Family Planning and the Dynamics of Fertility Preferences: Experimental Evidence from Urban Malawi*

Stefanie Pechar† Daniel Maggio‡ Mahesh Karra§

Abstract

Family planning (FP) programs are conventionally understood as helping women implement fertility preferences that exist prior to, and independently of, the program. Yet stated preferences are not fixed: women revise them as their circumstances, information, and reproductive options change—precisely the conditions FP programs are designed to alter. We use longitudinal data from a randomized expansion of FP services among women in urban Malawi to estimate the intervention's effect on both the level and the stability of stated fertility preferences. We first show that preferences are relatively stable in the aggregate yet vary substantially within women over time, in patterns inconsistent with reporting error. The intervention leaves average ideal family size unchanged in the full sample, but that null conceals an upward revision of 0.13 to 0.15 children among women postpartum at baseline. Within-woman variability, by contrast, falls by roughly 12 percent among women exposed during pregnancy, and within that group the reductions are largest among less-educated women and those living farthest from an FP facility. Our results establish that the stability of stated fertility preferences is itself policy-responsive, whether this reflects reduced uncertainty or reduced adaptive flexibility. This highlights a dimension of reproductive decision-making that cross-sectional measures overlook and challenges the conventional separation of fertility demand from its implementation.

Keywords: Fertility Preferences; Stability of Preferences; Family Planning; Malawi

*This trial was registered at the American Economic Association Registry for randomized controlled trials on May 7, 2015 (AEARCTR-0000697) and at the Registry for International Development Impact Evaluations (RIDIE) on May 28, 2015 (RIDIE-STUDY-ID-556784ed86956). This research uses original data collected by Canning and Karra with support from Innovations for Poverty Action (IPA) in Malawi. The authors would like to acknowledge the dedication and support of Carly Farver, Patrick Baxter, Bagrey Ngwira, Reginald Chunda, Viola Nyirongo, Violet Chitsulo, Macdonald Salamu, and the entire Malawi Family Planning Study team. This project was supported by two grants from the Hewlett Foundation and the Program for Women's Empowerment Research (POWER) at Columbia University. Ethical approval to conduct the study was received from the Harvard University IRB (protocol number IRB16-0421) and the Malawi National Health Sciences Research Committee (protocol number 16/7/1628). The findings, interpretations, and conclusions expressed in this study are entirely those of the authors. They do not represent the views of Innovations for Poverty Action and its affiliated organizations.

†**Corresponding Author.** Technical University of Munich, Germany. stefanie.pechar@tum.de

‡Department of Economics, Rutgers University, New Brunswick, NJ USA. daniel.maggio@rutgers.edu

§Mailman School of Public Health, Columbia University, New York, USA. mvk2109@cumc.columbia.edu

Introduction

Fertility preferences are central to how demographers explain fertility behavior and evaluate family planning (FP) programs. Reported ideal family size (IFS) is a key measure of fertility demand and the core quantity from which wanted and unwanted fertility are decomposed (Bongaarts, 1990; Casterline and El-Zeini, 2007). As such, it anchors a continuing debate over whether fertility decline in low- and middle-income countries is driven by a falling demand for children or by the fuller implementation of preferences that women already hold—a debate especially live across sub-Saharan Africa, where stated preferences have remained relatively high (Bongaarts, 2017; Ibitoye et al., 2022; Bongaarts, 2025). Both sides of that debate rest on the same premise: that stated preferences are established prior to, and separable from, the programs meant to help implement them. IFS enters as a stable expression of demand that conventional, access-oriented FP programs seek to satisfy rather than shape.

Yet preferences are not necessarily fixed. Longitudinal research has documented frequent revisions in reported IFS over the reproductive life course, in high- and low-income settings alike (Heiland et al., 2008; Liefbroer, 2009; Hayford, 2009; Mueller et al., 2022). Because these revisions largely offset across subgroups of women, population-level IFS can appear stable even as individual ideals move considerably. Longitudinal research therefore reveals something that a cross-sectional measure cannot: two women who report the same ideal at a given moment may differ substantially in the stability of that ideal over time. This temporal dimension need not be merely a property of measurement. It may reveal something about the nature of the preference itself—whether fertility goals are relatively settled, responsive to changing circumstances, or still being formed. A reported ideal, on this view, is not simply read off a quantity fixed in advance. It may instead be constructed or revised at the moment of elicitation, from the information a woman holds, the constraints she faces, and her current reproductive context (Bachrach and Morgan, 2013; Bhrolchin and Beaujouan, 2019). If preferences are formed and expressed in this way, then the conditions of the reproductive environment may enter preference formation itself, and not only the realization of an independently held goal.

To date, instability has mostly been documented descriptively: how often preferences change, in which direction, and at what points in the life course (Mueller et al., 2022; Kodzi et al., 2010). Where it has been explained, revisions to fertility preferences have been attributed to partnership transitions, childbearing experiences, and shifts in economic circumstance (Yeatman et al., 2013; Sennott and Yeatman, 2012; Trinitapoli and Yeatman, 2018). These explanations share a common form: instability is traced back to the woman herself and treated as a quantity to be predicted from her characteristics or her life-course transitions, rather than as one that might respond to conditions outside her. What has not been asked is whether instability responds to conditions that policy can set.

In this study, we ask that question directly. We use longitudinal data from a randomized expansion of postpartum FP services in urban Malawi to estimate whether an exogenous improvement in the reproductive environment moves both the level and the stability of stated fertility preferences. Stability enters as an outcome in its own right rather than as a nuisance property of the measure, and the randomization allows us to attribute movement in it to the program rather than to the individual circumstances with which it is usually explained. Program effects on the level of stated ideals have been examined (Herrera-Almanza and Seitz McCarthy, 2025); the stability with which women hold them has not.

FP programs are a natural setting for this question because they alter several of the conditions under which fertility goals are formed and implemented. Historically, FP programs have been conceived as instruments for helping women and couples to realize fertility goals that they already hold, by lowering the barriers that stand between their (latent) preferences and desired behavior (Freedman, 1997). Contemporary programs, however, rarely stop at commodity provision. Counseling, repeated individualized contact, and structured information about spacing and method choice engage women directly in reproductive decision-making (see for example, Ashraf et al., 2014; Dupas et al., 2025; Herrera-Almanza and Seitz McCarthy, 2025; Karra and Zhang, 2025). The intervention we study did both. It removed financial and travel barriers to FP services and paired access with structured counseling on birth spacing, family size, and longer-term reproductive goals. Prior work has established that the intervention changed behavior, raising contraceptive use and reducing the risk of short-interval pregnancies (Karra et al., 2022). Whether it also moved the preferences against which such behavior is conventionally assessed has not been examined.

Our data come from married women who were pregnant or within six months postpartum at enrollment and who were followed across three annual waves. We measure fertility preferences by reported IFS and operationalize stability as within-woman variability in these reports across waves, following Yeatman et al. (2013) but treating that variability as an outcome rather than as a trait to be correlated with individual characteristics. Stability in this sense is revealed by repeated measurement rather than self-assessed. We infer it from how a woman's answers move across separate interviews, not from asking her how certain she is of the IFS she has just stated. The two capture distinct properties, and we take up the distinction in the next section.

We report three sets of findings. First, reported IFS is relatively stable in aggregate but varies substantially within women over time. Second, the intervention leaves average IFS unchanged in the full sample, although this pooled null conceals a 0.13–0.15 child upward revision among women who were postpartum at baseline. Third, the intervention reduces within-woman variability in reported IFS by roughly 12 percent among women who were pregnant at baseline. This reduction is largest among women with less education and those living farther from an FP facility. These subgroup analyses describe where the stability effect is concentrated; they are not intended to identify the mechanisms producing it.

Taken together, these results support a view of fertility preferences as evolving constructs that emerge through the interaction between individual fertility aspirations and the institutional environments in which reproductive decisions are made. A program can leave the central tendency of stated fertility ideals untouched while measurably altering the processes that generate them. More fundamentally, our findings show that stability is not simply a descriptive feature of fertility preferences or a fixed characteristic of women, but a policy-responsive dimension of reproductive decision-making. This bears on how fertility decline is explained. If a program that improves implementation also moves the demand that it implements, then the two channels through which decline is conventionally decomposed are not independent (Pritchett, 1994; Bongaarts, 1994). These findings also bear on how preference-based measures should be interpreted. A growing literature has questioned such measures on grounds of reproductive autonomy and the normative content of classifying births as intended or unintended (Upadhyay et al., 2014; Senderowicz, 2020; Karp et al., 2025). Our argument is complementary: regardless of how preferences ought to be interpreted, an evaluation that uses stated preferences to assess a program stands on different footing once the program has moved those preferences. The concern is not that preference-based metrics are wrong, but that their behavior under policy exposure has not been treated as something to estimate.

Background

The Dynamics of Fertility Preferences and their Measurement

That fertility preferences are dynamic rather than fixed is by now well documented. Stated ideals and intentions are imperfect predictors of subsequent childbearing and are revised, often repeatedly, over the reproductive life course (Nair and Chow, 1980; Tan and Tey, 1994; Quesnel-Valle and Morgan, 2003; Hayford and Agadjanian, 2017; Cleland et al., 2020). Longitudinal studies have documented associations between these revisions and life-course events such as partnership formation and dissolution, childbearing, and changes in economic circumstance. That evidence came first from high-income settings (Heiland et al., 2008; Liefbroer, 2009; Hayford, 2009) and, more recently, from low- and middle-income settings, including several contexts in sub-Saharan Africa, where reported preferences are likewise revised and systematically patterned (Kodzi et al., 2010; Sennott and Yeatman, 2012; Yeatman et al., 2013; Hayford and Agadjanian, 2017).

This individual-level movement can coexist with stability in the aggregate. Because revisions run in largely offsetting directions, the population distribution of IFS changes little even as individual women move considerably within it (Mueller et al., 2022). The stability that makes aggregate preferences a convenient input thus conceals substantial within-woman revision. Any account of how fertility preferences are formed, or of what moves them, has to work at the level of the individual woman.

How this instability is characterized varies across the literature, and it is important to distinguish instability from uncertainty. Uncertainty concerns how a woman herself assesses the strength or reliability of her stated fertility goal at a given point in time. Instability, by contrast, is an observed property of the preference across time: it describes how much a woman's reported ideal changes across repeated elicitations. The two may be related, but they are not interchangeable. A woman may report the same ideal repeatedly despite feeling uncertain about it, or revise her ideal over time while feeling confident in each report when it is elicited.

The literature operationalizes these two properties in different ways. One strand measures preference uncertainty directly, eliciting the strength of a woman's certainty about her ideals through ranking exercises or graded confidence categories (Coombs, 1974; Morgan, 1981; Barker and Buber-Ennser, 2024; Badolato et al., 2025). These measures are informative about how women themselves regard their preferences, but a certainty report is elicited moments after the ideal itself, in the same interview and under the same conditions, so whatever moves the stated IFS at that moment can move the stated confidence with it. A separate, longitudinal strand instead infers stability from what women report over time, comparing a woman's answers across survey waves. Much of this work treats revision categorically, recording whether a woman revises her stated preference and in which direction (Kodzi et al., 2010; Sennott and Yeatman, 2012). A second approach by Yeatman et al. (2013) models its degree, using the residual variation in a woman's repeated reports, net of her own average and of common time effects, to describe how predictable her preference is. Repeated elicitation separates these influences in time: a shock specific to one interview appears as movement between waves rather than being absorbed into the within-interview association between an ideal and the confidence reported in it.

In each of these treatments, variability is a property of individual women, something to be documented or explained by her characteristics or by transitions in her own life course, rather than an outcome in its own right. We adopt the measurement strategy of Yeatman et al. (2013) but put it to a different use. Where that work uses the measure descriptively, we treat it as an outcome that may respond to the environment in which preferences are formed. This shifts the object of study from how variable women are to what makes them so and, in particular, to whether an exogenous change in the reproductive environment moves the stability of stated preferences at all.

Preference formation and the reproductive environment

Fertility preferences are commonly interpreted as reflecting underlying reproductive goals, yet there are different ways of conceptualizing the relationship between stated preferences and those underlying goals. One perspective views reported IFS as an observable expression of a relatively stable preference that exists prior to its measurement. From this perspective, changes in reported preferences reflect either genuine updating in response to new experiences or imperfect measurement of an underlying target. This interpretation underlies much of the demographic literature, where IFS is treated as a meaningful benchmark against which subsequent fertility behavior can be understood.

An alternative perspective emphasizes that fertility preferences are formed and revised through an ongoing process shaped by women's experiences and the contexts in which reproductive decisions are made (Bachrach and Morgan, 2013; Bhrolchin and Beaujouan, 2019). Rather than existing as fully specified goals awaiting measurement, preferences emerge through the interaction of aspirations, perceived opportunities, constraints, and available information. From this perspective, changes in reported IFS are not simply deviations from a latent preference. They provide evidence that preference formation continues over the reproductive life course, particularly under conditions of uncertainty where flexibility may itself be an adaptive response (Trinitapoli and Yeatman, 2018).

These perspectives are not mutually exclusive. Fertility preferences may exhibit considerable persistence while remaining responsive to changing circumstances. The key distinction lies in whether the reproductive environment is viewed primarily as influencing the realization of pre-existing preferences or as contributing to the process through which those preferences are formed, revised, and expressed over time. The latter perspective implies that changes in the reproductive environment may affect not only *what* women report as their fertility goal, but also *how persistently* they report it across time.

Viewing fertility preferences as an ongoing process of formation therefore highlights the importance of the reproductive environment. Reproductive environments provide information, shape the constraints and opportunities women consider, and determine how feasible it is to act on a given fertility goal. They can consequently affect both the formation and the expression of stated preferences. Family planning programs represent a particularly important component of this environment. Traditionally, they have been understood as enabling women and couples to realize existing fertility preferences by reducing barriers to contraceptive access (Freedman, 1997). However, interventions that combine service provision with counseling and information (see for example Ashraf et al., 2014; Herrera-Almanza and Seitz McCarthy, 2025) may also influence the informational and deliberative processes through which fertility preferences are developed and revised.

This framing identifies two dimensions of stated fertility preferences that a program might affect: their *level* and their *stability*. By relaxing access constraints, a program changes the opportunities and constraints that women face when considering and expressing fertility goals, not only their ability to implement an existing goal. By providing information and repeated counseling, it changes the informational inputs available to them as they consider spacing, family size, and longer-term reproductive plans. Either channel could therefore affect where a stated preference settles—its *level*—and how much it changes across repeated elicitations—its *stability*.

The framework does not predict the direction of either effect. The same information, or the same relaxation of constraints, could lead one woman to revise her ideal upward and another downward, depending on what she previously believed and what constraints she previously faced. Stability presents a further ambiguity. A program could reduce variability by resolving competing considerations or making a fertility goal more persistent; alternatively, exposure to new information could prompt women to reconsider previously stated goals and thereby increase variability. In both cases, the net effect on stability is an empirical question. Importantly, a change in stability need not be interpreted as a change in subjective certainty: greater stability could reflect greater resolution of uncertainty, but it could also reflect reduced flexibility in adapting fertility goals to changing circumstances.

These responses need not be uniform across women. Identical program components may produce different changes in reported IFS or its stability, depending on the circumstances in which women encounter them. Pregnancy or recent childbirth, for example, may alter how women evaluate future childbearing while also changing which constraints or information gaps are most relevant to them. More generally, the effect of a program depends on what was scarce, uncertain, or constraining before exposure. This provides a basis for expecting heterogeneity in both the level and stability of stated preferences.

This potential heterogeneity also has implications for how program effects are measured. If a program moves preferences in opposite directions across subgroups, or changes them only for a subset of women, a pooled estimate of mean IFS can register little or no effect even when stated preferences have changed. The same logic applies to stability: a program may alter how persistently women report their fertility goals without changing the population mean of those goals. Examining within-woman variation therefore captures a dimension of program response that a change in mean IFS alone cannot reveal.

Fertility Transition and Reproductive Preferences in Malawi

Our study takes place in Malawi, a setting where fertility has been in sustained decline for four decades, from seven births per woman in the early 1980s to 4.4 births in 2016, the year of our baseline (DHS Malawi, 2016), and has continued to decline since, reaching 3.7 births by 2024, well after our study concluded (DHS Malawi, 2024). This ongoing, unfinished decline in fertility makes Malawi a useful place to study how fertility preferences respond to a changing reproductive environment. At the time when we observe them, fertility preferences are actively evolving and subject to ongoing reassessment, rather than settled around a long-run equilibrium, which is the condition under which a program is most likely to leave a visible impact on how those preferences are held.

The fertility transition in Malawi also raises a related question: are women in this setting wanting fewer children than they are having, or is stated demand for children already closely tracking realized fertility? The premise that "stabilizing" preferences functions as a meaningful programmatic goal depends on there being a gap for a program to close. A simple comparison of IFS to the national total fertility rate (TFR) is not necessarily the right test here. Our baseline was fielded in 2016, when national TFR stood at 4.4, well above the mean IFS that we observe in our sample (3.19). We do not have a population-representative estimate of how this gap evolved after our 2018 endline, and we do not extrapolate to it. What our own panel data show directly is a within-sample comparison. Figure 4 plots mean ideal and actual family size by age, and actual fertility approaches stated ideals only gradually, converging around age 32; below that age, where most of our sample sits (given a baseline mean age of 24.6), realized fertility trails stated ideals rather than exceeding them. The gap between wanted and actual fertility in this population is therefore real during the years that we observe it, but it is also one that narrows and closes over the reproductive life course, rather than one that reads as structurally unbridgeable. We flag this observation here because it shapes how our level and stability results below should be interpreted. Specifically, this is not a setting where the demand for children is already fully realized; by the same token, we also may not be operating in a context where there is a large, unaddressed gap between fertility preferences and behavior.

Nationally, IFS has declined slowly and unevenly, and survey evidence from Malawi and similar settings documents considerable dispersion in stated preferences and substantial within-woman revision over time (Kodzi et al., 2010; Trinitapoli and Yeatman, 2018; Yeatman et al., 2013). Where fertility goals are still being formed and revised, programs that expand contraceptive access and provide counseling may affect not only whether women implement existing preferences but how those preferences are articulated and maintained. This makes Malawi a natural setting to study how the policy environment shapes both the level and stability of fertility preferences.

Data and Experimental Design

Study Setting and Intervention

We use longitudinal data from the Malawi Family Planning Study (MFPS), a randomized controlled trial conducted in Lilongwe, Malawi, between September 2016 and February 2019.[1] MFPS evaluated a multicomponent FP intervention designed to address key barriers to contraceptive access, including knowledge gaps and constraints related to distance, service availability, and cost. The study follows a cohort of married women aged 18 to 35 who were either pregnant or within six months postpartum at baseline. Women were interviewed annually over three survey waves.

Following the baseline survey, women were randomly assigned to a treatment or control arm. The treatment arm received a two-year multicomponent FP program designed to address informational and access-related barriers to contraceptive use, which comprised three main components. First, women in the treatment group received up to six individualized counseling visits delivered by trained FP counselors. These sessions provided comprehensive information on a full range of contraceptive methods, their potential side effects, and the health benefits of birth spacing, with the aim of supporting informed decision-making over time. Second, treatment women were offered free transportation to the Good Health Kauma Clinic, a high-quality provider offering a full range of FP services with reliable supply and short waiting times. Transportation was arranged through a private taxi driver, and women who utilized the service were accompanied by a female field manager in the taxi as a means to mitigate concerns about safety or stigma. Finally, all costs incurred for FP services received at the Kauma Clinic, including those associated with contraceptive methods, consultations, laboratory tests, and examination fees, were fully covered for the duration of the program.

Women assigned to the control arm received publicly available information on FP methods and on nearby clinics and were otherwise not recontacted until follow-up.

[1] A detailed protocol describing the study design and intervention can be found in Karra and Canning (2020)

Eligibility, Sample Size and Randomization

Eligibility required that participants were married; currently pregnant or within six months postpartum at baseline; between ages 18 and 35; permanent residents of Lilongwe; and had not been sterilized or undergone a hysterectomy. When multiple eligible women were present within a household, the youngest was selected to participate. To minimize potential spillovers between treatment and control groups, recruitment ensured that enrolled women resided at least five households apart. A total of 2,140 women were enrolled at baseline in 2016.

Following the baseline survey, women were randomly assigned to the treatment or control arm using stratified covariate-balanced randomization. Stratification was based on key baseline characteristics, including parity, current contraceptive use, age at sexual debut, educational attainment, work status, and residential neighborhood. The final allocation assigned 1,027 women to the treatment group and 1,113 to the control group. Table A1 presents key descriptive statistics for women in our study sample. Treatment and control groups are well balanced across observable baseline characteristics, with only minor differences in prior contraceptive use, and joint significance tests fail to reject the null of no systematic differences across arms at baseline. Additional details on the randomization protocol are provided in Karra and Canning (2020).

Of the 2,140 women enrolled at baseline, 1,672 were successfully followed up at endline, corresponding to a retention rate of 78 percent over the two-year follow-up. Table A2 compares baseline characteristics of retained and attrited women to assess whether attrition was selective. Reassuringly, attrition is uncorrelated with the two dimensions most consequential for our analysis: retained and attrited women are statistically indistinguishable in their treatment assignment (a difference of 0.027, $p > 0.10$) and in their baseline fertility preferences (3.18 versus 3.21, $p > 0.10$). The absence of differential attrition across arms preserves the internal validity of the experimental comparison, while balance in baseline preferences indicates that the analytic sample is not selected on the outcome of interest. Attrition is, however, related to some sociodemographic characteristics: relative to retained women, those lost to follow-up were on average younger, less educated, of lower parity, and somewhat more likely to desire additional children. We therefore interpret our estimates as representative of the retained sample.

Measuring Fertility Preferences

The MFPS data include detailed socioeconomic and demographic information, reproductive histories, contraceptive use, and a full set of conventional fertility preference measures, including whether a woman wants another child, the preferred timing of the next birth, and her IFS. In this study, we focus primarily on IFS, elicited consistently across waves using the following question, translated into Chichewa: *"If you could [go back to the time you did not have any children and could] choose exactly the number of children to have in your whole life, how many would that be?"* Since women in our sample are in the earlier stages of their reproductive lives and are unlikely to have exceeded their preferred fertility at baseline, we interpret IFS as desired lifetime fertility and use the terms IFS and desired family size interchangeably.

Leveraging the panel structure, we examine two complementary dimensions of fertility preferences: the level of desired family size and within-woman variation over time, which we interpret as a measure of preference stability.

Previous MFPS Findings

The Malawi Family Planning Study has been the basis for several analyses of the intervention's effects (Karra et al., 2022; Maggio et al., 2024; Canning et al., 2026). This work documents that the intervention was effective: two years after exposure, treated women were 5.9 percentage points more likely to use any contraceptive method (7.2 among the immediately postpartum), driven by long-acting reversible methods, and 43 percent less likely to have a subsequent pregnancy within 24 months. By improving access and enabling women to act on their reproductive goals, the intervention provides a well-suited context to examine how relaxing constraints shapes the level and stability of fertility preferences.

Fertility Preference Dynamics in Our Setting

That fertility preferences in sub-Saharan Africa are dynamic rather than fixed is well established. Longitudinal studies have shown that individual women revise their stated IFS over short horizons, even as aggregate distributions remain broadly stable (Kodzi et al., 2010; Sennott and Yeatman, 2012; Yeatman et al., 2013; Trinitapoli and Yeatman, 2018). We confirm that these dynamics hold in our urban Malawian setting and also show that the within-woman variation they describe is structured rather than mere reporting noise, and therefore a margin that a program could plausibly move. We organize this evidence around four patterns.

Pattern 1: Fertility preferences remain relatively stable in the aggregate

Consistent with prior evidence for this region (Sennott and Yeatman, 2012), women in our sample report broadly similar distributions of desired family size across survey waves. In each wave, approximately 90 percent of women state a desired family size between two and four children, and the mean rises only modestly, from 3.19 in the first wave to 3.36 in the third (Figure 1). From a population perspective, these patterns suggest a high degree of stability in fertility preferences.

At the same time, the distributions reveal small but systematic shifts across adjacent categories, notably between three and four children, alongside a small share of women who continue to report higher desired family sizes. This modest aggregate movement understates the extent of individual-level revision: because population distributions net out changes in opposing directions, they can conceal substantial offsetting movement by individual women.

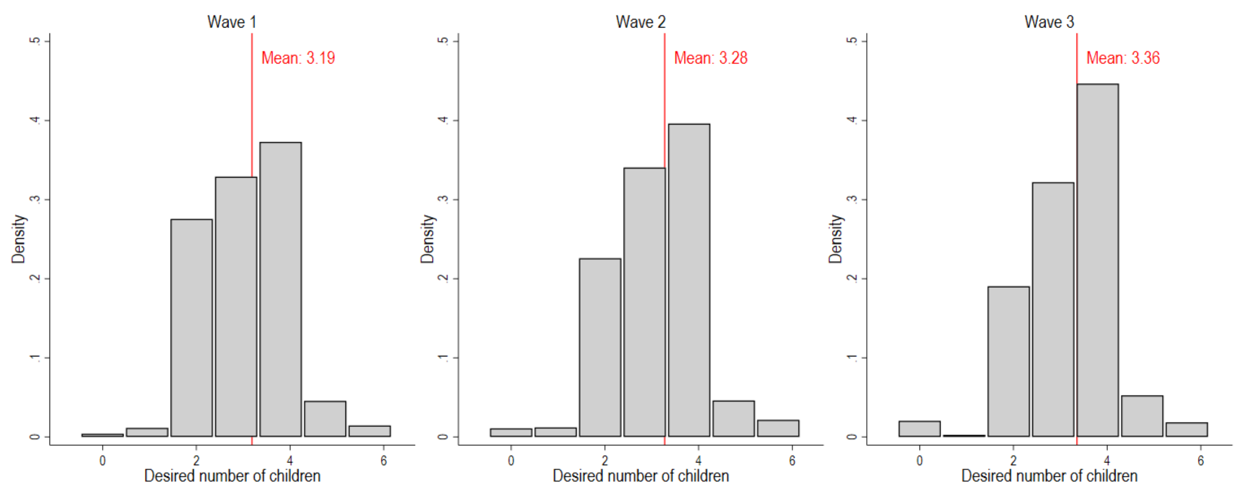


Figure 1. Desired Number of Children across Waves

Notes: The figure presents the distribution of women's reported desired number of children (ideal family size) in each of the three annual survey waves (20162018). Bars show densities; the red vertical lines denote the wave-specific sample mean (3.19, 3.28, and 3.36 in waves 13).

Pattern 2: Substantial within-woman variability in fertility preferences

While desired family size is stable in the aggregate, it exhibits substantial variation when examined longitudinally at the individual level, consistent with findings from rural Ghana (Kodzi et al., 2010) and among young women in southern Malawi (Yeatman et al., 2013; Trinitapoli and Yeatman, 2018). Figure 2 traces women's reported desired number of children across survey waves and reveals considerable movement between categories over time, indicating that many women revise their stated fertility preferences rather than maintaining a fixed target.

These dynamics are summarized in Table A3. Nearly 60 percent of women revise their IFS at least once during the study period, with similar rates across age groups and baseline parity. Among those who revise their preferences, most do so once, but a substantial share (38 percent) revise twice, indicating repeated reassessment of fertility goals over a relatively short time horizon. Revisions occur in both directions. Increases are more common overall, consistent with the modest upward drift in the aggregate mean documented above. Decreases are relatively more frequent among older women and those with higher baseline parity.

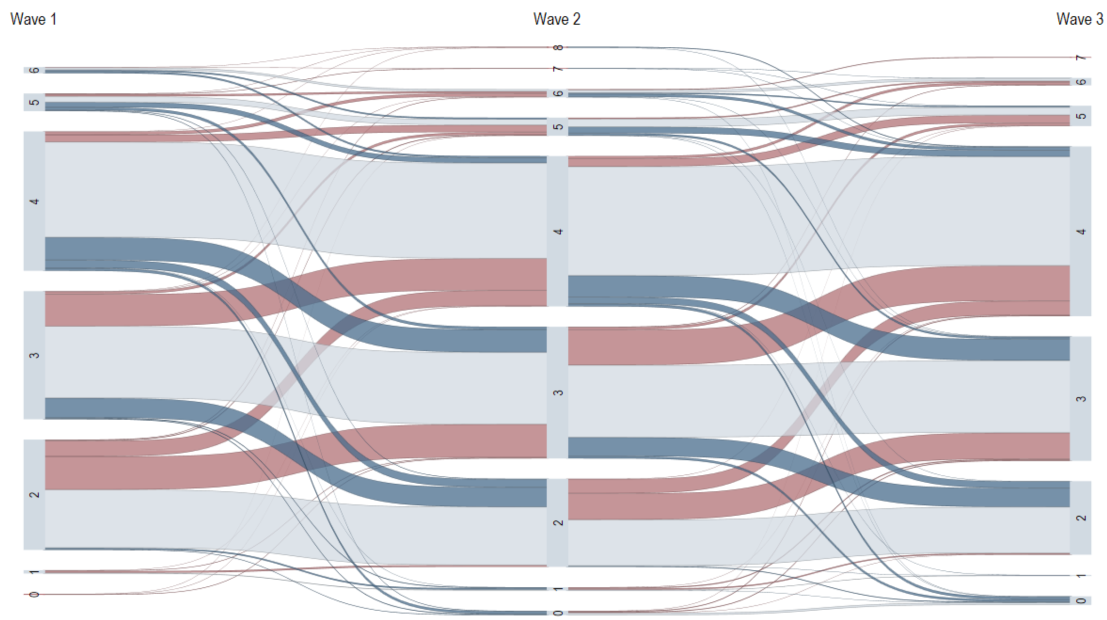


Figure 2. Within-Woman Variability in Fertility Preferences across Waves

Notes: The Sankey diagram presents individual-level transitions in ideal family size across survey waves. Flow widths are proportional to the number of women making each transition, nodes correspond to discrete ideal family size values at each wave, and colors indicate whether preferences increased, decreased, or remained stable between waves.

Preference revisions are often non-monotonic. Nearly one quarter of women who revise their preferences eventually return to their baseline value, and others revise in opposite directions across waves. Figure A1 shows that most adjustments are modest, typically by one child, while 11 percent of those who revise report larger revisions of two or more children. Within-woman variability therefore reflects both marginal adjustments as well as larger re-evaluations of preferences.

Finally, the direction of revisions is systematically related to baseline preferences. Women whose stated ideal is below the national total fertility rate (about four children) tend to revise upward and those at or above it revise downward. Dispersed initial preferences therefore converge on the national rate (Figure A2). This is not what regression to the mean would produce. Error-driven regression pulls observations toward the sample's own average, whereas the convergence here is on an external benchmark that differs from it. Classical error is also mean-preserving, yet the aggregate mean drifts upward monotonically across waves.

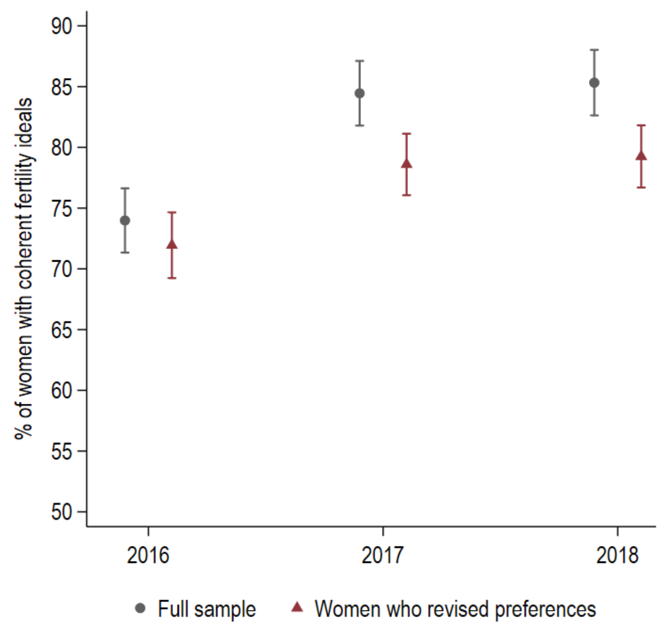


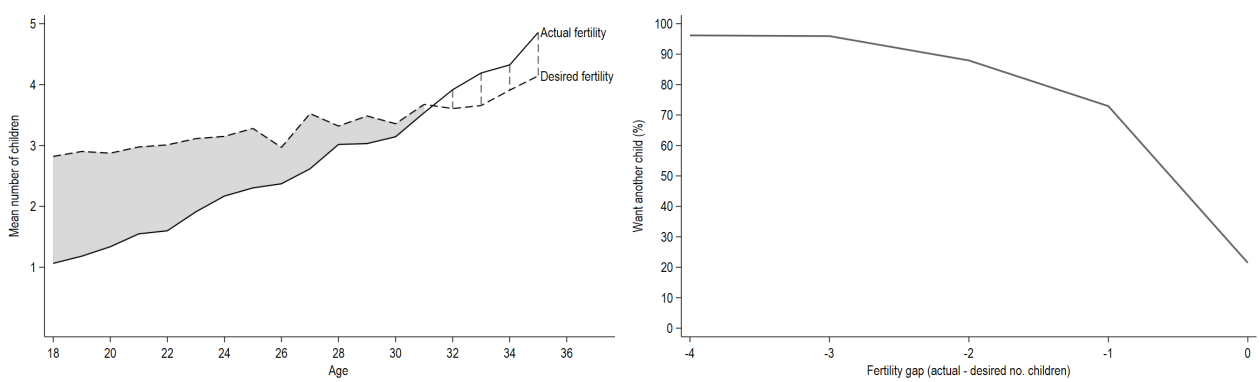
FIGURE 3. WITHIN-WOMAN CONSISTENCY IN FERTILITY PREFERENCES

Notes: The figure reports the share of women whose fertility intentions are consistent with their reported desired family size and current parity. Consistency is defined as women with fewer living children than their desired family size reporting that they want another child, and women who have reached or exceeded their desired family size reporting that they do not want additional children. Results are shown for the full sample and for women who revised their desired family size between survey waves. Points denote sample means, and error bars represent 95% confidence intervals.

Pattern 3: Revisions in fertility preferences are internally coherent and aligned with other fertility measures

Despite substantial within-woman variability in desired family size, fertility preferences remain internally coherent with other fertility measures. This indicates that revisions carry meaning rather than reflecting random reporting. We examine whether women's stated fertility intentions align with the gap between their living children and their contemporaneous IFS. Specifically, we examine whether women with fewer living children than their stated ideal report wanting another child, and whether those at or above it report wanting no more. Because intentions, IFS, and number of living children are elicited as separate survey items, their consistency is informative rather than mechanical.

Figure 3 shows that a large majority of women report intentions consistent with this preference–parity gap, rising from about 74 percent in 2016 to 85 percent in 2018. Crucially, coherence is only modestly lower among women who revise their preferences by at most six percentage points. If revisions were simply noise, coherence should deteriorate sharply among revisers; instead, at least three-quarters remain coherent in every wave. Revisions in IFS are therefore accompanied by corresponding adjustments in related measures rather than contradictory responses.



a) Desired and actual number of children by age   b) Probability of wanting another child

Figure 4. Fertility Preferences by Age and Realized Fertility

Notes: Panel A displays the evolution of mean actual (solid line) and desired (dashed line) number of children by age. Panel B shows the probability of wanting another child as a function of the gap between actual and desired fertility (defined as the number of living children minus desired family size). Negative values indicate that actual fertility falls below the desired level.

Pattern 4: Fertility preferences vary systematically with age and realized fertility

Fertility preferences and childbearing intentions evolve in structured ways over the life course, as documented for other sub-Saharan settings (Hayford, 2009; Hayford and Agadjanian, 2017). Panel A of Figure 4 plots mean ideal and actual family size by age. IFS is higher among older women, but actual fertility rises more steeply, so the gap between them closes around age 32. Beyond that age, realized fertility exceeds stated ideals. Because women are observed over a two-year window, this profile reflects differences across age groups rather than within-woman aging. Panel B shows that this narrowing gap predicts childbearing intentions. The probability of wanting another child falls sharply as the number of living children approaches IFS, from nearly universal among women well below their target to roughly one-fifth among those who have reached it.

Empirical Strategy

To estimate the causal effect of the FP intervention on both the *level* and *stability* of women's fertility preferences, we exploit the randomized design of the study and implement a two-step empirical strategy. We first estimate the intention-to-treat (ITT) effect on preference *levels* and then use the within-woman residual variance from this analysis to study preference *stability* in step two.

We estimate the effect on preference *levels* using two complementary specifications. First, exploiting random assignment, we estimate an analysis-of-covariance (ANCOVA) specification: we regress women's endline desired number of children on treatment assignment, baseline characteristics (age, age at sexual debut, number of living children, contraceptive use, educational attainment, work status, religion, and ethnicity), neighborhood fixed effects, and baseline fertility preferences. Because assignment was stratified, these controls include the variables used to stratify randomization.

Second, exploiting the panel, we estimate a difference-in-differences (DiD) specification with individual and survey-wave fixed effects:

$$Y_{it} = \alpha_i + \gamma_t + \beta(T_i \times P_t) + \varepsilon_{it}, \tag{1}$$

where Y_{it} denotes the desired number of children for woman i at wave t, α_i are individual fixed effects, γ_t are survey-wave fixed effects, and P_t equals one in the post-intervention waves (2017 and 2018). Because the individual fixed effects absorb all time-invariant characteristics, this specification includes no additional controls. The interaction $T_i \times P_t$ captures exposure to the intervention, and β is the ITT effect on preference levels, identified from within-woman changes relative to the control group. We cluster standard errors at the individual level. The residuals ε_{it} capture within-woman deviations net of individual means and common wave shocks and serve as the input to our analysis of preference stability.

In the second step, we examine within-woman instability in fertility preferences. Following prior work that has developed approaches for estimating longitudinal variation in fertility preferences (Culpepper, 2010; Yeatman et al., 2013), we extract the residuals $\hat{\varepsilon}_{it}$ from Equation 1 and define instability as their absolute value. Intuitively, these residuals capture the extent to which a woman's reported fertility preference in a given survey wave deviates from her predicted trajectory, after accounting for individual fixed effects, time effects, and treatment exposure. Taking the absolute value focuses on the magnitude of deviations, providing a measure of volatility over time. Because Equation 1 already removes the mean treatment effect on levels, this measure isolates the intervention's effect on the dispersion of preferences, net of any shift in their level.

We then estimate the effect of the intervention on instability using a Gamma generalized linear model with a log link to account for the distributional properties of the outcome:[2]

$$|\hat{\varepsilon}_{it}| \sim \text{Gamma}(\mu_{it}, \phi), \qquad \log(\mu_{it}) = \theta_0 + \theta_1 T_i. \tag{2}$$

The coefficient θ_1 captures the average difference in within-woman variability between treatment and control groups, with negative estimates indicating greater stability among treated women. Since the residuals are themselves estimated quantities from the first step, using them directly as outcomes in the second step discards the sampling variation they carry, resulting in standard errors with incorrect coverage properties. We therefore implement a bootstrap procedure that re-estimates both steps jointly across 500 replications, yielding standard errors with valid coverage. Together, this two-step approach allows us to separate treatment effects on the level of fertility preferences (step 1) from effects on their stability (step 2).

We also examine whether the intervention's effects vary across subgroups defined by baseline reproductive status, education, and distance to family planning services. These characteristics capture dimensions along which the effects of the intervention may plausibly differ. Because all three moderators are measured at baseline, treatment remains randomly assigned within subgroups, so within-subgroup comparisons retain their experimental interpretation. These analyses are exploratory; we motivate the choice of subgroups and interpret the results in the *Heterogeneity* section.

We complement this two-step approach with a simpler distributional comparison between groups of overall changes in fertility preferences during the study period. For each woman i, we compute the net change in desired family size between baseline and endline:

$$\Delta_i = y_{i,2018} - y_{i,2016}. \tag{3}$$

We then compare the variance of these net changes, Δ_i, across treatment and control groups:

$$H_1 : \ \text{Var}(\Delta_i \mid T_i = 1) < \text{Var}(\Delta_i \mid T_i = 0)$$

[2]We depart from standard OLS because the absolute deviations are non-negative, right-skewed, and heteroskedastic by construction. A Gamma specification accommodates these features by allowing the variance to proportionally increase with the mean, while the log link ensures non-negative predicted values and facilitates interpretation in proportional terms.

Under random treatment assignment, differences in dispersion can be attributed to the intervention, providing a complementary test of whether the program reduced or amplified overall shifts in preferences between baseline and endline. All analyses were carried out in Stata.[3]

Results

Intervention Effects on Preference Levels

Table 1 presents the estimated effects of the FP intervention on fertility preference *levels*. Columns 1–3 report results for the full sample; columns 4–6 and 7–9 report estimates separately for women who were pregnant and postpartum at baseline. Within each sample, the three specifications are OLS estimates without controls (column 1), OLS with baseline controls and neighborhood fixed effects (column 2), and a DiD estimate with individual and survey-wave fixed effects from Equation 1 (column 3).

Across the full sample, the estimated effect on preference levels is small and statistically insignificant in all three specifications. Subgroup results, however, reveal substantial heterogeneity by reproductive status. Among women who were pregnant at baseline (columns 4–6), the DiD specification, which leverages within-woman changes across all survey waves, yields a decrease of approximately 0.10 children, significant at the 10 percent level and equivalent to a 3 percent decline relative to the control mean. The OLS estimates for this group are negative but imprecise, and the DiD estimate does not survive restriction to the balanced panel (Table A4). In contrast, among women who were postpartum at baseline (columns 7–9), the intervention increased IFS by 0.13 to 0.15 children, about 4 percent relative to the control mean, with effects significant at the 5 percent level across all three specifications. These effects operate in opposite directions across subgroups, and the full-sample estimate is close to the sample-size-weighted average of the two. The two components do not rest on equal evidence: the postpartum increase holds across all three specifications, while the negative estimate among pregnant women appears in one and does not survive the balanced panel.

These magnitudes are modest in absolute terms but not negligible relative to the movement this population displays over the study period. The postpartum increase of 0.13 children is roughly three-quarters of the total drift in the sample mean across all three waves,

[3] We used Anthropic's Claude (Opus 4.8) to assist in debugging and revising portions of the Stata code implementing these analyses. The tool did not determine model specifications, generate data, or produce or interpret results; all analytical decisions were made by the authors, who reviewed, executed, and verified all code and output.

Table 1—Effects on Fertility Preference Levels

	Full sample			Pregnant (baseline)			Postpartum (baseline)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treatment	0.055	0.049		-0.039	-0.037		0.145**	0.135**	
	(0.047)	(0.046)		(0.066)	(0.065)		(0.067)	(0.065)	
Treatment × Post			0.012			-0.099*			0.127**
			(0.042)			(0.056)			(0.062)
Baseline Controls		√			√			√	
Neighborhood FE		√			√			√	
Woman FE			√			√			√
Year FE			√			√			√
Control mean	3.32	3.32	3.32	3.32	3.32	3.32	3.33	3.33	3.33
Number of observations	1664	1662	5276	839	838	2680	825	824	2596

Notes: The table reports the effects of the FP intervention on women's fertility preference levels for the full sample and for women who were pregnant or postpartum at baseline. Columns correspond to alternative specifications. Columns (1), (4), and (7) report simple OLS estimates. Columns (2), (5), and (8) report OLS estimates with baseline controls, including age, age at sexual debut, number of living children, contraceptive use, educational attainment, employment status, religion, ethnicity, neighborhood fixed effects, and baseline fertility preference levels. Columns (3), (6), and (9) report DiD estimates with individual and survey-wave fixed effects, with standard errors clustered at the individual level.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

from 3.19 to 3.36. The aggregate null is therefore a composition result rather than an absence of movement.

Since not all women are observed in all three survey waves, the DiD estimates include women with incomplete panel histories. Such attrition could introduce compositional differences across waves and bias the estimates, particularly if it were correlated with treatment status or fertility preferences. Table A2 shows that attrition is unrelated to both. We nevertheless re-estimate all specifications on a balanced panel of women observed in all three waves (Table A4). The results are broadly consistent with the main findings. The marginally significant DiD effect among pregnant women is no longer statistically significant, which is unsurprising given its imprecision and its appearance in the DiD specification alone. For postpartum women, the effect remains robust: the OLS estimates are essentially unchanged and remain significant at the 5 percent level, though the DiD estimate weakens to the 10 percent level.

Intervention Effects on Preference Stability

Figure 5 summarizes the effects of the intervention on preference *stability* using the residual-based measure from Equation 2. The outcome captures within-woman variability in stated fertility preferences after accounting for time-invariant characteristics and shifts in average preference levels. Negative estimates therefore indicate *greater* preference stability: less unexplained variability among treated women relative to controls.

In the full sample, the estimated effect is negative but small and not statistically significant. The subgroup estimates, however, reveal clear heterogeneity by reproductive status. Among women who were pregnant at baseline, the intervention reduces preference instability, an effect significant at the 10 percent level. The estimate from our main Gamma GLM specification implies a reduction of approximately 12 percent in within-woman variability: treated pregnant women show roughly one-eighth less unexplained fluctuation in stated fertility preferences than controls. In contrast, there is no evidence of an effect among women who were postpartum at baseline: estimates are close to zero and statistically insignificant.

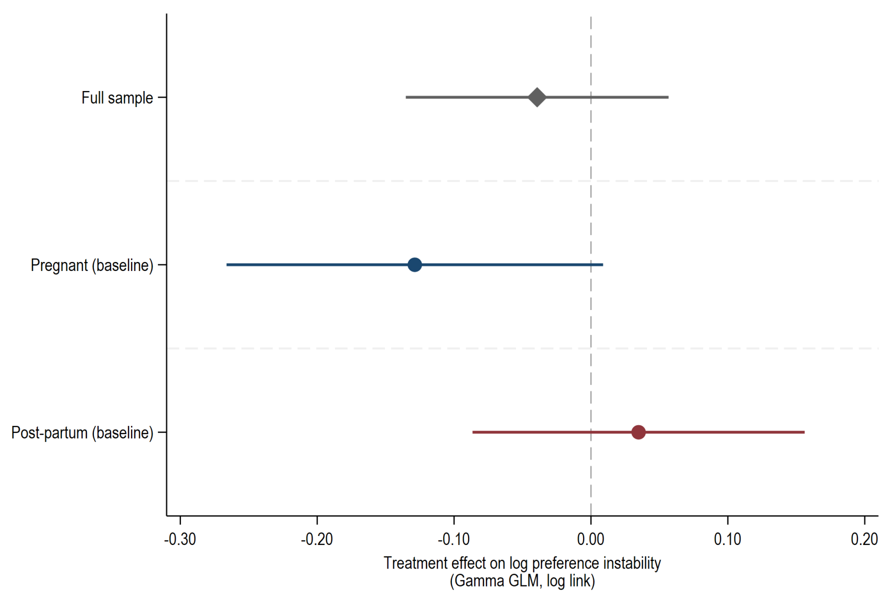


Figure 5. Effects on Fertility Preference Stability

Notes: The figure plots treatment coefficients from the second-stage Gamma GLM (Equation 2), where the outcome is the absolute value of first-stage residuals. Coefficients represent the proportional change in within-woman variability associated with the intervention, shown for the full sample and separately by baseline reproductive status. Confidence intervals are based on cluster bootstrap with 500 replications that resamples women rather than observations. Both estimation stages are re-run within each replication, so uncertainty in the first-stage residuals is reflected in the reported standard errors. The dashed line at zero indicates no treatment effect.

These patterns are robust across alternative specifications, including OLS, and across balanced and unbalanced samples (Table A5). Effect sizes differ across specifications because the models estimate different quantities, but the qualitative result is consistent: a significant reduction in preference instability emerges only for pregnant women, with no comparable effect for postpartum women.

Reassuringly, this conclusion does not depend on how we measure stability. As an independent check, we compare the dispersion of women's net preference changes between treatment and control directly, through a standard deviation (SD) test built on the baseline-to-endline net change in IFS, $\Delta_i = y_{i,2018} - y_{i,2016}$, and constructed without reference to the regression residuals. The two approaches measure related but distinct quantities: the residual-based estimate captures each woman's wave-to-wave deviation from her own trajectory across all three waves, whereas the SD test captures the dispersion across women in net revisions between two points. We would therefore not expect them to agree in precision, but they agree in direction and in where the effect falls. Table 2 reports the results. Among women pregnant at baseline, net preference changes are markedly less dispersed under treatment: their standard deviation falls from 1.15 to 0.96 children, a reduction of roughly 16 percent. Among postpartum women the two arms are statistically indistinguishable. In the pooled sample the difference remains significant, but it is arithmetically the pregnancy effect averaged with a null subgroup: a reduction large enough to survive dilution rather than one extending beyond the group in which it arises.

Taken together, the level and stability results reveal a single pattern: the intervention's effects differ by reproductive status and operate on different margins. For women exposed during pregnancy, the effect falls on stability: within-woman variability declines by roughly 12 percent, consistent in direction across our residual-based and distributional approaches, while the downward effect on the level is smaller and less robust. For postpartum women, the pattern reverses: IFS rises by about 4 percent while stability is unchanged. That the two groups respond on different margins is consistent with the framework's expectation that a woman's situation at exposure conditions what a program can change for her. Our design does not identify why they differ. Neither margin is visible in a pooled summary of stated preferences: the full-sample level estimate is a precise null, and the residual-based stability measure is likewise silent across the whole sample. An evaluation reading either quantity on its own would record no effect on preferences at all.

Table 2—SD Test of Fertility Preference Stability

	Control N=882	Treatment N = 781	Difference	P-value
	(1)	(2)	(1) - (2)	
Full sample				
Mean ($\bar{x}$)	0.157	0.202	-0.045	0.399
				$diff \neq 0$
Standard Deviation (SD)	1.121	1.039	0.082	0.014**
				$\frac{sd(1)}{sd(2)} > 1$
Pregnant at baseline				
Mean ($\bar{x}$)	0.198	0.127	0.071	0.338
				$diff \neq 0$
Standard Deviation (SD)	1.151	0.964	0.187	0.0001***
				$\frac{sd(1)}{sd(2)} > 1$
Postpartum at baseline				
Mean ($\bar{x}$)	0.114	0.275	-0.161	0.0353
				$diff \neq 0$
Standard Deviation (SD)	1.089	1.104	-0.015	0.607
				$\frac{sd(1)}{sd(2)} > 1$

Notes: The table presents the distribution of net changes in fertility preferences, Δ_i, by treatment status. Mean rows present average changes in Δ_i, while Standard Deviation (SD) rows report dispersion of Δ_i within each group. P-values in the Mean rows test for equality of means between the treatment and control groups, while p-values in the SD rows test whether dispersion is greater in the control group. A significant difference in SD indicates lower preference stability among control women. $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Heterogeneity in Stability Effects

The intervention combined two features that may be relevant to how women form and express fertility preferences: improved access to FP services and repeated individualized counseling. We therefore examine whether its effect on the stability of stated fertility preferences varies across baseline characteristics that capture differences in women's circumstances at the time of exposure. Specifically, we consider baseline education, distance to the nearest FP facility, and reproductive status. These characteristics may be associated with differences in women's access to information, access to services, or the circumstances in which fertility preferences are considered. Our purpose is not to identify the mechanisms through which the intervention affects stability, but to characterize whether the effects are concentrated among particular groups. Because all moderators are measured at baseline, treatment remains randomly assigned within subgroups, so within-subgroup comparisons retain their experimental interpretation.

We estimate the same two-stage stability model used in our main analysis on subgroups defined by these characteristics. For the information channel, we use women's highest completed level of education, at most primary versus secondary or higher, as a proxy for prior reproductive knowledge, and predict a larger stability effect among the lower-education subgroup. For the access channel, we proxy binding constraints with the neighborhood mean of self-reported distance to the nearest facility, split at the sample median.[4] The framework predicts a larger stability effect in high-distance neighborhoods. Because the framework also treats position in the reproductive cycle as moderating both channels, we report results pooled across the full sample and separately by baseline reproductive status.

Table 3 reports the results. Pooled across the full sample (Panel A), the treatment effects on stability are uniformly negative but small and statistically indistinguishable from zero across all four specifications. Among women pregnant at baseline (Panel B), however, the estimated effects are concentrated among women with at most primary education and those living in higher-distance neighborhoods. Among women with at most primary education, treatment reduces within-woman variability by 19.7 percent (-0.219, $p < 0.05$), while the estimated effect among women with secondary education or higher is close to zero. In higher-distance neighborhoods, the estimated effect is -0.161 ($p < 0.10$), corresponding to roughly a 15 percent reduction in within-woman variability. The estimate in low-distance neighborhoods is small and statistically indistinguishable from zero.

Among women postpartum at baseline (Panel C), the treatment effect on stability is

[4] We aggregate distance to the neighborhood level to address substantial individual-level missingness in the self-reported variable.

statistically indistinguishable from zero across the education and distance subgroups. This pattern mirrors the absence of an overall stability effect for this group in the main analysis.

TABLE 3—HETEROGENEITY IN TREATMENT EFFECTS ON PREFERENCE STABILITY

	Education level		Distance to health facility	
	(1)	(2)	(3)	(4)
	$\leq$ Primary	$>$ Primary	High distance	Low distance
Panel A: Full sample				
Treatment	-0.035	-0.030	-0.049	-0.018
	(0.063)	(0.067)	(0.060)	(0.074)
N	2942	2334	3420	1856
Panel B: Pregnant at baseline				
Treatment	-0.219**	0.031	-0.161*	-0.044
	(0.092)	(0.092)	(0.087)	(0.099)
N	1536	1144	1771	909
Panel C: Post-partum at baseline				
Treatment	0.144	-0.101	0.046	0.010
	(0.088)	(0.095)	(0.082)	(0.105)
N	1406	1190	1649	947

Notes: The table reports treatment effects on within-woman preference stability by education and access subgroups. The outcome is $|\hat{\varepsilon}_{it}|$, the absolute value of the residual from a first-stage DiD specification with woman and survey-wave fixed effects (Equation 1); coefficients are from a second-stage Gamma GLM with a log link, estimated separately for each cell. Education level (columns 1–2) is the woman's highest completed level of education, split into two categories: $\leq$ Primary refers to women with at most primary education; $>$ Primary refers to women with secondary or higher education. Distance to health facility (columns 3–4) is the neighborhood-mean of self-reported distance (in km) to the nearest family planning facility, split at the sample median. High distance indicates above-median neighborhoods, where access constraints are more binding; Low distance indicates neighborhoods at or below the median. Distance is measured at the neighborhood level to address individual-level missingness in the self-reported variable. $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.

Taken together, the subgroup estimates indicate that the intervention's effect on stability was concentrated among women who were pregnant at baseline and, within that group, among those with lower education and greater distance from family planning services. These patterns describe where the stability effect is strongest, but they do not by themselves establish why the effect differs across groups. In particular, because the intervention bundled improved access with counseling, the subgroup results cannot distinguish an information mechanism from an access mechanism, nor can they establish whether greater stability reflects reduced uncertainty, reduced flexibility, or another process.

Conclusion

This study examines how a comprehensive family planning intervention shapes both the level and the stability of fertility preferences in urban Malawi. As in other settings where preferences have been followed over time, our study population displays a sharp contrast between aggregate and individual dynamics: IFS is remarkably stable in the aggregate, yet a majority of women revise their own stated preference at least once. This within-woman variation is not reporting noise. It is consistent with other fertility preference measures and follows predictable life-cycle trajectories. What our setting adds is not this pattern but a way to intervene on it. Embedded in a randomized expansion of family planning services, it allows us to provide what is, to our knowledge, the first experimental estimate of the causal effect of a family planning intervention on the stability of preferences, distinct from its effect on their level. We measure stability as a longitudinal property inferred through observed variation across repeated reports.

The intervention shapes preferences along both dimensions, with effects that differ by baseline reproductive status. It raises IFS by 0.13 to 0.15 children among women who were postpartum at baseline, an effect that leaves no trace in the sample average. That pooled null is a composition result: it averages this movement with a weaker and less robust negative estimate among pregnant women. Its effect on stability falls on the other subgroup. Both the residual variance of the fixed effects model and a complementary distributional test show the intervention reducing within-woman variability among women pregnant at baseline. Postpartum women are unaffected on this margin, and the finding holds across specifications and across balanced and unbalanced samples. In the distributional test the reduction is pronounced enough to register across the full sample, though the residual-based measure is silent there. These stability effects among pregnant women concentrate among those whose informational and access constraints were plausibly most binding, the least-educated and those living farthest from a facility. The fact that the intervention affects different dimensions of stated preferences in different subgroups reinforces the point that a pooled change in mean IFS is an incomplete description of how policy can move preferences.

These results carry three broader implications. First, they show that the temporal stability of stated fertility preferences is itself policy-responsive. Preferences need not be treated as either fixed targets or wholly fluid constructs: they can exhibit persistence while also changing in systematic response to the reproductive environments in which they are formed and expressed. An intervention can leave the aggregate mean of stated ideals untouched while measurably altering their variability. This matters for the long-running debate over whether fertility decline reflects falling demand for children or the fuller implementation of preferences that women already hold (Ibitoye et al., 2022; Bongaarts, 2025). If a program that improves implementation can also move the demand it seeks to implement, then the two channels are not independent (Pritchett, 1994; Bongaarts, 1994). Our design was not built to adjudicate whether FP programs shape preferences as opposed to merely satisfying them, but it does show that the demand that a program is said to implement cannot always be treated as fixed independently of the program itself.

Second, the stability result has implications for preference-based measures of fertility and program impact. A decomposition of fertility into wanted and unwanted components requires a stated preference against which births are classified. But when a woman reports different ideals at different points in time, the choice of which report provides the benchmark matters. Changes in the stability of that benchmark can therefore affect classification even when its mean does not change. This is particularly relevant alongside the growing literature questioning preference-based indicators on grounds of reproductive autonomy and the normative content of classifying births as intended or unintended (Senderowicz, 2020; Upadhyay et al., 2014; Karp et al., 2025). Our point is complementary to that debate: whatever one takes a stated preference to mean, an evaluation that uses it to assess a program stands on different ground once the program has moved the preference. The concern is not that preference-based measures are inherently invalid, but that their temporal behavior under policy exposure has largely been left outside the object of evaluation.

Third, the preference findings are consistent with the intervention's effects on realized fertility outcomes in companion work using the same sample. Karra et al. (2026) find no effect of the intervention on a narrow, ex-ante measure of unwanted pregnancy which flags only pregnancies that are inconsistent with a woman's baseline preference for no more children, a result that they attribute to that measure's inability to register violations of spacing, as opposed to limiting, preferences. Once timing is incorporated and those pregnancies that arrive sooner than a woman's stated preferred spacing are also counted as unwanted (mistimed), the estimated intervention effect grows to a 42 to 49 percent relative reduction in unintended pregnancy over 24 months and is concentrated specifically among women who wanted another child at baseline rather than those who wanted no more. This pattern is consistent with an intervention that facilitates the timing of desired fertility more than it changes the desire for children itself: improved FP access appears to help women avoid pregnancies that arrive sooner than they would have wanted, without necessarily reducing the stated desire for additional children. We do not read this evidence as resolving the mechanisms behind the stability result, or as showing that greater stability causes reductions in unintended pregnancy. We cite it only for what it shows directly—that the intervention's effect on realized fertility outcomes is likely concentrated on timing, and not on the content of what women say they want.

Three limitations bound these conclusions. First, the intervention bundled free services, free transport, and individualized counseling, and our design estimates the effect of that package. The heterogeneity by baseline education and distance is consistent with both an informational and an access channel, but the components were never varied independently, so we cannot attribute the effects to either one. Second, we identify that the intervention altered the stability of stated preferences, but not why. Reduced variability is consistent with several processes: the resolution of genuine uncertainty, convergence toward goals that counseling makes salient, or a reduced tendency to revise in response to changing circumstances. Our data cannot distinguish among them. Nor can stability be equated with certainty: greater consistency across waves is not evidence that women felt more certain about their goals, and it is not evidence that the resulting preferences were better formed. Third, our estimates come from a single urban population and one program. Whether preferences prove equally responsive to other reproductive environments, and in which direction, remains open.

Finally, these findings do not recommend that programs be designed to stabilize preferences, nor do they treat greater stability as a desirable end. The welfare content of a more stable preference is precisely what our design cannot adjudicate. The contribution is more modest but, we argue, important: the stability of stated fertility preferences is not simply a fixed background characteristic of the women or a nuisance feature of measurement. It is an empirically observable temporal property that can itself respond to policy. Evaluations that rely on stated preferences should therefore treat their stability as something a program may affect—a quantity to monitor rather than assume. Stated preferences remain informative; our argument is for reading them as products of the reproductive environments in which they are elicited, and for recognizing that the programs evaluated against them are among the things that shape them.

References

Ashraf, N., E. Field, and J. Lee (2014). Household Bargaining and Excess Fertility: An Experimental Study in Zambia. *American Economic Review 104*(7), 2210–2237. 2, 6

Bachrach, C. A. and S. P. Morgan (2013). A Cognitive-Social Model of Fertility Intentions. *Population and development review 39*(3), 459–485. 1, 5

Badolato, L., S. R. Hayford, and K. B. Guzzo (2025). Multiple dimensions of uncertainty in fertility goals: recent trends and patterns in the United States. *Genus 81*(1), 14. 4

Barker, R. and I. Buber-Ennser (2024). Uncertainty and flexibility of fertility intentions. *Advances in Life Course Research 61*, 100618. 4

Bhrolchin, M. N. and . Beaujouan (2019). Do People Have Reproductive Goals? Constructive Preferences and the Discovery of Desired Family Size. In R. Schoen (Ed.), *Analytical Family Demography*, pp. 27–56. Cham: Springer International Publishing. 1, 5

Bongaarts, J. (1990). The Measurement of Wanted Fertility. *Population and Development Review 16*(3), 487. 1

Bongaarts, J. (1994). The impact of population policies: Comment. *Population and Development Review 20*(3), 616–620. 3, 25

Bongaarts, J. (2017). Africa's Unique Fertility Transition. *Population and Development Review 43*(S1), 39–58. 1

Bongaarts, J. (2025). Fertility Transitions in Low- and Middle-Income Countries: The Role of Preferences. *Population and Development Review 51*(1), 163–180. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/padr.12675. 1, 25

Canning, D., M. Karra, and D. Maggio (2026). Family planning, womens employment, and household labor supply. 10

Casterline, J. B. and L. O. El-Zeini (2007). The estimation of Unwanted Fertility. *Demography 44*(4), 729–745. 1

Cleland, J., K. Machiyama, and J. B. Casterline (2020). Fertility preferences and subsequent childbearing in Africa and Asia: A synthesis of evidence from longitudinal studies in 28 populations. *Population Studies 74*(1), 1–21. 3

Coombs, L. C. (1974). The Measurement of Family Size Preferences and Subsequent Fertility. *Demography 11*(4), 587–611. 4

Culpepper, S. A. (2010). Studying Individual Differences in Predictability With Gamma Regression and Nonlinear Multilevel Models. *Multivariate Behavioral Research 45*(1), 153–185. 15

Dupas, P., S. Jayachandran, A. Lleras-Muney, and P. Rossi (2025). The Negligible Effect of Free Contraception on Fertility: Experimental Evidence from Burkina Faso. *American Economic Review 115*(8), 2659–2688. 2

Freedman, R. (1997). Do Family Planning Programs Affect Fertility Preferences? A Literature Review. *Studies in Family Planning 28*(1), 1–13. 2, 6

Hayford, S. R. (2009). The Evolution of Fertility Expectations over the Life Course. *Demography 46*(4), 765–783. 1, 3, 14

Hayford, S. R. and V. Agadjanian (2017). Determined to stop? Longitudinal analysis of the desire to have no more children in rural Mozambique. *Population Studies 71*(3), 329–344. 3, 4, 14

Heiland, F., A. Prskawetz, and W. C. Sanderson (2008). Are Individuals Desired Family Sizes Stable? Evidence from West German Panel Data: La taille de famille dsire est-elle stable? Analyse de donnes de panel en Allemagne de lOuest. *European Journal of Population / Revue europeenne de Demographie 24*(2), 129–156. 1, 3

Herrera-Almanza, C. and A. Seitz McCarthy (2025). Strategic responses to disparities in spousal desired fertility: experimental evidence from rural Tanzania. *Journal of Population Economics 38*(4), 85. 2, 6

Ibitoye, M., J. B. Casterline, and C. Zhang (2022). Fertility Preferences and Contraceptive Change in Low and MiddleIncome Countries. *Studies in Family Planning 53*(2), 361–376. 1, 25

Karp, C., S. N. Wood, E. Omoluabi, F. M. OlaOlorun, P. Z. Akilimali, S. P. S. Kibira, F. Makumbi, G. Guiella, P. Gichangi, and C. Moreau (2025). Evaluating contraceptive empowerment among women and girls in sub-saharan africa: Validation of the women's and girls empowerment in sexual and reproductive health (wge-srh) index contraceptive empowerment subscale. *Studies in family planning 56*(3), 676–698. 3, 25

Karra, M. and D. Canning (2020). The Effect of Improved Access to Family Planning on Postpartum Women: Protocol for a Randomized Controlled Trial. *JMIR Research Protocols 9*(8), e16697. 8, 9

Karra, M., D. Maggio, M. Guo, B. Ngwira, and D. Canning (2022). The causal effect of a family planning intervention on womens contraceptive use and birth spacing. *Proceedings of the National Academy of Sciences 119*(22), e2200279119. 2, 10

Karra, M., M. C. Montás, and S. Vincent (2026). Family planning and unintended pregnancy: Experimental evidence from urban malawi. Working paper. 25

Karra, M. and K. Zhang (2025). User-centered approaches to contraceptive counseling: Experimental evidence from urban malawi. *Journal of Development Economics*, 103690.

2

Kodzi, I. A., J. B. Casterline, and P. Aglobitse (2010). The Time Dynamics of Individual Fertility Preferences Among Rural Ghanaian Women. *Studies in Family Planning 41*(1), 45–54. 1, 4, 8, 10, 11

Liefbroer, A. C. (2009). Changes in Family Size Intentions Across Young Adulthood: A Life-Course Perspective. *European Journal of Population / Revue europeenne de Demographie 25*(4), 363–386. 1, 3

Maggio, D., M. Karra, and D. Canning (2024). Family Planning and Childrens Human Capital: Experimental Evidence From Urban Malawi. *Demography 61*(5), 1667–1698. 10

Morgan, S. P. (1981). Intention and Uncertainty at Later Stages of Childbearing: The United States 1965 and 1970. *Demography 18*(3), 267–285. 4

Mueller, M. W., J. Hamory, J. Johnson-Hanks, and E. Miguel (2022). The illusion of stable fertility preferences. *Population Studies 76*(2), 169–189. 1, 4

Nair, N. K. and L. P. Chow (1980). Fertility Intentions and Behavior: Some Findings from Taiwan. *Studies in Family Planning 11*(7/8), 255. 3

National Statistical Office (NSO) [Malawi] and ICF (2017). Malawi demographic and health survey 2015–16. Technical report, NSO and ICF, Zomba, Malawi, and Rockville, Maryland, USA. 7

National Statistical Office (NSO) [Malawi], Utica International, and ICF (2025). Malawi demographic and health survey 2024: Final report. DHS Final Report FR398, NSO, Utica International, and ICF, Zomba, Malawi; Columbia, Maryland, USA; and Rockville, Maryland, USA. 7

Pritchett, L. H. (1994). Desired fertility and the impact of population policies. *Population and Development Review 20*(1), 1–55. 3, 25

Quesnel-Valle, A. and S. P. Morgan (2003). Missing the Target? Correspondence of Fertility Intentions and Behavior in the U.S. *Population Research and Policy Review 22*(5-6), 497–525. 3

Senderowicz, L. (2020). Contraceptive Autonomy: Conceptions and Measurement of a Novel Family Planning Indicator. *Studies in Family Planning 51*(2), 161–176. 3, 25

Sennott, C. and S. Yeatman (2012). Stability and Change in Fertility Preferences Among Young Women in Malawi. *International Perspectives on Sexual and Reproductive Health 38*(1), 34–42. 1, 4, 10, 11

Tan, P. C. and N. P. Tey (1994). Do Fertility Intentions Predict Subsequent Behavior? Evidence from Peninsular Malaysia. *Studies in Family Planning 25*(4), 222. 3

Trinitapoli, J. and S. Yeatman (2018). The Flexibility of Fertility Preferences in a Context of Uncertainty. *Population and Development Review 44*(1), 87–116. 1, 5, 8, 10, 11

Upadhyay, U. D., S. L. Dworkin, T. A. Weitz, and D. G. Foster (2014). Development and Validation of a Reproductive Autonomy Scale. *Studies in Family Planning 45*(1), 19–41. 3, 25

Yeatman, S., C. Sennott, and S. Culpepper (2013). Young Womens Dynamic Family Size Preferences in the Context of Transitioning Fertility. *Demography 50*(5), 1715–1737. 1, 2, 4, 8, 10, 11, 15

Online Appendix

Appendix A — Sample, Balance, and Attrition

Table A1—Baseline Balance Table

	Full sample (1)	Treatment (2)	Control (3)	Difference (3) - (2)
Background characteristics				
Woman's age (years)	24.58	24.66	24.51	-0.148
Primary education (1 = yes)	0.984	0.986	0.982	-0.004
Secondary education (1 = yes)	0.413	0.414	0.412	-0.001
Tertiary education (1 = yes)	0.023	0.019	0.028	0.009
Currently working (1 = yes)	0.096	0.099	0.093	-0.007
Religion	2.909	2.900	2.917	0.018
Ethnicity	1.857	1.875	1.840	-0.035
Age at first sex	18.86	18.90	18.81	-0.094
Currently pregnant (1 = yes)	0.515	0.515	0.516	0.001
Currently using contraception (1 = yes)	0.237	0.239	0.235	-0.003
Ever used contraception (1 = yes)	0.755	0.775	0.736	-0.039**
Number of children	1.738	1.771	1.708	-0.063
Fertility preferences				
Desired number of children	3.187	3.214	3.162	-0.052
Husband same preference (1 = yes)	0.605	0.587	0.622	0.035
Wants another child (1 = yes)	0.572	0.567	0.578	0.011
Preferred waiting time (months)	63.95	63.46	64.39	0.932
Number of observations	2140	1027	1113	2140
Joint p-value				0.124

Notes: Means are reported by group, with p-values testing for differences between treatment and control groups. The joint p-value tests for overall differences in baseline characteristics between groups.

Table A2—Tests for Selective Attrition: Baseline Characteristics by follow-up status

	Followed up (1)	Attrited (2)	Diff. (3)	SE (4)
Treatment assignment (1 = yes)	0.473	0.500	-0.027	(0.025)
Fertility preference	3.18	3.21	-0.028	(0.049)
Other characteristics				
Woman's age (years)	24.90	23.65	1.249***	(0.227)
Currently pregnant (1 = yes)	0.50	0.56	-0.058*	(0.025)
Primary education (1 = yes)	0.99	0.98	0.008	(0.006)
Secondary education (1 = yes)	0.45	0.30	0.150***	(0.024)
Tertiary education (1 = yes)	0.03	0.01	0.017*	(0.007)
Working (1 = yes)	0.10	0.08	0.026	(0.015)
Religion	2.87	3.03	-0.169*	(0.074)
Ethnicity	1.86	1.85	0.004	(0.065)
Age at first sex	18.95	18.57	0.379**	(0.138)
Currently using contraception (1 = yes)	0.24	0.22	0.028	(0.021)
Ever used contraception (1 = yes)	0.77	0.71	0.066**	(0.021)
Number of children	1.81	1.54	0.262***	(0.067)
Wants another child (1 = yes)	0.55	0.63	-0.079**	(0.025)
Preferred waiting time (months)	63.69	64.63	-0.947	(1.503)
Observations	1,672	468		

Notes: The table tests for selective attrition along the two dimensions most relevant to our analysis — treatment assignment and baseline fertility preferences — as well as on other baseline characteristics. Columns (1) and (2) report mean values within each group, where *Attrited* refers to women not observed in all panel waves. Column (3) reports the difference (Followed up − Attrited) from a two-sample t-test, with robust standard errors in column (4). Sample: women interviewed at the 2016 baseline wave. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix B — Evidence on Preference Dynamics

Table A3—Change Patterns in Fertility Preferences

	Full sample	Age Group			Baseline Parity			
		18-24	25-29	30-35	None	One	Two	≥ 3
Changed at least once	0.594	0.578	0.626	0.599	0.549	0.595	0.597	0.620
Times changed								
Once	0.615	0.638	0.609	0.563	0.722	0.625	0.601	0.558
Twice	0.385	0.362	0.391	0.437	0.278	0.375	0.399	0.442
Direction change								
Increased	0.660	0.697	0.627	0.604	0.641	0.700	0.701	0.585
Decreased	0.340	0.303	0.373	0.396	0.359	0.300	0.299	0.415
Timing								
W1 only	0.322	0.338	0.310	0.296	0.368	0.319	0.282	0.336
W1, reverted back	0.233	0.238	0.226	0.231	0.215	0.247	0.218	0.241
W2 only	0.293	0.300	0.299	0.266	0.354	0.306	0.319	0.223
W1 and W2	0.152	0.124	0.165	0.206	0.062	0.128	0.181	0.201

Notes: The table presents patterns of change in fertility preferences for the full sample and by age group and baseline parity. Percentages are conditional on having revised preferences at least once during the study period.

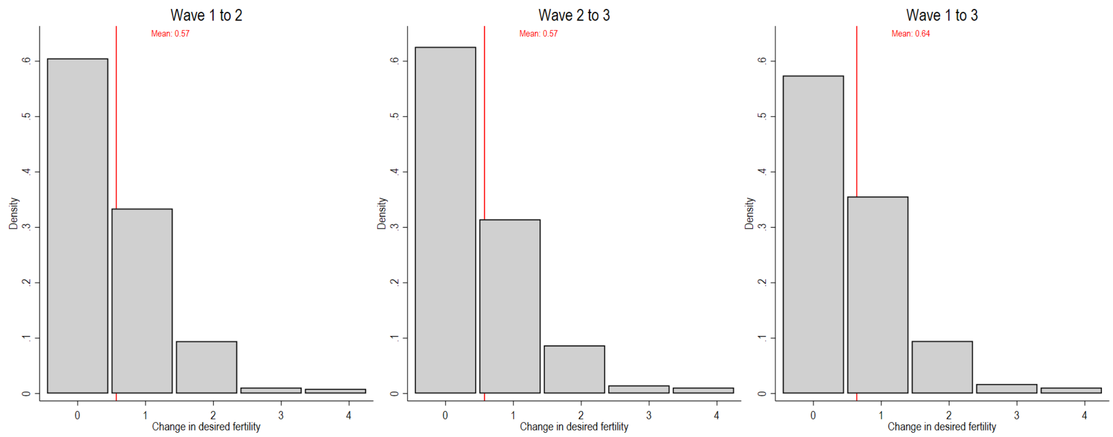


Figure A1. Absolute Change in Fertility Preferences across Waves

Notes: The figure displays the distribution of absolute changes in women's reported desired number of children between consecutive survey waves. Absolute changes capture the size of stated preference revisions, regardless of direction, with zero indicating no change and larger values indicating larger adjustments.

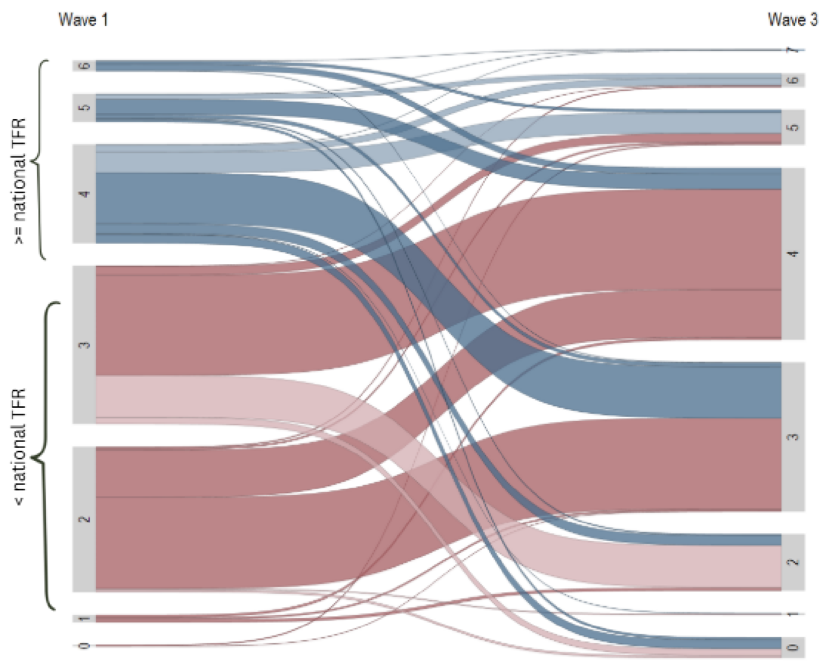


Figure A2. Net Revision Trajectories of Fertility Preferences

Notes: The Sankey diagram presents within-woman revisions in desired family size between Wave 1 and Wave 3 among women who revised their fertility preferences at least once during the study period. Nodes represent categories in each wave, and flows connect baseline to endline preference values, with widths proportional to the share of women making each transition. Colors and shading differentiate women with baseline preferences below versus at or above the national total fertility rate.

Appendix C — Robustness and Additional Estimates

Table A4—Effects on Fertility Preference Levels, Balanced Sample

	Full sample			Pregnant (baseline)			Postpartum (baseline)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treatment	0.062	0.056		-0.037	-0.036		0.155**	0.147**	
	(0.048)	(0.047)		(0.068)	(0.066)		(0.068)	(0.067)	
Treatment × Post			0.028			-0.072			0.127*
			(0.044)			(0.058)			(0.066)
Baseline Controls		√			√			√	
Neighborhood FE		√			√			√	
Woman FE			√			√			√
Year FE			√			√			√
Control mean	3.32	3.32	3.32	3.32	3.32	3.32	3.33	3.33	3.33
Number of observations	1590	1588	4770	796	795	2388	794	793	2382

Notes: The table reports the effects of the FP intervention on fertility preference levels for the full sample and by baseline reproductive status. These estimates replicate those presented in Table 1, restricted to the balanced sample of women observed in all three data waves.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A5—Effects on Fertility Preference Stability: Residual Variance Approach

	OLS		Gamma GLM	
	(1) Unbalanced	(2) Balanced	(3) Unbalanced	(4) Balanced
Panel A: Full sample				
Treatment	-0.015	-0.019	-0.039	-0.049
	(0.019)	(0.019)	(0.049)	(0.053)
Number of observations	5276	4770	5276	4770
Panel B: Pregnant (baseline)				
Treatment	-0.048*	-0.051*	-0.129*	-0.138*
	(0.026)	(0.026)	(0.070)	(0.074)
Number of observations	2680	2388	2680	2388
Panel C: Postpartum (baseline)				
Treatment	0.014	0.007	0.035	0.017
	(0.027)	(0.027)	(0.062)	(0.070)
Number of observations	2596	2382	2596	2382

Notes: The table reports estimates of the effects of the FP intervention on fertility preference stability using the residual-based measure from Equation 2. Columns (1)–(2) present OLS estimates, while Columns (3)–(4) present Gamma GLM estimates with a log link. Results are presented for balanced and unbalanced samples, for the full sample, and by baseline reproductive status. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A6—SD Test of Fertility Preference Stability, Balanced Sample

	Control N=838	Treatment N=752	Difference	P-value
	(1)	(2)	(1) - (2)	
Full sample				
Mean ($\bar{x}$)	0.150	0.205	-0.054	0.313
				$diff \neq 0$
Standard Deviation (SD)	1.107	1.034	0.073	0.028**
				$\frac{sd(1)}{sd(2)} > 1$
Pregnant at baseline				
Mean ($\bar{x}$)	0.199	0.130	0.069	0.355
				$diff \neq 0$
Standard Deviation (SD)	1.118	0.964	0.154	0.002***
				$\frac{sd(1)}{sd(2)} > 1$
Postpartum at baseline				
Mean ($\bar{x}$)	0.100	0.277	-0.177	0.023
				$diff \neq 0$
Standard Deviation (SD)	1.094	1.093	0.001	0.494
				$\frac{sd(1)}{sd(2)} > 1$

Notes: The table presents the distribution of net changes in fertility preferences, Δ_i, by treatment status for the balanced sample. Mean rows present average changes in Δ_i, while Standard Deviation (SD) rows report dispersion of Δ_i within each group. P-values in the Mean rows test for equality of means between the treatment and control groups, while p-values in the SD rows test whether dispersion is greater in the control group. A significant difference in SD indicates lower preference stability among control women. $^{*}p < 0.10$, $^{**}p < 0.05$, and $^{***}p < 0.01$.