 (0.065)	0.047 (0.051)	481.4 (207.2)	0.024 (0.050)	517.5 (484.1)	2.8 (2.2)	0.044 (0.054)	0.340 (0.114)
q -value, VHL = HL	0.954	0.954	0.954	0.954	--	--	--	--	--	--
q -value, V = C	1.000	1.000	1.000	1.000	0.125	0.625	0.446	0.446	0.461	0.038
FWER p -value, VHL = HL	0.954	0.954	0.954	0.954	--	--	--	--	--	--
FWER p -value, V = C	1.000	1.000	1.000	1.000	0.229	1.000	1.000	1.000	1.000	0.038
Panel C: Summary information										
Control mean	0	0	0	0.378	6,378	0	9,707	67.8	0	3.662
Control standard deviation	1	1	1	0.032	3,789	1	11,600	52.5	1	2.050
# observations in VHL	1,441	1,441	1,441	1,441	1,441	1,440	1,441	1,441	1,441	1,440
# observations in HL	1,366	1,366	1,366	1,366	1,366	1,365	1,366	1,366	1,366	1,365
# observations in V	1,389	1,389	1,389	1,389	1,389	1,389	1,389	1,389	1,389	1,389
# observations in C	1,481	1,481	1,481	1,481	1,481	1,479	1,481	1,481	1,481	1,480

Table 7. Mechanisms, 30-month survey

Panels A and B show treatment effect estimates relative to control. In Panel A, “Any V” refers to the “Values only” and “Values, Health, and Livelihood” treatment groups, and “Any HL” refers to the “Health and Livelihood only” and “Values, Health, and Livelihood” treatment groups. Dependent variables are indicated in the column title. Indexes have been coded so that more positive numbers are better. See Appendix and Section IX.B for details on variable construction. Standard errors clustered by community are in parentheses.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Social capital		Locus of control		Optimism			
	Trust index	Social safety net index	Community activities index	Powerful others index	Locus of control index	Expectations index	Grit index	Self-control index
Panel A: Pooled specification								
Any V	-0.021 (0.023)	0.038 (0.019)	-0.023 (0.023)	-0.047 (0.024)	-0.000 (0.021)	0.047 (0.022)	0.006 (0.020)	-0.014 (0.018)
Any HL	-0.027 (0.022)	0.032 (0.019)	-0.012 (0.023)	-0.004 (0.024)	0.007 (0.021)	0.016 (0.022)	-0.037 (0.020)	-0.017 (0.018)
<i>p</i> -value, Any V	0.354	0.046	0.324	0.047	0.989	0.034	0.761	0.458
Panel B: Disaggregated specification								
VHL	-0.047 (0.030)	0.068 (0.026)	-0.035 (0.031)	-0.050 (0.031)	0.008 (0.030)	0.062 (0.028)	-0.031 (0.028)	-0.031 (0.026)
HL	-0.083 (0.057)	0.035 (0.047)	-0.031 (0.053)	0.027 (0.069)	0.010 (0.058)	0.073 (0.049)	-0.053 (0.059)	-0.107 (0.063)
V	-0.067 (0.048)	0.054 (0.047)	-0.042 (0.055)	-0.017 (0.069)	0.013 (0.056)	0.116 (0.053)	-0.006 (0.063)	-0.101 (0.065)
<i>p</i> -value, VHL = HL	0.528	0.495	0.939	0.247	0.979	0.826	0.694	0.224
<i>p</i> -value, V = C	0.167	0.247	0.437	0.800	0.810	0.029	0.921	0.119
Panel C: Summary information								
Control mean	0	0	0	0	0	0	0	0
Control standard deviation	1	1	1	1	1	1	1	1
# observations in VHL	1,441	1,441	1,441	1,441	1,441	1,440	1,441	1,441
# observations in HL	1,366	1,366	1,366	1,366	1,366	1,365	1,366	1,366
# observations in V	1,389	1,389	1,389	1,389	1,389	1,388	1,389	1,389
# observations in C	1,481	1,481	1,481	1,481	1,481	1,479	1,481	1,481

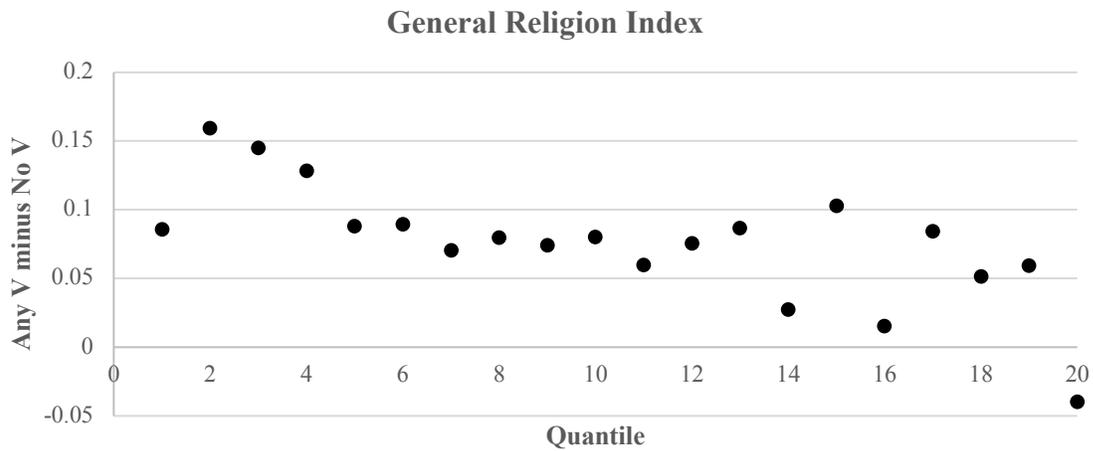
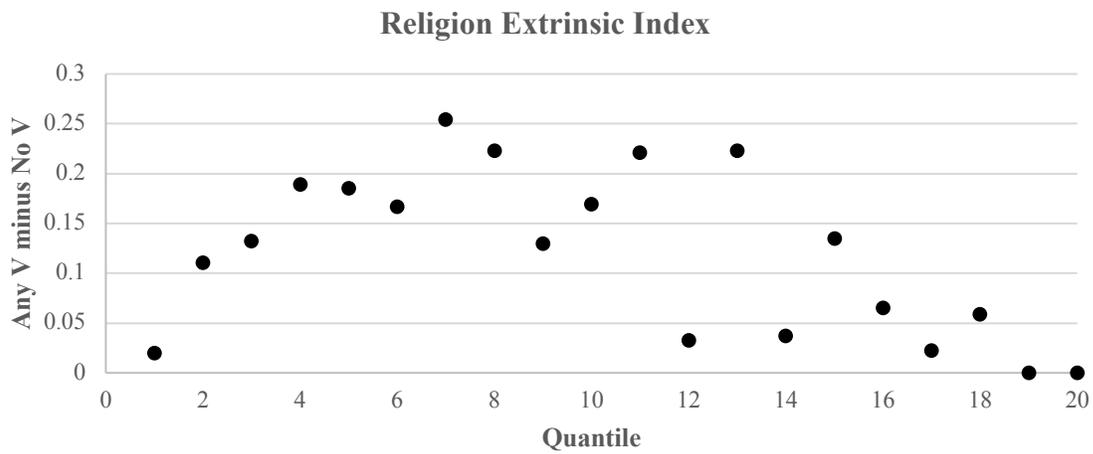
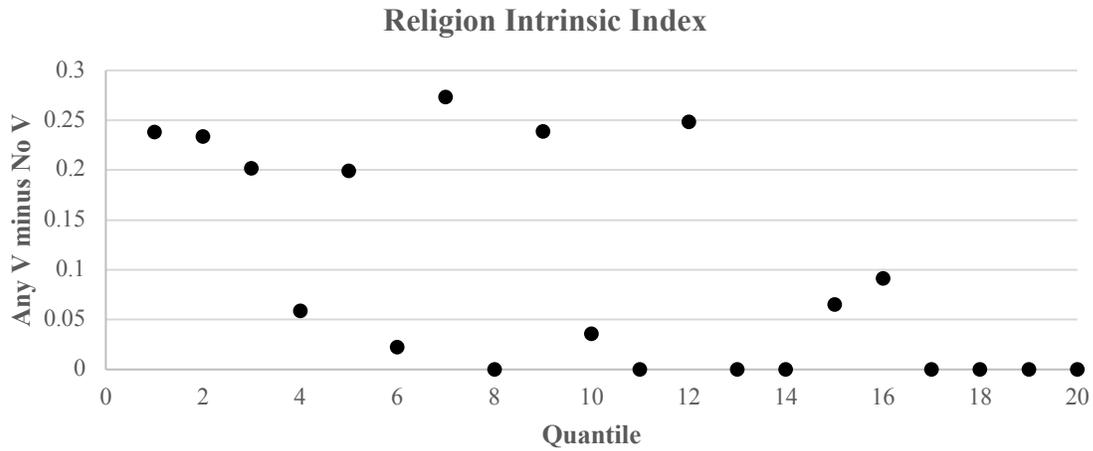


Figure 1. Average difference in religiosity measures between VHL and V groups versus HL and control groups within each 5-percentile-wide bucket, 6 months after treatment

Appendix Table 1. Six-Month Survey Outcome Variable Construction

Unless indicated otherwise in the table, the variable listed in the first column is created by summing its components listed in the second column. Some components are made up of sub-components, which are shown to the right of the components.

Variable	Components	Sub-components (if any)	Details	Possible answers
Panel A: Primary religious outcomes				
Religion intrinsic index	I enjoy thinking about my religion		From Gorsuch and McPherson (1989).	1 Strongly disagree - 5 Strongly agree
	It is important to me to spend time in private thought and prayer		Index formed by adding together responses without first normalizing.	1 Strongly disagree - 5 Strongly agree
	I have often had a strong sense of God's presence			1 Strongly disagree - 5 Strongly agree
	I try hard to live all my life according to my religious beliefs			1 Strongly disagree - 5 Strongly agree
	My whole approach to life is based on religion			1 Strongly disagree - 5 Strongly agree
	Although I am religious, I don't let it affect my daily life		This question not used in our main analysis	1 Strongly agree - 5 Strongly disagree
	It doesn't much matter what I believe so long as I am good		This question not used in our main analysis	1 Strongly agree - 5 Strongly disagree
	Although I believe in my religion, many other things are more important in life		This question not used in our main analysis	1 Strongly agree - 5 Strongly disagree
Religion extrinsic index	I go to religious services because it helps me to make friends		From Gorsuch and McPherson (1989). Index formed by adding together responses without first normalizing.	1 Strongly disagree - 5 Strongly agree
	I pray mainly to gain relief and protection			1 Strongly disagree - 5 Strongly agree
	What religion offers me most is comfort in times of trouble and sorrow			1 Strongly disagree - 5 Strongly agree
	Prayer is for peace and happiness			1 Strongly disagree - 5 Strongly agree
	I go to religious services mostly to spend time with my friends			1 Strongly disagree - 5 Strongly agree
	I go to religious services mainly because I enjoy seeing people there			1 Strongly disagree - 5 Strongly agree
General religion index	To what extent do you consider yourself a religious person?		From the Brief Multidimensional Measure of Religiousness/Spirituality (Fetzer Institute 1999)	1 Not religious at all - 4 Very religious
	In the last month, have you tried to convince anyone else to change the way they think about God?		From ICM survey	No = 0, Yes = 1
	How many people [have you tried to convince]?		Adapted from ICM survey	Integer ≥ 0

Variable	Components	Sub-components (if any)	Details	Possible answers	
Panel A: Primary religious outcomes					
Religion, list-randomized	How often do you go to religious services?			Daily = 365, More than once a week = 104, Once a week = 52, Once or twice a month = 18, Every month or so = 9, Once or twice a year = 1.5, Never = 0. Integer 0 – 7	
	In how many of the past 7 days did you pray privately in places other than at a place of worship?				
	How satisfied are you with your spiritual life right now?		From ICM survey		1 Not at all satisfied - 5 Very satisfied
	The Bible is accurate in all that it teaches		From ICM survey. These 3 responses are added together before standardizing, and then given triple weight when averaging the components to construct the general religion index. Asked only of Christians.		1 Strongly disagree - 5 Strongly agree
	I believe the Bible has decisive authority over what I say and do				1 Strongly disagree - 5 Strongly agree
	I believe the Christian God—Father, Son, and Holy Spirit—is the only true God				1 Strongly disagree - 5 Strongly agree
	I have made a personal commitment to Jesus Christ that is still important to me today		Adapted from ICM survey. Both questions elicited using list randomization. Outcome variable is average of two responses.		False = 0, True = 1
I have read or listened to the Bible in the past week				False = 0, True = 1	
Panel B: Primary non-religious outcomes					
Monthly consumption	Food consumption in the last week		Total amount spent in the last week on viand, rice/corn/beans/etc., bananas/cassava/potatoes/yams/starches/etc., fruits/vegetables, milk/eggs, non-alcoholic beverages. Multiplied by 30/7.	Amount in PHP (1 USD ≈ 45 PHP in 2015)	
	Non-food consumption in the last week		Total amount spent in the last week on alcoholic beverages, cigarettes, phone credit, transportation, clothing/shoes, soaps/cosmetics, gifts. Multiplied by 30/7.	Amount in PHP (1 USD ≈ 45 PHP in 2015)	
	Average monthly celebration spending in last six months		Total amount spent on weddings, funerals, festivals, anniversaries, and birthdays in the last six months divided by 6	Amount in PHP (1 USD ≈ 45 PHP in 2015)	

Variable	Components	Sub-components (if any)	Details	Possible answers
Panel B: Primary non-religious outcomes				
Food security index	No household member has gone to bed hungry in last six months		Constructed from question, “In the last 6 months, did you or any other person in this household ever go to bed hungry because there were not enough resources for food?”	No = 1; Yes = 0; Yes, but during lean season only = 0 [Lean season in the Philippines is usually July and August]
	No household member has gone to bed hungry in last six months outside of lean season		Constructed from question, “In the last 6 months, did you or any other person in this household ever go to bed hungry because there were not enough resources for food?”	No = 1; Yes = 0; Yes, but during lean season only = 1 [Lean season in the Philippines is usually July and August]
	Number of days where no household member has gone to bed hungry in past seven days		Constructed as 7 minus the number of days a member of the household has gone to bed hungry in past seven days	Integer 0 – 7
Monthly income	Total household payments received for agricultural labor on behalf of non-household member		Payments in the last 30 days	Amount in PHP (1 USD ≈ 45 PHP in 2015)
	Total household payments received for formal employment		Payments in the last 30 days	Amount in PHP (1 USD ≈ 45 PHP in 2015)
	Total household payments received for housework		Payments in the last 30 days	Amount in PHP (1 USD ≈ 45 PHP in 2015)
	Total household payments received for tending animals in an outside household		Payments in the last 30 days	Amount in PHP (1 USD ≈ 45 PHP in 2015)
	Total household payments received for operating business that is not the household’s		Payments in the last 30 days	Amount in PHP (1 USD ≈ 45 PHP in 2015)
	Total household payments for daily labor		Payments in the last 30 days	Amount in PHP (1 USD ≈ 45 PHP in 2015)
	Total household payments received for other work outside the household		Payments in the last 30 days	Amount in PHP (1 USD ≈ 45 PHP in 2015)
	Total profit from household businesses		In most recent month with normal sales	Amount in PHP (1 USD ≈ 45 PHP in 2015)

Variable	Components	Sub-components (if any)	Details	Possible answers
Panel B: Primary non-religious outcomes				
Adult weekly labor supply	Total hours spent in outside agricultural labor for non-household member		During past seven days, only household members age ≥ 17	Integer
	Total hours spent in formal employment		During past seven days, only household members age ≥ 17	Integer
	Total hours spent doing housework in an outside household		During past seven days, only household members age ≥ 17	Integer
	Total hours spent tending animals in an outside household during past seven days		During past seven days, only household members age ≥ 17	Integer
	Total hours spent operating business that is not the household's		During past seven days, only household members age ≥ 17	Integer
	Total hours spent on daily labor		During past seven days, only household members age ≥ 17	Integer
	Total hours spent on other work outside the household		During past seven days, only household members age ≥ 17	Integer
Life satisfaction index	Kessler K6 nonspecific distress scale	About how often during the past 30 days did you feel nervous?	From Kessler et al. (2002). Index formed by adding together responses without first normalizing.	1 All of the time - 5 none of the time
		About how often during the past 30 days did you feel hopeless?		1 All of the time - 5 none of the time
		About how often during the past 30 days did you feel restless or fidgety?		1 All of the time - 5 none of the time
		About how often during the past 30 days did you feel so depressed that nothing could you cheer you up?		1 All of the time - 5 none of the time
		About how often during the past 30 days did you feel that everything was difficult?		1 All of the time - 5 none of the time
		About how often during the past 30 days did you feel worthless?		1 All of the time - 5 none of the time

Variable	Components	Sub-components (if any)	Details	Possible answers
Panel B: Primary non-religious outcomes				
	Sum of 4 Gallup World Poll questions	Did you experience enjoyment during a lot of the day yesterday?		No = 0, Yes = 1
		Did you experience happiness during a lot of the day yesterday?		No = 0, Yes = 1
		Did you experience worry during a lot of the day yesterday?		No = 1, Yes = 0
		Did you experience sadness during a lot of the day yesterday?		No = 1, Yes = 0
	Did you smile or laugh a lot yesterday?		From Gallup World Poll	No = 0, Yes = 1
	How would you describe your satisfaction with life?		Elicited using Cantril's ladder	1 Very dissatisfied - 10 Very satisfied
	Taking all things together, would you say you are...		From World Values Survey	1 Not at all happy - 4 Very happy
Perceived relative economic status	Where would you place your household on the ladder in terms of economic status?		Elicited using Cantril's ladder	1 Poorest individuals of your community - 10 Best-off members of your community
Panel C: Mechanisms				
Trust index	In general, would you say that most people can be trusted or that most people cannot be trusted?			Most people can't be trusted = 0, Most people can be trusted = 1
	Do you think most people would try to take advantage of you if they got a chance, or would they try to be fair?		From World Values Survey	Try to take advantage of you = 0, Try to be fair = 1
	Would you say that most of the time people try to be helpful, or that they are mostly just looking out for themselves?		From General Social Survey	Looking out for themselves = 0, Try to be helpful = 1

Variable	Components	Sub-components (if any)	Details	Possible answers
Panel C: Mechanisms				
Social safety net index	In the case where someone in your household did not have 40 PHP available for an urgent need, how likely is it that you could access this 40 PHP from a source outside your household?			1 Very unlikely - 5 Very likely
	In the case where someone in your household did not have 1000 PHP available for an urgent need, how likely is it that you could access this 1000 PHP from a source outside your household?			1 Very unlikely - 5 Very likely
	Do you discuss personal issues with anyone outside your close family?			No = 0, Yes = 1
	How often do you usually speak to this person?			Daily = 365, A few times a week = 104, Weekly = 52, A few times a month = 24, Monthly = 12, Every month or so = 9, A few times a year = 6, Yearly = 1. If there is no such person, coded as 0.
Community activities index	Did anyone from the household receive any meals from another household in your local community? How many meals [were received]?		Top-coded at 99th percentile	No = 0, Yes = 1 Integer
	Did this household give any meals to anybody from another household in your local community? How many meals [were given]?		Top-coded at 99th percentile	No = 0, Yes = 1 Integer
	Did you attend any village leaders meetings in the last 6 months?			No = 0, Yes = 1
	In the past 6 months, have you participated in any community activities? How frequently did you participate in community activities?			No = 0, Yes = 1 Daily = 365, A few times a week = 104, Weekly = 52, A few times a month = 24, Monthly = 12, Every month or so = 9, A few times a year = 6, Yearly = 1. If the respondent did not participate, coded as 0.

Variable	Components	Sub-components (if any)	Details	Possible answers
Panel C: Mechanisms				
Perceived stress scale index	How often have you felt that you were unable to control the important things in your life?		From Cohen et al. (1983). Index formed by adding together responses without first normalizing.	1 Very Often - 5 Never
	How often have you felt confident about your ability to handle your personal problems?			1 Never - 5 Very Often
	How often have you felt that things were going your way?			1 Never - 5 Very Often
	How often have you felt difficulties were piling up so high that you could not overcome them?			1 Very Often - 5 Never
Powerful others index	I feel like what happens in my life is mostly determined by God		From Levenson (1981) Powerful Others scale, modified to apply to God's control of one's life. Index formed by adding together responses without first normalizing.	1 Strongly disagree - 5 Strongly agree
	Although I might have good ability, I will not be successful without appealing to God			1 Strongly disagree - 5 Strongly agree
	My life is chiefly controlled by God			1 Strongly disagree - 5 Strongly agree
	Getting what I want requires pleasing God			1 Strongly disagree - 5 Strongly agree
	Whether or not I have an accident and hurt myself physically depends mostly on God			1 Strongly disagree - 5 Strongly agree
	In order to have my plans work, I make sure that they fit with God's plan for me			1 Strongly disagree - 5 Strongly agree
Locus of control index	Internality subscale	Whether or not I am successful depends mostly on my ability	From Levenson (1981). Index formed by adding together responses without first normalizing.	1 Strongly disagree - 5 Strongly agree
		Whether or not I have an accident and hurt myself depends mostly on how careful I am on a daily basis		1 Strongly disagree - 5 Strongly agree
		When I make plans, I am almost certain to make them work		1 Strongly disagree - 5 Strongly agree
		How many friends I have depends on how nice a person I am		1 Strongly disagree - 5 Strongly agree

Variable	Components	Sub-components (if any)	Details	Possible answers
Panel C: Mechanisms				
		I can pretty much determine what will happen in my life		1 Strongly disagree - 5 Strongly agree
		I am usually able to protect my personal interests		1 Strongly disagree - 5 Strongly agree
		When I get what I want it's usually because I worked hard for it		1 Strongly disagree - 5 Strongly agree
		My life is determined by my own actions		1 Strongly disagree - 5 Strongly agree
	Chance subscale	To a great extent my life is controlled by accidental happenings	From Levenson (1981). Index formed by adding together responses without first normalizing.	1 Strongly agree - 5 Strongly disagree
		Often there is no chance of protecting my personal interests from bad luck happening		1 Strongly agree - 5 Strongly disagree
		When I get what I want, it is usually because I am lucky		1 Strongly agree - 5 Strongly disagree
		I have often found that what is going to happen will happen		1 Strongly agree - 5 Strongly disagree
		Whether or not I get into an accident and hurt myself physically is mostly a matter of luck		1 Strongly agree - 5 Strongly disagree
		It is not wise for me to plan too far ahead because many things turn out to be a matter of good or bad fortune		1 Strongly agree - 5 Strongly disagree

Variable	Components	Sub-components (if any)	Details	Possible answers
Panel C: Mechanisms				
		Whether or not I am successful depends on whether I am lucky enough to be in the right place at the right time		1 Strongly agree - 5 Strongly disagree
		It is chiefly a matter of fate whether or not I have a few friends or many friends		1 Strongly agree - 5 Strongly disagree
	World Values Survey locus of control	Which comes closest to your view on a scale on which (1) means "everything in life is determined by fate" and (10) means "people shape their fate themselves"?	From World Values Survey	1 fate - 10 people
Life orientation index	In uncertain times, I usually expect the best If something can go wrong for me, it will I'm always optimistic about my future I hardly ever expect things to go my way I rarely count on good things happening to me Overall, I expect more good things to happen to me than bad		From the Life Orientation Test – Revised index by Scheier et al. (1994). Index formed by adding together responses without first normalizing.	1 I disagree a lot - 5 I agree a lot 1 I agree a lot - 5 I disagree a lot 1 I disagree a lot - 5 I agree a lot 1 I agree a lot - 5 I disagree a lot 1 I agree a lot - 5 I disagree a lot 1 I disagree a lot - 5 I agree a lot
Expectations index	Which step [of the life satisfaction ladder] do you believe you will be on in 5 years? Where do you think you will be on this [relative economic status] ladder 5 years from now?		Elicited using Cantril's ladder Elicited using Cantril's ladder	1 Very dissatisfied - 10 Very satisfied 1 Poorest individuals - 10 Best-off members

Variable	Components	Sub-components (if any)	Details	Possible answers
Panel C: Mechanisms				
Optimism index	How optimistic are you in general, on a scale of 1 to 7?		From Scale Optimism-Pessimism-2 by Kemper et al. (2015). Pessimism scale shown to respondents had 1 be “not at all pessimistic” and 7 be “very pessimistic”	1 Not at all optimistic - 7 Very optimistic
	How pessimistic are you in general, on a scale of 1 to 7?			1 Very pessimistic - 7 Not at all pessimistic
Grit index	New ideas and projects sometimes distract me from previous ones		From the Short Grit Scale (Duckworth and Quinn 2009). Index formed by adding together responses without first normalizing.	1 Very much like me - 5 Not like me at all
	Setbacks don't discourage me			1 Not like me at all - 5 Very much like me
	I have been obsessed with a certain idea or project for a short time but later lost interest			1 Very much like me - 5 Not like me at all
	I am a very hard worker			1 Not like me at all - 5 Very much like me
	I often set a goal but later choose to pursue a different one			1 Very much like me - 5 Not like me at all
	I have difficulty maintaining my focus on projects that take more than a few months			1 Very much like me - 5 Not like me at all
	I finish whatever I begin			1 Not like me at all - 5 Very much like me
Self-control index	I am diligent		Subset of the Brief Self-Control Scale by Tangney, Baumeister, and Boone (2004). Index formed by adding together responses without first normalizing.	1 Not like me at all - 5 Very much like me
	I have a hard time breaking bad habits			1 Very much like me - 5 Not like me at all
	I get distracted easily			1 Very much like me - 5 Not like me at all
	I say inappropriate things			1 Very much like me - 5 Not like me at all
	I refuse things that are bad for me, even if they are fun			1 Not like me at all - 5 Very much like me
	People would say that I have very strong self-discipline			1 Not like me at all - 5 Very much like me
	Pleasure and fun sometimes keep me from getting work done			1 Very much like me - 5 Not like me at all

Variable	Components	Sub-components (if any)	Details	Possible answers
Panel C: Mechanisms				
	I'm good at resisting temptation			1 Not like me at all - 5 Very much like me
	I do things that feel good in the moment but regret later on			1 Very much like me - 5 Not like me at all
	Sometimes I can't stop myself from doing something, even if I know it's wrong			1 Very much like me - 5 Not like me at all
	I often act without thinking through all the alternatives			1 Very much like me - 5 Not like me at all
Panel D: Secondary outcomes				
Salvation by grace belief index	If I am good enough, God will cleanse me of my sins		Question asked only of Christians	1 Strongly agree - 5 Strongly disagree
	I follow God's laws so that I can go to heaven		Question asked only of Christians	1 Strongly agree - 5 Strongly disagree
	Which of the following best describes your belief about what happens after death?			There is no life after death = 0; I will go to heaven because I tried my best to be a good person and to live a good life = 0; I will go to heaven because I tried to be involved in my religion, pray, and live the way I think God wants me to = 0; I will go to hell = 0; I'm not sure if I will go to heaven or hell = 0; I will be reincarnated = 0; My belief is not well-described by any of these choices = 0; I will go to heaven because I have accepted Jesus Christ as my personal savior = 1
Assets index	Chance that you, or someone in your household, would have 40 PHP available for your use in this circumstance of urgent need?			1 Very unlikely - 5 Very likely
	Chance that you, or someone in your household, would have 1,000 PHP available for your use in this circumstance of urgent need?			1 Very unlikely - 5 Very likely

Variable	Components	Sub-components (if any)	Details	Possible answers
Panel D: Secondary outcomes				
	Number of productive assets acquired in last 6 months		Number of the following acquired in the last 6 months: tractors, sewing machines and farm tools. Top-coded at 99th percentile.	Integer ≥ 0
	Value of the productive assets in the household acquired in the last 6 months		Sum of the amount paid for the above categories of assets. Top-coded at 99th percentile.	Value of assets in PHP (1 USD \approx 45 PHP in 2015)
	Number of house assets acquired in last 6 months		Number of the following acquired in the last 6 months: TV, VTR/VHS/VCD/DVD player, radio/transistor/stereo, electric fan, refrigerator/freezer, telephone/mobile phone, sala set, bicycle or pedicab , motorcab or motorcycle, boat, washing machine, chair/stool, bed or cot, table, watch or clock, jewelry, gas stove. Top-coded at 99th percentile.	Integer ≥ 0
	Value of the house assets acquired in the last 6 months		Sum of the amount paid for the above categories of assets. Top-coded at 99th percentile.	Value of assets in PHP (1 USD \approx 45 PHP in 2015)
	Number of productive assets (level)		Number of tractors, sewing machines, and farm tools owned. Top-coded at 99th percentile.	Integer ≥ 0
	Value of productive assets (level)		Sum of the amount paid for the above assets. Top-coded at 99th percentile.	Value of assets in PHP (1 USD \approx 45 PHP in 2015)
	Number of house assets (level)		Number of the following owned: TV, VTR/VHS/VCD/DVD player, radio/transistor/stereo, electric fan, refrigerator/freezer, telephone/mobile phone, sala set, bicycle or pedicab , motorcab or motorcycle, boat, washing machine, chair/stool, bed or cot, table, watch or clock, jewelry, gas stove. Top-coded at 99th percentile.	Integer ≥ 0

Variable	Components	Sub-components (if any)	Details	Possible answers
Panel D: Secondary outcomes				
	Value of house assets (level)		Sum of the amount paid for the above assets. Top-coded at 99th percentile.	Value of assets in PHP (1 USD ≈ 45 PHP in 2015)
	How much money do you have set aside in savings?			Amount in PHP (1 USD ≈ 45 PHP in 2015)
Financial inclusion index	Do you or anyone in your household currently have money set aside as savings?			No = 0, Yes = 1
	Do you—by yourself or with other people—currently have an account at a bank?			No = 0, Yes = 1
	Have you made a deposit at a financial institution in the past 6 months?			No = 0, Yes = 1
Health index	Number of serious health events in the household (past 6 months)		We top-code at the 99th percentile and multiply by -1	Integer
	Total number of workdays missed by household members due to illness in past 30 days		We top-code each household member at 30 days and multiply by -1	Integer
	Number of household members that have suffered an illness that have kept them from working (last 30 days)		We code this as the negative of the response	Integer
Hygiene index, non-list randomized	Own or lease animals that are not kept in a separate stable			No = 1, Yes = 0
	At least one household member practices open defecation		Coded yes if primary latrine is forest, bushes, fields, bodies of water, hanging latrine, uncovered pit latrine, open pit	No = 1, Yes = 0
Hygiene, list-randomized	I treat my water before drinking it, for example by using solar disinfection, boiling it, or using a water filter		Both questions elicited using list randomization. Outcome variable is average of two components' responses	No = 0, Yes = 1
	I wash my hands after going to the bathroom			No = 0, Yes = 1

Variable	Components	Sub-components (if any)	Details	Possible answers
Panel D: Secondary outcomes				
House index	Are all rooms leak-free?			No = 0, Yes = 1
	Are at least some rooms leak-free?			No = 0, Yes = 1
	Are all rooms able to be safely locked?			No = 0, Yes = 1
	Are at least some rooms able to be safely locked?			No = 0, Yes = 1
	Primary source of energy for lighting is electricity			No = 0, Yes = 1
	Primary latrine is inside the house			No = 0, Yes = 1
Migration and remittance index	Number of migrators in the household		Number of household members who have slept outside the house for more than two consecutive nights for work in the past six months	Integer
	Number of days migrators in the household were gone in the last six months			Integer
	Number of migrators who sent remittances or brought money home to the household in the last six months			Integer
	Household had at least one migrator who sent remittances or brought cash home in the last six months			No = 0, Yes = 1
	Amount received in remittances or cash brought home by household migrators in the last six months			Amount in PHP (1 USD ≈ 45 PHP in 2015)
No discord index	During the last one month, did you have any major arguments with your spouse or partner over spending on major household items or assets?			No = 1, Yes = 0
	During the last one month, did you have any major arguments with your spouse or partner over saving decisions?			No = 1, Yes = 0
	During the last one month, did you have any major arguments with your spouse or partner over the behavior and disciplining of children?			No = 1, Yes = 0

Variable	Components	Sub-components (if any)	Details	Possible answers
Panel D: Secondary outcomes				
				No = 1, Yes = 0
				No = 1, Yes = 0
				No = 1, Yes = 0
No domestic violence, list randomized	Someone in my household is experiencing physical abuse		Question elicited using list randomization.	No = 1, Yes = 0
Child labor supply	Total hours spent in outside agricultural labor for non-household member		During past seven days, only household members age ≤ 16	Integer
	Total hours spent in formal employment		During past seven days, only household members age ≤ 16	Integer
	Total hours spent doing housework in an outside household		During past seven days, only household members age ≤ 16	Integer
	Total hours spent tending animals in an outside household during past seven days		During past seven days, only household members age ≤ 16	Integer
	Total hours spent operating business that is not the household's		During past seven days, only household members age ≤ 16	Integer
	Total hours spent on daily labor		During past seven days, only household members age ≤ 16	Integer
	Total hours spent on other work outside the household		During past seven days, only household members age ≤ 16	Integer
# children enrolled in school			Age ≤ 16	Integer