

SUPPLEMENTARY MATERIALS

Preaching Against the Choir: Religious Leaders and Reproductive Norms in Kenya

Eduard van der Merwe*

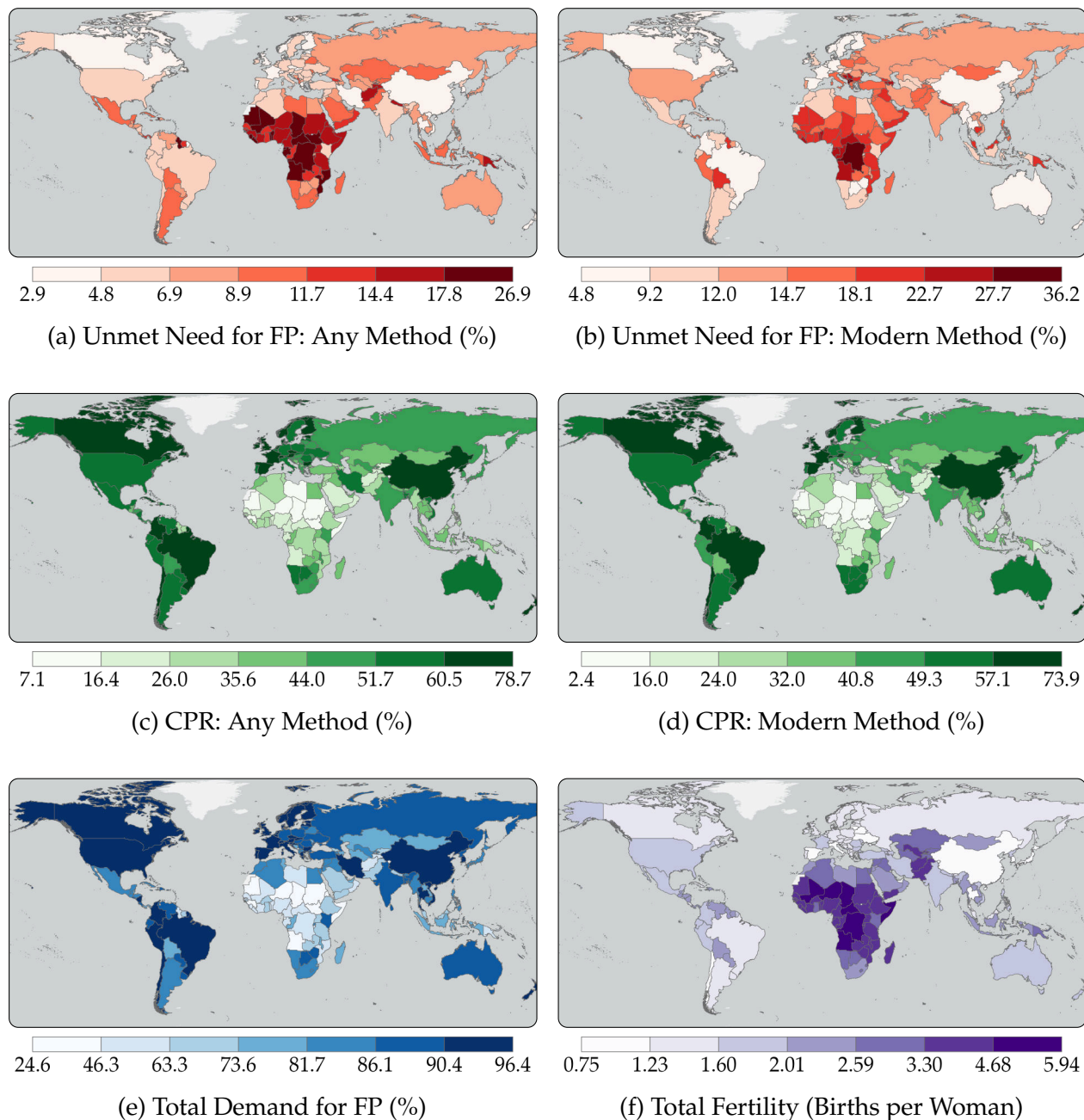
October 21, 2026

The latest version is available [here](#).

*Department of Economics, Rice University, Houston, Texas, United States. e.vdmerwe@rice.edu.

Appendix A: Motivational and Contextual Descriptives

Figure A1: Global Reproductive Health and Fertility Indicators



NOTE: Panels (a) and (b) show unmet need for family planning (FP), the share of women who want to stop or delay childbearing but use no method or no modern method. Panels (c) and (d) show the contraceptive prevalence rate (CPR) for any and for modern methods, panel (e) total demand for FP (users plus unmet need), and panel (f) the total fertility rate. Data are from UNDESA (2024).

Table A1: Health Facility Landscape in Kenya

	KEPH Level 2 (Dispensary)	KEPH Level 3 (Health Centre)	KEPH Level 4 (Sub-county Hosp.)	KEPH Level 5+ (Referral)	Total
<i>Panel A: All Kenya Health Facilities by Ownership</i>					
Government	5,125 (69.80%)	1,372 (77.50%)	397 (78.80%)	40 (82.50%)	6,934 (71.90%)
FBO	638 (39.70%)	249 (49.00%)	129 (48.80%)	10 (80.00%)	1,026 (43.50%)
Private	5,186 (49.70%)	2,124 (67.00%)	594 (61.60%)	25 (88.00%)	7,929 (55.40%)
Other	288 (35.10%)	60 (51.70%)	12 (58.30%)	0 (0.00%)	360 (38.60%)
Total	11,237 (57.90%)	3,805 (69.40%)	1,132 (66.20%)	75 (84.00%)	16,249 (61.30%)
<i>Panel B: CHAK Member Facilities by Denomination</i>					
ACK	32 (37.5%)	4 (100.0%)	5 (60.0%)	0 (0.0%)	41 (46.3%)
Methodist	9 (77.8%)	7 (28.6%)	0 (0.0%)	0 (0.0%)	16 (56.3%)
SDA	25 (80.0%)	8 (87.5%)	3 (100.0%)	0 (0.0%)	36 (83.3%)
AIC / AIPCA	37 (59.5%)	8 (100.0%)	4 (75.0%)	1 (0.0%)	50 (66.0%)
Presbyterian/PCEA	25 (56.0%)	5 (20.0%)	2 (50.0%)	1 (100.0%)	33 (51.5%)
Orthodox	3 (0.0%)	2 (50.0%)	0 (0.0%)	1 (0.0%)	6 (16.7%)
Other FBO	82 (68.3%)	35 (71.4%)	17 (64.7%)	2 (100.0%)	136 (69.1%)
Total	213 (61.5%)	69 (69.6%)	31 (67.7%)	5 (60.0%)	318 (63.8%)

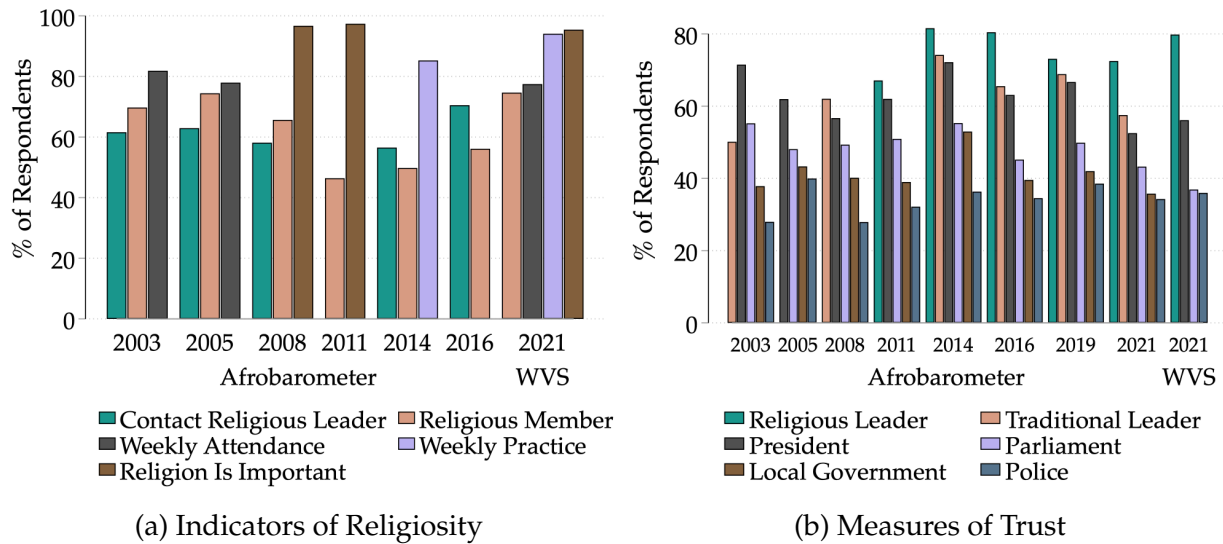
NOTE: KEPH = Kenya Essential Package for Health. Level 2 = Dispensary; Level 3 = Health Centre; Level 4 = Sub-county Hospital; Level 5+ = County and National Referral Hospitals. Percentages in parentheses show the share of facilities in each cell offering family planning services. FBO = Faith-Based Organization. CHAK = Christian Health Association of Kenya. Data are from the Kenya Master Health Facility Registry (KMHFR). Panel B includes only CHAK member facilities; CHAK facilities also appear in Panel A under the FBO category.

Table A2: Religiosity, Gender Roles, and Pronatalism in Kenya

	Share (%)			Change per SD of Religiosity Index (pp)	
	All	Other Kenyans	Most Devout	All	Women
<i>Panel A: Gender Roles (Agree)</i>					
Man has the final say in a marriage	43.4	45.2	40.9	-1.2 (1.1)	0.2 (1.6)
Men, not women, should provide for the family	37.5	38.5	36.1	-0.9 (1.2)	2.4 (1.8)
A man should always know his partner's movements	32.3	34.8	28.7	-1.4 (0.9)	-1.1 (2.3)
The family suffers when a woman works full time	33.3	34.5	31.5	0.5 (0.9)	3.9* (2.0)
University is more important for a boy than a girl	18.2	20.7	14.6	-4.2*** (1.0)	-5.3*** (1.5)
Men make better political leaders than women	42.4	45.0	38.6	-1.8** (0.9)	-3.0* (1.5)
Men make better business executives than women	32.3	37.3	25.0	-6.3*** (1.9)	-7.7*** (2.8)
Men have more right to a job when jobs are scarce	54.4	54.8	53.8	-1.6* (0.9)	-0.6 (2.2)
<i>Panel B: Fertility and Family Planning</i>					
Having children is a duty towards society	69.8	65.9	75.5	5.5*** (1.0)	9.1*** (2.0)
A woman is only a real woman after having a child	28.7	30.0	26.7	-0.8 (1.6)	-3.3 (2.2)
Contraceptives should be available to everyone	60.1	59.4	61.1	-2.8 (1.7)	-1.1 (2.3)
Contraception is never justifiable	14.1	12.0	17.1	2.9** (1.1)	2.8** (1.4)
Wants as many children as God gives	6.1	7.0	4.9	-0.9 (0.7)	0.2 (1.3)
Respondents	1266	752	514	1266	622

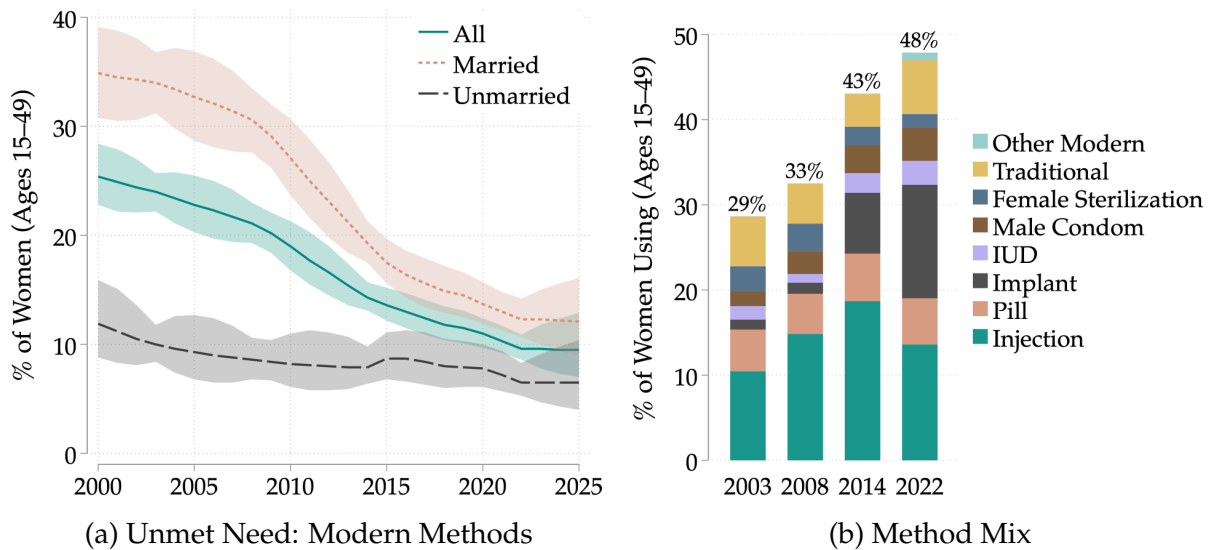
NOTE: Data are from the 2021 Kenya World Values Survey. The most devout give the most devout answer on all six religiosity items (weekly attendance, daily practice, active membership, religion very important, God rated 10 of 10, and self-described religious), and the religiosity index is the mean of the six standardized items. Columns 1–3 report the weighted share agreeing with each statement, except “Contraception is never justifiable” (1 on a 1–10 scale) and “Wants as many children as God gives” (preferred number of children). Columns 4–5 report the change in that share, in percentage points, per standard deviation of the index, from weighted linear probability models controlling for age, sex (column 4), education, residence, and religion, with standard errors clustered by enumeration area. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A2: Religiosity and Trust, All Survey Waves



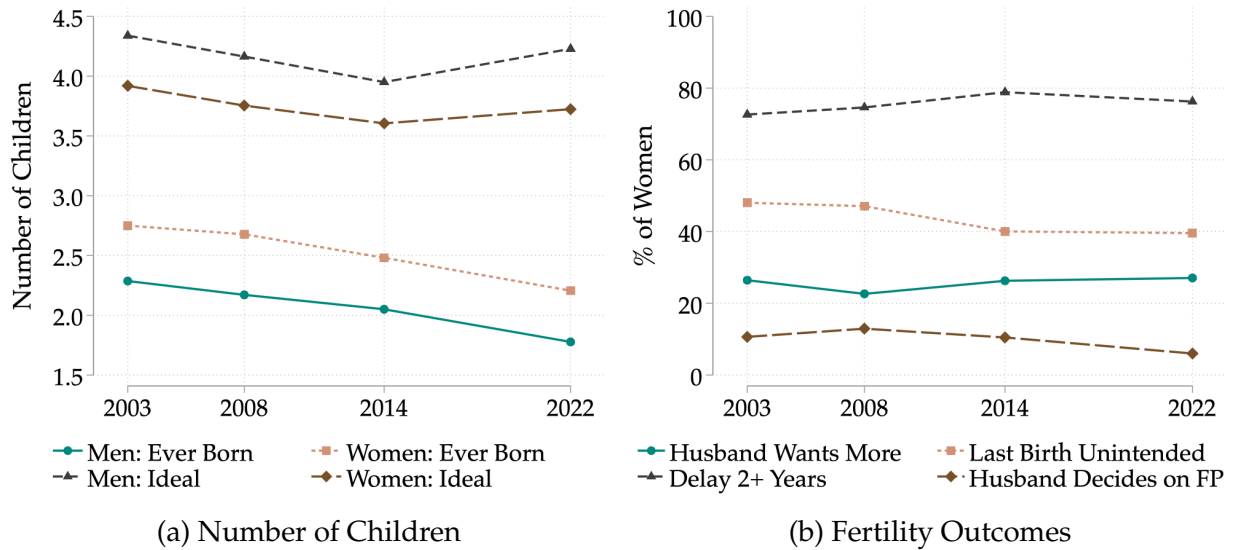
NOTE: As in main-text Figure 1, for every Afrobarometer and World Values Survey wave with available data. Panel (a) reports five dimensions of religiosity and panel (b) the share who trust each institution “somewhat” or “a lot.”

Figure A3: Women’s Reproductive Measures



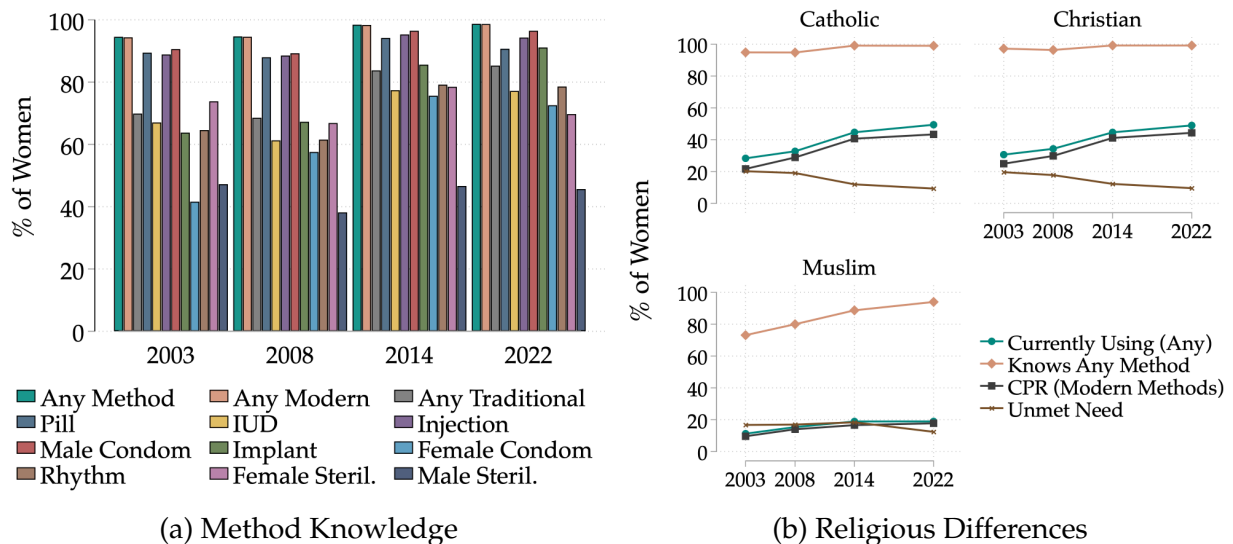
NOTE: Panel (a) shows model-based estimates of unmet need for modern methods among Kenyan women aged 15–49, by marital status, with 95% uncertainty intervals (UNDESA, 2024). Panel (b) shows the contraceptive method mix of Kenyan women aged 15–49, from the Kenya Demographic and Health Surveys (KDHS).

Figure A4: Fertility Preferences



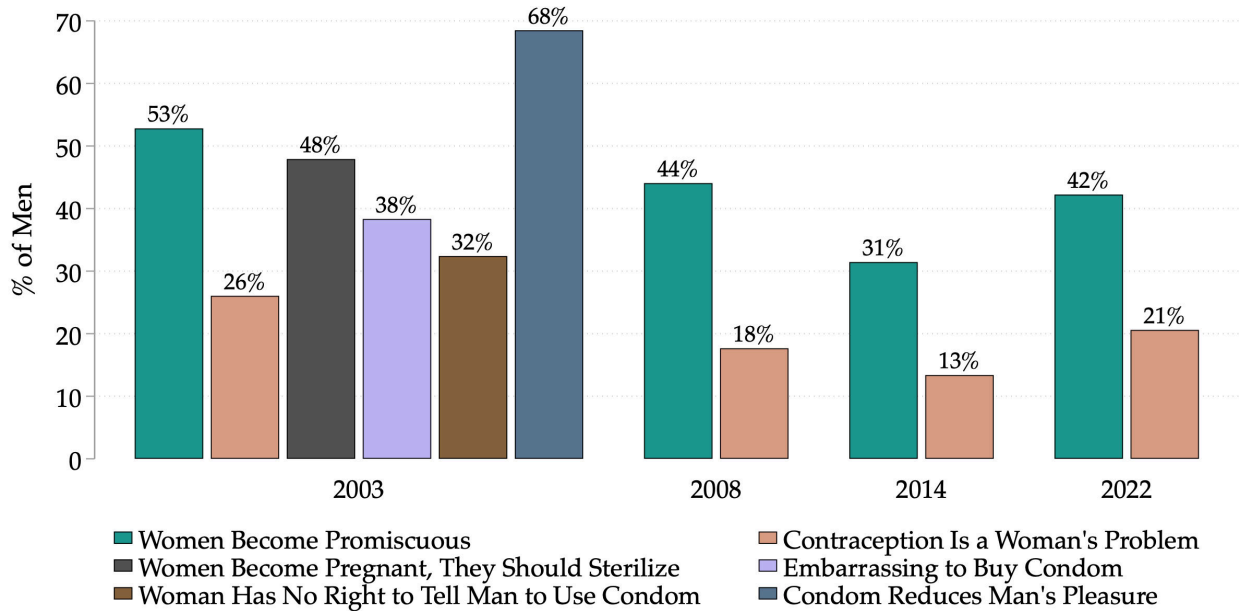
NOTE: Panel (a) compares children ever born with ideal family size, by gender, over all respondents, including those still bearing children. Panel (b) shows the share of women whose husband wants more children than they do, whose last birth was unintended, who want to delay the next birth by at least two years, and, among married users, whose husband mainly decides on contraception. Data are from the KDHS.

Figure A5: Contraceptive Knowledge and Religious Differences

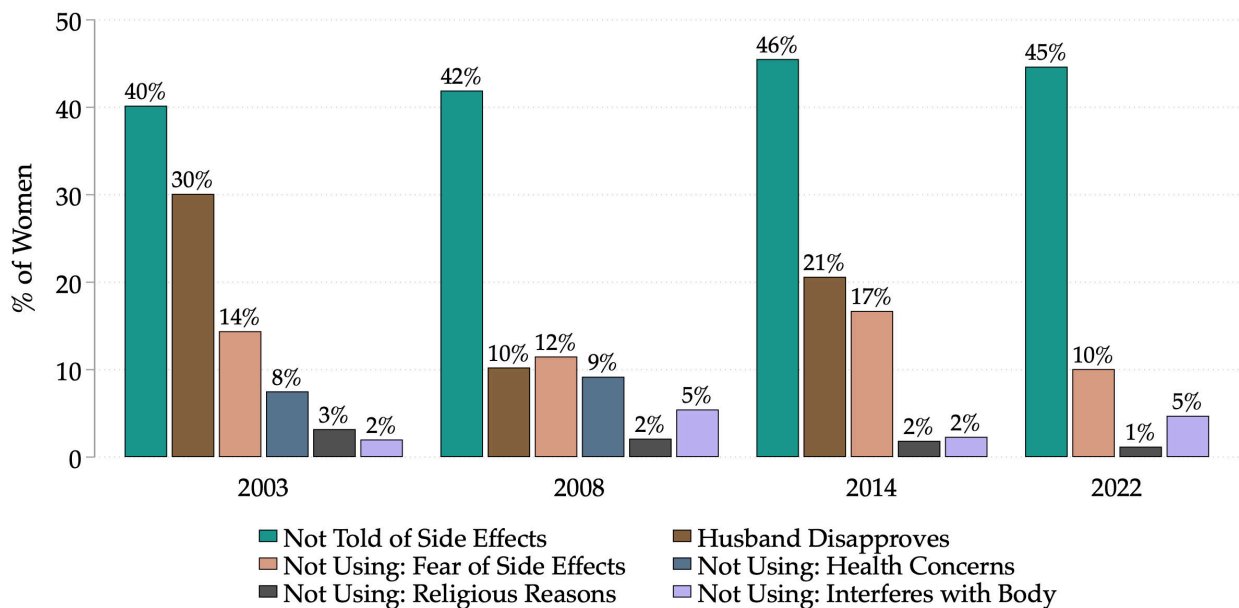


NOTE: Panel (a) shows the share of Kenyan women who know each contraceptive method, and panel (b) knowledge, current use, modern prevalence, and unmet need for Catholic, other Christian, and Muslim women, in the 2003, 2008, 2014, and 2022 KDHS.

Figure A6: Norms and Barriers Around FP



(a) Men's Attitudes Toward Contraception



(b) Barriers to Contraceptive Use

NOTE: Panel (a) shows the share of Kenyan men agreeing with each statement on contraception and condoms. The first two items were asked in all four KDHS waves and the rest only in 2003. Panel (b) shows the share of women reporting each barrier, where items prefixed "Not Using" are reasons for non-use. Husband disapproval was not asked in 2022, and health concerns only in 2003 and 2008.

Appendix B: Survey Instrument and Randomization

B.1: Sample

Table B1: Respondent Summary Statistics

	Mean	SD	N
<i>Panel A. Demographics</i>			
Age (years)	35.01	8.39	635
Living children	3.51	1.64	494
Age at first birth	19.29	3.50	494
Married or in union	0.81	0.40	635
Worked for pay (last 7 days)	0.56	0.50	635
Household income below KSh 3,000 a month	0.54	0.50	555
No formal schooling	0.07	0.25	635
Primary schooling	0.69	0.46	635
Secondary schooling or higher	0.25	0.43	635
<i>Panel B. Religion and trust</i>			
Catholic	0.08	0.27	635
Mainline Protestant	0.20	0.40	635
Pentecostal / Charismatic	0.21	0.40	635
African Instituted Church	0.15	0.36	635
Seventh-day Adventist	0.03	0.17	635
Other Christian	0.33	0.47	635
Attends services weekly or more	0.95	0.21	635
Trusts religious leaders (somewhat or more)	0.90	0.31	634
Scriptural orientation (stewardship +, pronatal -)	2.03	1.49	635
Perceived FP social sanctions (1–5)	2.19	0.79	635
<i>Panel C. Family planning</i>			
Knows a family planning method	1.00	0.04	635
Has ever used family planning	0.95	0.21	635
Currently uses a modern method	0.84	0.36	603
Approves of modern family planning	0.96	0.19	635
Husband approves of family planning	0.93	0.25	509
Ideal number of children	4.30	1.55	635
Currently pregnant	0.04	0.19	635
<i>Panel D. Perceived norms and religious leader</i>			
Women her age who use FP (of 10)	7.97	2.56	635
Women who would disapprove of FP use (of 10)	1.92	2.22	635
FP users who keep their use secret (of 10)	3.13	2.60	635
Has heard her leader speak about FP	0.43	0.50	635
Leader would say nothing if she used FP	0.82	0.38	627

NOTE: Sample of 635 reproductive-age Christian women in Meru County, Kenya. Indicator variables are proportions. Living children and age at first birth cover the 494 mothers, and household income the 555 women who reported it. Scriptural orientation is agreement (1–5) with *God gave us wisdom and stewardship to plan our families* minus agreement with *Children are a gift from God; we shouldn't limit their number*, from –4 (pronatal) to +4 (stewardship). Perceived FP social sanctions averages four items (1–5) on how likely a known FP user is to be gossiped about, called unfaithful, excluded from church activities, or have problems with her husband. Current modern use excludes 32 women whose method could not be classified. Husband approval covers the 509 partnered women who answered. Panel D asks how many of 10 village women her age use FP, would disapprove of a user, and keep their use secret.

B.2: The Vignette Stimuli

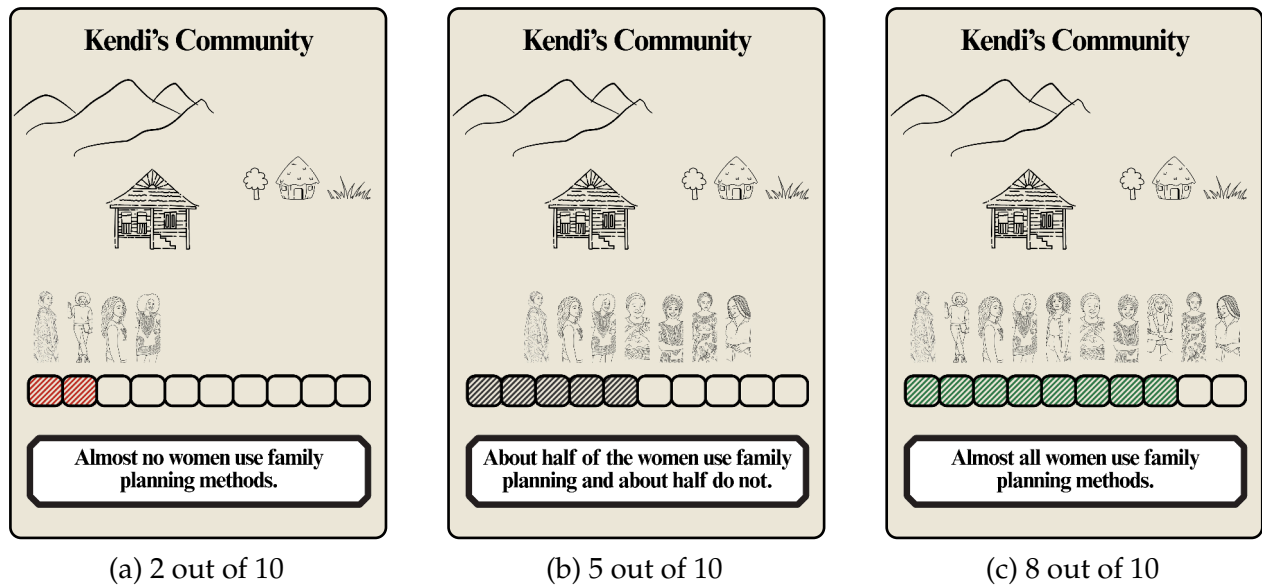
Vignette Cards This subsection reproduces the full set of cards used in the vignette survey of Section 3. Each respondent was shown four cards per story, one of the five Kendi profiles, which stayed fixed across her eight stories, and one level of each of the three randomized treatment factors. Figure B1 shows the five profiles. These are designed to represent a ladder of lifecycle stages, with each adjacent pair differing in a single attribute (age, marital

Figure B1: Kendi Lifecycle Profiles



NOTE: The five Kendi profiles. Each adjacent pair differs in one attribute, so the four adjacent contrasts identify the age, marriage, and parity effects in β .

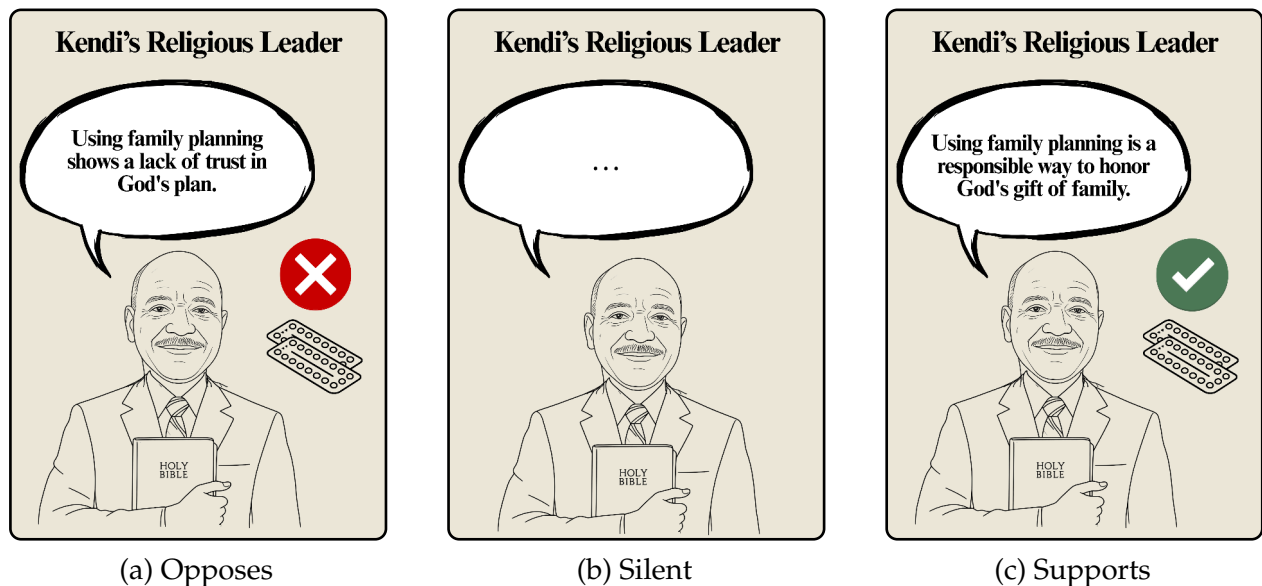
Figure B2: Social Norm Cards



NOTE: The social-norm card marks two, five, or eight of ten women as using family planning with shaded blocks.

status, or parity). Figures B2, B3, and B4 show the three levels of the social norm, the religious leader's message, and the scriptural frame. Each treatment factor has three levels, with the middle level representing a neutral or silent stance. The social-norm card shows two, five, or eight out of ten women using FP, the religious-leader card shows an opposing, silent, or supportive leader, and the scripture card shows a fruitfulness passage, no passage,

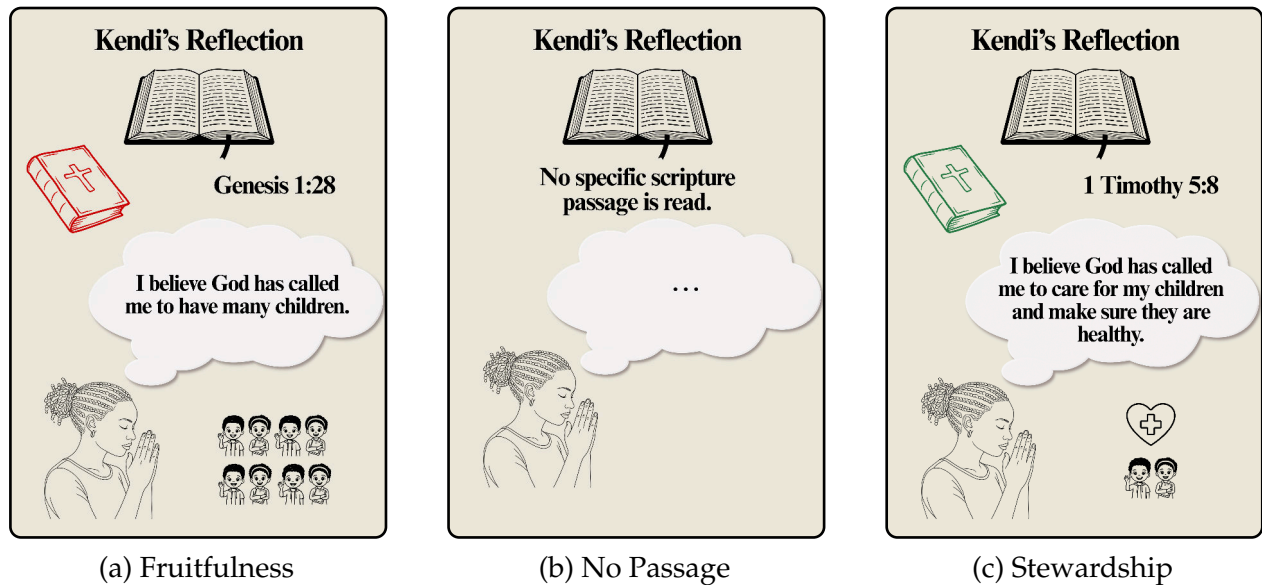
Figure B3: Religious Leader Cards



NOTE: The religious-leader card at its three levels. The leader teaches that family planning shows a lack of trust in God's plan, has not spoken about it, or teaches that it is a responsible way to honor God's gift of family.

or a stewardship passage.

Figure B4: Scripture Cards



NOTE: The scripture card at its three levels. Kendi reads Genesis 1:28 on being fruitful and multiplying, after which she believes God has called her to have many children, no particular passage, or 1 Timothy 5:8 on providing for one's household, after which she believes God has called her to care for her children and make sure they are healthy.

Elicitation Questions The survey was administered in Kimeru. The items below give the verbatim English source wording of the questions asked after each vignette.

- *Primary outcome.* "On a scale of 0 to 10, where 0 is no chance and 10 is definitely. What is the chance that Kendi will go to the clinic to receive a family planning method?"
- *Normative agreement.* "Based on this story, do you personally agree or disagree that Kendi should use a family planning method?"
- *Secret injection* (asked once per respondent). "Kendi can secretly receive a family planning injection (like Depo-Provera) at a clinic far from her village. The injection lasts three months, leaves no visible sign, and no one in her village, not even her husband, could find out. Do you think Kendi would get the injection?"

B.3: Randomization and Instrument Checks

I run two checks to verify that the randomization produced balanced groups. First, within each respondent, the tablet randomized the social norm, the leader's stance, and the scripture frame across her eight stories, as well as the order of the stories. Table B2 shows

Table B2: Orthogonality of the Randomized Vignette Factors

	Social norm (E)	Leader stance (R)	Scripture frame (S)
<i>Panel A. Pairwise correlations</i>			
Leader stance (R)	0.009		
Scripture frame (S)	0.013	-0.019	
Story slot (position 1–8)	-0.007	-0.005	0.009
<i>Demographics</i>			
Age (years)	-0.013	-0.001	-0.031
Living children	-0.007	-0.003	-0.028
Age at first birth	0.019	-0.012	0.009
Married or in union	0.011	-0.015	0.002
Worked for pay (last 7 days)	-0.001	0.002	-0.038
Household income below KSh 3,000 a month	0.037	0.002	0.012
No formal schooling	-0.010	-0.018	-0.019
Primary schooling	-0.005	0.010	-0.009
Secondary schooling or higher	0.011	-0.000	0.020
<i>Religion and trust</i>			
Catholic	-0.020	-0.003	-0.039
Mainline Protestant	-0.018	-0.008	0.005
Pentecostal / Charismatic	0.002	0.014	-0.009
African Instituted Church	0.023	-0.027	0.006
Seventh-day Adventist	0.005	-0.009	0.033
Other Christian	0.006	0.021	0.010
Attends services weekly or more	0.001	-0.010	-0.017
Trusts religious leaders (somewhat or more)	-0.011	-0.014	-0.003
Scriptural orientation (stewardship +, pronatal –)	0.008	0.007	0.004
Perceived FP social sanctions (1–5)	0.020	0.035	0.014
<i>Family planning</i>			
Knows a family planning method	0.014	0.001	-0.014
Has ever used family planning	0.006	0.012	0.012
Currently uses a modern method	-0.005	0.012	0.015
Approves of modern family planning	0.007	0.006	0.017
Husband approves of family planning	-0.036	0.016	0.019
Ideal number of children	-0.005	-0.007	-0.027
Currently pregnant	0.017	-0.015	0.012
<i>Perceived norms and religious leader</i>			
Women her age who use FP (of 10)	0.014	0.021	0.015
Women who would disapprove of FP use (of 10)	-0.002	0.020	-0.020
FP users who keep their use secret (of 10)	0.002	0.005	-0.009
Has heard her leader speak about FP	0.014	0.017	0.035
Leader would say nothing if she used FP	0.001	0.027	0.008
<i>Panel B. Joint test</i>			
Randomization p-value	0.264	0.190	0.216
R ²	0.014	0.016	0.016

NOTE: E is the social norm (–0.6, 0, +0.6), and R and S the leader and scripture cards (–1, 0, +1). Panel A reports correlations across the 5,080 stories between each factor and the other factors, the story slot, and each characteristic in Table B1. Panel B regresses each factor on story-slot indicators, all characteristics (missing values set to the mean and flagged), Kendi-profile indicators, and enumerator fixed effects, and reports a randomization-inference p -value for the joint F-statistic from 1,000 permutations of respondents' factor sequences. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B3: Balance of Respondent Characteristics Across Randomized Kendi Profiles

	Kendi profile					All	Equal means (p-value)
	A	B	C	D	E		
Panel A. Demographics							
Age (years)	36.10	33.96	34.20	35.18	35.75	35.01	0.17
Living children	3.63	3.60	3.24	3.67	3.45	3.51	0.28
Age at first birth	18.93	19.56	19.38	18.80	19.78	19.29	0.26
Married or in union	0.80	0.86	0.80	0.77	0.80	0.81	0.41
Worked for pay (last 7 days)	0.58	0.62	0.48	0.58	0.56	0.56	0.25
Household income below KSh 3,000 a month	0.57	0.52	0.58	0.52	0.51	0.54	0.73
No formal schooling	0.06	0.09	0.10	0.05	0.04	0.07	0.24
Primary schooling	0.77	0.65	0.60	0.70	0.73	0.69	0.03**
Secondary schooling or higher	0.17	0.26	0.30	0.25	0.23	0.25	0.16
Panel B. Religion and trust							
Catholic	0.06	0.14	0.10	0.09	0.02	0.08	<0.01***
Mainline Protestant	0.27	0.15	0.21	0.17	0.19	0.20	0.20
Pentecostal / Charismatic	0.22	0.26	0.22	0.17	0.16	0.21	0.28
African Instituted Church	0.16	0.12	0.12	0.17	0.20	0.15	0.32
Seventh-day Adventist	0.03	0.02	0.04	0.03	0.03	0.03	0.74
Other Christian	0.25	0.32	0.31	0.37	0.39	0.33	0.14
Attends services weekly or more	0.96	0.96	0.96	0.96	0.93	0.95	0.86
Trusts religious leaders (somewhat or more)	0.93	0.87	0.90	0.91	0.87	0.90	0.41
Scriptural orientation (stewardship +, pronatal -)	1.79	2.15	1.99	2.06	2.16	2.03	0.27
Perceived FP social sanctions (1-5)	2.20	2.11	2.26	2.14	2.22	2.19	0.52
Panel C. Family planning							
Knows a family planning method	1.00	1.00	1.00	1.00	0.99	1.00	0.32
Has ever used family planning	0.97	0.94	0.94	0.97	0.95	0.95	0.54
Currently uses a modern method	0.86	0.81	0.83	0.85	0.86	0.84	0.78
Approves of modern family planning	0.97	0.96	0.95	0.98	0.96	0.96	0.53
Husband approves of family planning	0.90	0.94	0.93	0.96	0.92	0.93	0.66
Ideal number of children	4.58	4.40	4.08	4.32	4.17	4.30	0.06*
Currently pregnant	0.05	0.03	0.01	0.07	0.04	0.04	0.02**
Panel D. Perceived norms and religious leader							
Women her age who use FP (of 10)	7.84	8.26	7.75	7.80	8.22	7.97	0.29
Women who would disapprove of FP use (of 10)	2.04	1.68	2.24	1.79	1.81	1.92	0.27
FP users who keep their use secret (of 10)	3.16	3.11	3.17	3.10	3.12	3.13	1.00
Has heard her leader speak about FP	0.43	0.41	0.45	0.47	0.41	0.43	0.82
Leader would say nothing if she used FP	0.84	0.81	0.81	0.77	0.86	0.82	0.46
Respondents	116	123	143	123	130	635	
All characteristics jointly (p-value)	0.70	0.03**	0.19	0.72	0.36		

NOTE: Means of each characteristic in Table B1 by the randomly assigned Kendi profile. The equal-means column is the p -value of an F-test, with robust standard errors, that the profile indicators are jointly zero. The final row regresses each profile indicator on all characteristics (missing values set to the mean and flagged) and reports a randomization p -value from 1,000 permutations of the profile labels. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

that the three factors are essentially uncorrelated with one another, with the position of the story, and with respondents' characteristics, as no correlation exceeds 0.04 in absolute value. A joint test asks whether the story position, all 31 characteristics, the Kendi profile, and the enumerator together predict each factor, and it fails to reject for all three treatment factors. Therefore, the within-respondent randomization was successful.

Second, between respondents, each woman was randomly assigned one of five Kendi profiles. Table B3 compares the women assigned to each profile. Three of the 31 characteristics differ at the 5% level (Catholic, primary schooling, and currently pregnant), compared with about 1.5 expected by chance, and the joint test rejects balance only for profile B, which does not survive a correction for testing five profiles. Therefore, the imbalance is of the size chance alone would produce. It also matters only for comparisons across profiles, since the weights on the norm, the leader, and scripture are identified within respondents. The one result that compares profiles, the lifecycle ladder, is unchanged when I control for the characteristics that differ, or for all 31, with the first-child step ranging from 0.49 to 0.51.

Table B4: Comprehension and Response Discrimination

<i>Panel A. Comprehension check</i>		
Factor checked	Respondents	Passed (%)
Religious leader	235	95.7
Scripture frame	189	92.1
Social norm	211	91.9
All	635	93.4
<i>Panel B. Variation across a respondent's eight answers</i>		
Same answer to all eight stories (%)	0.8	
Distinct values used, median (mean)	5 (5.2)	
Range of 5 points or more (%)	80.5	
Within-respondent standard deviation, mean	2.36	

NOTE: Panel A. After one random story, each respondent recalled the level of one random card, so the three rows are independent samples. The 42 women who failed are kept in every analysis. Panel B uses the eight elicited chances of each of the 635 respondents.

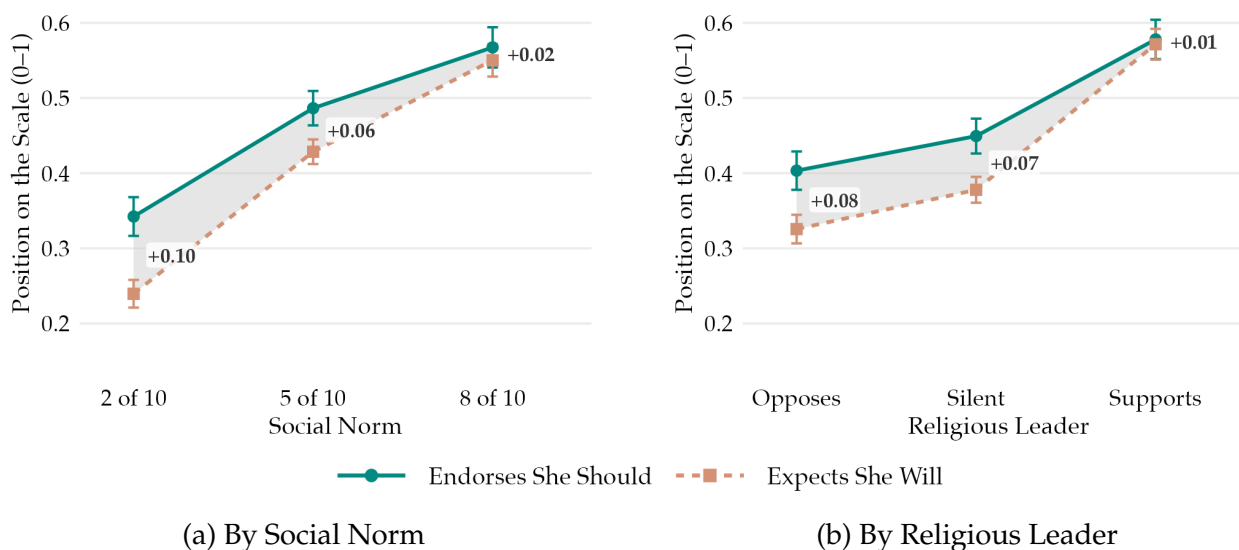
Appendix C: Additional Descriptive Evidence

Table C1: Hidden Demand: The Secret Minus Public Gap

	Public as 0–10 chance				Public as yes or no		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Secret – public (pp)	30.1*** (2.0)	30.1*** (1.9)	30.1*** (1.7)	30.1*** (1.7)	19.1*** (2.4)	26.5*** (2.4)	37.6*** (2.4)
Respondents	635	635	635	635	635	635	635
Public counted as yes if chance	–	–	–	–	≥ 4	≥ 5	≥ 6
Story card controls	No	Yes	Yes	Yes	No	No	No
Respondent controls	None	None	Core	All	None	None	None
Enumerator & profile FE	No	No	Yes	Yes	No	No	No

NOTE: One random story per respondent carries both questions. Public is the elicited chance that Kendi uses FP (mean 40.0%), and secret the share saying she would secretly obtain an injection at a distant clinic (70.1%). The outcome is the within-respondent gap in percentage points, with centered controls, so the constant is the average gap. Columns 5–7 recode the public chance to yes or no at three cutoffs. Core controls are age, union status, living children, and education, and all controls add the 31 characteristics of Table B1, with missing values set to the mean and flagged, so every specification keeps all 635 respondents. Robust standard errors, one observation per respondent. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure C1: Expectation Versus Endorsement



NOTE: Mean expectation (the elicited chance, rescaled to 0–1) and endorsement (agreement that Kendi should use FP, rescaled to 0–1) at each level of the social norm (panel a) and the leader's stance (panel b), with 95% confidence intervals. Means are taken within each respondent first. Labels give the gap between endorsement and expectation.

Table C2: Card Effects on the Public and the Secret Answer

	(1)	(2)	(3)	(4)	(5)
	Public	Secret	Secret – public	Secret	Secret
<i>Social norm</i> (vs. 5 of 10)					
2 of 10 use	-20.1*** (2.4)	-6.8 (4.6)	13.2*** (4.9)	-10.6*** (4.0)	-10.6*** (4.0)
8 of 10 use	10.4*** (2.9)	3.6 (4.3)	-6.9 (4.6)	-0.7 (3.8)	-0.8 (3.9)
<i>Religious leader</i> (vs. silent)					
Opposes	-7.1*** (2.6)	-3.9 (4.5)	3.3 (4.7)	0.1 (3.9)	-0.1 (4.0)
Supports	14.8*** (2.9)	-4.4 (4.7)	-19.1*** (4.9)	-2.9 (3.9)	-3.5 (4.1)
<i>Scripture</i> (vs. none)					
Fruitfulness (Gen 1:28)	-3.0 (2.6)	0.7 (4.4)	3.7 (4.6)	1.3 (4.0)	2.1 (4.0)
Stewardship (1 Tim 5:8)	13.3*** (3.0)	-1.7 (4.7)	-15.0*** (5.0)	-1.4 (4.0)	-1.1 (4.1)
Observations	635	635	1,270	635	635
Respondent controls	None	None	None	Core	All
Enumerator & profile FE	No	No	No	Yes	Yes
All six differences zero (<i>p</i>)			0.000		

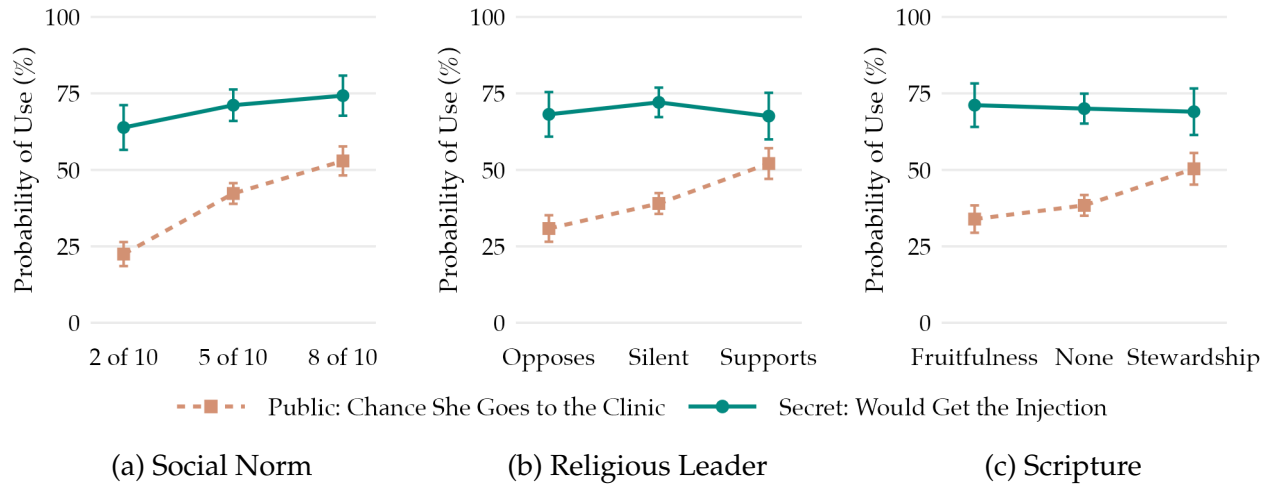
NOTE: The one random story per respondent that carries both questions. Public is the elicited chance that Kendi uses FP, and secret is 100 if the respondent says she would secretly obtain an injection and 0 otherwise. Coefficients are card effects relative to the neutral level, with the three cards entered jointly. Column 3 stacks the two answers and reports the secret minus the public effect, with standard errors clustered by respondent. Controls are as in Table C1. Robust standard errors elsewhere. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C3: Expectation Versus Endorsement by Card Level

	Social norm			Religious leader		
	2 of 10	5 of 10	8 of 10	Opposes	Silent	Supports
Endorses she <i>should</i>	0.342	0.487	0.567	0.403	0.449	0.578
Expects she <i>will</i>	0.240	0.429	0.550	0.326	0.378	0.572
Gap (<i>should</i> – <i>will</i>)	0.103*** (0.010)	0.058*** (0.009)	0.017* (0.010)	0.078*** (0.010)	0.071*** (0.009)	0.006 (0.010)
Lowest = highest level (<i>p</i>)		0.000			0.000	

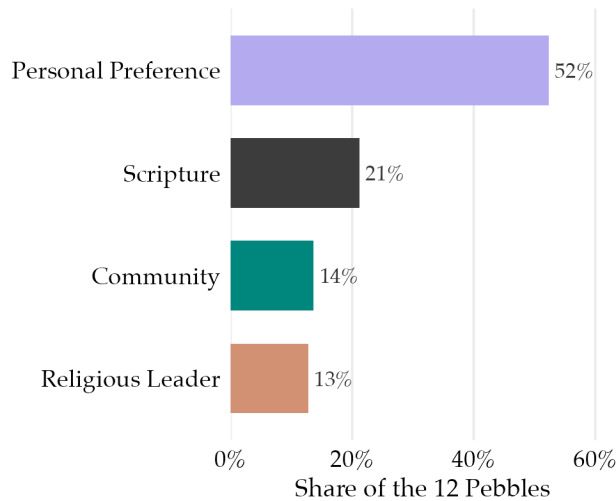
NOTE: Both outcomes are rescaled to 0–1, the elicited chance divided by 10 and endorsement as (Likert – 1)/4. Means are taken within each respondent at each level and then across respondents, as in Figure C1. Gaps come from regressing the within-respondent gap on level indicators, with standard errors clustered by respondent. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure C2: Public and Secret Answers by Card Level



NOTE: The one story per respondent that carries both questions. Public is the mean elicited chance that Kendi goes to the clinic, and secret the share saying she would secretly obtain the injection, at each level of the norm (panel a), leader (panel b), and scripture (panel c) cards, with 95% confidence intervals. Levels compare different women, and Table C2 adds controls and fixed effects.

Figure C3: The First-Person Pebble Game



NOTE: Mean share of the 12 pebbles that each of the 633 respondents allocated across four sources of influence on her own family planning decision. Figure C4 compares the shares of the three social forces with the vignette responses.

Table C4: Effects of Each Card on the Elicited Chance

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Social norm</i> (vs. 5 of 10)						
2 of 10 use	-2.05*** (0.14)	-1.83*** (0.08)	-1.82*** (0.08)	-1.83*** (0.08)	-1.81*** (0.08)	-1.81*** (0.08)
8 of 10 use	1.88*** (0.17)	1.25*** (0.09)	1.25*** (0.08)	1.25*** (0.08)	1.24*** (0.08)	1.27*** (0.08)
<i>Religious leader</i> (vs. silent)						
Opposes	-0.48*** (0.14)	-0.47*** (0.07)	-0.43*** (0.07)	-0.44*** (0.07)	-0.45*** (0.07)	-0.47*** (0.07)
Supports	2.05*** (0.14)	1.87*** (0.09)	1.84*** (0.08)	1.85*** (0.08)	1.86*** (0.08)	1.86*** (0.08)
<i>Scripture</i> (vs. none)						
Fruitfulness (Gen 1:28)	-0.20 (0.14)	-0.20** (0.09)	-0.21** (0.09)	-0.22** (0.09)	-0.26*** (0.09)	-0.31*** (0.09)
Stewardship (1 Tim 5:8)	1.50*** (0.15)	1.33*** (0.09)	1.34*** (0.09)	1.35*** (0.09)	1.39*** (0.08)	1.35*** (0.08)
Stories	2,540	5,080	5,080	5,080	5,080	4,744
Respondents	635	635	635	635	635	593
Respondent controls	None	None	Core	All	None	None
Enumerator & profile FE	No	No	Yes	Yes	No	No
Story position FE	No	No	Yes	Yes	Yes	Yes
Respondent FE	No	No	No	No	Yes	Yes
Comprehension check	All	All	All	All	All	Passed
<i>Symmetry test, low level = -(high level), p-value</i>						
Norm, clustered by respondent	0.438	0.000	0.000	0.000	0.000	0.000
Leader, clustered by respondent	0.000	0.000	0.000	0.000	0.000	0.000
Scripture, clustered by respondent	0.000	0.000	0.000	0.000	0.000	0.000
Norm, village wild bootstrap	0.422	0.000	0.000	0.000	0.000	0.000
Leader, village wild bootstrap	0.000	0.000	0.000	0.000	0.000	0.000
Scripture, village wild bootstrap	0.000	0.000	0.000	0.000	0.000	0.000

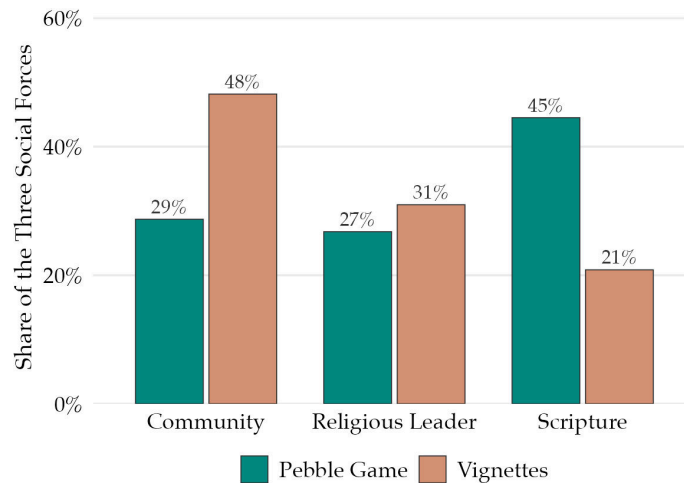
NOTE: OLS of the elicited chance (0–10) on card-level indicators, relative to each card’s neutral level. Column 1 uses Story 1 and the single-card Stories 2–4, so each coefficient is the gap from the all-neutral story (mean 3.5), as in Figure 8. Columns 2–6 pool all stories with the three cards entered jointly, and column 6 keeps the respondents who passed the comprehension check. Controls are as in Table C1. The symmetry test asks whether the against-FP level moves the chance by as much as the for-FP level. The wild cluster bootstrap resamples the 27 villages with 9,999 Rademacher draws. Standard errors clustered by respondent. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C5: The Leader's Effect at Each Level of the Social Norm

Social norm	Mean chance of FP use (pp)			Leader's effect vs. silent		
	Opposes	Silent	Supports	Supports (pp)	Opposes (pp)	Supports (ratio)
2 of 10 use	17.4 (1.2)	20.3 (1.1)	38.5 (1.7)	18.1*** (1.8)	-2.9** (1.3)	1.89*** (0.12)
5 of 10 use	34.9 (1.1)	38.4 (0.9)	57.8 (1.2)	19.4*** (1.1)	-3.5*** (1.1)	1.50*** (0.03)
8 of 10 use	44.7 (1.7)	53.0 (1.4)	69.9 (1.7)	16.9*** (1.9)	-8.3*** (1.8)	1.32*** (0.04)
8 of 10 minus 2 of 10				-1.2 (2.7)	-5.4** (2.3)	-0.57*** (0.13)
with respondent FE				-3.5* (2.1)	-7.3*** (1.9)	

NOTE: Mean elicited chance, in percentage points, in each cell of the norm and leader cards, pooled over the scripture frames and all 5,080 stories, as in Figure 9. The last three columns give the leader's effect relative to a silent leader at the same norm, as a difference and, for a supportive leader, as a ratio (stars test a ratio of one). The 8 of 10 minus 2 of 10 row tests whether that effect differs across norms, and the final row adds respondent, scripture, and story-position fixed effects. Standard errors clustered by respondent, delta method for the ratios. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure C4: Stated Versus Vignette Influence of the Social Forces



NOTE: Share of each social force in the first-person pebble game (excluding personal preference) and in the vignettes, where a force's vignette share is the change in the mean elicited chance from the lowest to the highest level of its card, with the other cards neutral, divided by the sum over the three forces. Only the ordering is comparable, because the vignette shares depend on how far apart the card levels are set, the two instruments ask about different women, and the normalization is a convention.

Table C6: The Pull of the Norm When an Authority Speaks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Norm, 2 to 8 of 10, no leader or scripture	3.93*** (0.20)	3.82*** (0.18)	3.82*** (0.18)	3.78*** (0.17)	5.11*** (0.26)	1.83*** (0.10)	1.11*** (0.06)
Norm $\times$ leader takes a stance	-0.39** (0.19)	-0.40** (0.17)	-0.41** (0.17)	-0.56*** (0.15)	-0.67*** (0.24)	-0.18* (0.09)	-0.14*** (0.05)
Norm $\times$ scripture passage read	-1.31*** (0.20)	-1.11*** (0.18)	-1.08*** (0.18)	-0.91*** (0.17)	-1.84*** (0.27)	-0.67*** (0.10)	-0.38*** (0.06)
Stories	5,080	5,080	5,080	5,080	5,080	5,080	5,080
Estimator	OLS	OLS	OLS	OLS	Tobit	Frac. logit	OLS
Outcome scale	0–10	0–10	0–10	0–10	0–10	0–1	Half log-odds
Respondent controls	None	Core	All	None	None	None	None
Enumerator, profile & position FE	No	Yes	Yes	No	No	No	No
Respondent FE	No	No	No	Yes	No	No	No
Norm $\times$ leader, village wild bootstrap (p)	0.082	0.048	0.046	0.004	–	–	–
Norm $\times$ scripture, village wild bootstrap (p)	0.000	0.000	0.000	0.000	–	–	–

NOTE: The first row is the change in the elicited chance as the norm moves from 2 to 8 of 10 women using FP when neither authority speaks. The interactions give how much that change shrinks when the leader takes either stance or either passage is read, so a negative value means the norm pulls less. All columns control for the leader and scripture levels. Column 5 treats 0 and 10 as censoring points, column 6 fits the chance divided by 10 with endpoints at 0 and 1, and column 7 uses half the log-odds with 0 and 10 set to 0.5 and 9.5, the scale of the structural model. Controls are as in Table C1. The wild cluster bootstrap resamples the 27 villages with 9,999 Rademacher draws. Standard errors clustered by respondent. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table C7: The First-Person Pebble Allocation for Family Planning

<i>Panel A. Mean share of the 12 pebbles (%)</i>				
	Own preference	Scripture	Community	Religious leader
Share	52.4 (1.0)	21.2 (0.7)	13.7 (0.5)	12.8 (0.5)
<i>Panel B. Paired differences</i>				
	Difference (pp)	Standard error	p -value	Holm p -value
Own preference – scripture	31.1	1.5	0.000	0.000
Own preference – community	38.7	1.3	0.000	0.000
Own preference – religious leader	39.6	1.3	0.000	0.000
Scripture – community	7.5	1.0	0.000	0.000
Scripture – religious leader	8.5	0.8	0.000	0.000
Community – religious leader	0.9	0.7	0.193	0.193

NOTE: The 633 respondents who completed the game for their own FP decision, dividing 12 pebbles across four boxes as in Figure C3. Panel A reports the mean share with its standard error. Panel B reports paired t -tests of each difference between two forces within respondent, with p -values also adjusted for the six comparisons by the Holm step-down method.

Table C8: Stated Versus Vignette Shares of the Social Forces

	Stated (pebble game)	Revealed (vignette responses)
Community	28.7% [26.7, 30.7]	48.2% [44.0, 52.6]
Religious leader	26.8% [25.1, 28.4]	31.0% [26.7, 34.9]
Scripture	44.5% [42.4, 46.8]	20.8% [16.6, 25.3]
Bootstrap draws in which the ranking reverses: 100.0%		

NOTE: Shares of the three social forces, each summing to 100, as in Figure C4. Stated is the mean pebble share, and revealed is each card's single-card span, the mean chance at its highest minus its lowest level with the other cards neutral. Brackets give 95% percentile intervals from 1,000 bootstrap draws over respondents. The ranking reverses in a draw when scripture ranks above the community when stated and below it when revealed.

Appendix D: Proofs and Equilibrium

This appendix proves the results stated in Section 4. Proposition 1 characterizes when the equilibrium map has one or several equilibria and which of them are stable. Proposition 2 derives the social multiplier. Proposition 3 establishes existence and uniqueness once the conformity weight differs across women with the cues they hear. Corollary 1 derives the effect of the share of women who hear a supportive leader and shows why that effect is largest against the norm. Throughout, an equilibrium is stable if it is a stable rest point of the adjustment $\bar{y}_{t+1} = g(\bar{y}_t)$, in which women respond to the average they last observed. Since every response curve here is increasing, an equilibrium $\bar{y}^*$ is stable when $g'(\bar{y}^*) < 1$ and unstable when $g'(\bar{y}^*) > 1$, because $\bar{y}_{t+1} - \bar{y}^* \approx g'(\bar{y}^*)(\bar{y}_t - \bar{y}^*)$ near $\bar{y}^*$.

✕ **Proof of Proposition 1** By (3), an equilibrium is a zero of $\phi(\bar{y}) = g(\bar{y}) - \bar{y}$ on $[-1, 1]$, where $g(\bar{y}) = \tanh(h + \alpha\bar{y})$ and $\alpha \geq 0$.

Existence. g is continuous with values in $(-1, 1)$, so $\phi(-1) = g(-1) + 1 > 0$ and $\phi(1) = g(1) - 1 < 0$. By the intermediate value theorem, ϕ has at least one zero.

Shape. Since $\frac{d}{dz} \tanh(z) = 1 - \tanh^2(z)$, the slope of the curve is $g'(\bar{y}) = \alpha(1 - g(\bar{y})^2)$. It lies in $(0, \alpha]$ and attains its maximum α where $g = 0$. Hence $\phi'(\bar{y}) = \alpha(1 - g(\bar{y})^2) - 1$.

Part 1, $\alpha \leq 1$. Then $\phi'(\bar{y}) \leq \alpha - 1 \leq 0$, with equality only if $\alpha = 1$ and $g(\bar{y}) = 0$, which holds at a single point because g is strictly increasing. So ϕ is strictly decreasing and has exactly one zero $\bar{y}^*$. At that zero $g'(\bar{y}^*) = \alpha(1 - \bar{y}^{*2}) < 1$, except in the knife-edge case $\alpha = 1, h = 0$, so the equilibrium is stable.

Part 2, $\alpha > 1$. Let $u = \sqrt{1 - 1/\alpha} \in (0, 1)$. Then $\phi'(\bar{y}) > 0$ if and only if $|g(\bar{y})| < u$. Because g is strictly increasing, this set is an interval (a, b) , with $g(a) = -u$ and $g(b) = u$, that is

$$a = \frac{-\tanh^{-1}(u) - h}{\alpha}, \quad b = \frac{\tanh^{-1}(u) - h}{\alpha}.$$

So ϕ decreases on $[-1, a]$, increases on $[a, b]$, and decreases on $[b, 1]$. Together with $\phi(-1) > 0 > \phi(1)$, it has at most three zeros, with a local minimum at a and a local maximum at b . Evaluating ϕ at these points,

$$\phi(a) = -u - a = \frac{h - (\alpha u - \tanh^{-1}(u))}{\alpha} = \frac{h - h^*}{\alpha}, \quad \phi(b) = u - b = \frac{h + h^*}{\alpha},$$

where $h^* \equiv \alpha u - \tanh^{-1}(u)$. There are three zeros, one in each interval, when $\phi(a) < 0 < \phi(b)$, that is when $|h| < h^*$. If $h > h^*$, then $\phi(a) > 0$, so ϕ is positive on $[-1, b]$ and has a single zero in $(b, 1)$. If $h < -h^*$, then $\phi(b) < 0$ and the single zero lies in $(-1, a)$. If $|h| = h^*$, one of the local extrema touches zero, giving a double root and two equilibria.

The tipping point is positive. Writing $\alpha = 1/(1 - u^2)$, $h^* = k(u) \equiv u/(1 - u^2) - \tanh^{-1}(u)$. Since $k(0) = 0$ and $k'(u) = \frac{1+u^2}{(1-u^2)^2} - \frac{1}{1-u^2} = \frac{2u^2}{(1-u^2)^2} > 0$ for $u \in (0, 1)$, $h^* > 0$ whenever $\alpha > 1$.

Stability. With three equilibria, the outer zeros lie in $(-1, a)$ and $(b, 1)$, where $\phi' < 0$, so $g' < 1$ there and both are stable. The middle zero lies in (a, b) , where $\phi' > 0$, so $g' > 1$ and it is unstable. With a single zero for $|h| > h^*$, the zero lies where $\phi' < 0$ and is stable. $\square$

✂ **Proof of Proposition 2** Let $\bar{y}$ be an equilibrium with $\alpha(1 - \bar{y}^2) \neq 1$. By the implicit function theorem, $\bar{y}$ is a differentiable function of $\bar{R}$ near this point. Differentiate the equilibrium condition $\bar{y} = \tanh(h + \alpha\bar{y})$ with respect to $\bar{R}$, using $\partial h / \partial \bar{R} = \gamma$ and $\frac{d}{dz} \tanh(z) = 1 - \tanh^2(z)$, which equals $1 - \bar{y}^2$ at the equilibrium. This gives

$$\frac{\partial \bar{y}}{\partial \bar{R}} = (1 - \bar{y}^2) \left(\gamma + \alpha \frac{\partial \bar{y}}{\partial \bar{R}} \right).$$

Collecting terms yields $\frac{\partial \bar{y}}{\partial \bar{R}} [1 - \alpha(1 - \bar{y}^2)] = \gamma(1 - \bar{y}^2)$, so

$$\frac{\partial \bar{y}}{\partial \bar{R}} = \gamma (1 - \bar{y}^2) M(\bar{y}; \alpha), \quad M(\bar{y}; \alpha) = \frac{1}{1 - \alpha(1 - \bar{y}^2)}.$$

Repeating the argument with $\partial h / \partial \bar{S} = \delta$ gives $\partial \bar{y} / \partial \bar{S} = \delta(1 - \bar{y}^2) M(\bar{y}; \alpha)$. At a stable equilibrium, $\alpha(1 - \bar{y}^2) = g'(\bar{y}) < 1$, and with $\alpha \geq 0$ it follows that $1 - \alpha(1 - \bar{y}^2) \in (0, 1]$. Hence $M \geq 1$, and $M \rightarrow \infty$ as $\alpha(1 - \bar{y}^2) \rightarrow 1$, the tangency at which an equilibrium appears or disappears in Proposition 1. When $\alpha < 1$, every equilibrium is stable, so $M \geq 1$ holds for all $\bar{y}$. The multiplier is the sum of the rounds of feedback: a shift of one unit in the index moves the norm, which moves the index by $\alpha(1 - \bar{y}^2)$, and so on, and $\sum_{k \geq 0} [\alpha(1 - \bar{y}^2)]^k = M$ whenever $\alpha(1 - \bar{y}^2) < 1$. $\square$

✂ **Proof of Proposition 3** Write the equilibrium condition as $\bar{y} = G(\bar{y})$, where

$$G(\bar{y}) = \int \tanh(h(\tau) + \alpha(\tau) \bar{y}) dF(\tau), \quad \alpha(\tau) = \alpha_0 + \alpha_R |R| + \alpha_S |S| \geq 0,$$

and $h(\tau)$ is the deterministic utility of type τ , including the asymmetric leader and scripture

effects $\gamma^\pm$ and $\delta^\pm$. G is the population-weighted average of the response curves of the different types of women.

Existence. Each integrand lies in $(-1, 1)$ and F is a probability distribution, so G maps $[-1, 1]$ into $(-1, 1)$. G is continuous by dominated convergence, since the integrand is bounded by one. Hence $G(-1) + 1 > 0$ and $G(1) - 1 < 0$, and the intermediate value theorem gives a $\bar{y}$ with $G(\bar{y}) = \bar{y}$. This holds for any distribution of stances and frames, so no shift or flattening of the curve can remove every equilibrium.

Uniqueness. Differentiating under the integral, which is permitted because the derivative of the integrand is bounded by $\alpha(\tau)$ and $\int \alpha(\tau) dF = \bar{\alpha} < \infty$,

$$G'(\bar{y}) = \int \alpha(\tau) \left(1 - \tanh^2(h(\tau) + \alpha(\tau)\bar{y})\right) dF(\tau) \leq \int \alpha(\tau) dF(\tau) = \bar{\alpha},$$

because $\alpha(\tau) \geq 0$ and $1 - \tanh^2(\cdot) \in (0, 1]$. The weight averages to

$$\bar{\alpha} = \alpha_0 + \alpha_R \mathbb{E}|R| + \alpha_S \mathbb{E}|S| = \alpha_0 + \alpha_R(f_R^+ + f_R^-) + \alpha_S(f_S^+ + f_S^-).$$

If $\bar{\alpha} < 1$, then $G'(\bar{y}) - 1 < 0$ for all $\bar{y}$, so $G(\bar{y}) - \bar{y}$ is strictly decreasing and has exactly one zero. Equivalently, suppose two equilibria $\bar{y}_1 < \bar{y}_2$ existed. Both lie on the 45-degree line, so $G(\bar{y}_2) - G(\bar{y}_1) = \bar{y}_2 - \bar{y}_1$, while the mean value theorem gives $G(\bar{y}_2) - G(\bar{y}_1) \leq \bar{\alpha}(\bar{y}_2 - \bar{y}_1) < \bar{y}_2 - \bar{y}_1$, a contradiction. The argument uses only the slope of G , so it holds for any deterministic utilities $h(\tau)$: shifts in the curve cannot create a second equilibrium.

Stability. At the equilibrium, $G'(\bar{y}) \leq \bar{\alpha} < 1$, so it is stable. Moreover $|G(a) - G(b)| \leq \bar{\alpha}|a - b|$ for all a, b , so G is a contraction and the adjustment $\bar{y}_{t+1} = G(\bar{y}_t)$ converges to the equilibrium from any starting point.

Substitution. If $\alpha_R, \alpha_S \leq 0$, then $\bar{\alpha} \leq \alpha_0$ for every distribution of stances and frames, so $\alpha_0 < 1$ is sufficient for uniqueness in every scenario. The condition $\bar{\alpha} < 1$ is sufficient but not necessary. With $\bar{\alpha} > 1$, the equilibrium can still be unique if the deterministic utilities are far enough from zero, as in Proposition 1. The mean-field map $\bar{y} = \tanh(h + \bar{\alpha}\bar{y})$ is the special case of a single type, for which the condition is also necessary. $\square$

✂ **Proof of Corollary 1** Consider the representative-member form of the extended model, $\bar{y} = \tanh(h + \bar{\alpha}\bar{y})$, with

$$h = \bar{x}\beta + \gamma^+ f_R^+ + \gamma^- f_R^- + \delta^+ f_S^+ + \delta^- f_S^-, \quad \bar{\alpha} = \alpha_0 + \alpha_R(f_R^+ + f_R^-) + \alpha_S(f_S^+ + f_S^-).$$

A rise in the share of women who hear a supportive leader, f_R^+ , moves both objects,

$$\frac{\partial h}{\partial f_R^+} = \gamma^+, \quad \frac{\partial \bar{\alpha}}{\partial f_R^+} = \alpha_R.$$

Differentiating the equilibrium condition, with $\frac{d}{dz} \tanh(z) = 1 - \bar{y}^2$ at the equilibrium,

$$\frac{\partial \bar{y}}{\partial f_R^+} = (1 - \bar{y}^2) \left(\gamma^+ + \alpha_R \bar{y} + \bar{\alpha} \frac{\partial \bar{y}}{\partial f_R^+} \right).$$

Collecting terms and writing $M = 1 / (1 - \bar{\alpha}(1 - \bar{y}^2))$ yields

$$\frac{\partial \bar{y}}{\partial f_R^+} = (\gamma^+ + \alpha_R \bar{y}) (1 - \bar{y}^2) M.$$

The same steps give $(\gamma^- + \alpha_R \bar{y})(1 - \bar{y}^2)M$ for f_R^- , and $(\delta^\pm + \alpha_S \bar{y})(1 - \bar{y}^2)M$ for $f_S^\pm$. By Proposition 3, $\bar{\alpha} < 1$ ensures M is finite and positive.

The economic content lies in the salience term $\alpha_R \bar{y}$. Under substitution ($\alpha_R < 0$), it is positive when the norm runs against FP ($\bar{y} < 0$), so a supportive leader's push and the loosened norm reinforce one another. It is negative when the norm already favors FP ($\bar{y} > 0$), so loosening the norm works against the leader's message. For an opposing leader the signs reverse: the loosened norm adds to the push γ^- when the norm favors FP. In both directions, an authority moves behavior most when speaking against the norm.

Exact aggregation. The counterfactuals average exactly over women who do and do not hear a leader instead of using the representative member. With a share f of women hearing a supportive leader, the equilibrium is $\bar{y} = f \tanh(h_0 + \gamma^+ + (\alpha_0 + \alpha_R)\bar{y}) + (1 - f) \tanh(h_0 + \alpha_0 \bar{y}) \equiv G(\bar{y}; f)$. Differentiating with respect to f ,

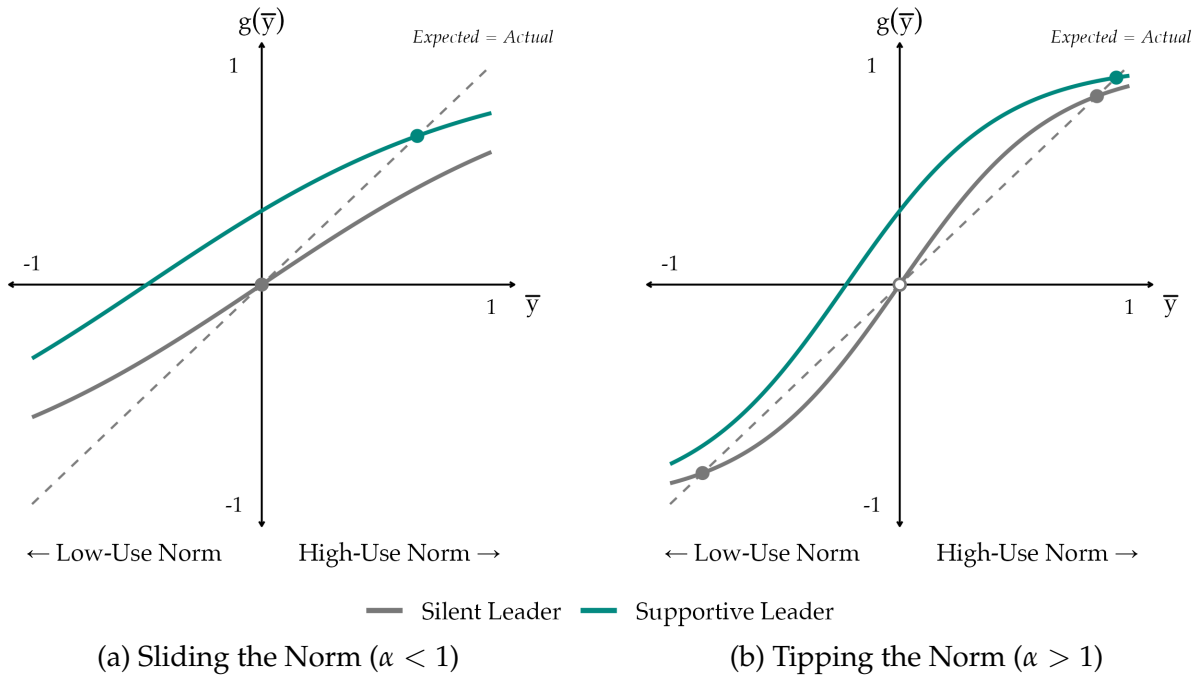
$$\frac{\partial \bar{y}}{\partial f} = \frac{\tanh(h_0 + \gamma^+ + (\alpha_0 + \alpha_R)\bar{y}) - \tanh(h_0 + \alpha_0 \bar{y})}{1 - G'(\bar{y}; f)}.$$

The numerator is the gap between how a woman who hears her leader and one who does not behave at the current norm. Their indices differ by $\gamma^+ + \alpha_R \bar{y}$, the same effective push as in the representative form, and the multiplier of the mixture is the inverse of the denominator, which is finite because $G' \leq \bar{\alpha} < 1$ by Proposition 3. $\square$

In Figure D1, I plot the simplified equilibrium outcomes, with a silent leader ($h = 0$) and a supportive one ($h = 0.35$). Both panels show that any support from religious leaders

shifts the curve up. In panel (a), the leader slides the equilibrium from 50% use to around 84% use. However, once the leader stops advocating, the curve shifts down and the use decreases back to equilibrium use at 50%. In panel (b), a strong upward shift, caused by religious advocacy, can remove the low-use equilibrium and the community settles at the high-use equilibrium. When leader advocacy stops, the curve shifts back down, but the norm is self-reinforcing, so the high-use equilibrium remains. A religious leader moves behavior by shifting the payoff of using FP. When conformity is below one, this slides the equilibrium smoothly, and when it is above one, it can tip it.

Figure D1: Sliding Versus Tipping the Norm



NOTE: Each panel plots $g(\bar{y}) = \tanh(h + \alpha\bar{y})$ against the expected norm $\bar{y}$, with a silent (gray) and a supportive (teal) leader. Filled dots are stable equilibria and the hollow dot is unstable. Panel (a) sets $\alpha = 0.70$ and panel (b) $\alpha = 1.5$. Values are illustrative.

Table D1: Vignette Treatment Factors

Factor	Symbol	Levels	Parameter Identified
Social Norm	E_k	-0.6, 0, +0.6	α_0
Leader Stance	R_k	-1, 0, +1	γ^+, γ^-
Scripture Frame	S_k	-1, 0, +1	δ^+, δ^-

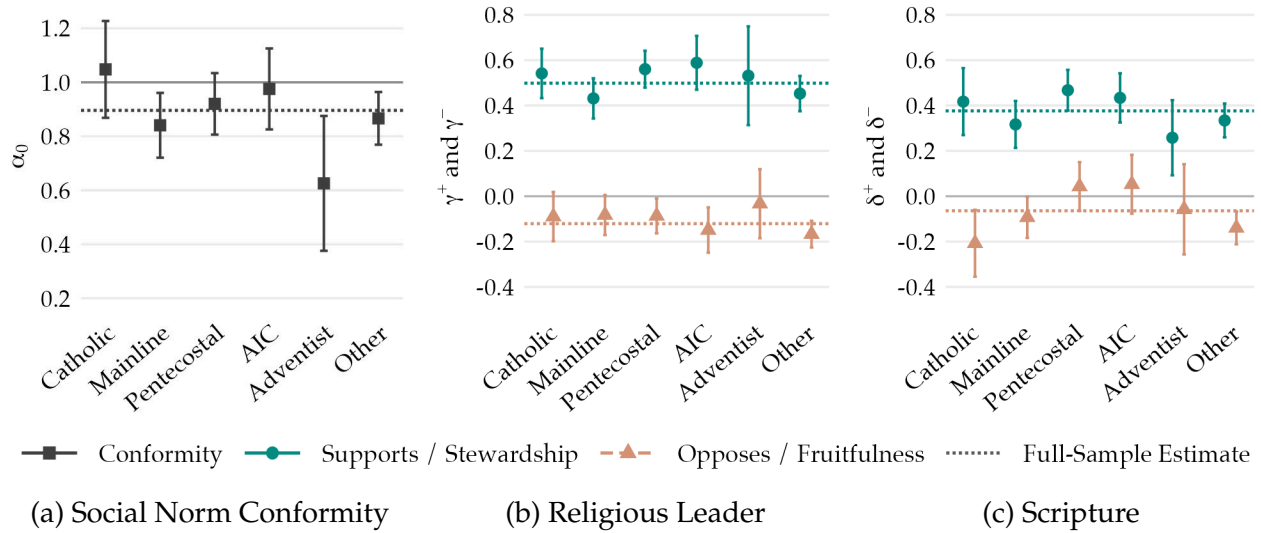
Appendix E: Heterogeneity and Robustness

This appendix collects the evidence behind the heterogeneity and robustness results in Sections 5.1 to 5.3.

E.1: Heterogeneity

Table E1 reports who responds to religious authority, the moderators of the leader and scripture effects discussed in Section 5.1. Table E2 lets the vacuum conformity weight and the leader and scripture loadings of the extended model differ by household income and religious tradition, holding the salience terms common, as described in Section 4.4. Figure E1 plots the estimates by religious tradition. Table E3 does the same across Kendi's five life stages, and Figure E2 plots the interaction of each weight with respondent traits. Appendix F.3 re-estimates the weights for women who perceive a conservative village and for the sample reweighted to low-use and high-use villages, to test how far the counterfactuals of Section 6 carry.

Figure E1: Structural Weights by Religious Tradition



NOTE: Extended-model weights by religious tradition, with 95% confidence intervals, from Panel B of Table E2. Panel (a) plots α_0 against the tipping threshold of one, and panels (b) and (c) the leader ($\gamma^\pm$) and scripture ($\delta^\pm$) effects. Dotted lines mark the full-sample estimates. Pentecostal includes Charismatic churches, AIC combines the African Instituted Churches with the Africa Inland Church (9 of its 98 respondents), and Adventist denotes the Seventh-day Adventists ($n = 20$). The intervals are too wide to rank traditions.

Table E1: Who Responds to Religious Authority

	(1)	(2)	(3)
<i>Panel A. Scripture by Her Own Scriptural View</i>			
	Observed Chance All Stories	Observed Use (pp) Secret Stories	Secret Use (pp) Secret Stories
Stewardship Passage	0.376*** (0.022)	13.99*** (2.80)	-0.81 (4.16)
Fruitfulness Passage	-0.065*** (0.023)	-3.01 (2.56)	1.29 (4.10)
Stewardship $\times$ Her View	0.072*** (0.021)	6.57** (2.85)	1.00 (4.21)
Fruitfulness $\times$ Her View	0.025 (0.023)	-0.76 (2.64)	2.94 (4.11)
Leader: Supports $\times$ Her View	-0.001 (0.020)		
Leader: Opposes $\times$ Her View	-0.020 (0.019)		
Authority $\times$ View, Joint p -value	0.008		
Respondents	635	635	635
Stories	5,080	635	635
<i>Panel B. Authority by Her Leader's Expected Reaction</i>			
	Expects Discouragement	Expects Silence	Expects Encouragement
Leader: Supports (γ^+)	0.597*** (0.076)	0.517*** (0.023)	0.255*** (0.067)
Leader: Opposes (γ^-)	-0.153** (0.063)	-0.117*** (0.021)	-0.167*** (0.052)
Scripture: Stewardship (δ^+)	0.370*** (0.079)	0.366*** (0.024)	0.485*** (0.064)
Scripture: Fruitfulness (δ^-)	-0.147** (0.066)	-0.057** (0.026)	-0.085 (0.077)
Respondents	52	514	61
Stories	416	4,112	488
<i>Panel C. Other Dispositions</i>			
	Trust in Religious Leaders	Ideal Family Size	Household Assets
Leader: Supports $\times$ Z	0.045** (0.020)	0.034 (0.021)	-0.010 (0.021)
Leader: Opposes $\times$ Z	0.012 (0.018)	0.014 (0.020)	-0.027 (0.019)
Scripture: Stewardship $\times$ Z	0.012 (0.020)	0.016 (0.024)	0.013 (0.022)
Scripture: Fruitfulness $\times$ Z	-0.009 (0.022)	0.038 (0.024)	-0.004 (0.026)
Authority $\times$ Z, Joint p -value	0.206	0.365	0.640
Respondents	634	635	635
Stories	5,072	5,080	5,080

NOTE: Extended-model estimates as in Table 2, column (2), with standard errors clustered by respondent, unless stated otherwise. Panel A, column (1), and Panel C interact one standardized disposition with the norm and the four direct authority effects, per standard deviation on the utility scale. Panel A, columns (2) and (3), are linear probability models on the secret-injection story, in percentage points, with leader, norm, and enumerator controls and robust standard errors. Panel B splits the authority effects by how the respondent expects her own leader to react to her using FP. Ideal family size is winsorized at 2 and 10 children. All results are exploratory. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table E2: Extended-Model Weights by Household Income and Religious Tradition

	Conformity α_0	$\alpha_0 = 1$ (p)	Leader supports γ^+	Leader opposes γ^-	Scripture pro δ^+	Scripture anti δ^-
<i>Panel A. Monthly Household Income (KSh)</i>						
No household income ($N = 43$)	0.96 (0.10)	0.687	0.61 (0.07)	0.00 (0.08)	0.41 (0.07)	0.01 (0.10)
Under 3,000 ($N = 258$)	0.94 (0.05)	0.216	0.51 (0.03)	-0.13 (0.03)	0.41 (0.03)	-0.04 (0.04)
3,000–6,000 ($N = 150$)	0.86 (0.06)	0.015	0.50 (0.04)	-0.10 (0.03)	0.33 (0.04)	-0.08 (0.05)
6,000–10,000 ($N = 60$)	0.72 (0.09)	0.001	0.56 (0.08)	-0.13 (0.06)	0.50 (0.09)	0.06 (0.07)
10,000 or more ($N = 44$)	0.51 (0.09)	0.000	0.50 (0.10)	-0.14 (0.06)	0.16 (0.08)	0.00 (0.08)
Equal across groups (p)	0.000		0.674	0.562	0.002	0.361
Linear trend across bands (p)	0.000		0.539	0.181	0.082	0.787
Common salience: $\alpha_R = -0.13$ (0.04), $\alpha_S = -0.23$ (0.04); common across groups, $p = 0.469$						
<i>Panel B. Religious Tradition</i>						
Catholic ($N = 52$)	1.05 (0.09)	0.601	0.54 (0.06)	-0.09 (0.06)	0.42 (0.08)	-0.21 (0.08)
Mainline Protestant ($N = 125$)	0.84 (0.06)	0.009	0.43 (0.05)	-0.08 (0.04)	0.32 (0.05)	-0.09 (0.05)
Pentecostal or Charismatic ($N = 131$)	0.92 (0.06)	0.170	0.56 (0.04)	-0.09 (0.04)	0.47 (0.05)	0.04 (0.05)
African Instituted (AIC) ($N = 98$)	0.98 (0.08)	0.749	0.59 (0.06)	-0.15 (0.05)	0.43 (0.06)	0.05 (0.07)
Seventh-day Adventist ($N = 20$)	0.63 (0.13)	0.003	0.53 (0.11)	-0.03 (0.08)	0.26 (0.08)	-0.06 (0.10)
Other Christian ($N = 209$)	0.87 (0.05)	0.007	0.45 (0.04)	-0.17 (0.03)	0.33 (0.04)	-0.14 (0.04)
Equal across groups (p)	0.004		0.055	0.456	0.039	0.000
Common salience: $\alpha_R = -0.15$ (0.03), $\alpha_S = -0.24$ (0.04); common across groups, $p = 0.717$						

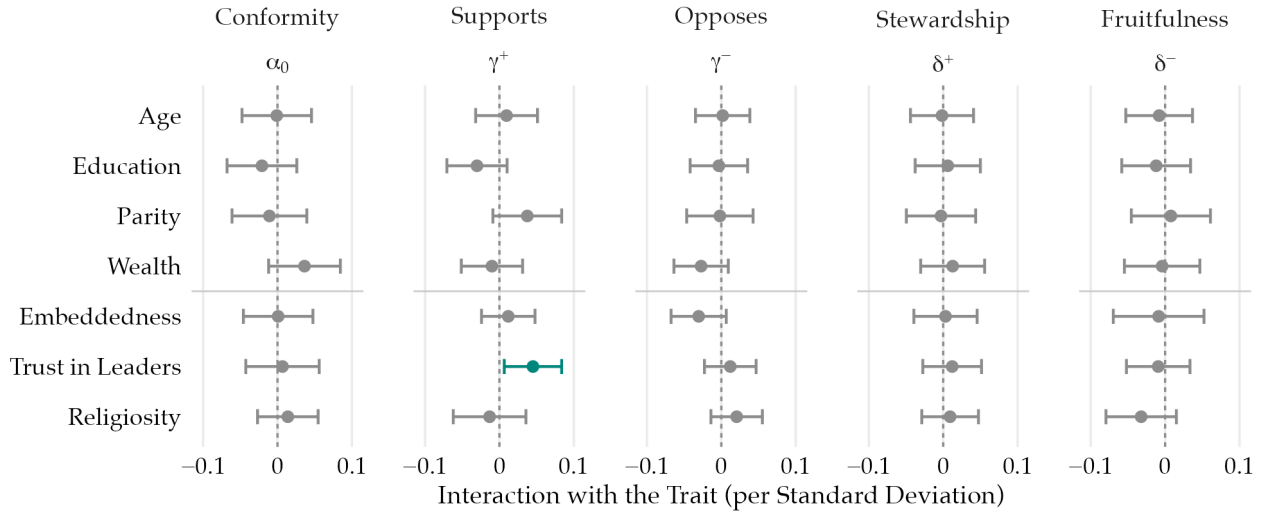
NOTE: Extended-model estimates as in Table 2, column (2), one model per panel, in which α_0 , $\gamma^\pm$, and $\delta^\pm$ differ by group and the salience terms α_R and α_S are common. Group sizes N are under each name, and the second column tests α_0 against the tipping threshold of one. Equal across groups is a likelihood-ratio test, and linear trend a Wald test with the income bands scored from -2 to 2 . The salience line tests that α_R and α_S do not differ across groups. Panel A covers the 555 respondents who reported household income and Panel B all 635. AIC combines the African Instituted Churches with the Africa Inland Church.

Table E3: Extended-Model Weights by Kendi's Life Stage

	Conformity α_0	$\alpha_0 = 1$ (p)	Leader supports γ^+	Leader opposes γ^-	Scripture pro δ^+	Scripture anti δ^-
A: age 16, unmarried, no children ($N = 116$)	0.81 (0.05)	0.000	0.49 (0.05)	-0.12 (0.04)	0.21 (0.05)	-0.05 (0.05)
B: age 22, unmarried, no children ($N = 123$)	0.90 (0.07)	0.149	0.57 (0.05)	-0.10 (0.04)	0.42 (0.05)	-0.04 (0.05)
C: age 22, married, no children ($N = 143$)	0.84 (0.06)	0.008	0.50 (0.04)	-0.06 (0.04)	0.38 (0.05)	-0.01 (0.05)
D: age 22, married, one child ($N = 123$)	0.89 (0.06)	0.094	0.42 (0.05)	-0.19 (0.04)	0.41 (0.05)	-0.11 (0.06)
E: age 28, married, one child ($N = 130$)	1.04 (0.06)	0.529	0.52 (0.05)	-0.16 (0.04)	0.44 (0.04)	-0.13 (0.05)
Equal across profiles (p)	0.002		0.154	0.147	0.001	0.246
Common salience: $\alpha_R = -0.15$ (0.03), $\alpha_S = -0.25$ (0.04); common across profiles, $p = 0.508$						

NOTE: Extended-model estimates as in Table 2, column (2), in which α_0 , $\gamma^\pm$, and $\delta^\pm$ differ by Kendi's profile and the salience terms are common. N is under each profile. Equal across profiles is a likelihood-ratio test, and the salience line tests that α_R and α_S do not differ across profiles.

Figure E2: Structural Weights by Respondent Trait



NOTE: Interaction of each extended-model weight with a standardized respondent trait, one model per trait, with 95% confidence intervals. A positive value means the weight is larger for women one standard deviation higher on the trait. Each model adds the trait and its interactions with the norm and the four authority terms, as in Table E1, Panel C. Religiosity is the mean of four standardized practice items.

E.2: Robustness

Table E4 re-estimates the structural weights on the normative endorsement outcome alongside the behavioral one. Table E5 varies the estimator and the recoding of boundary answers, and Table E6 varies the controls and the sample. Table E7 reports the nested

specification ladder. Table E8 varies the treatment of the 0–10 answer itself, with random perturbations, interval regressions, and ordinal and yes/no versions of the outcome.

Table E4: Structural Weights on Expectation versus Endorsement

	Would use (expectation)	Should use (endorsement)
<i>Panel A. Social and Religious Weights</i>		
Social Norm Conformity (α_0)	0.90*** (0.04)	0.68*** (0.04)
Leader: Supports (γ^+)	0.50*** (0.02)	0.36*** (0.03)
Leader: Opposes (γ^-)	-0.12*** (0.02)	-0.09*** (0.02)
Scripture: Stewardship (δ^+)	0.38*** (0.02)	0.27*** (0.02)
Scripture: Fruitfulness (δ^-)	-0.06*** (0.02)	-0.10*** (0.02)
<i>Panel B. Authority Salience</i>		
Leader Salience (α_R)	-0.15*** (0.03)	-0.14*** (0.04)
Scripture Salience (α_S)	-0.24*** (0.04)	-0.17*** (0.05)
Respondents	635	635
Stories	5,080	5,080

NOTE: Extended-model weights (atanh scale) for the elicited chance that Kendi *would* use family planning and for agreement that she *should* (five-point Likert, rescaled to the same interval), estimated as in Table 2, column (2). The two outcomes are on different scales, so only the signs and relative sizes of the weights compare across columns.
 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table E5: Robustness of the Structural Estimates to Estimator and Recoding

	(1)	(2)	(3)	(4)	(5)
<i>Panel A. Social and Religious Weights</i>					
Social Norm Conformity (α_0)	0.90*** (0.04)	1.05*** (0.05)	1.25*** (0.06)	0.81*** (0.04)	0.88*** (0.02)
Leader: Supports (γ^+)	0.50*** (0.02)	0.56*** (0.02)	0.63*** (0.03)	0.45*** (0.02)	0.50*** (0.01)
Leader: Opposes (γ^-)	-0.12*** (0.02)	-0.13*** (0.02)	-0.14*** (0.03)	-0.11*** (0.02)	-0.12*** (0.01)
Scripture: Stewardship (δ^+)	0.38*** (0.02)	0.43*** (0.02)	0.50*** (0.03)	0.34*** (0.02)	0.39*** (0.01)
Scripture: Fruitfulness (δ^-)	-0.06*** (0.02)	-0.07*** (0.03)	-0.08*** (0.03)	-0.04* (0.02)	-0.06*** (0.01)
<i>Panel B. Authority Salience</i>					
Leader Salience (α_R)	-0.15*** (0.03)	-0.20*** (0.04)	-0.25*** (0.05)	-0.09** (0.04)	-0.12*** (0.03)
Scripture Salience (α_S)	-0.24*** (0.04)	-0.30*** (0.04)	-0.37*** (0.06)	-0.27*** (0.04)	-0.24*** (0.03)
Estimator	RE MLE	RE MLE	RE MLE	Frac. logit	Binom. GLMM
Outcome	$\tanh^{-1}$	$\tanh^{-1}$	$\tanh^{-1}$	Share	Count
Boundary recode	0.05	0.025	0.01	None	None
Respondents	635	635	635	635	635
Stories	5,080	5,080	5,080	5,080	5,080

NOTE: Each column reports the seven structural weights on the atanh scale, and column (1) is the headline. Columns (1) to (3) recode answers of 0 and 10 to the boundary value shown and one minus it. Column (4) is a fractional logit on the raw share and column (5) a binomial(10) logit with a respondent random intercept. Neither needs a recode, and their coefficients are halved to the atanh scale. Delta-method standard errors. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table E6: Robustness of the Structural Estimates to Specification and Sample

	(1)	(2)	(3)	(4)	(5)	(6)
<i>Panel A. Social and Religious Weights</i>						
Social Norm Conformity (α_0)	0.90*** (0.04)	0.90*** (0.04)	0.90*** (0.04)	0.89*** (0.04)	0.89*** (0.04)	0.91*** (0.04)
Leader: Supports (γ^+)	0.50*** (0.02)	0.50*** (0.02)	0.50*** (0.02)	0.50*** (0.02)	0.50*** (0.02)	0.50*** (0.02)
Leader: Opposes (γ^-)	-0.12*** (0.02)	-0.12*** (0.02)	-0.12*** (0.02)	-0.12*** (0.02)	-0.12*** (0.02)	-0.12*** (0.02)
Scripture: Stewardship (δ^+)	0.38*** (0.02)	0.38*** (0.02)	0.38*** (0.02)	0.38*** (0.02)	0.38*** (0.02)	0.37*** (0.02)
Scripture: Fruitfulness (δ^-)	-0.06*** (0.02)	-0.06*** (0.02)	-0.06*** (0.02)	-0.07*** (0.02)	-0.06*** (0.02)	-0.08*** (0.02)
<i>Panel B. Authority Salience</i>						
Leader Salience (α_R)	-0.15*** (0.03)	-0.15*** (0.03)	-0.15*** (0.03)	-0.16*** (0.04)	-0.14*** (0.03)	-0.16*** (0.04)
Scripture Salience (α_S)	-0.24*** (0.04)	-0.24*** (0.04)	-0.25*** (0.04)	-0.23*** (0.04)	-0.24*** (0.04)	-0.26*** (0.04)
Respondent effect	Random	Random	Random	Fixed	Random	Random
Enumerator FE	Yes	No	Yes	Absorbed	Yes	Yes
Respondent controls	No	No	Yes	No	No	No
Story-order FE	No	No	No	No	Yes	No
Passed comprehension check only	No	No	No	No	No	Yes
Respondents	635	635	635	635	635	593
Stories	5,080	5,080	5,080	5,080	5,080	4,744

NOTE: Each column reports the seven structural weights on the atanh scale, and column (1) is the headline. Respondent controls are age, living children, education, low household income, religious tradition, and trust in religious leaders, with missing values set to the mean and flagged. Column (4) uses respondent fixed effects, which absorb the enumerator effect. Story-order FE are fixed effects for the story slot. Column (6) keeps the respondents who pass the comprehension check. Standard errors clustered by respondent. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table E7: Specification Ladder for the Structural Model

Nested comparison	χ^2	df	p
Add salience mains $ R , S $ (needed)	255.1	2	0.000
Add direction bilinears $E \times R, E \times S$ (not needed)	2.05	2	0.359
Parsimonious vs. 27-cell saturated	52.6	19	0.000

NOTE: Likelihood-ratio tests between nested random-effects models fit by maximum likelihood.

Table E8: Robustness to the Answer Scale

		Perturbation		Interval Regression				Latent Scale	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Headline	± 1 All	Heaping Jitter	Censored Ends	Ends at 1%	Ends at 2.5%	Wide Ends	Ordered Logit	Yes/No Logit
Panel A. Social and Religious Weights									
Social Norm Conformity (α_0)	0.90*** (0.04)	0.85*** (0.04)	0.84*** (0.04)	1.15*** (0.05)	1.06*** (0.05)	0.98*** (0.04)	0.98*** (0.04)	3.22*** (0.15)	3.20*** (0.22)
Leader: Supports (γ^+)	0.50*** (0.02)	0.49*** (0.02)	0.48*** (0.02)	0.60*** (0.03)	0.56*** (0.02)	0.54*** (0.02)	0.54*** (0.02)	1.77*** (0.08)	2.18*** (0.12)
Leader: Opposes (γ^-)	-0.12*** (0.02)	-0.12*** (0.02)	-0.12*** (0.02)	-0.14*** (0.02)	-0.13*** (0.02)	-0.13*** (0.02)	-0.13*** (0.02)	-0.38*** (0.06)	-0.57*** (0.12)
Scripture: Stewardship (δ^+)	0.38*** (0.02)	0.36*** (0.02)	0.36*** (0.02)	0.47*** (0.03)	0.44*** (0.02)	0.41*** (0.02)	0.41*** (0.02)	1.34*** (0.07)	1.61*** (0.12)
Scripture: Fruitfulness (δ^-)	-0.06*** (0.02)	-0.06*** (0.02)	-0.06*** (0.02)	-0.07*** (0.03)	-0.07*** (0.03)	-0.07*** (0.02)	-0.06*** (0.02)	-0.19*** (0.08)	-0.35*** (0.13)
Panel B. Authority Salience									
Leader Salience (α_R)	-0.15*** (0.03)	-0.13*** (0.04)	-0.13*** (0.03)	-0.22*** (0.04)	-0.19*** (0.04)	-0.17*** (0.04)	-0.17*** (0.04)	-0.53*** (0.12)	0.03 (0.21)
Scripture Salience (α_S)	-0.24*** (0.04)	-0.23*** (0.04)	-0.23*** (0.04)	-0.34*** (0.05)	-0.30*** (0.04)	-0.27*** (0.04)	-0.28*** (0.04)	-0.91*** (0.13)	-1.08*** (0.22)
Panel C. Scale-Free Ratios									
Leader Premium ($\gamma^+ / \gamma^- $)	4.12*** (0.69)	4.06*** (0.72)	4.07*** (0.68)	4.37*** (0.80)	4.27*** (0.75)	4.18*** (0.71)	4.21*** (0.72)	4.70*** (0.84)	3.84*** (0.83)
Salience Share (α_R / α_0)	-0.17*** (0.03)	-0.15*** (0.04)	-0.16*** (0.04)	-0.19*** (0.03)	-0.18*** (0.03)	-0.17*** (0.03)	-0.17*** (0.03)	-0.17*** (0.03)	0.01 (0.06)
Conformity = 1, p -value	0.007			0.004	0.212	0.711	0.716		

NOTE: Extended-model estimates under nine treatments of the 0–10 answer, otherwise as in Table 2, column (2). Column (1) is the headline, recoding 0 and 10 to 0.05 and 0.95. Column (2) moves every answer at random by one point up, down, or not at all, within 0 to 10, and column (3) only the focal answers 0, 5, and 10, each averaged over 50 replications. Columns (4) to (7) are interval regressions that read an answer k as a chance between $(k - 0.5) / 10$ and $(k + 0.5) / 10$, with a 0 below 0.05 in column (4), between 0.01 and 0.05 in (5), between 0.025 and 0.05 in (6), and below 0.10 in (7), and a 10 as the mirror image. Column (8) is an ordered logit and column (9) a logit of a chance of 6 or more, both on their own latent scales, so only signs and the ratios in Panel C compare across columns. Ratios use delta-method standard errors, and conformity = 1 is tested where its level is comparable. All columns use the 635 respondents and 5,080 stories. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix F: Counterfactual Simulations

This appendix documents the counterfactual simulations of Section 6. It sets out how each simulation is computed, shows how the one-in-five gain depends on where leaders stand today, and asks whether the weights estimated in Meru carry to villages where far fewer, or far more, women use family planning.

F.1: Method

Each simulation pushes the estimated weights through the equilibrium map of Section 4.2. A community's modern contraceptive prevalence p maps to the mean action $\bar{y} = 2p - 1$, and a community at rest with every leader silent satisfies $\bar{y}_0 = \tanh(h_0 + \alpha_0 \bar{y}_0)$. Given the estimated α_0 , this fixes the private-preference intercept at the baseline,

$$h_0 = \operatorname{atanh}(\bar{y}_0) - \alpha_0 \bar{y}_0. \quad (\text{F1})$$

Meru's prevalence of 50.8 percent in the 2022 KDHS gives $\bar{y}_0 = 0.016$. The simulations that vary the baseline recalibrate h_0 at every point on the horizontal axis, so they hold the weights fixed and change only where the community starts.

When a share f of leaders supports family planning, the women they reach hear the full cue, a shift of γ^+ and a conformity weight lowered to $\alpha_0 + \alpha_R$, and the remaining women hear silence. All women conform to the same village norm, so the new equilibrium solves

$$\bar{y} = f \tanh(h_0 + \gamma^+ + (\alpha_0 + \alpha_R) \bar{y}) + (1 - f) \tanh(h_0 + \alpha_0 \bar{y}). \quad (\text{F2})$$

This reads f as the share of leaders who speak, with each leader reaching an equal number of women, and applies α_R as the vignette identified it, to the woman who hears her own leader. An opposing leader replaces γ^+ with γ^- , and with $f = 1$ every woman hears the leader.

The decomposition of Figure 12 builds the gain in three cumulative steps. The direct push holds the village norm at $\bar{y}_0$ and the conformity weight at α_0 , so reached women move to $\tanh(h_0 + \gamma^+ + \alpha_0 \bar{y}_0)$. Substitution lowers their conformity weight to $\alpha_0 + \alpha_R$ at the same norm. The equilibrium then lets neighbors respond, solving Equation (F2). The first two steps reach only the women who hear a leader, so at reach f they are scaled by f . Near an even split $\bar{y}_0$ is close to zero, so weakening the weight on the norm changes little and the first two lines nearly coincide at Meru. Below an even split the norm pulls against use and

weakening it adds to the push, while above an even split it takes away from it.

The durability exercise starts from the equilibrium of a full-reach campaign and lets the leaders fall silent, iterating $\bar{y} \leftarrow \tanh(h_0 + \alpha_0 \bar{y})$ to its resting point. With α_0 below one that point is the starting equilibrium. With α_0 above one the community can instead remain at a second, high equilibrium. The bands in Section 6 draw 1,000 vectors from the estimated joint sampling distribution of the seven weights, re-solve each simulation for every draw, and report the 2.5th and 97.5th percentiles. Table F4 uses 500 draws for each set of weights.

F.2: Leaders' Starting Stance

The headline calibration assumes that every leader in Meru is silent today. The survey supports this as a first approximation. Asked how her own leader would react if he learned she used family planning, about 82 percent of women expect him to say nothing, about 10 percent expect encouragement, and 8 percent discouragement, mostly in private. Table F1 recalibrates Meru to that reported mix before adding supportive leaders. The one-in-five gain falls from 20.7 percentage points to between 16.8 and 18.9, and netting the two stances against each other gives 20.3. The item records direction rather than a stance as salient as the vignette's, so these rows bound the headline from below.

Table F1: Leaders' Starting Stance and the One-in-Five Gain

	Leaders supporting today	Leaders opposing today	Gain (pp)
All leaders silent (headline)	0	0	+20.7
Net stance only	2	0	+20.3
Reported mix, one in five of all leaders added	10	8	+18.9
Reported mix, one in five of the silent leaders added	10	8	+16.8

NOTE: Gain in Meru's equilibrium modern prevalence, in percentage points, when one in five leaders begins to support family planning, under four assumptions about where leaders stand today (shares in percent). Each row recalibrates the private-preference intercept to Meru's 50.8 percent. The headline assumes every leader is silent. The reported mix uses how women expect their own leader to react to their FP use (10 percent encouragement, 8 percent discouragement, 82 percent silence), and the net row nets support against opposition. In the third row the added supporters are one in five of all leaders, drawn from the silent, and in the fourth one in five of the silent leaders. Point estimates at the headline weights.

F.3: Conservative and Liberal Communities

The decomposition in Figure 12 carries Meru's weights to villages where few or most women use family planning, yet the survey was fielded only in Meru. Three checks ask whether the result depends on that. Table F2 reports the weights behind each check,

Table F4 the gain from a supportive leader under each set of weights, and Figure F1 the full curves.

Table F2: Structural Weights Behind the Extrapolation Checks

	Headline	Norm-side split	Perceived use ≤ 5	Perceived use > 5	Perceived use < 5	Low-use villages	High-use villages
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Conformity (α_0)	0.90 (0.04)		0.82 (0.07)	0.92 (0.04)	0.80 (0.09)	1.02 (0.08)	0.99 (0.10)
below neutral card (α_0^-)		1.03 (0.05)					
above neutral card (α_0^+)		0.77 (0.05)					
Leader supports (γ^+)	0.50 (0.02)	0.50 (0.02)	0.44 (0.04)	0.52 (0.02)	0.48 (0.07)	0.50 (0.03)	0.49 (0.04)
Leader opposes (γ^-)	-0.12 (0.02)	-0.12 (0.02)	-0.08 (0.04)	-0.13 (0.02)	-0.04 (0.05)	-0.10 (0.04)	-0.15 (0.04)
Scripture pro (δ^+)	0.38 (0.02)	0.38 (0.02)	0.35 (0.05)	0.38 (0.02)	0.39 (0.06)	0.38 (0.04)	0.35 (0.05)
Scripture anti (δ^-)	-0.06 (0.02)	-0.06 (0.02)	0.05 (0.05)	-0.10 (0.03)	0.10 (0.08)	-0.06 (0.04)	-0.04 (0.05)
Leader salience (α_R)	-0.15 (0.03)	-0.15 (0.03)	-0.15 (0.03)	-0.15 (0.03)	-0.15 (0.03)	-0.16 (0.06)	-0.25 (0.08)
Scripture salience (α_S)	-0.24 (0.04)	-0.24 (0.04)	-0.24 (0.04)	-0.24 (0.04)	-0.25 (0.04)	-0.32 (0.09)	-0.24 (0.10)
Conformity = 1 (p)	0.007	0.507	0.009	0.032	0.025	0.760	0.937
Conformity equal (p)			0.039		0.072		
All five equal (p)			0.000		0.010		
Salience common (p)			0.217		0.001		
Respondents	635	635	145	490	67	635	635
Effective sample size						174	141

NOTE: Extended-model estimates as in Table 2, column (2), which column (1) repeats. Column (2) lets conformity differ below and above the neutral norm card. Columns (3) and (4) come from one model that separates women who perceive at most five of ten village women using family planning from the rest, with common salience terms, and column (5) draws the line below five. Columns (6) and (7) reweight the respondents to the low-use and high-use villages of Table F3 and treat the weights as fixed. Conformity = 1 uses α_0^- in column (2). The equality rows are likelihood-ratio tests of the subgroup model against equal weights.

The Range of the Norm The experiment did vary the norm far from Meru's. The norm cards told respondents that two, five, or eight of ten women in Kendi's village use family planning, so the conformity weight is identified from how women expect a woman to act in villages at 20 and 80 percent use. The figures in Section 6 shade the baselines outside that range. What the experiment cannot vary is who lives in such a village, which the next two checks address. The specification test of Section 5.3 adds one qualification. Table F2, Column 2, shows that women respond more strongly to the two-in-ten card than the linear model allows, with a conformity weight of 1.03 below the neutral card and 0.77 above it.

Under that split the full-reach gains barely move, at 60.9 rather than 62.0 points in a village at 25 percent and 36.3 rather than 39.2 in Meru. At one-in-five reach the split raises the gain in a low-use village, to 35.0 from 23.4 points, and lowers it in Meru, to 14.5 from 20.7, because Meru sits just above an even split, where the smaller conformity weight to the norm applies.

Table F3: Reweighting the Survey to Low-Use and High-Use Villages

	Survey (Meru)	Low-use villages	Survey reweighted	High-use villages	Survey reweighted
Age (years)	35.0	32.6	32.6	33.6	33.6
Living children	2.7	3.2	3.2	2.8	2.8
No schooling	0.07	0.20	0.20	0.01	0.01
Primary	0.70	0.38	0.38	0.44	0.44
Secondary	0.21	0.24	0.37	0.37	0.47
Higher	0.02	0.19	0.06	0.18	0.08
In a union	0.81	0.82	0.82	0.82	0.82
Catholic	0.08	0.23	0.23	0.18	0.18
Household assets (of 8)	2.5	3.2	3.2	3.7	3.7
Electricity	0.22	0.44	0.34	0.62	0.42
Radio	0.57	0.58	0.75	0.74	0.82
Television	0.21	0.45	0.37	0.63	0.55
Mobile phone	0.90	0.92	0.91	0.97	0.97
Refrigerator	0.01	0.17	0.02	0.12	0.03
Bicycle	0.04	0.16	0.12	0.19	0.18
Car or truck	0.02	0.12	0.05	0.10	0.09
Solar panel	0.50	0.33	0.60	0.35	0.66
Modern CPR, women in a union (%)	68.9	22.9		79.8	
Modern CPR, all women (%)	50.8	22.8		56.1	
Women (villages)	635 (27)	2,224 (183)		5,193 (437)	
Effective sample size			174		141
Largest weight			15.9		17.3

NOTE: Means for the survey respondents and two KDHS 2022 targets, the ever-married Christian women aged 18 to 49 in Christian-majority clusters with at least ten interviewed women, where modern prevalence among women in a union is at most 35 percent (low-use villages) or at least 70 percent (high-use villages). Entropy-balancing weights match the survey to each target on age, living children, schooling, union status, Catholic affiliation, and household assets, and the remaining rows are not balanced. The all-women prevalence places each target on the baseline axis of the counterfactuals. The effective sample size is $(\sum_i w_i)^2 / \sum_i w_i^2$ out of 635, and the largest weight is relative to a mean of one.

Women Who See a Conservative Village One in four respondents (145 women) perceives at most five of ten women her age in her village using family planning, on average 3.9 of ten against 9.2 among the rest. They resemble the rest in age, schooling, number of children, household assets, and income, but fewer of them use a modern method themselves (75 against 87 percent). Their conformity weight is lower, 0.82 against 0.92, and a supportive leader moves them somewhat less, 0.44 against 0.52. A community of such women would gain less from a supportive leader, not more. At full reach in a village at 25 percent the gain

is 55.7 points against 63.4 for the rest, and at one-in-five reach 14.6 against 27.4, although neither difference is precise. The narrower group below five of ten (67 women, on average 2.5 of ten) gives the same picture, though there the salience terms also differ from the rest.

Reweighting to Low-Use and High-Use Villages The third check changes who is in the sample. Table F3 describes two target populations from KDHS 2022, the ever-married Christian women aged 18 to 49 in Christian-majority clusters where at most 35 percent, or at least 70 percent, of women in a union use a modern method. The low-use villages average 22.9 percent use among married women and are concentrated in West Pokot, Samburu, and Turkana. The high-use villages average 79.8 percent. Compared with either

Table F4: A Supportive Leader's Gain Under Alternative Weights

Baseline modern CPR	Full reach			One in five leaders		
	25%	50.8%	75%	25%	50.8%	75%
<i>Panel A. Equilibrium Gain (pp)</i>						
(1) Headline	62.0 [58.6, 64.6]	39.2 [37.4, 40.8]	17.1 [15.8, 18.2]	23.4 [16.2, 38.9]	20.7 [16.9, 24.8]	5.8 [5.2, 6.4]
(2) Norm-side split	60.9 [57.7, 63.5]	36.3 [33.9, 38.5]	16.2 [14.9, 17.5]	35.0 [27.7, 43.7]	14.5 [11.5, 18.4]	4.9 [4.4, 5.5]
(3) Perceived use at most 5 of 10	55.7 [45.2, 63.2]	35.9 [30.3, 39.9]	15.4 [12.3, 17.7]	14.6 [9.5, 32.1]	15.8 [11.0, 23.2]	4.8 [3.5, 6.0]
(4) Perceived use above 5 of 10	63.4 [60.2, 66.0]	40.0 [38.1, 41.6]	17.6 [16.3, 18.8]	27.4 [18.0, 43.5]	22.1 [18.2, 26.3]	6.1 [5.5, 6.9]
Difference, (3) minus (4)	-7.6 [-17.4, 0.2]	-4.1 [-9.1, -0.1]	-2.2 [-5.1, -0.1]	-12.8 [-28.5, 1.9]	-6.3 [-11.9, 0.8]	-1.3 [-2.6, -0.0]
(5) Perceived use below 5 of 10	56.1 [40.5, 66.4]	36.5 [28.5, 42.1]	16.1 [11.5, 19.2]	14.6 [8.3, 39.6]	15.4 [9.2, 25.4]	5.0 [3.2, 7.0]
(6) Low-use villages	66.2 [62.1, 69.2]	41.2 [38.5, 43.3]	17.7 [15.7, 19.3]	49.1 [21.7, 62.2]	27.9 [20.2, 34.3]	6.9 [5.6, 8.3]
(7) High-use villages	63.0 [56.9, 67.1]	38.7 [35.7, 41.1]	15.8 [13.2, 17.9]	42.2 [15.4, 60.9]	25.2 [16.5, 33.2]	5.9 [4.8, 7.2]
<i>Panel B. Shift Outlasts the Campaign (% of Draws)</i>						
(1) Headline	0	0	0			
(2) Norm-side split	7	0	0			
(3) Perceived use at most 5 of 10	0	1	0			
(4) Perceived use above 5 of 10	0	1	0			
(5) Perceived use below 5 of 10	0	1	0			
(6) Low-use villages	26	61	0			
(7) High-use villages	17	45	0			

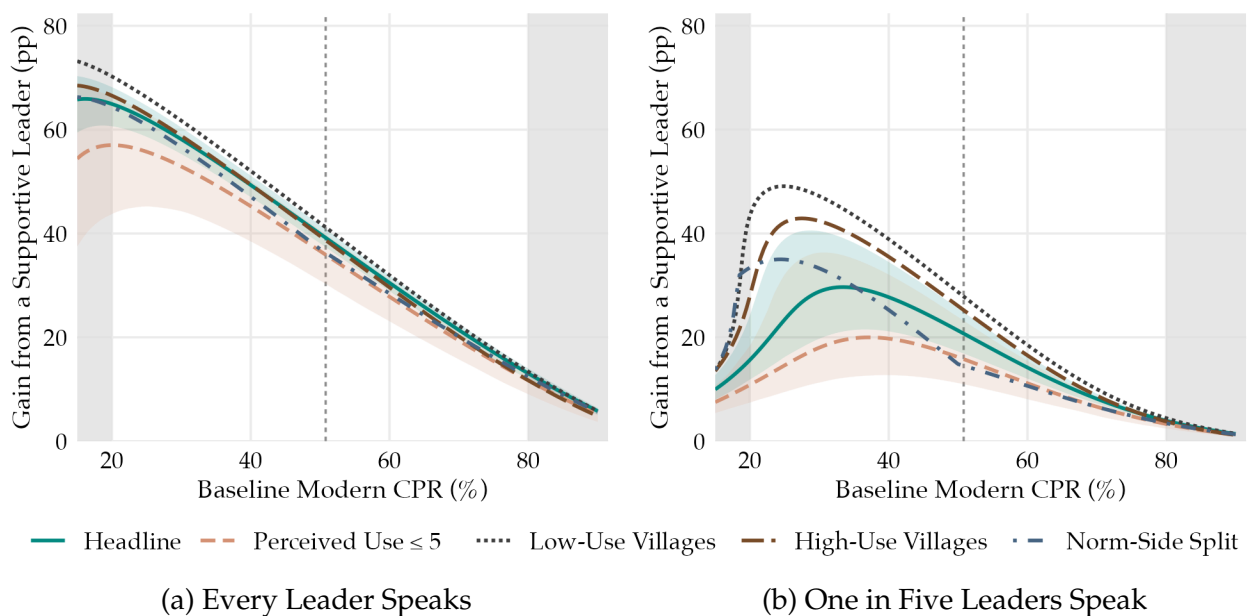
NOTE: Panel A reports the gain in equilibrium modern prevalence from a supportive leader, in percentage points, at the baseline in the column head, with 95% bootstrap intervals in brackets (500 draws over each column's weights in Table F2). Each row recalibrates the private-preference intercept to the baseline. Full reach has every leader speaking and one in five a fifth of leaders, with all women conforming to one village norm. Under the norm-side split, conformity is α_0^- below an even split and α_0^+ above it, and the difference row uses joint draws of the subgroup model. Panel B reports the share of draws in which a full-reach campaign leaves the village more than one point above its baseline after the leaders fall silent, which requires conformity above one. Near an even split such a baseline is itself unstable, so the middle column overstates what a campaign could lock in.

target, the survey's women are older, less often Catholic, less schooled, and own fewer assets.

Entropy balancing reweights the respondents to match each target on age, children, education, union status, Catholic affiliation, and assets, leaving effective samples of 174 and 141. Table F2, Columns 6 and 7, show the reweighted conformity weights rise toward one, to 1.02 and 0.99, while the authority loadings barely move. Table F3 shows that the reweighting nearly triples the weight on Catholic respondents, whose conformity weight is close to one in Table E2. With these weights the gain in a low-use village is larger, 66.2 points at full reach and 49.1 at one-in-five reach. Table F4, Panel B, displays that because their conformity weights reach one, a full-reach campaign outlasts itself in a quarter of the draws in a village at 25 percent.

Across all seven sets of weights, the full-reach gain is 56 to 66 points in a village at 25 percent, 36 to 41 in Meru, and 15 to 18 in a village at 75 percent. Therefore, that a leader is worth most where the norm is most hostile does not rest on Meru's weights. What does depend on them is the size of the gain at partial reach in low-use villages, which is imprecise within each check and differs across them, and whether a campaign there could outlast itself, which turns on whether conformity in such villages sits above one.

Figure F1: A Supportive Leader's Gain Under Alternative Weights



NOTE: Equilibrium gain from a supportive leader against baseline use, as in Figure 12, under the weights of Table F2, namely the headline (column 1), the norm-side split (2), perceived use ≤ 5 (3), and low-use and high-use villages (6 and 7). Bands are 95% bootstrap intervals for the headline and perceived-use weights. Gray regions lie outside the norm cards, and the vertical line marks Meru.

Appendix G: Quasi-Experimental Design, Data, and Results

This appendix section documents the full faith-based-program evidence that Section 7 presents in compressed form. The programs and institutional context, the data sources, the construction of the religious moderators, the estimating equations, the full county and distance result tables, and the additional robustness are all reported here, complementing the main-text summary.

G.1: Institutional Context and Programs' Overview

I evaluate two faith-based family planning (FP) interventions in Kenya: (1) the CHAK-Packard Program (2013–2017) and (2) the Christian Advocacy for Family Planning in Africa (CAFPA) Program (2014–2021). Both were implemented by the Christian Health Association of Kenya (CHAK) with technical support from Christian Connections for International Health (CCIH), engaging religious leaders as primary FP advocates using evidence-based training rooted in scriptural interpretation (Bormet et al., 2021; Ruark et al., 2019). Despite this shared foundation, the programs differed in what accompanied the leaders' message and in location. Figure 14a in the main text shows the counties of implementation. CHAK-Packard was concentrated in western Kenya, around Lake Victoria. CAFPA's Phase 1–2 counties, Kiambu, Murang'a, and Meru, lie in central Kenya, and its Phase 3 counties, Kilifi, Narok, and Machakos, in the Coast, Rift Valley, and Eastern regions. Due to data privacy constraints, this analysis does not track individual religious leaders but instead anchors treatment to the church-affiliated health facility at which religious leader (RL) training took place, as identified from CCIH and CHAK program records. Figure G5 maps all CHAK facilities in Kenya along with places of worship and the Ministry of Health facilities.

CHAK-Packard was a bundled intervention. RL engagement was its primary aim, complemented by community health volunteer (CHV) community-based commodity distribution and facility-level supply reinforcement. CAFPA was a dual-channel program. That is, religious leaders preached to their congregations and communities while simultaneously lobbying county governments for FP budget allocations, FBO inclusion in supply chains, and costed implementation plans. What CAFPA lacked relative to CHAK-Packard was CHV community-based commodity distribution. CHVs were available during RL messaging events for demonstrations and health advice, though not deployed for commodity distribution. The religious leaders' advocacy secured county commitments to supply the 33 participating faith-based facilities, and the share of these facilities reporting stock-outs

Table G1: CHAK-Packard Versus CAFPA

	CHAK-Packard	CAFPA
Funder	David and Lucile Packard Foundation	Bill and Melinda Gates Foundation
Dates	January 2013 – December 2017	October 2014 – December 2021
Primary Mechanism	RL community messaging + CHV commodity distribution + facility supply reinforcement	RL community messaging + county government advocacy, including FBO commodity supply
Counties	Siaya, Kakamega, Vihiga, Homa Bay	Kiambu, Meru, Murang'a; Phase 3: Kilifi, Narok, Machakos
Religious Leaders	62 (30 Phase 1 + 32 Phase 2)	55 (14 Phases 1–2 + 41 Phase 3)
CHVs	112 (60 Phase 1 + 52 Phase 2)	At messaging events only; no commodity distribution
Selection Criterion	High TFR, unmet need, low age at first birth	BMGF priority counties
Concurrent Programs	CHAK-E2A (Nyanza, 2014–16) APHIAplus Western Kenya (2011–16) Reversing the Stall (Siaya, 2009–14)	CHAK-E2A (Upper Eastern, 2014–16) APHIAplus Kamili (Kiambu, Meru, Murang'a, 2011–17) DESIP (Narok, 2018–) Afya Pwani (Kilifi, 2016–21) ACHAP AFYA (Kilifi, 2020–23)

NOTE: RL is religious leader, CHV community health volunteer, FBO faith-based organization, TFR total fertility rate, and BMGF the Bill and Melinda Gates Foundation. CHAK-E2A worked in eight CHAK facilities in Nyanza and Upper Eastern, which are not individually identified. Facility and program records were shared by Mona Bormet and CHAK, and information stems from Ruark et al. (2019) and Bormet et al. (2021).

fell from 97% in 2014 to 20% in 2017 (Bormet et al., 2021).

Before implementation, CHAK and CCIH worked with church bodies to establish written FP position statements to ensure RL credibility within their own congregations.¹ Both programs used the CCIH definition of family planning, “enabling couples to determine the number and timing of pregnancies, including the voluntary use of methods for preventing

¹For example, the following were used as support for FP regarding the education of women and girls:

By avoiding unintended pregnancies, family planning enables girls and women to stay in school, enabling them to increase lifetime earnings and fulfill their potential - “Hold on to instruction, do not let it go; guard it well, for it is your life.” (Proverbs 4:13)

pregnancy—not including abortion—harmonious with their values and religious beliefs.” FP messaging was integrated into existing religious functions, including sermons, weddings, and *Barazas* (communal forums), reducing informational barriers and the social cost of contraception within faith communities.

CHAK-Packard Program The CHAK-Packard program (2013–2017) selected counties on the basis of high fertility rates, high unmet need for contraception, and low median age at first birth, explicit criteria reflecting targeted, non-random assignment that motivates a DiD design. The program’s primary innovation was RL engagement, with 62 trained religious leaders delivering FP messages through sermons, community dialogue days, women’s and youth groups, and direct client referrals. See Figure G1 for precise details and the implementation timeline. Complementing the RL component, 112 community health volunteers distributed pills, condoms, and CycleBeads at community level, reducing the transaction cost of obtaining contraception without a facility visit, while supply reinforcement (buffer stock of IUDs and implants; on-the-job LARC training for nurses) ensured that referrals from RLs and CHVs reached stocked facilities.

Figure G1: CHAK-Packard Program Implementation



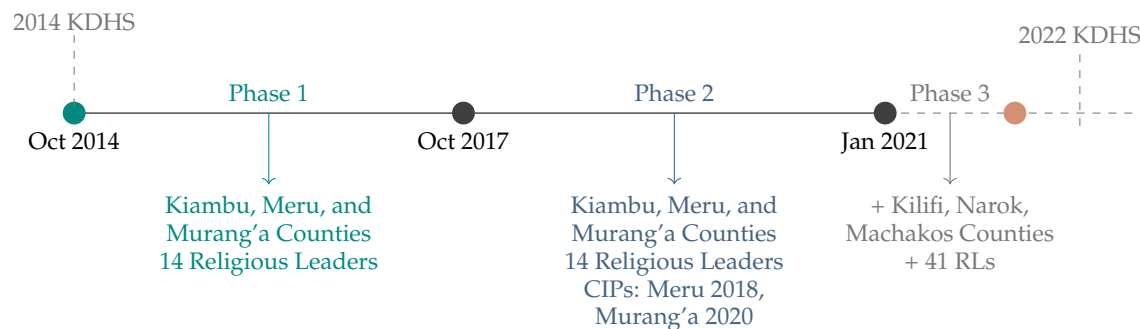
NOTE: ● program start; ● program end; ● phase transition. Gray dashed lines mark the 2014 and 2022 KDHS survey waves. Namasoli Health Centre in Kakamega was introduced in Phase 1 and remained active through Phase 2; no new Kakamega facility was added in Phase 2.

By program end, religious leaders had reached 675,000 people and generated 85,810 client referrals, with FP visits increasing sixfold at Phase 1 facilities (Ruark et al., 2019). The concurrent CHAK-E2A project (Pathfinder/USAID; 8 other CHAK facilities in Nyanza and Upper Eastern regions, September 2014–June 2016) creates a county-level FP culture spillover risk. The six-year gap between E2A’s end and the 2022 KDHS outcome wave mitigates this risk. Ruark et al. (2019) provides a full program overview.

CAFPA Program CAFPA (2014–2021) operated through two complementary channels. First, trained religious leaders delivered community messaging in their congregations (sermons, youth forums, community events, and radio and media outreach), reaching

an estimated 14,500 people across program counties in 2018–2019 (Bormet et al., 2021). Second, these same leaders engaged in county government policy advocacy, lobbying for FP budget allocations, FBO inclusion in supply chains, and formal costed implementation plans (CIPs). There was no community-based commodity distribution and no program-run supply intervention at facilities.

Figure G2: CAFPA Program Implementation



NOTE: Drawn to scale. ● program start; ● program end; ● phase transitions. Phase 3 (gray, dashed) is excluded from the primary analysis because of concurrent FP programs in its counties (Afya Pwani and ACHAP AFYA in Kilifi, DESIP in Narok). Gray dashed lines mark the 2014 and 2022 KDHS waves.

Figure G2 outlines the implementation timeline. Phases 1 and 2 (October 2014 – January 2021) involved the same 14 trained religious leaders in Kiambu, Meru, and Murang'a, with advocacy deepening through the Meru CIP (2018) and Murang'a CIP (December 2020). Phase 3 (January–December 2021) expanded to Kilifi, Narok, and Machakos, but concurrent programs in those counties, namely Afya Pwani (Kilifi, 2016–2021), ACHAP AFYA (Kilifi, 2020–2023), and DESIP (Narok, 2018–), preclude credible identification. Given these concerns, the primary analysis restricts to Phases 1 and 2. Bormet et al. (2021) provides commentary on the program.

G.2: Data

The outcome data are the 2003, 2008, 2014, and 2022 waves of the Kenya Demographic and Health Survey (KDHS), repeated cross-sections of women aged 15–49 with information on fertility, contraceptive use, and demand for FP. Because clusters are resampled between waves, identification comes from geographic variation in program exposure across survey rounds rather than from following the same women. The KDHS records GPS coordinates for each survey cluster, displaced for privacy by up to 2 km in urban areas and 5 km in rural areas (1% of rural clusters by up to 10 km). Treatment is built from CHAK and CCIH program records. I geolocate the 21 church-affiliated health facilities that hosted religious-leader training, cross-reference them with the Kenya Master Health Facility

Registry (KMHFR), and compute the straight-line distance from each cluster to the nearest facility of each program, using each facility’s year of program inception. County and ward boundaries are from geoBoundaries.

The moderators come from three sources. Ward-level trust in religious leaders, religious affiliation, and religious practice are small-area estimates from Afrobarometer rounds 2–8 (2003–2021), poststratified on the 2019 Kenya Population and Housing Census (IPUMS International), with WorldPop population rasters allocating census cells to wards and satellite-derived covariates (night-time lights, travel time to cities, conflict deaths, and vegetation) from AidData’s GeoQuery (see Appendix G.3). Ward counts of CHAK, other faith-based, and Ministry of Health facilities come from the KMHFR and CHAK’s facility records, and the Protestant share of places of worship from OpenStreetMap. The World Values Survey is used only descriptively, in Appendix A.

G.3: Baseline Religiosity via Hierarchical Small-Area Estimation

To construct the baseline religious moderators M_w that enter the moderated difference-in-differences specification of Equation (G7), I implement a hierarchical framework that progresses from direct survey estimation at the national level, through multilevel regression with poststratification (MRP) at coarser administrative units, to hybrid small-area estimation (SAE) at the ward level (Claassen & Traunmüller, 2020; Gelman & Little, 1997; Park et al., 2004). It integrates Afrobarometer data, the 2019 Kenya Population and Housing Census ledger, and satellite-derived spatial proxies while preserving national representativeness across Kenya’s 47 counties (Admin 1), 290 sub-counties (Admin 2), and $\approx 1,450$ wards (Admin 3). The framework is built to overcome two challenges in the Afrobarometer data: (1) the absence of observations in some geographic units in particular waves, and (2) the high variance of small-unit estimates where respondent counts are low.

Survey Specification and National Baseline The primary data source is the Afrobarometer (waves 2–8), which measures indicators of religiosity from 2003 to 2021. The sample is representative at the national level, and national proportions for each religious indicator r at time t are obtained via design-weighted direct estimation,

$$\bar{r}_t = \frac{\sum_i w_i r_{it}}{\sum_i w_i},$$

where w_i is the Afrobarometer sampling weight.

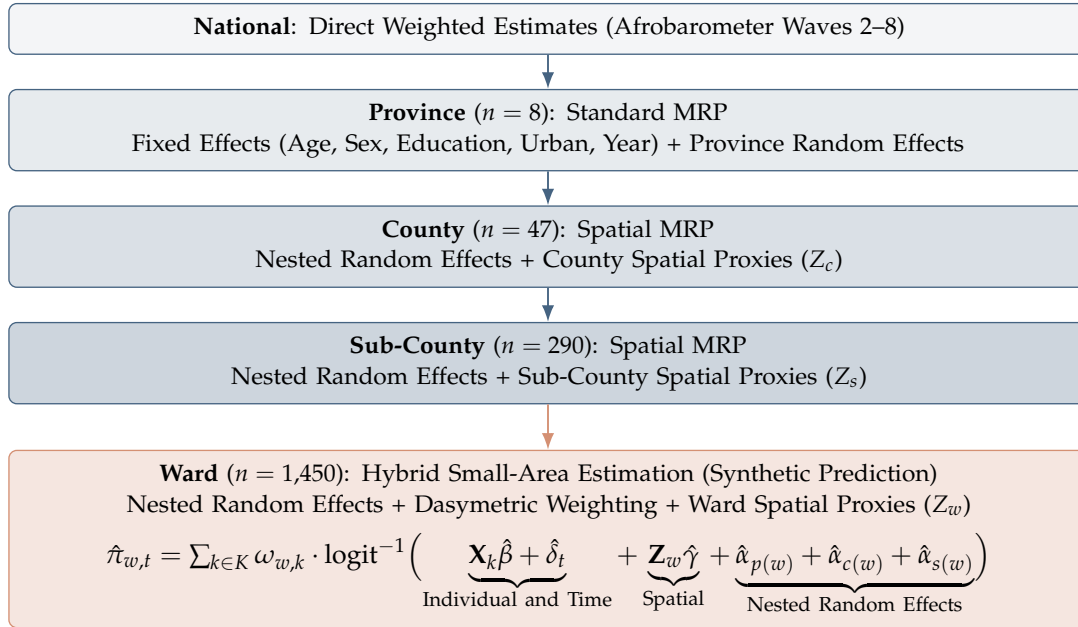
Multilevel Model Specification To move beyond national averages, I estimate a multi-

level logistic regression relating religiosity to individual demographics, temporal trends, and geographic heterogeneity. The probability that individual i in demographic cell k , located in sub-county s , county c and province p , holds religious indicator r at time t is modeled as

$$\text{logit}(\Pr(r_{it} = 1)) = \beta_0 + \mathbf{X}_k\beta + \delta_t + \mathbf{Z}_l\gamma + \alpha_p + \alpha_c + \alpha_s \quad (\text{G1})$$

where $\mathbf{X}_k$ is a vector of individual-level fixed effects for age category, sex, education, and urban/rural status. Wave fixed effects δ_t absorb common time trends, including genuine shifts in religiosity and any wave-specific sampling variation. Level-specific spatial covariates, such as night-time lights, travel time to the nearest city, conflict mortality, and vegetation indices, are denoted $\mathbf{Z}_l$ ($l = c, s$) and capture environmental and economic drivers of religiosity. The terms $\alpha_p, \alpha_c, \alpha_s$ are a nested hierarchy of random intercepts with $\alpha_l \sim N(0, \sigma_l^2)$. These random intercepts capture latent geographic variation not explained by observables and let the model borrow strength across the hierarchy, so that a ward with few or no respondents is shrunk toward its sub-county mean, which is in turn informed by county and regional trends.

Figure G3: Hierarchical Spatial-MRP Framework



NOTE: Each level borrows strength from the level above, so that sparsely sampled units are shrunk toward their parent unit. Ward estimates are synthetic predictions from the fitted model, using ward spatial covariates and census cells allocated to wards with WorldPop population rasters.

Poststratification and the Population Ledger Using the estimated coefficients from Equation (G1), I poststratify onto a population ledger derived from the 2019 Kenya Population and Housing Census. The ledger cross-tabulates the population into K demographic cells

(for example, “primary-educated females, aged 18–29, in rural sub-county s ”). For each administrative unit, the MRP estimate is the population-weighted average of the predicted probabilities for all cells k within it,

$$\hat{\pi}_{\text{unit},t} = \sum_{k \in K} \omega_{\text{unit},k} \cdot \text{logit}^{-1}(\hat{\eta}_{kt}),$$

where $\omega_{\text{unit},k}$ is the census share of cell k within the unit and $\hat{\eta}_{kt}$ is the fitted linear predictor of Equation (G1) for cell k in year t .

Constraints and Normalization Because religious affiliations are mutually exclusive, independent binomial models for each category ($j = 1, \dots, J$) need not yield predicted probabilities that sum to unity ($\sum \hat{p} \neq 1$). To ensure logical consistency in the spatial dataset, I apply a post-estimation proportional normalization,

$$\tilde{p}_{jt} = \frac{\hat{p}_{jt}}{\sum_{m=1}^J \hat{p}_{mt}}, \quad (\text{G2})$$

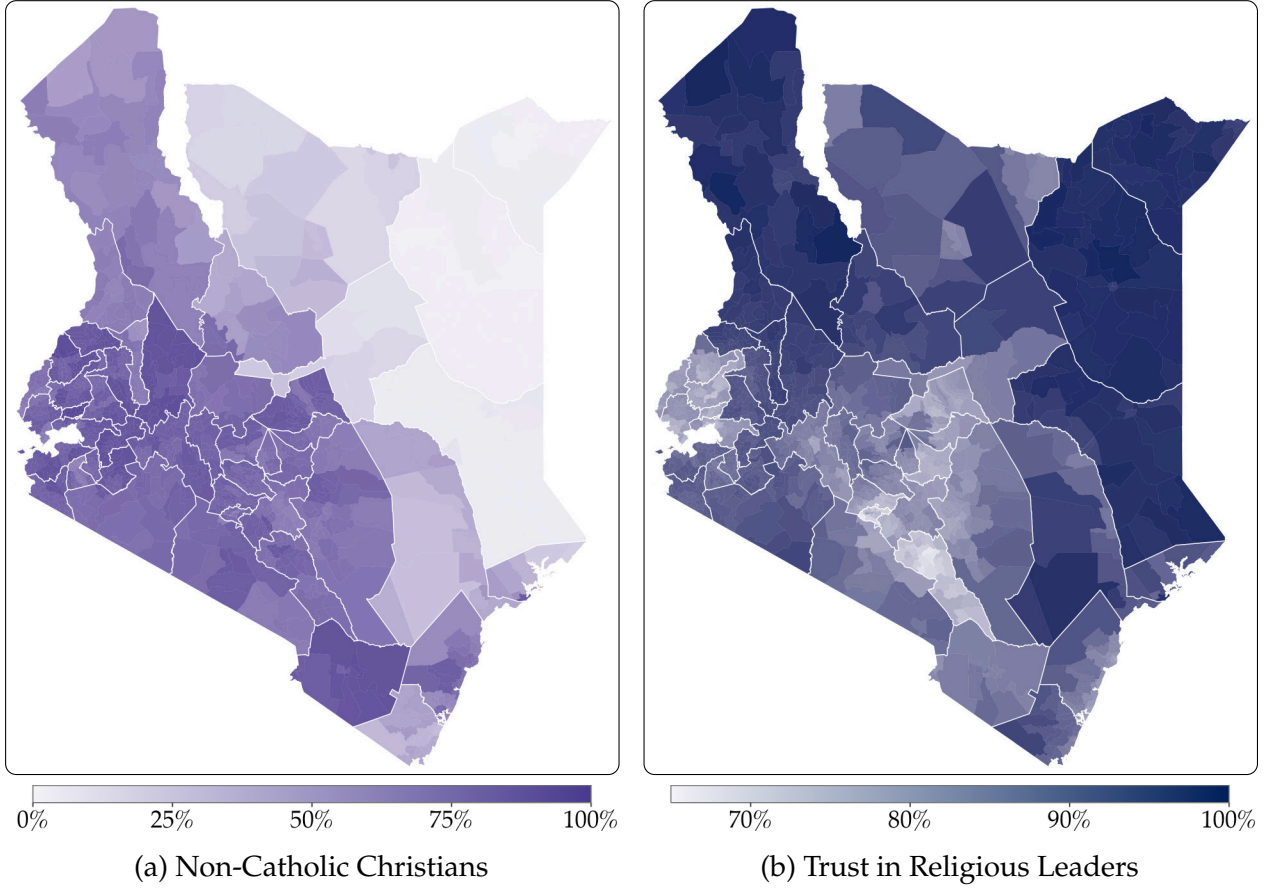
where $\hat{p}_{jt}$ is the raw predicted proportion for category j and $\tilde{p}_{jt}$ the constrained estimate used for mapping. Predictions are generated for every administrative unit in every survey year, including units with no Afrobarometer respondents.

Ward-Level Hybrid SAE The finest resolution unit, the ward ($n = 1,450$), presents a unique challenge, as no census ledger exists at this resolution and many units are unsampled in the Afrobarometer. I resolve this through dasymetric weighting and synthetic prediction. First, I map IPUMS-derived census demographics from the sub-county boundaries to the current GeoBoundaries ward polygons, using WorldPop raster data (100m resolution) to compute high-precision disaggregation weights. The ward-level estimate is then

$$\hat{\pi}_{w,t} = \sum_{k \in K} \omega_{w,k} \cdot \text{logit}^{-1}(\mathbf{x}_k \hat{\beta} + \hat{\delta}_t + \mathbf{Z}_w \hat{\gamma} + \hat{\alpha}_{p(w)} + \hat{\alpha}_{c(w)} + \hat{\alpha}_{s(w)}), \quad (\text{G3})$$

where $\mathbf{Z}_w$ are spatial proxies measured at the ward level. For unsampled wards, the model relies on the Best Linear Unbiased Predictions from the nested random effects of the parent sub-county and county, refined by the ward’s spatial characteristics, so that ward-level estimates are anchored in regional realities and local environmental contexts even without direct survey data. This hybrid approach is summarized in Figure G3, and the resulting ward-level estimates for the proportion of non-Catholic Christians and for trust in religious leaders appear in Figure G4.

Figure G4: Ward-Level Estimates



NOTE: Estimated 2014 ward-level shares of adults who are non-Catholic Christians (panel a) and who trust religious leaders at least somewhat (panel b), from the small-area estimator described in this subsection. Data are from the Afrobarometer and the 2019 Kenya Population and Housing Census.

G.4: Empirical Strategy

To estimate the impact of the faith-based FP programs, I employ a generalized two-way fixed effects (TWFE) difference-in-differences framework. The KDHS provides four usable waves (2003, 2008, 2014, 2022), but reliable county-level representativeness is only achieved from 2014 onward. The identification therefore collapses to a 2×2 design with 2014 as the pre-period and 2022 as the post-period. I estimate two complementary specifications, a county-level binary treatment and an individual-level continuous distance exposure, followed by religious moderator extensions that allow treatment effects to vary with the local religious context and supply-side infrastructure.

County-Level TWFE DiD The primary county-level specification is

$$Y_{it} = \beta_c (\text{Treat}_c \times \text{Post}_t) + \gamma_c + \delta_t + \mathbf{X}'_{it} \lambda + \varepsilon_{it} \quad (\text{G4})$$

where Y_{it} is an outcome for woman i in year t : birth spacing (the mean preceding birth interval, in months, for births in the past five years, defined for women with at least one such interval), the five-year birth indicator, unmet need for contraception, or modern contraceptive use. $\text{Treat}_c \in \{0, 1\}$ is a county-level program indicator equal to one if the county contains at least one health facility anchoring a faith-based RL training program. $\text{Post}_t \in \{0, 1\}$ equals one for the 2022 survey wave. The county fixed effects γ_c absorb all time-invariant county-level characteristics including baseline outcome levels, long-run cultural norms, and pre-existing health infrastructure. Year fixed effects δ_t absorb national trends in family planning policy and survey-wave differences. $\mathbf{X}_{it}$ is a vector of individual-level controls, including age, age², education, marital status, wealth index score, and urban indicator. The coefficient β_c captures the average treatment effect on the treated (ATT).

Individual Distance-Decay DiD The continuous-exposure specification measures treatment intensity as a function of each household cluster's Euclidean distance to the nearest program facility:

$$Y_{it} = \beta_d \left(\text{Exp}_{i,p} \times \text{Post}_t \right) + \text{Exp}_{i,p} + \gamma_c + \delta_t + \mathbf{X}'_{it} \lambda + \varepsilon_{it} \quad (\text{G5})$$

where the continuous exposure index is

$$\text{Exp}_{i,p} = \frac{-\ln(\max(d_{i,p}, 0.1)) - \mu_{\text{Exp}}}{\sigma_{\text{Exp}}} \quad (\text{G6})$$

$d_{i,p}$ is the Euclidean distance (km) from DHS cluster i 's centroid to the nearest program (p) facility, and $\text{Post}_t \in \{0, 1\}$ again equals one for the 2022 wave. The floor at 0.1 km prevents singularity for near-coincident clusters. $\text{Exp}_{i,p}$ is monotonically increasing in proximity and is standardized using the full-sample 2014 mean μ_{Exp} and standard deviation σ_{Exp} , so that β_d measures the effect of a 1 SD increase in log-proximity. Because DHS clusters are resampled across waves, cluster fixed effects are infeasible and the county fixed effects do not absorb within-county proximity. The exposure main effect $\text{Exp}_{i,p}$ is therefore included so that it nets out the permanent (pre-period) spatial gradient and β_d identifies the *change* in the proximity–outcome gradient from 2014 to 2022 rather than a post-period cross-sectional correlation. Because the county fixed effects absorb only level differences, β_d pools proximity variation within and across program and control counties.

Heterogeneity by Religious Moderators To investigate whether program effects vary with the local religious context and faith-based infrastructure, I extend the county specifi-

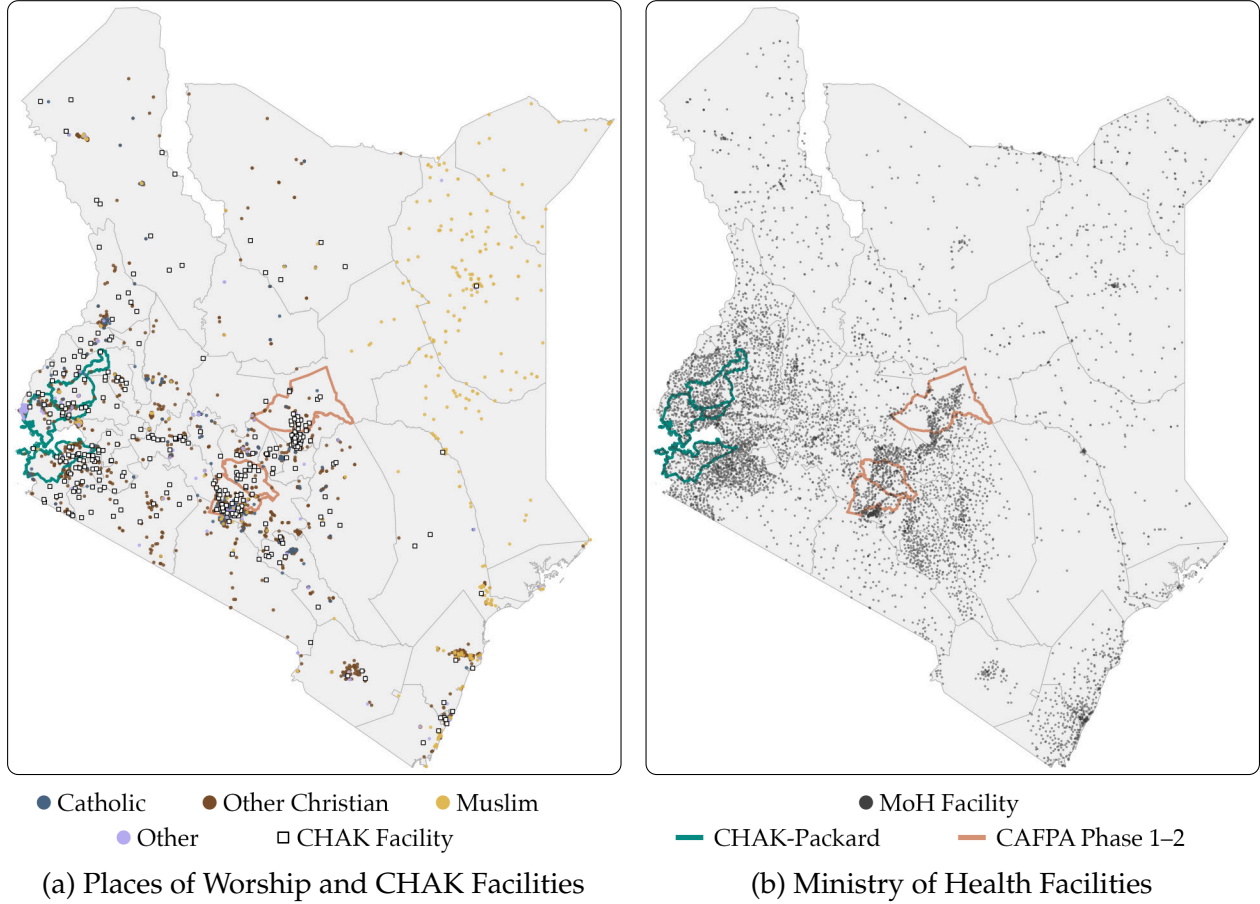
cation with a moderator M_w :

$$Y_{it} = \beta_1 (T_i \times \text{Post}_t) + \beta_2 M_w + \beta_3 (\text{Post}_t \times M_w) + \beta_4 (T_i \times M_w) + \beta_5 (T_i \times \text{Post}_t \times M_w) + \gamma_c + \delta_t + \mathbf{X}'_{it} \lambda + \varepsilon_{it} \quad (\text{G7})$$

where $T_i = \text{Treat}_c$ is the county program indicator. Each moderator enters a separate regression and, apart from the individual Protestant indicator, is standardized on the full-sample distribution. The attitudinal moderators are measured at the 2014 baseline from the small-area estimator of Appendix G.3: (1) an individual Protestant indicator; (2) the ward share of adults who trust religious leaders; (3) the ward share of non-Catholic Christians; and (4) ward religious practice intensity. The structural moderators describe ward-level supply: (5) the count of CHAK health facilities; (6) the Protestant and Evangelical share of all places of worship (faith-network density); (7) the count of faith-based-organization facilities in the KMHFR; and (8) the count of Ministry of Health facilities. Including (8) provides a falsification test: if the faith-based interactions reflect faith-network mechanisms, the Ministry of Health interaction should be null. β_1 captures the program effect at the moderator's reference value, and β_5 , the coefficient of interest, is the differential program effect per standard deviation of the moderator. The interaction $\text{Post}_t \times M_w$ controls for differential pre-post outcome trends correlated with the moderator. Figure G5 maps the facilities and places of worship behind the structural counts.

Inference and Identification Standard errors are clustered at the county level throughout, which is the level at which program assignment and the dominant policy and health-administration shocks operate. Because only three counties are treated under CAFPA and four under CHAK-Packard, conventional cluster-robust standard errors are unreliable, and I therefore supplement them with three checks drawn from the recent difference-in-differences literature (Baker et al., 2026; Roth et al., 2023): (1) a synthetic difference-in-differences estimator that reweights control units and pre-periods to match the treated pre-trend before differencing (Arkhangelsky et al., 2021); (2) a wild cluster bootstrap to correct small-cluster inference; and (3) for CHAK-Packard's staggered Phase 1 (2013) and Phase 2 (2015) rollout, the Callaway–Sant'Anna group-time estimator (Callaway & Sant'Anna, 2021). In the tables, stars use the conventional standard errors and wild cluster bootstrap p -values are reported in brackets. Identification in all specifications rests on the parallel-trends assumption that treated and control units would have followed common outcome paths absent the programs. I assess it with a joint test of the 2003 and 2008 event-study leads in main-text Table 3, the event-study coefficients in Figure G7, and the relative-magnitude bounds of Rambachan and Roth (2023) in Table G9. With four treated

Figure G5: Faith-Based and Government Health Infrastructure



NOTE: Panel (a) shows places of worship by religion and the facilities of the Christian Health Association of Kenya (CHAK). Panel (b) shows Ministry of Health (MoH) facilities, the falsification counterpart of the faith-based moderators in main-text Table 3. OpenStreetMap coverage of places of worship is incomplete, especially outside towns. Data are from the Kenya Master Health Facility Registry and OpenStreetMap.

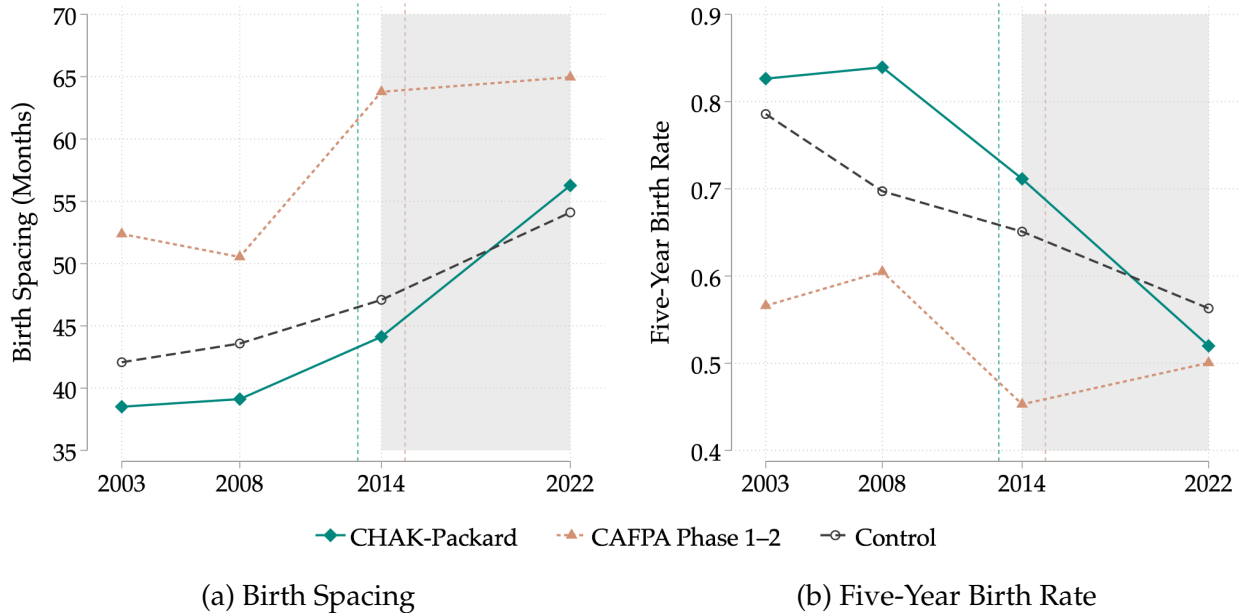
counties and two sparse pre-treatment waves, the pre-trend test has low power.

G.5: Results

This section reports the full results behind main-text Table 3. CHAK-Packard is the program I evaluate, and I report CAFPA for completeness. Both programs trained and mobilized religious leaders through the same CHAK/CCIH architecture, but they differed in what accompanied that message. CHAK-Packard paired religious-leader engagement with community health volunteers and the reinforced supply of its own network of faith-based facilities, while CAFPA paired the same religious-leader engagement with county-government policy advocacy, without a parallel community-distribution arm. All estimates use the county-level 2×2 design. Stars follow the conventional standard errors and the

tables also report wild cluster bootstrap p -values, so I read the estimates with caution. I report the CAFPA estimates in full, but I do not read them causally, for the reasons in the last paragraph of this section.

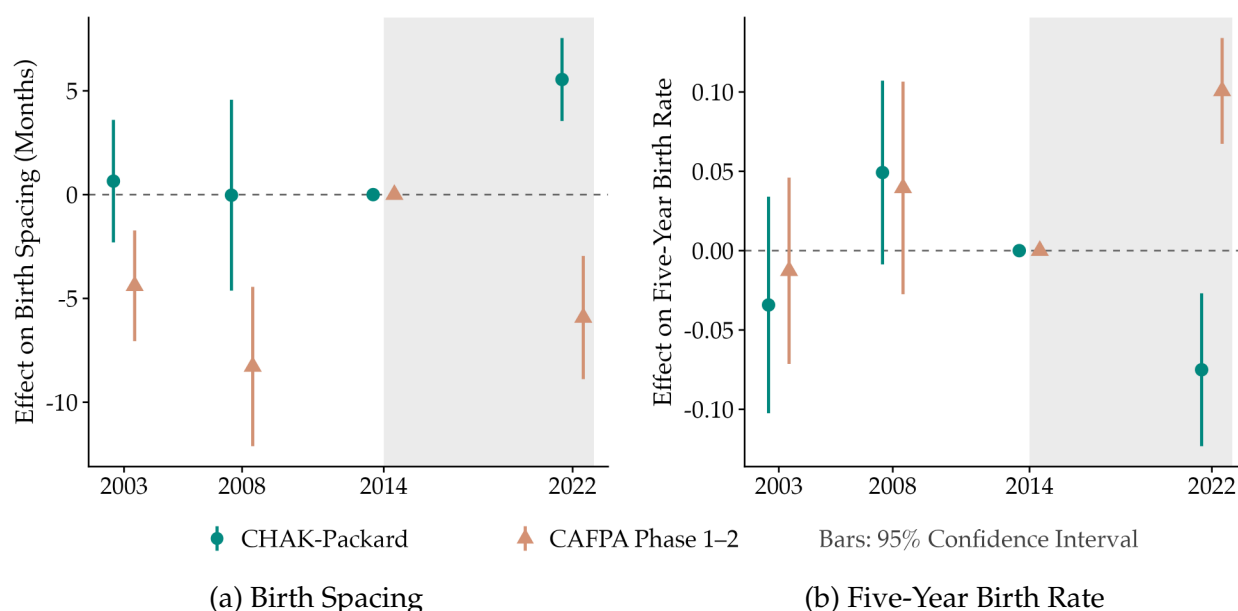
Figure G6: Birth Spacing and Births by Program Group, 2003–2022



NOTE: Survey-weighted county means of (a) the mean preceding birth interval, in months, and (b) the share of women with a birth, both over the past five years, averaged within the program and control groups for counties with at least 25 interviewed women in every wave. Dashed lines mark program inception and the gray band the 2014–2022 difference-in-differences window. Data are from the KDHS.

Average Effects Table G2 reports the county-level estimates for all outcomes, and Figures G6 and G7 plot the underlying group trajectories and the event-study coefficients. Where CHAK operated, the mean birth interval in the five years before the survey lengthens by roughly 5.5 months. Against a control mean of 46.9 months, births are spaced about 12 percent further apart. The synthetic DiD estimator that reweights the pre-period before differencing gives 4.1 months, and the Callaway–Sant’Anna estimator that handles the staggered rollout gives 2.3 months, which is not significant. CHAK-Packard also reduced the five-year birth rate by 7.7 percentage points, about 12 percent of the control mean, and by 9.6 and 11.4 points under the synthetic DiD and Callaway–Sant’Anna estimators. Fewer births within five years is what longer intervals imply, so I read the two results together. Unmet need does not move under the two-way fixed-effects estimator and falls under the other two, but its pre-trend test rejects, so I place no weight on it. Modern contraceptive use and the use of any method do not change. Current use is a snapshot of one day, while spacing and births accumulate over five years, and the survey cannot separate the two further.

Figure G7: Event-Study Estimates for Birth Spacing and Births



NOTE: County TWFE event-study coefficients of program $\times$ survey wave, relative to 2014, with 95% confidence intervals clustered by county. The 2003 and 2008 leads underlie the pre-trend test in main-text Table 3, and the gray band marks the difference-in-differences window.

Timing and Family Size Table G2 also shows that, where CHAK operated, the share of women whose last child was not wanted at the time of birth falls by 5.3 percentage points, the signature of women gaining the means to space births. Table G3 shows which other margins of fertility moved. Ideal family size, age at first birth, and the method mix do not change. The share of women who want no more children rises by 3.5 percentage points from a base of 42 percent, so part of the reduction in births may also come from limiting. Taken together, the program changed when women had children more than how many children they wanted.

Where the Effect Is Larger Tables G4 and G5 report the moderated estimates for all eight moderators. Baseline trust in religious leaders is the attitudinal moderator that behaves most consistently. Where trust is one standard deviation higher, CHAK's effect on birth spacing is 4.1 months larger, and this interaction survives the wild cluster bootstrap. Trust also deepens the reduction in births and in unmet need, though neither interaction survives the bootstrap. Faith-based delivery points the same way. Where CHAK-affiliated facilities are denser, the reduction in births is larger and births are spaced further apart. Denser faith-based health infrastructure also strengthens the program's effect on modern contraceptive use, though not under the bootstrap. The falsification test passes for every outcome, as the interaction with the count of Ministry of Health facilities is small and insignificant. The

Table G2: County Difference-in-Differences Estimates for All Outcomes

	CHAK-Packard	CAFPA Phase 1–2	2014 Control	Observations	
	Estimate (1)	Estimate (2)	Mean (3)	CHAK (4)	CAFPA (5)
<i>Panel A. Main Outcomes</i>					
Birth Spacing (Months)	5.54*** (0.99) [0.03]	-6.02*** (1.47) [0.03]	46.9	20,017	19,166
Births (Five Years)	-0.077*** (0.025) [0.06]	0.102*** (0.017) [0.03]	0.654	52,900	51,037
Unmet Need	-0.003 (0.028) [0.92]	0.031** (0.013) [0.08]	0.136	26,519	25,572
Modern CPR	-0.007 (0.023) [0.75]	-0.038* (0.020) [0.22]	0.369	52,900	51,037
<i>Panel B. Contraceptive Use, Demand, and Norms</i>					
Any Method (CPR)	-0.003 (0.031) [0.93]	-0.006 (0.024) [0.83]	0.405	52,900	51,037
Heard of FP in the Media	-0.002 (0.039) [0.96]	-0.017 (0.081) [0.85]	0.767	26,530	25,583
Last Child Not Wanted at Birth	-0.053** (0.024) [0.05]	0.036 (0.022) [0.26]	0.391	15,008	14,405
Non-Use Reason: Religious Prohibition	0.013 (0.011) [0.32]	0.035 (0.037) [0.56]	0.023	6,294	5,898
Non-Use Reason: Fear of Side Effects	-0.024 (0.053) [0.74]	0.135*** (0.046) [0.05]	0.173	6,294	5,898
Husband Wants More Children	0.029 (0.078) [0.64]	0.058 (0.052) [0.38]	0.286	13,019	12,646

NOTE: Each cell is the County TWFE estimate of main-text Table 3, estimated separately for each program against the never-treated counties, with the same inference. Panel A repeats the County TWFE rows of Table 3 and Appendix Table G6. Non-use reasons are asked of women not using contraception, and the last child not wanted refers to the most recent birth. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

relevant infrastructure is the faith-based network specifically, and general health-system density does not matter. The remaining moderators, religious practice and the Protestant

Table G3: Fertility Intentions, Birth Timing, and Method Mix

	CHAK-Packard		CAFPA Phase 1–2		2014 Control
	Estimate (1)	Pre-Trend p (2)	Estimate (3)	Pre-Trend p (4)	Mean (5)
<i>Panel A. Fertility and Intentions</i>					
Births (Three Years)	-0.026 (0.025) [0.49]	0.27	0.050*** (0.013) [0.05]	0.11	0.389
Wants No More Children	0.035* (0.019) [0.06]	0.46	0.014 (0.017) [0.59]	<0.01	0.420
Ideal Number of Children	-0.16 (0.11) [0.19]	0.22	0.00 (0.10) [0.98]	0.08	3.8
<i>Panel B. Birth Timing</i>					
Age at First Birth (Years)	0.03 (0.08) [0.76]	0.45	0.16 (0.16) [0.73]	0.95	19.4
Birth Spacing, All Births (Months)	0.56 (0.65) [0.47]	0.10	-1.66** (0.63) [0.05]	<0.01	41.4
<i>Panel C. Current Method</i>					
Traditional Method	0.003 (0.010) [0.76]	0.70	0.026*** (0.008) [0.06]	0.56	0.034
Injectable	-0.003 (0.011) [0.82]	0.24	-0.013 (0.020) [0.80]	<0.01	0.188
Implant	0.017 (0.015) [0.34]	<0.01	0.026 (0.020) [0.31]	0.12	0.068
IUD	-0.003 (0.007) [0.77]	0.55	-0.001 (0.003) [0.79]	0.01	0.021
Pill	-0.005 (0.004) [0.28]	0.71	-0.024*** (0.003) [0.03]	0.62	0.044
Male Condom	0.017 (0.015) [0.39]	0.20	-0.014 (0.014) [0.38]	0.03	0.027

NOTE: County TWFE estimates as in main-text Table 3, with the same inference and pre-trend test. Wants no more children includes sterilized and infecund women, and method indicators are shares of all women. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

and non-Catholic shares, are mixed in sign and mostly do not survive the bootstrap, so I do not rest any claim on them. With only four treated counties, these interactions are

suggestive. Religious leaders move behavior on the ground when their message is bundled with the means to act on it.

Table G4: CHAK-Packard: Moderated Effects by Religious Context and Infrastructure

	Birth Spacing (1)	Births (2)	Unmet Need (3)	Modern CPR (4)
<i>Panel A. Attitudinal Moderators</i>				
Protestant (Individual) $\times$ DiD	-4.98* (2.63) [0.19]	-0.031 (0.027) [0.40]	0.013 (0.026) [0.67]	-0.106*** (0.019) [0.05]
Trust in Leaders $\times$ DiD	4.05*** (1.28) [0.04]	-0.046** (0.017) [0.25]	-0.049*** (0.014) [0.58]	0.015 (0.017) [0.77]
Non-Catholic Christian Share $\times$ DiD	-4.23 (2.81) [0.44]	0.065*** (0.022) [0.06]	-0.009 (0.036) [0.80]	0.018 (0.034) [0.68]
Religious Practice $\times$ DiD	7.09* (3.63) [0.08]	-0.137*** (0.032) [0.16]	-0.168** (0.072) [0.56]	0.086 (0.060) [0.53]
<i>Panel B. Structural Moderators</i>				
CHAK Facilities $\times$ DiD	2.13** (1.01) [0.09]	-0.033*** (0.011) [0.04]	-0.006 (0.009) [0.46]	-0.015 (0.013) [0.42]
Protestant Share of Places of Worship $\times$ DiD	1.00 (0.94) [0.31]	-0.011 (0.012) [0.40]	0.001 (0.011) [0.94]	-0.021*** (0.005) [0.03]
FBO Facilities $\times$ DiD	1.92 (1.22) [0.16]	-0.017 (0.014) [0.27]	-0.033** (0.016) [0.30]	0.028** (0.012) [0.37]
MoH Facilities $\times$ DiD	-0.63 (1.00) [0.56]	-0.032 (0.031) [0.60]	-0.008 (0.020) [0.73]	-0.009 (0.023) [0.72]
Observations	20,017	52,900	26,519	52,900

NOTE: Each row is a separate regression of Appendix Equation (G7), the County TWFE estimate interacted with one moderator, and reports the triple interaction per standard deviation of the moderator, or Protestant against non-Protestant women for the individual indicator. Attitudinal moderators are 2014 ward-level small-area estimates, and structural moderators are ward counts and shares from the KMHFR, CHAK records, and OpenStreetMap. The trust, CHAK-facility, FBO, and MoH rows are Panel B of main-text Table 3. Inference as in main-text Table 3, with only four treated counties. Observations are for the trust regressions. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

CAFFA Table G6 reports the CAFFA estimates. CAFFA shows the opposite sign, with five-year births up 10.2 percentage points, birth intervals 6.0 months shorter, unmet need up 3.1 points, and a borderline decline in modern contraceptive prevalence of 3.8 points.

Table G5: CAFPA Phase 1–2: Moderated Effects by Religious Context and Infrastructure

	Birth Spacing (1)	Births (2)	Unmet Need (3)	Modern CPR (4)
<i>Panel A. Attitudinal Moderators</i>				
Protestant (Individual) $\times$ DiD	-5.45 (5.28) [0.40]	-0.037 (0.037) [0.65]	0.022 (0.028) [0.74]	0.090* (0.045) [0.23]
Trust in Leaders $\times$ DiD	-0.63 (2.35) [0.85]	0.038** (0.017) [0.18]	0.012 (0.011) [0.26]	0.011 (0.015) [0.59]
Non-Catholic Christian Share $\times$ DiD	-6.37 (16.76) [0.90]	-0.049 (0.118) [0.60]	0.118* (0.064) [0.17]	-0.168*** (0.053) [0.40]
Religious Practice $\times$ DiD	-3.43*** (1.17) [0.23]	0.011 (0.013) [0.48]	0.012 (0.010) [0.39]	0.039*** (0.010) [0.06]
<i>Panel B. Structural Moderators</i>				
CHAK Facilities $\times$ DiD	-0.67 (2.74) [0.73]	-0.012 (0.014) [0.46]	-0.001 (0.011) [0.93]	-0.006 (0.013) [0.71]
Protestant Share of Places of Worship $\times$ DiD	-0.21 (2.02) [0.88]	-0.040*** (0.012) [0.05]	0.021 (0.013) [0.18]	-0.016 (0.011) [0.18]
FBO Facilities $\times$ DiD	3.75* (1.87) [0.31]	-0.028** (0.013) [0.25]	0.005 (0.011) [0.68]	-0.003 (0.015) [0.87]
MoH Facilities $\times$ DiD	0.16 (2.37) [0.94]	0.007 (0.021) [0.74]	-0.007 (0.014) [0.68]	-0.007 (0.012) [0.66]
Observations	19,166	51,037	25,572	51,037

NOTE: Each row is a separate regression of Appendix Equation (G7), the County TWFE estimate interacted with one moderator, and reports the triple interaction per standard deviation of the moderator, or Protestant against non-Protestant women for the individual indicator. Attitudinal moderators are 2014 ward-level small-area estimates, and structural moderators are ward counts and shares from the KMHFR, CHAK records, and OpenStreetMap. The trust, CHAK-facility, FBO, and MoH rows are Panel B of Appendix Table G6. Inference as in main-text Table 3, with only three treated counties. Observations are for the trust regressions. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The increase in births persists across the same estimators, including the synthetic DiD that reweights the pre-period, at 8.5 points. The reason for caution on CAFPA lies instead in the baseline. Its birth rate dips below trend in 2014 and partially recovers by 2022, and its birth spacing jumps in 2014 and stays flat after, as Figure G6 shows. In the event study of Figure G7, the CAFPA spacing coefficients sit far below zero before 2014, and in

Table G6: CAFPA Phase 1–2 Effects on Fertility and Its Channels

	Birth Spacing (1)	Births (2)	Unmet Need (3)	Modern CPR (4)
<i>Panel A. Average Effects</i>				
County TWFE	-6.02*** (1.47) [0.03]	0.102*** (0.017) [0.03]	0.031** (0.013) [0.08]	-0.038* (0.020) [0.22]
Synthetic DiD	-4.58** (1.80)	0.085** (0.033)	0.016 (0.014)	-0.036 (0.025)
Callaway–Sant’Anna	-6.46*** (1.22)	0.136*** (0.020)	0.030** (0.014)	-0.037 (0.023)
Distance to Facility (per SD of Exposure)	-0.52 (0.82) [0.63]	0.037*** (0.008) [<0.01]	0.010 (0.006) [0.27]	-0.007 (0.006) [0.34]
<i>Panel B. Moderated Effects</i>				
Trust in Leaders × DiD (Authority)	-0.63 (2.35) [0.85]	0.038** (0.017) [0.18]	0.012 (0.011) [0.26]	0.011 (0.015) [0.59]
CHAK Facilities × DiD (Program Network)	-0.67 (2.74) [0.73]	-0.012 (0.014) [0.46]	-0.001 (0.011) [0.93]	-0.006 (0.013) [0.71]
FBO Facilities × DiD (All Faith-Based)	3.75* (1.87) [0.31]	-0.028** (0.013) [0.25]	0.005 (0.011) [0.68]	-0.003 (0.015) [0.87]
MoH Facilities × DiD (Falsification)	0.16 (2.37) [0.94]	0.007 (0.021) [0.74]	-0.007 (0.014) [0.68]	-0.007 (0.012) [0.66]
2014 Control Mean	46.9	0.654	0.136	0.369
Observations	19,166	51,037	25,572	51,037
Counties (Treated / Control)	3 / 36	3 / 36	3 / 36	3 / 36
Pre-Trend Test, <i>p</i> -Value	<0.01 [0.05]	0.32 [0.61]	0.47 [0.58]	<0.01 [0.12]

NOTE: County TWFE is the 2×2 difference-in-differences (2014 against 2022), with standard errors clustered by county in parentheses and wild cluster bootstrap *p*-values in brackets (Rademacher weights, 999 replications). Stars use the conventional standard errors. Synthetic DiD follows (Arkhangelsky et al., 2021), and Callaway–Sant’Anna reports the overall average treatment effect on the treated (Callaway & Sant’Anna, 2021). Distance to Facility replaces the program indicator with the standardized log inverse distance to the nearest program facility. Panel B interacts treatment and post with a standardized 2014 ward-level moderator. The bottom rows refer to the County TWFE sample, and the pre-trend test jointly tests the 2003 and 2008 event-study coefficients relative to 2014. Birth spacing is the mean preceding birth interval, in months, for births in the last five years, and births is the five-year birth indicator. Nairobi is excluded. **p* < 0.10, ***p* < 0.05, ****p* < 0.01.

Figure G6 the gap in births between CAFPA and control counties moves as much between the pre-treatment waves as after the program. This movement is consistent with the

unwinding of a concurrent facility-strengthening program, APHIAplus Kamili, which operated in the same three counties through 2017. Dropping the contested 2014 wave from the fit entirely softens the increase in births, and with only the sparse 2003 and 2008 waves remaining, mean reversion cannot be separated from a true increase. Among women not using contraception, CAFPA is associated with a 13.5 percentage point increase in citing side effects as the reason for non-use, consistent with a message that raised the salience of side effects without a delivery arm to address them. CHAK plausibly lowered fertility, and the CAFPA estimate I do not read causally.

G.6: Additional Robustness

Estimators Across Designs The county estimates are robust across estimators, as main-text Table 3 and Table G6 show, and they reappear as a distance gradient. Table G7 reports the continuous-exposure DiD. For CHAK, each standard deviation of proximity lengthens birth spacing by 3.2 months and lowers the five-year birth rate by 4.9 percentage points, and both survive the wild cluster bootstrap. The pre-treatment interactions are small, especially for spacing, where the 2003 and 2008 coefficients are close to zero. Because county fixed effects absorb only level differences, the gradient pools proximity variation within and across program and control counties. Within the four CHAK-Packard counties alone, the gradient is insignificant for both outcomes, close to zero for births and of the opposite sign for spacing. The distance design therefore confirms that the effect is concentrated in program areas, and it does not show the effect falling with distance from a facility.

Staggered Timing Table G8 decomposes the CHAK two-way fixed-effects estimate for birth spacing and births. Almost all of the weight, 97.6%, falls on clean treated-versus-never-treated comparisons, which give 4.7 months for spacing and 9.0 percentage points fewer births. The “forbidden” comparison that uses already-treated counties as controls receives under one percent of the weight. For births it carries the same sign as the rest, and for spacing it carries the opposite sign, which its small weight makes immaterial. The design is therefore effectively a 2×2 , and two-way fixed effects is not meaningfully biased by staggered timing.

Parallel Trends A strong pre-trend test is infeasible in this setting, and Table G9 makes the limitation explicit. The KDHS provides only two pre-treatment waves, 2003 and 2008. They are widely and unevenly spaced, their samples of women and their geographic coverage are markedly smaller than in 2014 and 2022, and 2014 is itself mid-treatment for CHAK Phase 1. Under the relative-magnitude restriction of Rambachan and Roth (2023) the

Table G7: Distance-to-Facility Difference-in-Differences

	Birth Spacing (1)	Births (2)	Unmet Need (3)	Modern CPR (4)
<i>Panel A. CHAK-Packard</i>				
Exposure $\times$ Post	3.25*** (0.63) [<0.01]	-0.049*** (0.010) [<0.01]	-0.008 (0.007) [0.31]	0.006 (0.008) [0.50]
2003 $\times$ Exposure (Pre)	-0.15 (0.62)	-0.006 (0.016)	0.025** (0.010)	-0.033** (0.012)
2008 $\times$ Exposure (Pre)	-0.13 (0.69)	0.024* (0.014)	0.005 (0.009)	-0.017 (0.012)
Pre-Trend Test, p -Value	0.97 [0.97]	0.08 [0.13]	0.05 [0.05]	0.03 [0.08]
Within Program Counties: Exposure $\times$ Post	-1.15 (1.23)	-0.005 (0.011)	-0.015 (0.010)	-0.004 (0.028)
<i>Panel B. CAFPA Phase 1–2</i>				
Exposure $\times$ Post	-0.52 (0.82) [0.63]	0.037*** (0.008) [<0.01]	0.010 (0.006) [0.27]	-0.007 (0.006) [0.34]
2003 $\times$ Exposure (Pre)	-0.89 (0.70)	-0.021* (0.011)	-0.008 (0.009)	0.030*** (0.009)
2008 $\times$ Exposure (Pre)	-1.13 (1.11)	-0.025 (0.021)	0.008 (0.007)	0.013 (0.012)
Pre-Trend Test, p -Value	0.41 [0.62]	0.17 [0.20]	0.29 [0.32]	<0.01 [0.06]
Within Program Counties: Exposure $\times$ Post	2.76 (1.75)	-0.002 (0.010)	0.013 (0.020)	0.009 (0.023)

NOTE: Exposure is the standardized log inverse distance from a woman's survey cluster to the nearest program facility (Appendix Equation (G6)). *Exposure $\times$ Post* is the Distance to Facility row of main-text Table 3 and Appendix Table G6, pooling program and control counties. The 2003 and 2008 rows are exposure-by-wave coefficients relative to 2014, and the pre-trend test is their joint test. The last row restricts the sample to the program counties and restandardizes exposure, so it uses only within-county variation, and with three or four clusters its standard errors are unreliable. Inference as in Table 3. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

robust bounds are correspondingly wide. Birth spacing holds up best. The CHAK effect on spacing stays positive until the allowed post-period violation reaches the largest pre-period deviation, while the CHAK reduction in births admits zero once the allowed violation reaches half of it. The pre-trend test in main-text Table 3 cannot reject parallel trends for spacing or births, but with four treated counties and two sparse pre-treatment waves it has low power, and Figure G7 plots the event-study coefficients behind it. This sparse and uneven pre-period is exactly why I present the program evidence as corroborating the survey, without claiming independent identification.

Table G8: Goodman–Bacon Decomposition of the CHAK-Packard TWFE Estimate

Comparison Type	Estimate		
	Birth Spacing (1)	Births (2)	Weight (3)
Treated Versus Never-Treated	4.66	-0.090	0.976
Earlier- Versus Later-Treated (Permitted)	-0.01	-0.074	0.016
Later- Versus Earlier-Treated (“Forbidden”)	-2.12	-0.070	0.008

NOTE: Goodman–Bacon decomposition of the two-way fixed-effects estimate into its 2×2 comparisons, on the balanced county-year panel with CHAK-Packard’s Phase 1 (2013) and Phase 2 (2015) cohorts and never-treated controls, without individual controls. The weights depend only on treatment timing and are the same for both outcomes. Birth spacing is in months, and births is the five-year birth indicator.

Table G9: HonestDiD Sensitivity of the Program Effects to Pre-Trend Violations

	CHAK-Packard		CAFPA Phase 1–2	
	Birth Spacing (1)	Births (2)	Birth Spacing (3)	Births (4)
Original ($\bar{M} = 0$)	[3.61, 7.48]	[-0.122, -0.028]	[-8.79, -3.05]	[0.068, 0.133]
$\bar{M} = 0.5$	[2.71, 8.34]	[-0.155, 0.031]	[-13.72, 0.07]	[0.027, 0.158]
$\bar{M} = 1.0$	[0.98, 10.07]	[-0.220, 0.104]	[-19.35, 4.80]	[-0.026, 0.206]
$\bar{M} = 1.5$	[-0.88, 11.89]	[-0.294, 0.180]	[-25.01, 10.16]	[-0.080, 0.260]
$\bar{M} = 2.0$	[-2.76, 13.72]	[-0.368, 0.255]	[-30.67, 15.80]	[-0.135, 0.315]

NOTE: 95% robust confidence sets of Rambachan & Roth (2023) for the 2022 program effect from the County TWFE event study, under the relative-magnitude restriction that the post-treatment violation of parallel trends is at most $\bar{M}$ times the largest pre-treatment violation. The first row is the event-study confidence interval. Birth spacing is in months, and births is the five-year birth indicator.

References

- Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., & Wager, S. (2021). Synthetic Difference-in-Differences. *American Economic Review*, 111(12), 4088–4118.
- Baker, A., Callaway, B., Cunningham, S., Goodman-Bacon, A., & Sant’Anna, P. H. C. (2026). Difference-in-Differences Designs: A Practitioner’s Guide. *Journal of Economic Literature*, 64(2), 498–557.
- Bormet, M., Kishoyian, J., Siame, Y., Ngalande, N., Erb, K., Parker, K., Huber, D., & Hardee, K. (2021). Faith-Based Advocacy for Family Planning Works: Evidence From Kenya and Zambia. *Global Health: Science and Practice*, 9(2), 254–263.
- Callaway, B., & Sant’Anna, P. H. (2021). Difference-in-Differences with multiple time periods. *Journal of Econometrics*, 225(2), 200–230.
- Claassen, C., & Traunmüller, R. (2020). Improving and Validating Survey Estimates of Religious Demography Using Bayesian Multilevel Models and Poststratification. *Sociological Methods & Research*, 49(3), 603–636.
- Gelman, A., & Little, T. C. (1997). Poststratification Into Many Categories Using Hierarchical Logistic Regression. *Survey Methodology*, 23(2), 127–135. Retrieved August 7, 2026, from <https://sites.stat.columbia.edu/gelman/research/published/poststrat3.pdf>
- Park, D. K., Gelman, A., & Bafumi, J. (2004). Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls. *Political Analysis*, 12(4), 375–385.
- Rambachan, A., & Roth, J. (2023). A More Credible Approach to Parallel Trends. *The Review of Economic Studies*, 90(5), 2555–2591.
- Roth, J., Sant’Anna, P. H., Bilinski, A., & Poe, J. (2023). What’s Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature. *Journal of Econometrics*, 235(2), 2218–2244.
- Ruark, A., Kishoyian, J., Bormet, M., & Huber, D. (2019). Increasing Family Planning Access in Kenya Through Engagement of Faith-Based Health Facilities, Religious Leaders, and Community Health Volunteers. *Global Health: Science and Practice*, 7(3), 478–490.
- UNDESA. (2024). Model-Based Estimates and Projections of Family Planning Indicators. Retrieved September 11, 2025, from <https://population.un.org/dataportal>