The Causal Effects of Counseling Quality: Experimental Evidence from Urban Malawi

Marie C. Montás[a*] Sarah Vincent[b] Mahesh Karra[b]

June 19, 2026

Preliminary and Incomplete.
Please do not cite without author permission.
Click here for the latest draft.

Abstract

Across many primary health services in low- and middle-income countries, counseling is often the first point of entry and engagement with the health system; yet, the empirical literature offers little causal evidence on how its quality, as distinct from access, shapes health care utilization and outcomes. We exploit variation in first counselor assignment by means of a randomized controlled trial in urban Malawi, in which 901 postpartum and pregnant women were each randomly allocated to one of six family planning counselors at their first home visit session. We construct a judge-design instrument from each counselor's leave-one-out average of patient-reported *first-session* quality on two dimensions: satisfaction with the counselor and the breadth of topics covered. Higher-quality counseling compresses session dose and expands per-session content for both subgroups. Family planning take-up rises sharply, with particularly large effects on injectable contraception around delivery. Reproductive outcomes diverge in opposite directions by baseline pregnancy status and concentrate among women closest to the peri-partum window: recent postpartum women experience higher birth and pregnancy rates within 24 months, while pregnant women in their third trimester experience lower pregnancy rates. We discuss several readings consistent with this heterogeneity, including a timing-misalignment hypothesis under which post-delivery engagement aligns differently with the fertile horizon across subgroups. Further, we document that the reproductive divergence rests on a single counselor: it does not survive dropping the highest-quality counselor, whereas the family-planning take-up effects remain positive under every leave-one-counselor-out check.

Keywords: family planning, counseling quality, judge design, postpartum contraception, Malawi.

JEL Codes: I12, I15, J13, O15.

[*a]Department of Global Health and Population, Harvard T.H. Chan School of Public Health. [b]Heilbrunn Department of Population and Family Health, Columbia University Mailman School of Public Health. [*]Montás, corresponding author: mariemontas@fas.harvard.edu. Vincent: sv2890@cumc.columbia.edu; Karra: mvk2109@cumc.columbia.edu.

1 Introduction

In low- and middle-income countries (LMICs), counseling is the gateway to care for many primary health services: for most people, it is the first point of entry and engagement with the health system, and the interaction through which programs deliver information, care options, and follow-up (Bruce, 1990; Jain and Hardee, 2018). In the case of family planning (FP), the literature has has had surprisingly little to say about how the *quality* of counseling, distinct from access to the service, shapes contraceptive uptake and reproductive outcomes. Recent experimental evidence from sub-Saharan Africa makes the question urgent. Dupas et al. (2025) document that three years of free modern contraception in Burkina Faso produced no measurable effect on either contraceptive use or fertility, with confidence intervals tight enough to rule out modest changes. Miller et al. (2025) show structurally, in Mozambique, that correcting women's beliefs about pregnancy risk absent contraception would raise contraceptive use considerably, while removing supply-side constraints would not. Bau et al. (2024) report that dispelling the misconception that contraceptives cause infertility in a Zambian sample, combined with a small visit incentive, produces a permanent increase in modern-method take-up at six months, while the incentive alone produces only a temporary one. If access and price are not the binding constraints in much of sub-Saharan Africa, which features of the patient-provider interaction actually move behavior?

A central empirical obstacle to answering this question is that counseling quality is endogenous to woman-level engagement. Women who engage more with counseling, by attending more sessions, asking more questions, or reporting higher satisfaction, likely differ from those who do not. Their fertility preferences, partner dynamics, baseline knowledge, and household bargaining position plausibly correlate with both their engagement and the downstream outcomes counseling is meant to shift. Naive OLS regressions of contraceptive outcomes on quality measures therefore confound the causal effect of counseling content with selection on woman characteristics. The judge-design literature (Doyle, 2007; Chan et al., 2022; Chyn et al., 2025) has built tools for precisely this kind of selection problem when patients can be randomly allocated to providers who differ systematically in their approach to similar cases. The design has been deployed widely in criminal justice, disability adjudication, and clinical diagnosis, but to our knowledge has not been applied to FP counseling quality in a low-resource setting.

We close this gap by exploiting random first counselor assignment in the Malawi Family Planning Study (Karra et al., 2022; Maggio et al., 2024), an RCT conducted in Lilongwe between 2016 and 2018. Within the intervention arm, 901 pregnant and postpartum women were allocated at their first home visit to one of six trained female family-planning counselors. Section 3.1 describes the counselor training and materials, and the allocation procedure that drives our identification. We construct a Counselor Quality Tendency (CQT) instrument that captures the persistent quality each counselor tends to deliver. For each woman, the CQT is the leave-one-out, scheduling-residualized average of first-session quality among the other women assigned to the same counselor. Quality is measured along two patient-reported dimensions with substantial between-counselor variation:

satisfaction with the counselor and the breadth of family-planning topics covered in the session. The first stage is strong ($F = 869$), and the CQT is not predicted by baseline covariates and is orthogonal to endline attrition. Because our setting has only six counselors who plausibly differ in counseling skill rather than purely in preferences, we engage directly with recent guidance on what a judge-design IV identifies under skill variation (Chan et al., 2022; Chyn et al., 2025). We report a leave-one-counselor-out exercise on six headline outcomes in the main text, with the full 19-outcome version in the appendix.

A conceptual literature on family-planning counseling quality develops frameworks for what quality means (Bruce, 1990; Jain and Hardee, 2018; Kruk et al., 2018) and validates patient-reported scales for measuring it (Cavallaro et al., 2020; Holt et al., 2019; Sudhinaraset et al., 2018). A more recent experimental literature manipulates counseling protocols and tests the effect on contraceptive behavior (Wagner et al., 2025; Athey et al., 2023; Dieye et al., 2026). To our knowledge, no prior work has identified the causal effect of within-program variation in family-planning counseling quality. Our paper addresses this gap by using patient satisfaction and the breadth of topics covered, two dimensions that vary substantially across our six counselors, as IV-identified quality measures while holding the counseling protocol fixed. Section 2 develops these connections.

Our results group into two mechanism findings and two outcome findings. On mechanism, higher counseling quality compresses session dose, with women receiving fewer total sessions and a lower share of sessions with their first counselor. Higher-quality counselors also expand per-session content sharply, particularly on pregnancy, birth spacing, partner engagement, and child health, with the largest content gains among pregnant women before delivery. On outcomes, family-planning take-up rises sharply for both subgroups: postpartum women gain +19.4 percentage points on any family-planning service received and +0.71 additional services per woman, while pregnant women gain +15.5 percentage points on any injectable contraceptive service and +10.6 percentage points on post-delivery injectable coverage. These take-up gains are robust to dropping any of the six counselors. Reproductive outcomes, in contrast, diverge sharply by baseline pregnancy status: postpartum women exposed to higher-quality counselors experience +14.4 percentage points more births and +11.8 percentage points more pregnancies within 24 months of the index birth, while pregnant women experience the opposite pattern. We treat this divergence cautiously: it is identified primarily from the contribution of one of the six counselors. We present alternative readings, including a timing-misalignment hypothesis under which the peri-partum engagement window aligns differently with the fertile horizon across subgroups.

The paper makes three contributions. Our methodological contribution is, to our knowledge, the first application of the judge design to family-planning counseling quality in a low-resource setting. We engage explicitly with the skill-versus-preferences interpretation of agent-IV estimates (Chan et al., 2022) and with practitioner guidance on examiner designs (Chyn et al., 2025). Substantively, we anchor patient satisfaction and topic coverage as measurable quality dimensions in a causal-identification framework. For policy, our results suggest that investing in counseling quality could

expand peri-partum injectable coverage among both pregnant and postpartum women. We caution that downstream reproductive effects in this setting are sensitive to one counselor's contribution and should not be generalized without further evidence.

2 Literature Review

This paper contributes to four strands of literature: the measurement of family-planning counseling quality, the applied-econometrics literature on examiner and judge designs, the recent experimental literature on the take-up–continuation distinction in contraceptive use, and the literature on peri-partum family planning in sub-Saharan Africa.

Our first contribution is to the measurement of family-planning counseling quality. A conceptual literature traces to Bruce (1990)'s six-element framework (choice of methods, information given to clients, technical competence, interpersonal relations, follow-up and continuity mechanisms, and appropriate constellation of services), revised in a rights-based direction by Jain and Hardee (2018), and later absorbed into the broader health-systems quality agenda of the Lancet Global Health Commission on High-Quality Health Systems (Kruk et al., 2018). Building on these frameworks, a measurement literature has developed validated patient-reported instruments specific to family-planning counseling: Chang et al. (2019) document that responses to the standard Method Information Index are often inconsistent with the contraceptive method women actually received, and that adjusting for this inconsistency reduces MII scores by up to 50 percent in Pakistan and 30 percent in Uganda; Sudhinaraset et al. (2018), Holt et al. (2019), and Dehlendorf et al. (2018) develop and validate person-centered quality-of-counseling scales in India and Kenya, Mexico, and the United States, respectively. A more recent experimental literature manipulates counseling content directly: Wagner et al. (2025) produce "more comprehensive counselling" through a bias-reduction intervention for FP providers in Tanzania, Burkina Faso, and Pakistan; Athey et al. (2023) use a tablet-based shared-decision-making tool in Cameroon to triple long-acting reversible contraceptive take-up at full price (from 11 to 35 percent), an effect as large as a substantial price subsidy. What this literature does not provide, to our knowledge, is a causally-identified estimate of the effect of *within-program* variation in counseling quality, distinct from the effect of swapping one counseling protocol for another. We close this gap by using random first-counselor assignment as the source of identifying variation, holding the counseling protocol fixed.

Our second contribution is to the applied-econometrics literature on examiner and judge designs. The strategy of using random assignment of cases to decision-makers as an instrument for the decision-maker's typical behavior originated in Kling (2006) on incarceration length and Doyle (2007) on foster-care placement, with the leave-one-out construction popularized by Dahl et al. (2014) and applied to bail-setting and pretrial detention by Dobbie et al. (2018). Frandsen et al. (2023) provide empirical tests of the identifying assumptions distinguishing average from pairwise monotonicity; Chan et al. (2022) show that judge-IV estimates can be misleading when agents differ in diagnostic skill rather than purely in preferences, with skill variation explaining 39 percent of the

variation in radiologist diagnoses in their application. Chyn et al. (2025) and Goldsmith-Pinkham et al. (2025) consolidate the literature into parallel practitioner's guides covering identification, estimation, and inference; the latter argues that heteroskedasticity-robust standard errors are the appropriate choice when leniency assignment is at the individual level. Closest to our setting in spirit is Chan et al. (2022): our six counselors plausibly differ in counseling skill rather than purely in preferences, and our index is designed to capture that skill dimension. Our paper is, to our knowledge, the first application of the judge design to family-planning service-delivery quality in a low-resource setting; we differ from the canonical design in having a small number of decision-makers ($G = 6$) and many cases per counselor, the opposite of the canonical many-judges-few-cases setup, and we report a leave-one-counselor-out exercise on six headline outcomes in the main text, with the full 19-outcome version in the appendix.

Our third contribution is to the literature on the take-up–continuation distinction in contraceptive use. A recent experimental literature finds that access is often not the binding constraint on family-planning behavior: Dupas et al. (2025) randomize a full versus small (10 percent) price subsidy across 14,545 Burkinabè households and find no measurable effect on either contraceptive use or fertility over three years; Miller et al. (2025) estimate a structural model of contraceptive choice in Mozambique and find that correcting women's beliefs about pregnancy risk would raise use considerably while removing supply-side constraints would yield only modest gains (about 1.1 percentage points). On the continuation margin, Ali et al. (2012) pool 60 DHS surveys across 25 countries and document that roughly 38 percent of reversible-method episodes are discontinued within 12 months; Castle and Askew (2015) synthesize evidence on the drivers of discontinuation, including side-effect experience, partner dynamics, and provider follow-up. Closest to our paper is Dieye et al. (2026), who randomize 4,751 Kenyan pregnant women into SMS-based postpartum counseling via the PROMPTS platform and find that the intervention raises knowledge, intentions, and perceived counseling quality but produces a precise null on modern contraceptive use through six months postpartum, a divergence between counseling-quality outcomes and behavioral outcomes that motivates our take-up–continuation framing. We differ in setting (in-person counseling rather than SMS), instrument (a judge-design IV rather than cross-randomization), and unit of variation (counselor identity rather than counseling-protocol assignment), and we extend the literature by separating the take-up margin (extensive contraceptive engagement) from the continuation margin (sustained use and downstream reproductive outcomes) within a single causal-identification framework.

Our fourth contribution is to the literature on peri-partum family planning in sub-Saharan Africa. Postpartum and pregnant women are simultaneously the highest-need and hardest-to-reach segment of the FP user population in low- and middle-income countries (Cleland et al., 2012; Pasha et al., 2015; Tully et al., 2017), with Pasha et al. (2015) documenting postpartum unmet need ranging from 25 to 96 percent across five LMIC sites and Cleland et al. (2012) showing that children born within two years of an elder sibling are roughly 60 percent more likely to die in infancy. According to the National Statistical Office [Malawi] and ICF (2017), the modern

contraceptive prevalence rate among married women in Malawi is 58 percent, with injectables the single most prevalent modern method (used by roughly 30 percent of married women, or about half of all modern-method users). The Malawi Family Planning Study (Karra et al., 2022) was designed to test whether a bundled intervention combining counseling, transport subsidies, and clinic reimbursement could increase contraceptive use and reduce unintended pregnancy among pregnant and immediate-postpartum women in Lilongwe; Maggio et al. (2024) extend the analysis to child outcomes and document that children of intervention-arm women were 0.28 to 0.34 standard deviations taller for their age within one year of exposure and scored 0.17 to 0.20 standard deviations higher on a caregiver-reported measure of cognitive development after two years. Our paper holds the bundled intervention fixed and exploits within-treatment variation generated by random first-counselor assignment to identify the contribution of counseling quality, specifically, to the bundle's overall effects.

3 Methods

We exploit random first-counselor assignment within a bundled family-planning RCT in urban Malawi as an instrument for realized counseling quality, with the goal of identifying the causal effect of family-planning counseling quality on contraceptive take-up and downstream reproductive outcomes. The conceptual framework in Section 3.2.1 predicts that higher-quality counseling shifts behavior through three channels of the perceived-payoff function: information acquisition, belief updating, and partner involvement. A separate timing-misalignment prediction follows from the alignment of the peri-partum counseling window with the post-index fertile horizon. The remainder of this section describes the setting and data sources, the quality and outcome measures, the analytic sample, and the estimation strategy.

3.1 Setting and Data

3.1.1 The Malawi Family Planning Study and Intervention Bundle

We use data from the Malawi Family Planning Study (MFPS), a randomized controlled trial conducted in Lilongwe, Malawi between 2016 and 2018 (Karra et al., 2022; Maggio et al., 2024). The MFPS enrolled 2,140 married women aged 18–35 who were either currently pregnant or had given birth within the previous six months at baseline. The sample was randomized into a treatment arm that received a bundled family-planning intervention over the following two years and a pure control arm that received no intervention.

Women in the treatment arm received four bundled services (Karra et al., 2022): (i) an in-home family-planning counseling component of up to six private visits with a trained female family-planning counselor at or near the woman's home; (ii) free taxi transportation to a designated partner clinic; (iii) free family-planning services at the partner clinic, with full method coverage (pills, injectables, IUDs, implants, and clinical insertion and removal of long-acting methods),

together with financial reimbursement for services received at any other clinic; and (iv) free over-the-phone doctor consultations and reimbursement for the costs of treating any side effects. Our paper holds the bundle fixed and exploits within-treatment variation in the counseling component, using random first-session counselor assignment as the source of identifying variation.

3.1.2 Data Sources

We combine four data sources: per-session counseling records, three waves of household interviews, and two administrative data streams.

Counseling component surveys. For every counseling session attended by a treatment-arm woman, a structured session record was completed at the end of the visit. The record captures session date and duration, who was present (woman alone, husband, other household members), the woman's satisfaction with the counselor on a four-point ordinal scale, the breadth of FP topics discussed from a checklist of nine pre-specified topics (FP methods, side effects, birth spacing, pregnancy, partner engagement, sexually transmitted infections, child health, breastfeeding, and FP services), and whether the counselor answered all of the woman's questions during the session. We use these per-session records to construct both the session-level quality index and the woman-level cumulative quality measure described in Section 3.1.5.

Baseline, midline, and endline household surveys. The full MFPS sample was interviewed at three waves. The baseline survey (Wave 1, conducted September 2016–January 2017) collected demographic and fertility-history information, baseline contraceptive use and knowledge, marriage and partner characteristics, education and work status, and intentions for the coming year. The midline survey (Wave 2, conducted 2017) and the endline survey (Wave 3, conducted 2018) collected updated information on contraceptive use, knowledge, and decision-making; pregnancy and birth histories; husband's awareness of the woman's FP use; and a module on pregnancy and postnatal care that we use to construct downstream reproductive outcomes. Of the 901 treatment-arm women with a first-counselor assignment at baseline, 786 (87.2 percent) were interviewed at midline and 727 (80.7 percent) at endline. Attrition by counselor and its implications for identification are discussed in Section 3.2.7.

Administrative data: FP reimbursement and transport logs. Two administrative data streams track program-side service receipt. The FP reimbursement log records every family-planning service received by a treatment-arm woman through either the designated partner clinic or an external clinic that submitted a reimbursement claim, together with the service date, type (pills, injectables, IUDs, implants, sterilization, side-effect management), and unit cost. We use the reimbursement log to construct the four FP-coverage outcomes (any FP service received, any injectable service received, total number of FP services received, and the fraction of the post-index fertile window covered by injectable contraception). The transport voucher log records every taxi voucher used by a treatment-arm woman, together with date and destination clinic; we use it, together with the reimbursement log, to construct ancillary-component take-up indicators whose relationship to the

CQT instrument we examine in Section 3.2.7 and Appendix Table A9.

3.1.3 Outcome Variables

We distinguish between behavioral outcomes that capture the woman's family-planning and reproductive trajectory after counseling, described here, and mechanism outcomes that capture the counseling process itself, described in Section 3.1.4. The behavioral outcomes fall into two groups, family-planning coverage and reproductive outcomes, all measured at endline (2018).

Family-planning coverage is captured by five outcomes constructed from program administrative records and the endline survey. *Any FP service received* is an indicator that the woman received at least one family-planning service (pills, injectables, IUDs, implants, or sterilization) through the program's reimbursement channel between baseline and endline. *Any injectable service* narrows this margin to the dominant peri-partum method, and the *total number of FP services received* counts services per woman. *Injectable coverage fraction* is the share of the woman's post-index fertile window covered by injectable protection, computed as the number of injectable services received times the standard 90-day protection window, divided by the days from her index birth to endline interview. *Husband knows woman's FP use* is a binary indicator from the endline survey that the woman's partner is aware of her current family-planning use; this outcome is restricted to women who report current contraceptive use at endline.

Reproductive outcomes are captured by four measures constructed from the endline survey's pregnancy and postnatal-care module, all anchored at the woman's index birth date (her own most recent birth at baseline for postpartum women, or the delivery of the baseline pregnancy for pregnant-at-baseline women). *Birth within 24 months* is an indicator for any subsequent live birth within 24 months of the index birth. *Pregnancy within 12 months* and *Pregnancy within 24 months* are analogous indicators for any subsequent pregnancy event over the corresponding horizons, counting every recorded pregnancy regardless of whether it ended in a live birth, miscarriage, abortion, or stillbirth, and measured on the same at-risk population as the birth outcomes. *Mistimed pregnancy* is an indicator equal to one if the woman had a subsequent pregnancy or birth that she reported as occurring earlier than she had wanted, and zero otherwise (including women with no subsequent pregnancy), defined over the full reproductive sample.

3.1.4 Mechanism Variables

The mechanism outcomes capture the counseling process itself and provide the channels through which higher counseling quality is hypothesized to affect behavioral outcomes (Section 3.2.1). All mechanism outcomes are constructed as woman-level summaries across all counseling sessions attended during the two-year follow-up. We group them into counseling continuity and counseling knowledge.

Counseling continuity is captured by four measures of how counseling was delivered. *Total sessions* counts the in-home counseling sessions the woman attended (up to six). *Share of sessions*

with the same counselor is the fraction of attended sessions conducted by the woman's first-assigned counselor and proxies the continuity of the counselor–client relationship. *Counseling sessions per month* normalizes total sessions by the months elapsed between the woman's first and last attended session, capturing visit density. *Husband present at any session* is an indicator that the woman's husband attended at least one counseling session, proxying partner involvement at the visit level.

Counseling knowledge is captured by six binary indicators for whether the woman ever requested each of six pre-specified topics in any counseling session at or after her second session: *FP information* (methods and services), *birth spacing*, *pregnancy*, *partner engagement*, *child health*, and *breastfeeding*. Restricting these indicators to sessions two and onward breaks the mechanical correlation with the first-session topic count that enters our quality index (Section 3.2.3).

3.1.5 Explanatory Variable

Our explanatory variable of interest is counseling quality. We measure quality at the session level along two patient-reported dimensions recorded at the end of every attended session. *Satisfaction with the counselor* is a four-point ordinal scale reporting the woman's overall satisfaction with the session. *Engaged topic count* is the count of the nine pre-specified family-planning topics covered in the session, set to zero if the counselor did not answer all of the woman's questions during the session; this engagement gate ensures that nominal topic coverage in unresponsive sessions does not enter the index as substantive content. The two components have substantial between-counselor variation in our sample ($\eta^2 = 0.60$ for satisfaction and $\eta^2 = 0.90$ for the engaged topic count) and load on a single factor with Cronbach's $\alpha = 0.74$, supporting their joint use as a scalar quality index. Section 3.2.3 develops the formal construction of the session-level quality index, the woman-level cumulative quality measure that serves as the endogenous variable in our IV regressions, and the leave-one-out instrument constructed from it. Appendix Table A5 reports session-level topic coverage by counselor.

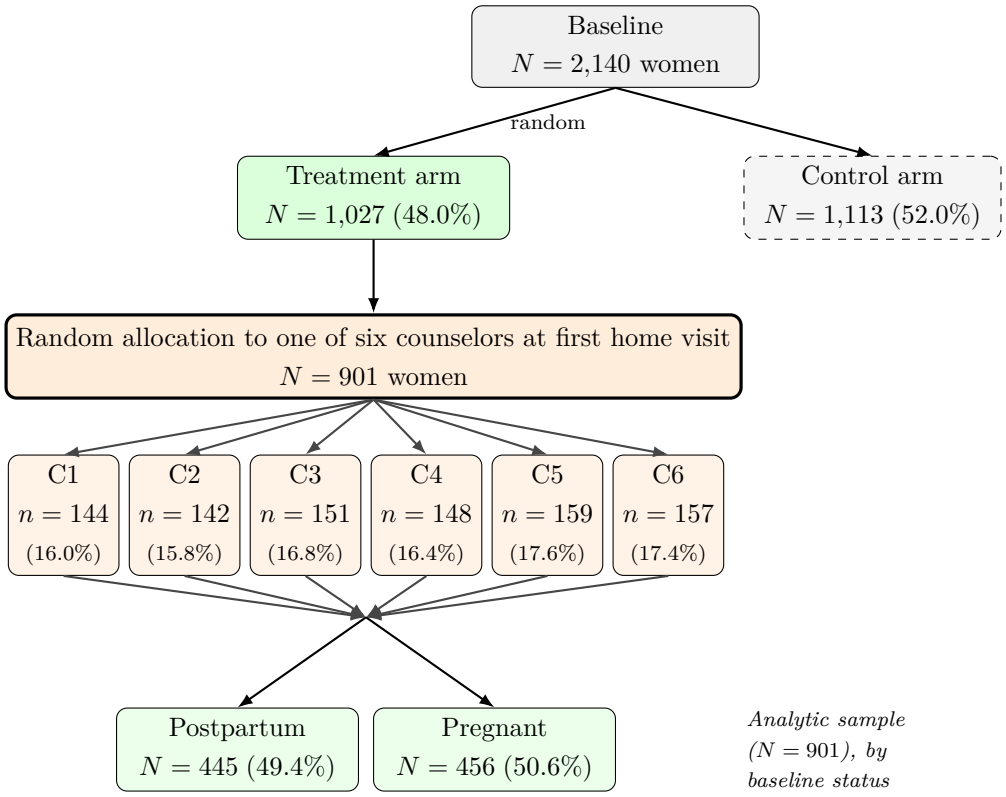
3.1.6 Sample Construction

The analytic sample is constructed from the 1,027 treatment-arm women in our merged dataset through the following steps. Of the 1,027 randomized, 901 (87.7 percent) attended at least one counseling session and received a first-counselor assignment. Of these 901, 786 (87.2 percent) were interviewed at midline and 727 (80.7 percent) at endline; the endline analytic cross-section restricts further to the 724 women with all 12 baseline covariates non-missing. Per-outcome sample sizes within this endline universe range from 591 (the spousal-knowledge outcome, restricted to current contraceptive users) to 724, with the four reproductive outcomes all measured on the same at-risk population ($N = 701$, or 702 for mistimed pregnancy), since we count every subsequent pregnancy regardless of how it ended. Panel A and B outcomes summarize each woman's full course of counseling and do not vary across the panel waves; we therefore estimate them at the woman level, on one observation per woman, so that the unit of analysis matches the woman-level instrument

and endogenous variable. The corresponding estimation samples range from 781 women (husband present at any session) to 896 women, depending on outcome-specific missingness.

Figure 1 summarizes the design diagrammatically. Reading the figure top to bottom, the 2,140 baseline women were randomly assigned with equal probability to the treatment arm or the pure control arm; within the treatment arm, baseline pregnancy status (postpartum or pregnant) was fixed at randomization and forms the basis of the heterogeneity analysis. The 901 treatment-arm women who attended at least one counseling session were then allocated to one of six counselors at their first home visit. This first-counselor allocation, indexed by the per-counselor cell counts, is the random variation our paper uses to instrument for realized counseling quality across all attended sessions; the figure shows the analytic sample at the bottom, split by baseline pregnancy status (445 postpartum, 456 pregnant). Descriptive statistics for the analytic sample, separately by baseline pregnancy status, are reported in Section 4.1.

Figure 1. Study Design and Source of Identifying Variation



Notes: MFPS RCT design as represented in our analytic dataset of $N = 2{,}140$ baseline women. Treatment-versus-control percentages are shares of the 2,140 baseline total; per-counselor percentages and the postpartum/pregnant split shown at the bottom are shares of the 901 women with a first session (our analytic sample).

3.2 Estimation Strategy

Our empirical strategy follows the literature that uses the quasi-random assignment of agents to cases to identify the causal effect of a decision-maker's typical behavior on case outcomes, often referred to as the "judges design" or "examiner design." Papers in this literature exploit variation in a decision-maker's leniency tendency, constructed from her average outcome across other cases, as an instrument for the case-specific decision she renders. The design originated in labor economics with Kling (2006) on sentencing length and Doyle (2007) on child-protection caseworkers, and has since been applied to bail-setting and pretrial detention (Dobbie et al., 2018), disability adjudication (Dahl et al., 2014), and medical diagnosis (Chan et al., 2022); Chyn et al. (2025) and Goldsmith-Pinkham et al. (2025) synthesize identification, estimation, and inference for the design.

We adapt this design to identify the causal effect of family-planning counseling quality. Within the bundled MFPS intervention, the only source of variation in realized counseling quality across treatment-arm women is the identity of the counselor to whom they are first assigned. We exploit this assignment with the Counselor Quality Tendency (CQT) instrument, defined as the leave-one-out, scheduling-residualized average of first-session quality realized by the other women assigned to a given counselor. Because first-counselor assignment is random and counselors differ persistently in the quality they deliver (first-stage Kleibergen-Paap $F = 869$), the two-stage least squares estimator recovers the local average treatment effect of being assigned a counselor whose realized average quality typically sits above the within-treatment median, restricted to the subset of women whose realized quality status is shifted by their first-counselor assignment (Imbens and Angrist, 1994).

3.2.1 Conceptual Framework

Our conceptual framework adapts the discrete-choice model of postpartum family-planning choice in Dieye et al. (2026), which in turn builds on the long tradition of modeling contraceptive choice as a random-utility discrete choice over methods in which behavior is governed by *subjective* rather than objective payoffs (Delavande, 2008; Miller et al., 2025). We adopt this framework, retain its notation, and extend it in two ways specific to our setting: we add a third, partner-involvement channel, and we introduce a timing/exposure-window mechanism that is the basis of our central reproductive-outcome predictions. We model a woman's family-planning decision as a static discrete choice over contraceptive options, indexed by $j \in \{0, 1, \ldots, J\}$, where $j = 0$ denotes non-use and $j \geq 1$ indexes available modern methods (pills, injectables, IUDs, implants). The true payoff to woman i from option j is

$$V_{ij} \;=\; B_{ij}(j) \;-\; C_{\text{now},ij}(j) \;-\; \beta_i\, C_{\text{future},ij}(j) \;+\; \varepsilon_{ij}, \tag{1}$$

where $B_{ij}(j)$ is the current benefit of protection against an unintended pregnancy, $C_{\text{now},ij}(j)$ collects the immediate costs of initiating method j (time, transport, fees, immediate side effects,

relational frictions), $C_{\text{future},ij}(j)$ collects the deferred costs of sustaining j (maintenance, later side effects, removal, switching), $\beta_i \in (0,1]$ is woman i's discount factor on future costs, and ε_{ij} is an idiosyncratic shock realized at the time of choice. Women choose the option that maximizes their *perceived* payoff $\widehat{V}_{ij}$, obtained by replacing B, C_{now}, and C_{future} with their perceived counterparts $\widehat{B}$, $\widehat{C}_{\text{now}}$, $\widehat{C}_{\text{future}}$:

$$\widehat{V}_{ij} \;=\; \widehat{B}_{ij}(j) \;-\; \widehat{C}_{\text{now},ij}(j) \;-\; \beta_i\, \widehat{C}_{\text{future},ij}(j) \;+\; \varepsilon_{ij}, \qquad j_i^* \;=\; \arg\max_{j\in\{0,\ldots,J\}} \widehat{V}_{ij}. \tag{2}$$

The perceived–true wedge $\widehat{V}_{ij} - V_{ij}$ captures behavioral frictions that depress modern-method take-up in low-resource settings: misperceptions about pregnancy risk and method efficacy (Dupas, 2011; Miller et al., 2025), side-effect uncertainty, planning costs, and information gaps about side-effect management and method availability (Wagner et al., 2025).

We model counseling as a scalar quality input q delivered by the woman's counselor that simultaneously shifts three arguments of the perceived-payoff function. Treating q as scalar is supported by the fact that our two patient-reported components, satisfaction and the engaged topic count, load on a single factor with Cronbach's $\alpha = 0.74$. Before these payoff channels operate, quality also shapes the *delivery* of counseling itself, namely the quantity, continuity, and intensity of sessions, which is the vehicle through which information and belief-updating reach the woman and the locus of a quality-quantity trade-off; we map this engagement margin to the counseling-continuity outcomes in Panel A. First, an *information channel*: higher-q counseling delivers more accurate information about method effectiveness, side effects, and how to manage them, moving $\widehat{B}_{ij}(j)$, $\widehat{C}_{\text{now},ij}(j)$, and $\widehat{C}_{\text{future},ij}(j)$ toward their true counterparts (Wagner et al., 2025; Athey et al., 2023). Second, a *belief-updating channel*: higher-q counseling corrects misperceptions about the woman's own pregnancy risk absent contraception. Because $\widehat{B}_{ij}(j)$ is the perceived benefit of protection against an unintended pregnancy, a higher perceived risk of pregnancy without a method raises $\widehat{B}_{ij}(j)$ for every modern method $j \geq 1$, and thereby raises the perceived value of all modern-method options relative to non-use $\widehat{V}_{i0}$ (Dupas, 2011; Miller et al., 2025). This is precisely the margin Miller et al. (2025) identify structurally: correcting beliefs about pregnancy risk absent contraception, rather than easing supply constraints, is what moves modern-method demand. Third, a *partner-involvement channel*: higher-q counseling expands joint discussion of family-planning preferences with the woman's husband, shifting a partner-approval argument that enters the woman's perceived payoff and, in our data, maps directly onto the spousal-knowledge outcome. Partner approval is a first-order choice attribute in this literature: Delavande (2008) finds partner disapproval among the most important determinants of method choice, and Miller et al. (2025) include partner approval as a method attribute; joint spousal decision-making is emphasized by Ashraf et al. (2014). Following Dieye et al. (2026), we hold the discount factor β_i fixed: counseling shifts information and beliefs, not time preferences.

Timing enters the framework as a property of the *exposure window* rather than as a property of q itself. Counseling is delivered during a narrow peri-partum window, but the fertile horizon

over which contraceptive protection matters extends well beyond that window. For a pregnant-at-baseline woman in her third trimester, the engagement window aligns with the start of her post-delivery fertile horizon; for a postpartum-at-baseline woman who delivered six months before baseline, the engagement window opens after her fertile horizon has already begun. Asymmetric coverage of the fertile horizon by the engagement window is the basis of the timing-misalignment hypothesis we develop in Section 5.

The framework yields four testable hypotheses linking each channel of q to a specific empirical panel. H1 (engagement channel): higher-q counseling changes the *dose* of counseling, that is, the number, continuity, and intensity of sessions through which the information and belief-updating channels reach the woman. Because higher quality may substitute for the sheer quantity of contact, this is where we test for a quality-quantity trade-off, measured by the counseling-continuity outcomes in Panel A. H2 (information channel): higher-q counseling broadens the range of family-planning topics a woman discusses in her later sessions, measured by the knowledge outcomes in Panel B. H3 (belief-updating channel): higher-q counseling raises a woman's take-up of modern contraception, measured by the family-planning services she receives in the program's administrative records (Panel C). H4 (partner-involvement channel): higher-q counseling increases a husband's awareness of his wife's contraceptive use by reorganizing the couple's information environment around family planning. We further hypothesize a timing-misalignment prediction (developed in Section 5): the downstream reproductive outcomes in Panel D diverge between postpartum and pregnant women because the peri-partum engagement window covers their post-delivery fertile horizons asymmetrically, not because counseling quality affects the two groups differently in absolute terms.

3.2.2 Counselor Assignment as the Source of Identifying Variation

The counseling component of the bundle was delivered by six trained female family-planning counselors. The counselors were trained by a master trainer from the Malawi Ministry of Health's Reproductive Health Directorate, and counseling materials, including demonstration kits, contraceptive samples, and flip charts in Chichewa, were provided by the Ministry of Health and Population Services International (PSI). Of the 901 treatment-arm women who attended at least one counseling session, each was allocated to one of these six counselors at her first home visit. Allocation was randomly assigned through a block-by-block procedure indexed by the woman's randomly-generated caseid; women, study staff, and counselors had no input into the first-session assignment. We use this random first-session counselor assignment as the source of identifying variation for counselor-quality differences.

3.2.3 Quality Index Construction

Counseling quality is measured along two patient-reported dimensions with substantial between-counselor variation: satisfaction with the counselor (four-point ordinal scale) and the engaged topic

count (count of nine pre-specified family-planning topics, gated on whether the counselor answered all of the woman's questions). The two components map onto the two domains Sudhinaraset et al. (2018) identify in their validated patient-centered family-planning scale: (i) autonomy, respect, and communication; and (ii) health-facility environment. The session-level engagement gate parallels the gating of patient-centered counseling sub-indices in Wagner et al. (2025): a session in which the counselor does not engage with the woman's questions does not deliver substantive content even if the checklist was nominally covered. The choice of two components rather than a broader battery responds to Afulani et al.'s (2023) call for standardized patient-reported measurement of person-centered sexual and reproductive health. Following the standardized-index methodology of Kling et al. (2007) and Anderson (2008), we z-score each component on the first-session treatment-sample distribution and average the two standardized components with equal weights:

$$Q_{is} = \tfrac{1}{2}\left(\tilde{S}_{is} + \tilde{T}^{\text{eng}}_{is}\right), \quad \tilde{S}_{is} = \frac{S_{is} - \bar{S}^{(1)}}{\sigma^{(1)}_S}, \quad \tilde{T}^{\text{eng}}_{is} = \frac{T_{is} \cdot \mathbf{1}\{A_{is} = 1\} - \bar{T}^{\text{eng},(1)}}{\sigma^{\text{eng},(1)}_T}, \tag{3}$$

where S_{is} is the four-point satisfaction at session s for woman i, T_{is} is the topic count at that session, A_{is} is an indicator that all the woman's questions were answered at that session, and $\bar{S}^{(1)}$, $\sigma^{(1)}_S$, $\bar{T}^{\text{eng},(1)}$, and $\sigma^{\text{eng},(1)}_T$ are the first-session sample mean and standard deviation of each component. The two z-scored components are winsorized at the 1st and 99th percentiles before averaging. The engagement gate $\mathbf{1}\{A_{is} = 1\}$ sets the topic contribution to zero if the counselor did not answer all questions at session s, parallel to the construction of patient-centered counseling sub-indices in Wagner et al. (2025).

Evidence has found patient satisfaction alone a weak proxy for clinical quality: across eight low-income countries, Kruk et al. (2018) find that 85 percent of family-planning clients report being very satisfied with care in consultations in which providers completed fewer than half of the essential clinical actions. Pairing satisfaction with the breadth of topics actually covered guards against this ceiling, since the topic-coverage component captures substantive content that high reported satisfaction can mask. Two design choices warrant brief discussion. First, although patient-reported counseling quality has traditionally been summarized by the binary Method Information Index (Chang et al., 2019), the all-or-nothing aggregation produces a measure that is inconsistent with the method women actually received in roughly half of cases in low- and middle-income settings; Chang et al. recommend enriching, rather than discretizing, the underlying measure. Our topic-coverage component retains the underlying count, and our satisfaction component preserves the four-point ordinal scale. Second, we exclude two additional candidate components, session duration and a binary "answered all questions" indicator entering the index in its own right, because both have essentially no between-counselor variation in our sample ($\eta^2 = 0.004$ and 0.033, respectively, compared with $\eta^2 = 0.60$ for satisfaction and $\eta^2 = 0.90$ for the engaged topic count). A variable that does not differ systematically across counselors cannot identify a counselor-quality contrast in a judge-design IV even in principle. Appendix Table A1 reports the formal component diagnostics, including the between-counselor analysis of variance for each of the four candidate components,

the pairwise correlations between them, and the standardized Cronbach's α for the primary two-component index ($\alpha = 0.74$) and the four-component alternative ($\alpha = 0.33$); Figures B1 and B2 display the by-counselor distributions and pairwise correlations. As a sensitivity check, Appendix Tables A2 and A3 re-estimate every main coefficient using a four-component index that adds session duration and the questions-answered indicator: estimates are within 0.01 to 0.04 of every main coefficient, with identical sign and significance patterns.

3.2.4 Endogenous Variable and Instrument

The endogenous variable in the IV regression is the woman-level average of her per-session quality realizations, dichotomized at the within-treatment median:

$$\mathrm{HQ}_i = \mathbf{1}\Big\{\bar{Q}_i > \mathrm{median}(\bar{Q}) \,\Big|\, \text{treatment} = 1\Big\}, \quad \bar{Q}_i = \frac{1}{n_i}\sum_{s=1}^{n_i} Q_{is}, \tag{4}$$

where n_i is the number of counseling sessions woman i attended (up to six over the two-year horizon). $\bar{Q}_i$ integrates over the woman's entire counseling experience and is the natural object of interest for a paper about counseling quality.

Let $c(i)$ denote the counselor woman i is first assigned to, $Q_i^{(1)}$ her first-session quality realization, and $\tilde{Q}_i^{(1)}$ the residual from a regression of $Q_i^{(1)}$ on first-session scheduling fixed effects (day of week, calendar month, and hour of day). The Counselor Quality Tendency (CQT) instrument is

$$\mathrm{CQT}_i \;=\; \frac{1}{N_{c(i)} - 1} \sum_{i' \neq i:\, c(i') = c(i)} \tilde{Q}_{i'}^{(1)}, \tag{5}$$

where $N_{c(i)}$ is the number of women whose first counselor is $c(i)$. The scheduling residualization removes the portion of counselor identity attributable to when sessions happen, and the leave-one-out construction removes woman i's own realization from her counselor's quality measure. This is the canonical examiner-design construction (Kling, 2006; Dahl et al., 2014; Chyn et al., 2025; Goldsmith-Pinkham et al., 2025). By design, the instrument is built from first-session quality only, while the endogenous variable averages quality across all attended sessions: this separation between an exogenous shifter (first-session, set by random counselor assignment) and the woman's realized experience (cumulative across sessions, the policy-relevant quantity) is standard in the leniency-design literature (Goldsmith-Pinkham et al., 2025).

The IV system consists of a first stage and a second stage, with a heterogeneity specification that interacts both with the pregnant-at-baseline indicator.

First stage. The first stage projects the endogenous variable HQ_i on the instrument and a vector X_i of baseline covariates,

$$\mathrm{HQ}_i \;=\; \pi_0 \;+\; \pi_1\,\mathrm{CQT}_i \;+\; X_i'\pi_2 \;+\; v_i. \tag{6}$$

The vector X_i includes the woman's age group (five-year bins), sex of her youngest child, total number of alive children, an indicator for ever-use of contraception, an indicator for completed primary education, work status, religion, *neighborhood fixed effects* (indicators for the seven Lilongwe catchment areas of residence), ethnicity, an indicator for pregnant at baseline, an indicator for current contraceptive use at baseline, and the counseling-location indicator, all measured at baseline. All first-stage, second-stage, and reduced-form regressions reported in the paper include this full X_i vector, and hence neighborhood fixed effects, unless otherwise noted. The strength of the instrument is summarized by the Kleibergen-Paap rk Wald F-statistic on π_1.

Second stage. The second stage estimates the local average treatment effect of higher counseling quality on outcome Y_i by replacing HQ_i with its predicted value from (6),

$$Y_i \;=\; \alpha \;+\; \beta\,\widehat{\mathrm{HQ}}_i \;+\; X_i'\gamma \;+\; \varepsilon_i. \tag{7}$$

The coefficient β is the local average treatment effect of higher-quality counseling among complier women, that is, women whose above-median realized quality status is shifted by their first-counselor assignment (Imbens and Angrist, 1994).

Heterogeneity by baseline pregnancy status. Because the conceptual framework predicts that effects of counseling quality may differ between postpartum-at-baseline and pregnant-at-baseline women, most sharply via the timing-misalignment mechanism for downstream reproductive outcomes, we estimate an interacted two-stage least squares specification alongside the pooled IV system. We treat *both* the endogenous quality indicator HQ_i and its interaction with the pregnant-at-baseline indicator, $\mathrm{HQ}_i \times P_i$, as endogenous, and instrument them with the corresponding pair of excluded instruments CQT_i and $\mathrm{CQT}_i \times P_i$; the order condition is exactly met, with two instruments for two endogenous regressors. Each endogenous regressor is projected on the full instrument set in its own first-stage equation,

$$\mathrm{HQ}_i = \pi_0 + \pi_1 \mathrm{CQT}_i + \pi_2(\mathrm{CQT}_i \times P_i) + X_i'\pi_3 + v_{1i}, \tag{8}$$

$$\mathrm{HQ}_i \times P_i = \rho_0 + \rho_1 \mathrm{CQT}_i + \rho_2(\mathrm{CQT}_i \times P_i) + X_i'\rho_3 + v_{2i}, \tag{9}$$

and the second stage uses the fitted values from *both* first stages,

$$Y_i = \alpha + \delta P_i + \beta_{\mathrm{PP}}\widehat{\mathrm{HQ}}_i + \beta_{\Delta}(\widehat{\mathrm{HQ} \times P})_i + X_i'\gamma + \varepsilon_i, \tag{10}$$

where $(\widehat{\mathrm{HQ} \times P})_i$ is the fitted value of the interaction from its own first stage (9) (not the product $\widehat{\mathrm{HQ}}_i \times P_i$), and $P_i \in X_i$ enters as an included exogenous control. Here β_{PP} is the postpartum-subgroup LATE (the reference category, recovered directly as $\hat{\beta}_{\mathrm{PP}}$) and β_{Δ} is the interaction coefficient that captures the cross-subgroup *difference* in LATEs. The pregnant-subgroup LATE is recovered as the linear combination $\hat{\beta}_{\mathrm{PP}} + \hat{\beta}_{\Delta}$. Both subgroup estimates throughout the paper are recovered from this single pooled interacted system, as linear combinations of the estimated coefficients, not from split-sample regressions; the subgroup-specific first-stage F-statistics reported

in Table 1 re-estimate the simple first stage within each subsample and are reported only as evidence of instrument strength within each subgroup. The test $H_0 : \beta_\Delta = 0$ tests equality of the postpartum and pregnant subgroup LATEs and is reported alongside both subgroup estimates as $p_{\text{PP=Preg}}$. Every results table in this paper reports both the pooled and subgroup estimates side by side; we pre-commit to this layout to avoid post-hoc subgroup selection.

Standard errors throughout are heteroskedasticity-robust, which Goldsmith-Pinkham et al. (2025) argue is the appropriate choice when leniency assignment is at the individual level rather than at a cluster level. The LATE we recover is a convex combination of woman-level treatment effects across complier types, in the sense of Imbens and Angrist (1994) and Frandsen et al. (2023).

3.2.5 First-Stage Results

Table 1 reports estimates of (6) on the maximum first-stage analytic sample ($N = 727$ in the unadjusted specification and $N = 724$ in the covariate-adjusted specification, with 370 postpartum and 354 pregnant women in the two subgroup columns). The first stage is strong: the Kleibergen-Paap rk Wald F-statistic is 921 in the unadjusted pooled specification, 869 in the covariate-adjusted pooled specification, 359 within the postpartum subgroup, and 457 within the pregnant subgroup, all well above the Stock-Yogo 10 percent maximal-IV-size critical value of 16.38. The CQT coefficient is 0.346 in the covariate-adjusted pooled specification, 0.338 among postpartum women, and 0.365 among pregnant women, with no statistically detectable difference across the unadjusted, covariate-adjusted, and subgroup-stratified specifications.

Table 1. First-Stage Estimates: CQT Instrument on High-Quality Counseling

	(1) Unadj. pooled	(2) Adj. pooled	(3) Adj. Postpartum	(4) Adj. Pregnant
CQT (sat+topics)	0.342***	0.346***	0.338***	0.365***
Robust SE	(0.011)	(0.012)	(0.018)	(0.017)
Kleibergen-Paap rk LM χ^2	200.88	201.59	94.62	108.44
Kleibergen-Paap rk Wald F	921.41	869.47	359.06	456.89
Cragg-Donald F	286.25	290.54	132.80	158.44
R^2	0.283	0.308	0.306	0.342
N	727	724	370	354

Notes: First-stage OLS regression of the binary endogenous variable (an above-median indicator for the woman's average satisfaction-and-topics quality across all attended sessions) on the leave-one-out, scheduling-residualized Counselor Quality Tendency (CQT) of the woman's first counselor. Column (1) reports the unadjusted pooled regression; columns (2)–(4) include baseline covariates. Columns (3) and (4) restrict to postpartum and pregnant women respectively. Heteroskedasticity-robust standard errors in parentheses. The Stock-Yogo (2005) 10

3.2.6 Balance

Figure 2 shows the between-counselor variation in the CQT instrument and in realized average counseling quality that the design exploits. Panel A plots each counselor's CQT, the leave-one-

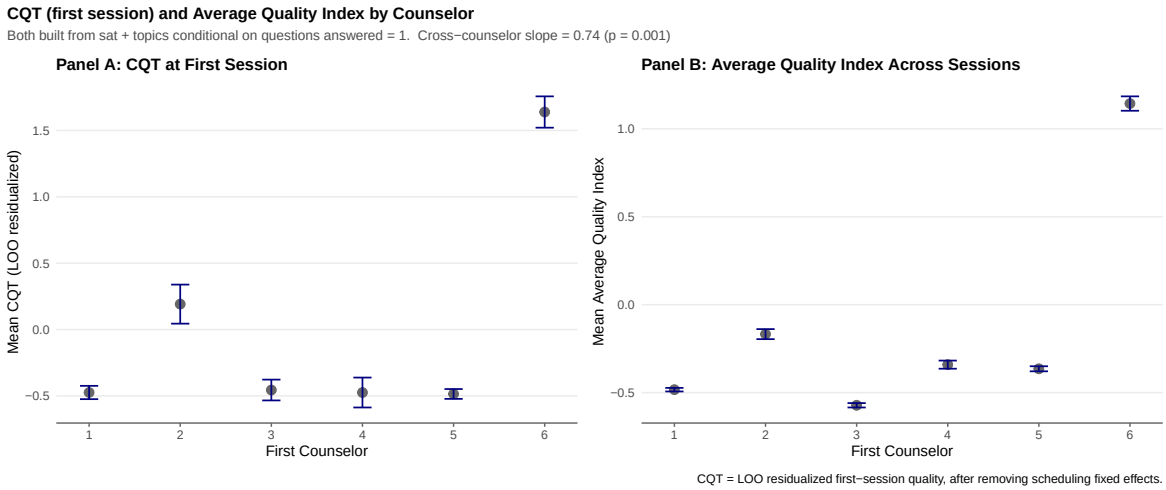
out scheduling-residualized mean of first-session quality among the other women assigned to her. Panel B plots each counselor's mean woman-level average quality index across all attended sessions. Counselors who deliver higher cumulative quality on average (Panel B) are also those whose other-woman first-session residualized mean is higher (Panel A); the cross-counselor slope reported in the subtitle captures the first stage at the counselor level. Table 2 reports the corresponding formal balance test: each baseline covariate's mean by first-assigned counselor, together with a one-way ANOVA F-test of equality of means across the six counselors. No covariate's F-test is significant at the 5 percent level. As a more direct test of instrument exogeneity, we additionally regress the CQT and the cumulative quality measure on the full baseline control set used in the IV specification and fail to reject the joint null that all covariate coefficients are zero (CQT: $F(19, 704) = 0.79$, $p = 0.72$; cumulative quality: $F(19, 704) = 0.60$, $p = 0.91$, both $N = 724$, heteroskedasticity-robust). Endogeneity in this setting therefore runs primarily through unobserved channels (fertility preferences, partner dynamics, baseline knowledge) that an observable-covariate balance test cannot detect, but that the IV is constructed to address.

Table 2. Balance Table: Baseline Covariates by First Assigned Counselor

	Counselor							ANOVA	
Covariate	1	2	3	4	5	6	Total	F	p
Age ≤20	0.24 (0.43)	0.20 (0.41)	0.18 (0.38)	0.22 (0.41)	0.17 (0.38)	0.24 (0.43)	0.21 (0.41)	0.74	0.60
Age 21–25	0.40 (0.49)	0.40 (0.49)	0.44 (0.50)	0.37 (0.49)	0.44 (0.50)	0.42 (0.50)	0.41 (0.49)	0.48	0.79
Age 26–30	0.24 (0.43)	0.24 (0.43)	0.19 (0.40)	0.28 (0.45)	0.23 (0.42)	0.19 (0.39)	0.23 (0.42)	1.02	0.40
Age 31–35	0.13 (0.34)	0.15 (0.36)	0.19 (0.39)	0.13 (0.34)	0.16 (0.37)	0.15 (0.36)	0.15 (0.36)	0.51	0.77
Age at first sex	18.83 (2.79)	18.73 (2.71)	19.35 (3.06)	18.70 (2.71)	19.07 (2.95)	19.26 (2.70)	18.99 (2.83)	1.43	0.21
Children alive	2.15 (1.19)	2.36 (1.30)	2.33 (1.32)	2.39 (1.25)	2.41 (1.29)	2.08 (1.17)	2.29 (1.26)	1.79	0.11
FP ever use	0.72 (0.45)	0.79 (0.41)	0.80 (0.40)	0.76 (0.43)	0.84 (0.37)	0.74 (0.44)	0.78 (0.42)	1.75	0.12
Education: primary+	0.59 (0.49)	0.61 (0.49)	0.58 (0.50)	0.60 (0.49)	0.52 (0.50)	0.54 (0.50)	0.57 (0.50)	0.72	0.61
Working	0.10 (0.30)	0.06 (0.25)	0.14 (0.35)	0.08 (0.27)	0.09 (0.28)	0.13 (0.33)	0.10 (0.30)	1.37	0.23
Religion (Christian)	0.84 (0.37)	0.84 (0.37)	0.88 (0.33)	0.82 (0.39)	0.79 (0.41)	0.84 (0.37)	0.83 (0.37)	1.08	0.37
Neighborhood area indicators[a]									
Area 7	0.09 (0.29)	0.07 (0.26)	0.09 (0.29)	0.09 (0.28)	0.09 (0.29)	0.09 (0.29)	0.09 (0.28)	0.14	0.98
Area 8	0.08 (0.28)	0.09 (0.29)	0.07 (0.25)	0.07 (0.26)	0.07 (0.26)	0.11 (0.31)	0.08 (0.28)	0.51	0.77
Area 23	0.02 (0.14)	0.01 (0.12)	0.01 (0.12)	0.03 (0.18)	0.01 (0.11)	0.03 (0.16)	0.02 (0.14)	0.54	0.74
Area 24	0.01 (0.12)	0.01 (0.12)	0.02 (0.14)	0.03 (0.16)	0.03 (0.18)	0.01 (0.08)	0.02 (0.14)	0.72	0.61
Area 50	0.35 (0.48)	0.44 (0.50)	0.37 (0.49)	0.35 (0.48)	0.39 (0.49)	0.35 (0.48)	0.37 (0.48)	0.74	0.59
Area 51	0.14 (0.35)	0.11 (0.32)	0.15 (0.35)	0.16 (0.37)	0.13 (0.33)	0.13 (0.33)	0.14 (0.34)	0.38	0.86
Area 56	0.31 (0.46)	0.26 (0.44)	0.29 (0.46)	0.26 (0.44)	0.28 (0.45)	0.29 (0.46)	0.28 (0.45)	0.23	0.95
Ethnicity (Chewa)	0.40 (0.49)	0.46 (0.50)	0.40 (0.49)	0.41 (0.49)	0.41 (0.49)	0.40 (0.49)	0.41 (0.49)	0.33	0.89
Pregnant at BL	0.51 (0.50)	0.49 (0.50)	0.52 (0.50)	0.54 (0.50)	0.46 (0.50)	0.53 (0.50)	0.51 (0.50)	0.54	0.74
Current FP use	0.19 (0.39)	0.24 (0.43)	0.24 (0.43)	0.19 (0.39)	0.30 (0.46)	0.27 (0.45)	0.24 (0.43)	1.73	0.13
Counseled last preg	0.04 (0.18)	0.04 (0.20)	0.03 (0.16)	0.06 (0.24)	0.07 (0.26)	0.04 (0.19)	0.05 (0.21)	0.93	0.46
N	144	142	151	148	159	157	901		

Notes: Each row reports a baseline covariate that enters the control set X_i in the IV regressions. Row entries are counselor-specific means with standard deviations in parentheses; the Total column reports the full-sample mean (SD); the ANOVA F-statistic and p-value test the null that means are equal across all six counselors. None of the F-tests are significant at the 5% level, supporting random counselor assignment. [a]Neighborhood area is a categorical variable; the seven rows are dummy indicators for the seven distinct area codes observed in the sample. The Pearson chi-square test of joint independence between area and counselor yields $\chi^2(30) = 13.97$, $p = 0.994$. As a complementary test of instrument exogeneity, heteroskedasticity-robust regressions of the CQT instrument and of the cumulative quality measure on the full covariate set fail to reject the joint null that all coefficients are zero (CQT: $F(19, 704) = 0.79$, $p = 0.72$; cumulative quality: $F(19, 704) = 0.60$, $p = 0.91$), estimated on the $N = 724$ endline analytic sample.

Figure 2. Counselor Quality Tendency and Realized Quality by Counselor



Notes: Panel A plots each counselor's CQT instrument: the leave-one-out, scheduling-residualized mean of first-session quality among the other women assigned to her. Panel B plots each counselor's mean woman-level average quality index across all attended sessions. Both objects are built from the satisfaction-plus-engaged-topics index in which the topic count enters only when the counselor answered all of the woman's questions at the session. Points are counselor-level means. In Panel A, the vertical bars are 95 percent confidence intervals for each counselor's mean residualized first-session quality, computed from the underlying first-session residualized quality scores (standard error equal to the within-counselor standard deviation of those scores divided by $\sqrt{N_c}$, with a t-distribution on $N_c - 1$ degrees of freedom); they are deliberately *not* computed from the dispersion of the leave-one-out CQT values, whose within-counselor variance is mechanically a factor $(N_c - 1)^{-2}$ smaller. In Panel B, the bars are the analogous intervals for each counselor's mean woman-level average quality, which is a genuine per-woman quantity. Counselors are indexed 1 through 6.

3.2.7 Threats to Identification

There are four key threats to our identification strategy. First, our IV requires that first-counselor assignment be *random* with respect to woman-level potential outcomes. As described in Section 3.2.2, allocation was randomly assigned through a block-by-block procedure indexed by the woman's randomly-generated caseid, with no input from women, study staff, or counselors. As an additional safeguard, we residualize first-session quality on day-of-week, calendar-month, and hour-of-day scheduling fixed effects before constructing the CQT instrument, and the balance test in Section 3.2.6 confirms the CQT is not predicted by observable baseline covariates. The exclusion restriction further requires that counselor identity affect outcomes only through the realized counseling quality channel we measure. Because the non-counseling components of the bundle (transport vouchers, free or reimbursed clinic-based services, side-effect coverage) are *offered* identically to all treatment-arm women, the most plausible alternative channel is the relationship the counselor builds with the woman during home visits, which is what the quality index is constructed to capture. Take-up of these components is realized after counselor assignment and is itself plausibly an outcome of counseling quality, so conditioning on it in the main specification would amount

to controlling for a post-treatment variable that lies on the causal pathway of interest (Angrist and Pischke, 2009). Two features support reading the relationship between the CQT and take-up as part of the counseling channel rather than as an exclusion-restriction violation. The transport voucher is the only ancillary component that does not enter the construction of any outcome, so it provides the cleanest test: the CQT is only marginally associated with transport take-up, raising it by about 9 percentage points across the range of counselor quality ($p = 0.06$; Appendix Table A9). Clinic-reimbursement take-up, by contrast, is drawn from the same administrative records used to construct our Panel C FP-coverage outcomes, so its significant association with the CQT ($p = 0.04$) restates the headline take-up result rather than indicating an independent channel. The positive sign throughout is what the counseling-quality mechanism predicts, since higher-quality counselors lead more women to act on the free transport and clinic services they are all offered. We therefore do not condition on take-up in our main specification; nonetheless, Appendix Tables A10 and A11 confirm that re-estimating with an additional control for ancillary-component take-up leaves the counseling-continuity, knowledge, and reproductive coefficients within 0.06 of the main estimates.

Second, monotonicity. The local average treatment effect interpretation of our IV requires that a counselor with higher CQT produces weakly higher quality for every woman she sees (Frandsen et al., 2023). Chan et al. (2022) show that judge-IV estimates can be misleading when agents differ in skill rather than purely in preferences, because skill variation can violate the monotonicity condition. Our setting is closer to the skill-variation case: the six counselors share a common counseling protocol and we expect them to differ primarily in counseling skill, which is the dimension our index is intended to capture. We therefore interpret the LATE as the effect of being assigned a counselor whose realized counseling quality, integrated over the woman's full course of counseling sessions, typically sits above the within-treatment median, rather than as the effect of a uniform shift in a one-dimensional preference parameter. The first-stage estimates do not detect a statistically significant difference in the CQT slope across the postpartum and pregnant subgroups (Section 3.2.5), consistent with the average-monotonicity condition holding within each subgroup of interest.

Third, differential attrition by counselor. The IV identification breaks down if women assigned to higher-CQT counselors are systematically more (or less) likely to remain in the analytic sample at endline. The CQT instrument does not predict endline attrition after controlling for baseline covariates ($p = 0.764$). Individual counselors do differ in endline retention (Table 3; $F(5, 895) = 2.45$, $p = 0.032$, with counselors 1 through 3 retaining approximately 85 percent of their assigned women through endline and counselors 4 through 6 retaining 75 to 78 percent), but baseline-to-midline differential attrition is not statistically detectable ($F(5, 895) = 0.70$, $p = 0.622$), and the raw counselor-level variation washes out in the leave-one-out, scheduling-residualized construction of the CQT, which is itself orthogonal to attrition. Covariate balance is also maintained in the endline analytic sample of $N = 724$ women: no individual baseline covariate differs significantly across counselors (minimum per-covariate $p = 0.14$), and the Fisher combined test of all 12 covariates yields $\chi^2(24) = 19.09$, $p = 0.747$.

Fourth, small G. Our setting has only six counselors, which is at the small end of the examiner-

design literature. With six decision-makers, one counselor's residualized quality tendency contributes disproportionately to the identifying contrast, raising the concern that a single counselor's behavior could drive any main result. We address this concern directly with a complete leave-one-counselor-out exercise reported in Section 4.3: for each main outcome, we re-estimate the IV after sequentially dropping each of the six counselors. The exercise establishes that the FP-coverage findings in Panel C are stable across every leave-one-out specification, while the reproductive heterogeneity in Panel D does not survive excluding counselor 6: the postpartum estimates collapse toward zero or reverse sign, with the mistimed-pregnancy estimate turning negative and significant. Because conventional cluster-robust standard errors over-reject when the number of clusters is small (Cameron et al., 2008), our inference rests primarily on woman-level robust standard errors and the leave-one-counselor-out exercise; we additionally report a per-panel summary index (Appendix Table A12) and wild-cluster bootstrap p-values clustering on the six counselors (Section 4.2; Appendix Table A13) (MacKinnon and Webb, 2017) as complements. This inferential strategy is consistent with the recommendation that examiner designs demonstrate robustness to the set of examiners included (Chyn et al., 2025). We treat this contrast as a structural feature of examiner designs with few decision-makers and reflect it in our discussion of policy implications.

Table 3. Attrition from Baseline to Midline and Endline by First Assigned Counselor

	Counselor						
	1	2	3	4	5	6	Total
N baseline (2016)	144	142	151	148	159	157	901
N midline (2017)	127	125	136	131	134	133	786
N endline (2018)	123	121	129	113	119	122	727
Attrition rate BL→ML	0.118	0.120	0.099	0.115	0.157	0.153	0.128
Attrition rate BL→EL	0.146	0.148	0.146	0.236	0.252	0.223	0.193

Notes: Sample restricted to treatment-arm women with a first-counselor assignment ($N = 901$). Attrition is defined as present at baseline but absent at the indicated wave. **Baseline to midline:** no statistically significant differential attrition by counselor, $F(5, 895) = 0.70$, $p = 0.622$. **Baseline to endline:** significant differential attrition, $F(5, 895) = 2.45$, $p = 0.032$, driven by counselors 4, 5, and 6, who have endline attrition rates of 22 to 25 percent compared with approximately 15 percent for counselors 1–3. The CQT instrument does not predict endline attrition after controlling for baseline covariates ($p = 0.764$), and covariate balance is maintained in the endline analytic sample ($N = 724$): no individual baseline covariate differs significantly across counselors (minimum per-covariate $p = 0.14$), and the Fisher combined test of all 12 covariates yields $\chi^2(24) = 19.09$, $p = 0.747$. All main IV regressions include the same baseline covariates as controls. Numbers are computed under the current satisfaction-plus-engaged-topics specification.

4 Results

This section reports descriptive statistics for the analytic sample (Section 4.1), the IV estimates of higher counseling quality on the four panels of outcomes (Section 4.2), and a robustness exercise that examines the contribution of each counselor to the identifying variation (Section 4.3). We treat the mechanism channels (counseling continuity and counseling knowledge) as describing how the treatment operates, and the behavioral outcomes (FP coverage and reproductive outcomes) as

describing its consequences for the woman herself.

4.1 Descriptive Findings

Table 4 reports descriptive statistics for the analytic sample of $N = 901$ treatment-arm women with a first-counselor assignment, separately for the 456 pregnant-at-baseline women and the 445 postpartum-at-baseline women. The two subgroups are similar on the headline pre-treatment covariates entering X_i: their distributions of age, education, work status, religion, ethnicity, and number of alive children differ by no more than a few percentage points, and the formal balance test in Section 3.2.6 cannot reject that baseline covariates are jointly unrelated to the CQT instrument. The two systematic baseline differences both reflect the construction of the subgroups themselves: postpartum women are mechanically more likely to have ever used contraception (85 percent vs. 70 percent), and current FP use at baseline is structurally zero among pregnant women.

The analytic sample is demographically young and modestly educated. Forty-one percent of women are between 21 and 25 years old at baseline, 23 percent are aged 26 to 30, and the mean age at first intercourse is 19 years. Just over half (57 percent) have completed primary education; 10 percent report any current paid work; 83 percent are Christian; and 41 percent identify as ethnically Chewa, broadly tracking the demographic profile of urban Lilongwe (Karra et al., 2022). Average parity at baseline is 2.3 alive children. Only 5 percent of women report having received any FP counseling during a prior pregnancy, so the program is for most women their first structured family-planning interaction in a clinical setting. The counseling-intensity profile is moderate by program standards: women attend 4.6 sessions on average over the two-year follow-up window, about one session every five months across the full window. Counseling is concentrated rather than evenly spaced, however: over each woman's active counseling window, from her first to her last attended session, the visit density averages 0.38 sessions per month, or roughly one session every two to three months. On average, 87 percent of attended sessions are delivered by the woman's originally assigned counselor. Partner engagement is markedly low at the household level: only 13 percent of women have a husband present at any of their counseling sessions, foreshadowing the spousal-knowledge results we report in Section 4.2.

The mechanism outcomes summarized in Panels B and C of Table 4 are similar across the two subgroups. Two topic outcomes display visible PP/Preg gaps that we will revisit in the IV regressions below: *pregnancy* is requested by 26 percent of pregnant women versus 15 percent of postpartum women, and *child health* by 65 percent of pregnant women versus 73 percent of postpartum women, both consistent with each subgroup placing more weight on the topic most immediately relevant to her own reproductive stage. The behavioral outcomes summarized in Panels D and E are also broadly similar in their raw means: about 23 percent of women in each subgroup receive an administrative FP service through the program, and reproductive-event rates are low (2–7 percent across the four endline outcomes), with postpartum women slightly more likely to experience a subsequent pregnancy than pregnant-at-baseline women. The low base rates on the

reproductive outcomes limit statistical power for any individual outcome considered in isolation, motivating the four-outcome panel structure that follows. The IV estimates document substantial within-subgroup contrasts that the raw means alone do not reveal.

Table 4. Descriptive Statistics for the Analytic Sample

	Total		Pregnant		Postpartum	
Characteristic	Mean	SD	Mean	SD	Mean	SD
Panel A: Baseline covariates						
Pregnant at baseline[d]	0.506	0.500	1.000	0.000	0.000	0.000
Age ≤ 20	0.206	0.405	0.235	0.424	0.178	0.383
Age 21–25	0.413	0.493	0.421	0.494	0.404	0.491
Age 26–30	0.228	0.419	0.197	0.398	0.258	0.438
Age 31–35	0.153	0.360	0.147	0.354	0.160	0.367
Age at first sex	18.99	2.83	18.98	2.85	19.01	2.81
Children alive	2.29	1.26	2.25	1.24	2.33	1.27
Ever used contraception	0.776	0.417	0.702	0.458	0.852	0.356
Completed primary education	0.572	0.495	0.568	0.496	0.575	0.495
Working	0.100	0.300	0.118	0.323	0.081	0.273
Religion (Christian)	0.834	0.373	0.814	0.390	0.854	0.354
Ethnicity (Chewa)	0.410	0.492	0.393	0.489	0.428	0.495
Current FP use at baseline[a]	0.240	0.427	0.000	0.000	0.485	0.500
Counseled at prior pregnancy	0.046	0.209	0.048	0.215	0.043	0.203
Panel B: Counseling continuity outcomes						
Share same counselor across sessions	0.869	0.179	0.875	0.177	0.862	0.182
Total sessions	4.64	1.69	4.54	1.74	4.74	1.64
Counseling sessions per month[b]	0.383	0.149	0.385	0.166	0.381	0.131
Husband present at any session	0.134	0.341	0.139	0.347	0.129	0.335
Panel C: Counseling knowledge outcomes						
Ever requested FP information	0.872	0.334	0.860	0.348	0.885	0.319
Ever requested birth spacing	0.483	0.500	0.485	0.500	0.481	0.500
Ever requested pregnancy info	0.205	0.404	0.261	0.440	0.148	0.356
Ever requested partner engagement	0.330	0.470	0.331	0.471	0.328	0.470
Ever requested child health	0.688	0.464	0.649	0.478	0.728	0.445
Ever requested breastfeeding	0.543	0.498	0.575	0.495	0.510	0.500
Panel D: FP coverage outcomes						
Any FP service received	0.243	0.429	0.234	0.424	0.253	0.435
Any injectable service	0.149	0.356	0.135	0.342	0.161	0.368
Number of FP services	0.571	1.407	0.515	1.261	0.624	1.534
Injectable coverage fraction	0.051	0.161	0.052	0.166	0.051	0.157
Husband knows woman's FP use[c]	0.983	0.129	0.983	0.131	0.984	0.127
Panel E: Reproductive outcomes						
Birth within 24 months of index birth	0.028	0.166	0.018	0.133	0.038	0.191
Pregnancy within 12 months	0.024	0.154	0.025	0.155	0.024	0.154
Pregnancy within 24 months	0.053	0.224	0.040	0.196	0.065	0.246
Mistimed pregnancy within 24 months	0.034	0.182	0.024	0.155	0.043	0.203
N (Panels A–C)	901		456		445	
N (Panels D–E, endline)	593–727		287–355		306–372	

Notes: Sample restricted to treatment-arm women with a first-counselor assignment ($N = 901$). Panel A reports the baseline covariates entering X_i in the IV regressions. Panels B and C are woman-level summaries computed across all sessions attended during the two-year follow-up (e.g., total sessions = count across all attended sessions; ever requested topic x = indicator that the woman raised topic x in any session at or after session 2). Panels D and E are measured at endline (year 2018). [a]Current FP use at baseline is structurally zero among pregnant women. [b]Counseling sessions per month is missing for 85 women with no session timing data (total $N = 816$). [c]Husband knows FP is asked of current FP users only. [d]Pregnant at baseline defines the two subgroups; it equals one (zero) for pregnant (postpartum) women by construction. IV estimation samples are marginally smaller than the descriptive counts above because the IV specification additionally requires all 12 baseline covariates to be non-missing (for example, $N = 811$ for sessions per month and $N = 591$ for husband knows FP in the IV regressions).

4.2 Main IV Results

This section reports the IV estimates of higher counseling quality on all 19 main outcomes, organized into four panels: counseling continuity (Panel A), counseling knowledge (Panel B), FP coverage (Panel C), and reproductive outcomes (Panel D). Panels A and B describe how the treatment operates (the mechanism channels); Panels C and D describe its behavioral consequences. Because the outcomes within a panel are multiple indicators of a single underlying channel rather than independent tests, we read each panel as one economic hypothesis and interpret its estimates jointly, rather than treating the nineteen outcomes as nineteen separate tests. As a formal check against multiple testing, Appendix Table A12 aggregates each panel into a single standardized summary index (Kling et al., 2007; Anderson, 2008): the continuity, topic, and FP-coverage indices (Panels A–C) are each significant in the pooled sample, and the reproductive index (Panel D) reproduces the postpartum-versus-pregnant divergence. We also subject the estimates to wild-cluster bootstrap inference clustering on the six counselors (Appendix Table A13): the FP-coverage take-up effects in Panel C remain significant under this conservative small-sample procedure, while the counselor-level mechanism channels in Panels A and B, identified off only six clusters, are not separately distinguishable from cluster-level noise under it even where their point estimates are large. Table 5 reports the full IV estimates, with the mechanism panels in the upper block and the behavioral-outcome panels in the lower block. Figure 3 displays the mechanism-channel estimates (counseling continuity and knowledge) graphically, and Figure 4 displays the FP-coverage and reproductive estimates; both report 95 percent confidence intervals.

Counseling continuity (Panel A). Higher-quality counseling compresses session dose for both subgroups. Pooled, women exposed to higher-quality counselors receive -0.42 fewer total sessions per woman ($p < 0.05$) and have a -0.17 lower share of sessions with the same counselor ($p < 0.01$) over the two-year horizon. The session reduction is largest among postpartum women (-0.62 sessions, $p < 0.05$) and smaller and not statistically significant among pregnant women (-0.25 sessions). High-quality counselors do not run longer or more numerous sessions; they achieve their results by delivering more content per session. The pooled effect on counseling sessions per month is small and statistically insignificant ($+0.021$), and the subgroup pattern is only suggestive: postpartum women receive directionally fewer visits per month under higher-quality counselors (-0.025, not significant) while pregnant women receive directionally more ($+0.061$, not significant), with a subgroup difference that does not reach conventional significance ($p_{\text{PP=Preg}} = 0.071$). This pattern is consistent with higher-quality counselors front-loading engagement during the antenatal window and reducing it during the postpartum window, but we read it as suggestive rather than statistically established, and return to it in Section 5 as part of the timing-misalignment hypothesis.

Counseling knowledge (Panel B). We measure knowledge coverage as the probability that the woman ever requested each of six FP-related topics in any session beyond her first, breaking the mechanical correlation with the first-session topic count that enters the quality index.[1] Higher-quality counseling expands per-session content sharply for both subgroups. Five of the six topics show large pooled effects, all significant at the 1 percent level: birth spacing (+0.61), pregnancy (+0.98), partner engagement (+0.77), child health (+0.18), and breastfeeding (+0.34). FP information is the exception: it is near zero on average and not individually significant in either subgroup, though the postpartum and pregnant estimates differ marginally (−0.06 versus +0.08, $p_{\text{PP=Preg}} = 0.095$). The largest subgroup contrast is on child health, where pregnant women gain +0.28 versus a not-significant +0.07 for postpartum women ($p_{\text{PP=Preg}} = 0.019$): higher-quality counselors who serve pregnant women appear to expand the topic set substantially toward child-health content that prepares the woman for the postpartum period.

FP coverage (Panel C). The FP-coverage outcomes measure the behavioral take-up consequences of the mechanism documented above, using endline administrative records of family-planning service receipt through the program's reimbursement channel. Higher-quality counseling raises FP take-up sharply for both subgroups. Pooled, women exposed to higher-quality counselors are +13.2 percentage points more likely to receive any FP service through the program ($p < 0.05$) and receive +0.69 additional services per woman ($p < 0.01$). The extensive margin is largest among postpartum women (+19.4 percentage points on any FP service, $p < 0.05$); the intensive margin is symmetric across subgroups, with postpartum women receiving an extra +0.71 services and pregnant women receiving an extra +0.67 services on average. The fraction of post-index fertile time covered by injectables rises by +8.3 percentage points pooled ($p < 0.01$), with the gain concentrated among pregnant-at-baseline women (+10.6 percentage points, $p < 0.05$). The spousal-knowledge outcome (whether the husband knows that the woman uses FP) is precisely zero on average but flips by pregnancy status: postpartum husbands are 8.2 percentage points less likely to know ($p < 0.10$) and pregnant husbands are 3.1 percentage points more likely to know ($p < 0.10$), with $p_{\text{PP=Preg}} = 0.021$. These FP-coverage estimates attenuate mechanically when we control for ancillary-component take-up (Section 3.2.7), because the ancillary indicator is nearly collinear with the administrative FP records used to construct the outcomes; the mechanism and reproductive coefficients are unchanged under the same control.

Reproductive outcomes (Panel D). We report IV effects on four reproductive outcomes measured at endline: pregnancy within 12 months; birth within 24 months; pregnancy within 24 months; and mistimed pregnancy. All four are measured on the same at-risk population: each pregnancy outcome counts any subsequent pregnancy regardless of whether it ended in a live birth, miscarriage,

[1] The postpartum pregnancy-information estimate (+1.06, s.e. 0.08) lies just above the unit bound for a binary outcome; because a binary-outcome complier LATE is bounded in $[-1, 1]$, we read this as finite-sample noise in the linear-probability IV (within one standard error of one), not a treatment effect outside the feasible range (Imbens and Angrist, 1994; Frandsen et al., 2023).

abortion, or stillbirth, which avoids conditioning the sample on a post-treatment event. The pooled estimates are modestly positive: +5.6 percentage points on birth within 24 months ($p < 0.10$), +3.6 percentage points on pregnancy within 24 months (not significant), and +4.8 percentage points on mistimed pregnancy (not significant). The subgroup contrast is sharp. Postpartum women exposed to higher-quality counselors experience +14.4 percentage points more births within 24 months ($p < 0.05$), +11.8 percentage points more pregnancies within 24 months ($p < 0.10$), and +12.0 percentage points more mistimed pregnancies ($p < 0.05$). Pregnant-at-baseline women experience the opposite direction, with point estimates of −3.3 percentage points on 24-month birth ($p < 0.10$), −4.3 percentage points on pregnancy within 12 months ($p < 0.05$), and −4.8 percentage points on 24-month pregnancy (not significant); the 12-month pregnancy result foreshadows the timing-misalignment evidence in Section 5. The subgroup differences are sharp for 24-month birth ($p_{\text{PP=Preg}} = 0.004$), for 24-month pregnancy ($p_{\text{PP=Preg}} = 0.023$), and for mistimed pregnancy ($p_{\text{PP=Preg}} = 0.033$).

The +12.0 percentage-point postpartum increase in mistimed pregnancies is an aggregate effect on a low-frequency outcome (postpartum mean ≈ 4 percent), and we read it cautiously, considering three alternative interpretations. First, the result is consistent with a real behavioral failure of peri-partum counseling: higher-quality counselors may inadvertently reduce sustained method continuation among postpartum women whose contraceptive needs evolve over the 24 months following delivery, even as initial take-up rises. Second, it is consistent with a selection-in-reported-wantedness channel: women who engage more actively with FP services may have stronger ex ante preferences for avoiding pregnancy, raising the probability of reporting any subsequent pregnancy as mistimed conditional on its occurrence. Third, it is consistent with noise in a rare-event outcome: the estimate is identified off a small set of complier-margin birth and non-birth flips against a small base rate, and the reproductive panel proves the part of our results most sensitive to the exclusion of any single counselor, as Section 4.3 documents. The FP-coverage outcomes in the preceding paragraph, by contrast, are robust to dropping any of the six counselors.

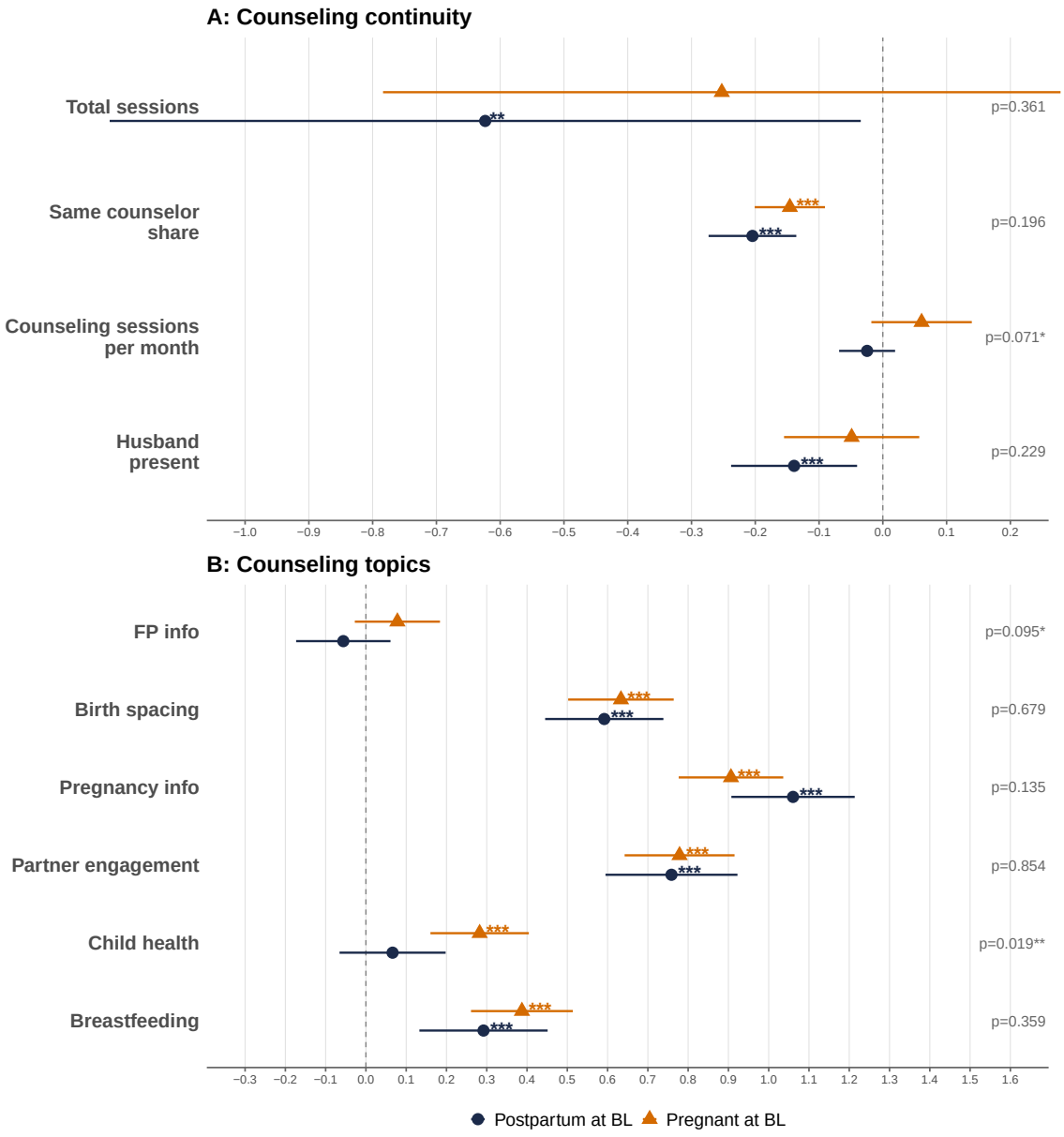
Table 5. Main IV Results: Counseling Quality on Mechanism Channels and Behavioral Outcomes

	Panel A: Counseling continuity				*Panel B: Counseling topics (post-1st-session)*					
	Total sessions	Same counselor	Couns sessions per month	Husband present	FP information	Birth spacing	Pregnancy info	Partner engagement	Child health	Breastfeeding
Pooled										
Estimate	-0.422**	-0.173***	0.021	-0.093**	0.017	0.614***	0.977***	0.770***	0.183***	0.344***
	(0.201)	(0.022)	(0.023)	(0.036)	(0.040)	(0.050)	(0.050)	(0.054)	(0.046)	(0.051)
Postpartum (preg=0)										
Estimate	-0.623**	-0.204***	-0.025	-0.139***	-0.056	0.592***	1.060***	0.759***	0.066	0.292***
	(0.300)	(0.035)	(0.022)	(0.050)	(0.060)	(0.075)	(0.078)	(0.084)	(0.067)	(0.081)
Pregnant (preg=1)										
Estimate	-0.253	-0.146***	0.061	-0.049	0.078	0.633***	0.906***	0.779***	0.282***	0.387***
	(0.271)	(0.028)	(0.040)	(0.054)	(0.054)	(0.067)	(0.066)	(0.070)	(0.062)	(0.065)
p(PP=Preg)	0.361	0.196	0.071	0.229	0.095	0.679	0.135	0.854	0.019	0.359
N	896	896	811	781	896	896	896	896	896	896
Mean DV (pooled)	4.643	0.865	0.382	0.133	0.873	0.483	0.204	0.328	0.688	0.543

	Panel C: FP coverage					*Panel D: Reproductive*			
	Any FP service	Any injectable	Number of FP services	Injectable coverage	Husband knows FP	Birth within 24 months	Pregnancy within 12 months	Pregnancy within 24 months	Mistimed pregnancy
Pooled									
Estimate	0.132**	0.129**	0.689***	0.083***	-0.027	0.056*	-0.003	0.036	0.048
	(0.062)	(0.054)	(0.238)	(0.030)	(0.024)	(0.031)	(0.023)	(0.037)	(0.034)
Postpartum (preg=0)									
Estimate	0.194**	0.102	0.712**	0.061*	-0.082*	0.144**	0.036	0.118*	0.120**
	(0.094)	(0.079)	(0.346)	(0.036)	(0.045)	(0.059)	(0.040)	(0.065)	(0.060)
Pregnant (preg=1)									
Estimate	0.072	0.155**	0.667**	0.106**	0.031*	-0.033*	-0.043**	-0.048	-0.026
	(0.082)	(0.075)	(0.314)	(0.045)	(0.018)	(0.017)	(0.019)	(0.034)	(0.030)
p(PP=Preg)	0.328	0.623	0.921	0.418	0.021	0.004	0.071	0.023	0.033
N	724	724	724	701	591	701	701	701	702
Mean DV (pooled)	0.230	0.141	0.536	0.051	0.983	0.028	0.024	0.053	0.033

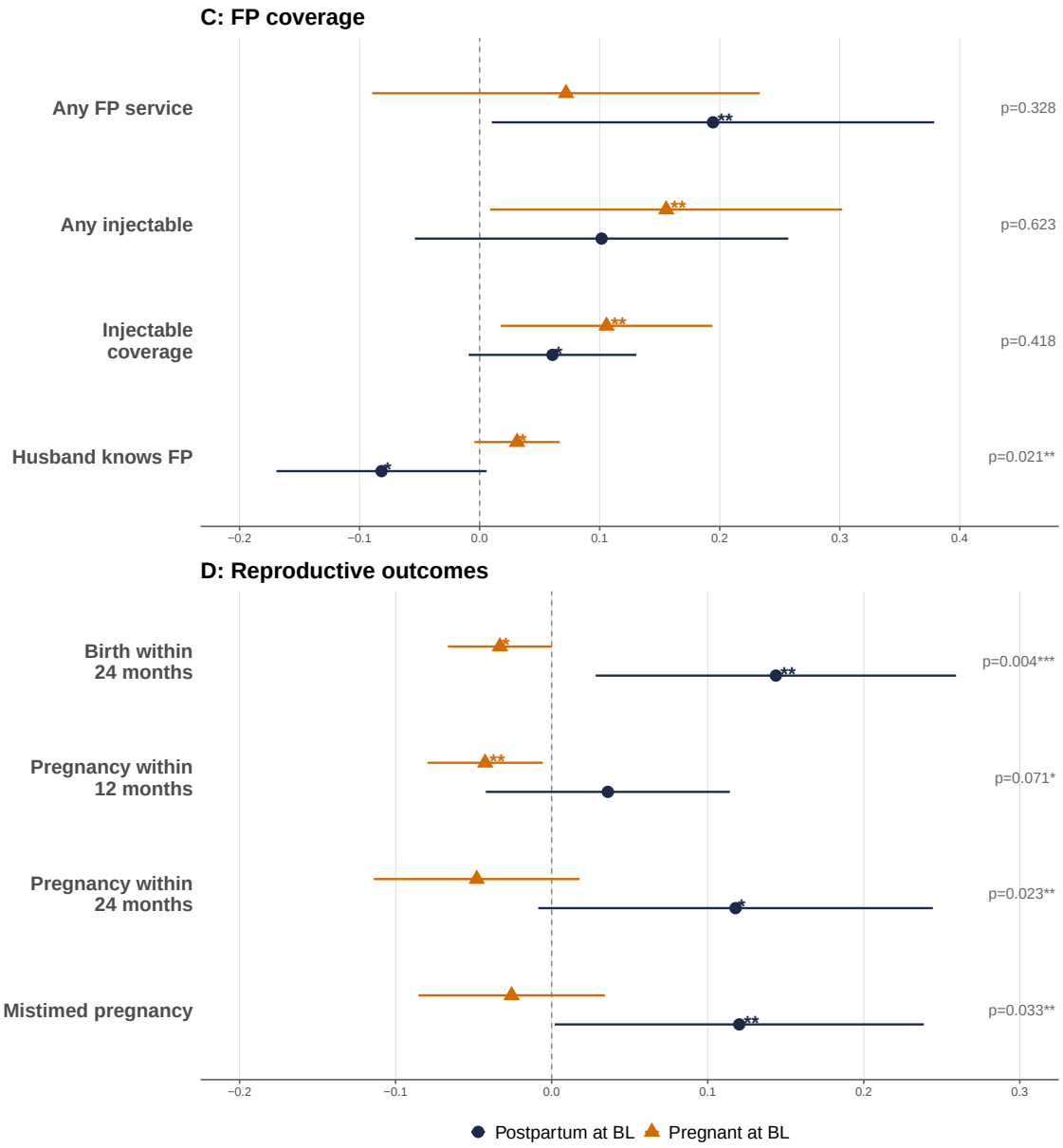
Notes: Two-stage least squares estimates of higher counseling quality on each outcome. The upper block reports the mechanism-channel panels (Panels A and B); the lower block reports the behavioral-outcome panels (Panels C and D). Endogenous variable: an above-median indicator for the woman's average satisfaction-and-topics quality across all attended sessions. Instrument: the leave-one-out, scheduling-residualized Counselor Quality Tendency (CQT) of the woman's first counselor. Heteroskedasticity-robust standard errors in parentheses below each estimate. The row *p(PP=Preg)* reports the p-value testing equality of the postpartum and pregnant subgroup effects. Stars: * p<0.10, ** p<0.05, *** p<0.01.

Figure 3. Heterogeneous IV Effects on Mechanism Channels by Pregnancy Status



Notes: 2SLS estimates of higher counseling quality on counseling continuity and knowledge outcomes, with 95 percent confidence intervals, separately for postpartum (navy) and pregnant-at-baseline (orange) subgroups. Stars indicate two-sided significance at the 10, 5, and 1 percent levels. Right-margin labels report the p-value for the test of equality of postpartum and pregnant subgroup effects.

Figure 4. Heterogeneous IV Effects on Behavioral Outcomes (FP Coverage and Reproductive)



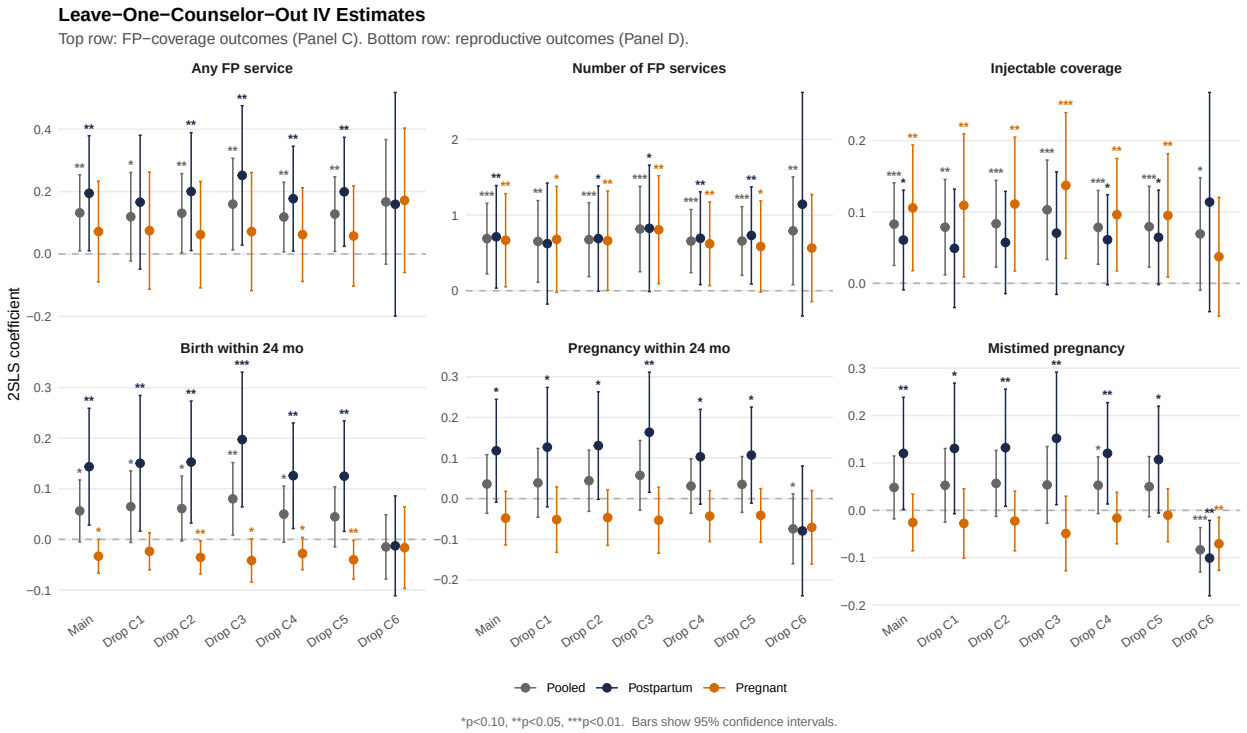
Notes: 2SLS estimates of higher counseling quality on FP coverage and reproductive outcomes by pregnancy status at baseline. Postpartum women in navy; pregnant women in orange. 95 percent confidence intervals; stars indicate two-sided significance at the 10, 5, and 1 percent levels.

4.3 Robustness: Leave-One-Counselor-Out

A central concern in any examiner design with few decision-makers is that one decision-maker contributes disproportionately to the identifying contrast. Figure 5 reports the main IV estimate for six outcomes, three drawn from FP coverage (Panel C) and three from reproductive outcomes (Panel D), alongside the six leave-one-counselor-out estimates. The top row, showing FP-coverage outcomes, demonstrates that the headline take-up results are directionally stable across every leave-one-out specification: the point estimates retain their positive sign in every drop, but their statistical significance is not uniform. The pooled estimates for the number of family-planning services received and for injectable coverage remain significant under all six drops, whereas the pooled estimate for receiving any family-planning service, along with several of the postpartum and pregnant subgroup estimates, loses significance when counselor 6 is excluded. The bottom row, showing reproductive outcomes, shows a different and more fragile pattern. For postpartum women, the +14.4 percentage-point effect on births within 24 months, the +11.8 percentage-point effect on pregnancies within 24 months, and the +12.0 percentage-point effect on mistimed pregnancies all turn negative when counselor 6 is excluded (to −1.3, −7.9, and −10.1 percentage points, respectively); the first two become statistically insignificant, while the mistimed-pregnancy estimate remains significant at the 5 percent level.

We interpret this contrast as a structural feature of examiner designs with few decision-makers. With six counselors, one counselor's residualized quality tendency contributes disproportionately to the identifying variation, and outcomes with low base rates (the reproductive panel is measured on outcomes with means near 4 percent) are particularly sensitive to that contribution. The FP-coverage mechanism, by contrast, is identified off outcomes with substantially higher base rates and substantially larger compliance margins. Our reading of the LOO evidence is that the mechanism through which counseling quality operates is robust and policy-relevant, while any policy claim running through the reproductive panel alone is not supported by the data once counselor 6 is excluded. Appendix Tables A6, A7, and A8 report the full 19-outcome leave-one-counselor-out exercise for the pooled, postpartum, and pregnant samples, respectively.

Figure 5. Leave-One-Counselor-Out IV Estimates: FP-Coverage (Stable) vs. Reproductive Outcomes (C6-Driven)



Notes: 2SLS coefficients in the main specification (column 1) and after sequentially dropping each of the six counselors (columns 2–7). Top row: three FP-coverage outcomes (Panel C); bottom row: three reproductive outcomes (Panel D). Pooled (grey), postpartum (navy), and pregnant (orange) subgroup estimates with 95 percent confidence intervals. Stars indicate two-sided significance at the 10, 5, and 1 percent levels.

5 A Timing-Misalignment Hypothesis

5.1 Hypothesis and Timing Mechanism by Baseline Proximity to Delivery

One hypothesis consistent with the reproductive heterogeneity documented in Section 4.2 is a timing misalignment between the engagement window in which higher-quality counseling produces FP take-up, concentrated within roughly three months of delivery for both subgroups, and the fertile window over which protection actually matters. For postpartum women, that engagement window opens shortly before re-exposure to pregnancy; for pregnant-at-baseline women in the third trimester, the same engagement window aligns with a shorter post-delivery follow-up over which protection mechanically applies. We do not formally test this mechanism and emphasize that other explanations, including differential method continuation, behavioral spillovers to partners, selection in reported wantedness, and noise in low-frequency reproductive events, are not ruled out by these estimates.

Table 6 decomposes the main IV estimates by baseline proximity to delivery within each sub-

group. Postpartum women are split at the median of months-since-delivery at baseline ($\approx$ 2.9 months) into recent-PP (delivered within the past 2.9 months) and older-PP (delivered between 2.9 and 6 months before baseline). Pregnant women are split at the median of months-until-delivery at baseline ($\approx$ 2.6 months) into third-trimester Preg (delivery imminent within 2.6 months) and earlier-gestation Preg (delivery 2.6 to 8 months away).

The pattern is consistent with peri-partum concentration of the effects. Recent-PP women experience a +31.8 percentage-point effect on any FP service ($p < 0.05$) compared with +9.3 percentage points for older-PP women (not significant), and a +14.2 percentage-point effect on 24-month births ($p < 0.10$) compared with +12.6 percentage points for older-PP women (not significant). Third-trimester pregnant women see a +24.4 percentage-point effect on any injectable service ($p < 0.05$) compared with +11.0 percentage points for earlier-gestation pregnant women (not significant), a +10.7 percentage-point increase in injectable coverage of the post-index fertile window ($p < 0.05$) compared with +7.7 percentage points (not significant), and a −8.0 percentage-point effect on 24-month pregnancies ($p < 0.05$) compared with a non-significant −2.2 percentage points for earlier-gestation pregnant women. The injectable-uptake and FP-coverage effects therefore concentrate sharply among women whose index birth date is closest to baseline, consistent with the visit-density flip documented in the counseling-continuity panel above.

A formal decomposition of the 24-month pregnancy window into four non-overlapping six-month intervals, estimated as an interval-hazard IV, is reported in Appendix Table A4 and Appendix Figure B3. The dominant component is a +10.5 percentage-point catch-up in pregnancy occurrence among postpartum women in the 12–18 month window after the index birth, consistent with the peri-partum injectable protection initiated near delivery having expired by the start of that window.

Table 6. Timing of FP Coverage: IV Heterogeneity by Baseline Reproductive Timing

	Any FP service	Any injectable	Number of FP services	Injectable coverage	Inj. protected days	Birth within 24mo *or* Preg within 24mo (panel-specific)
Panel A: Postpartum women, by months postpartum at baseline (last column = Birth within 24 months)						
Recent-PP (≤2.9 mo): Estimate	0.318***	0.178	0.916*	0.067	51.2	0.142*
	(0.123)	(0.109)	(0.516)	(0.053)	(38.9)	(0.077)
Older-PP (>2.9 mo): Estimate	0.093	0.037	0.360	0.043	33.2	0.126
	(0.118)	(0.105)	(0.377)	(0.041)	(33.1)	(0.078)
Mean DV (Recent-PP (≤2.9 mo))	0.246	0.139	0.599	0.046	33.4	0.032
Mean DV (Older-PP (>2.9 mo))	0.263	0.183	0.651	0.056	45.7	0.043
Panel B: Pregnant women, by months until delivery at baseline (last column = Pregnancy within 24 months)						
3rd-tri Preg (≤2.6 mo): Estimate	0.164	0.244**	0.720**	0.107**	68.9**	-0.078**
	(0.128)	(0.112)	(0.358)	(0.047)	(29.7)	(0.035)
Earlier-gest Preg (>2.6 mo): Estimate	0.058	0.110	0.575	0.077	44.2	-0.021
	(0.112)	(0.098)	(0.415)	(0.058)	(32.5)	(0.059)
Mean DV (3rd-tri Preg (≤2.6 mo))	0.302	0.154	0.598	0.044	28.8	0.036
Mean DV (Earlier-gest Preg (>2.6 mo))	0.188	0.127	0.491	0.059	32.8	0.042

Notes: Two-stage least squares estimates of higher counseling quality (sat+topics) on each outcome, run separately within each baseline-timing subgroup. Median splits within each pregnancy-status group: 2.9 months postpartum at baseline for Panel A; 2.6 months until delivery for Panel B. Heteroskedasticity-robust standard errors in parentheses below each estimate. "Mean DV" rows report the within-subgroup mean of the outcome among treatment women. The last outcome column is panel-specific: *birth within 24 months* in Panel A, *pregnancy within 24 months* in Panel B. Sample sizes per subgroup (same across columns except for the last reproductive column due to small differences in missingness): Panel A Recent-PP $N = 186$, Older-PP $N = 184$; Panel B 3rd-tri Preg $N = 168$ (Pregnancy-within-24-months $N = 168$), Earlier-gest Preg $N = 163$ (Pregnancy-within-24-months $N = 163$). Stars: * p<0.10, ** p<0.05, *** p<0.01.

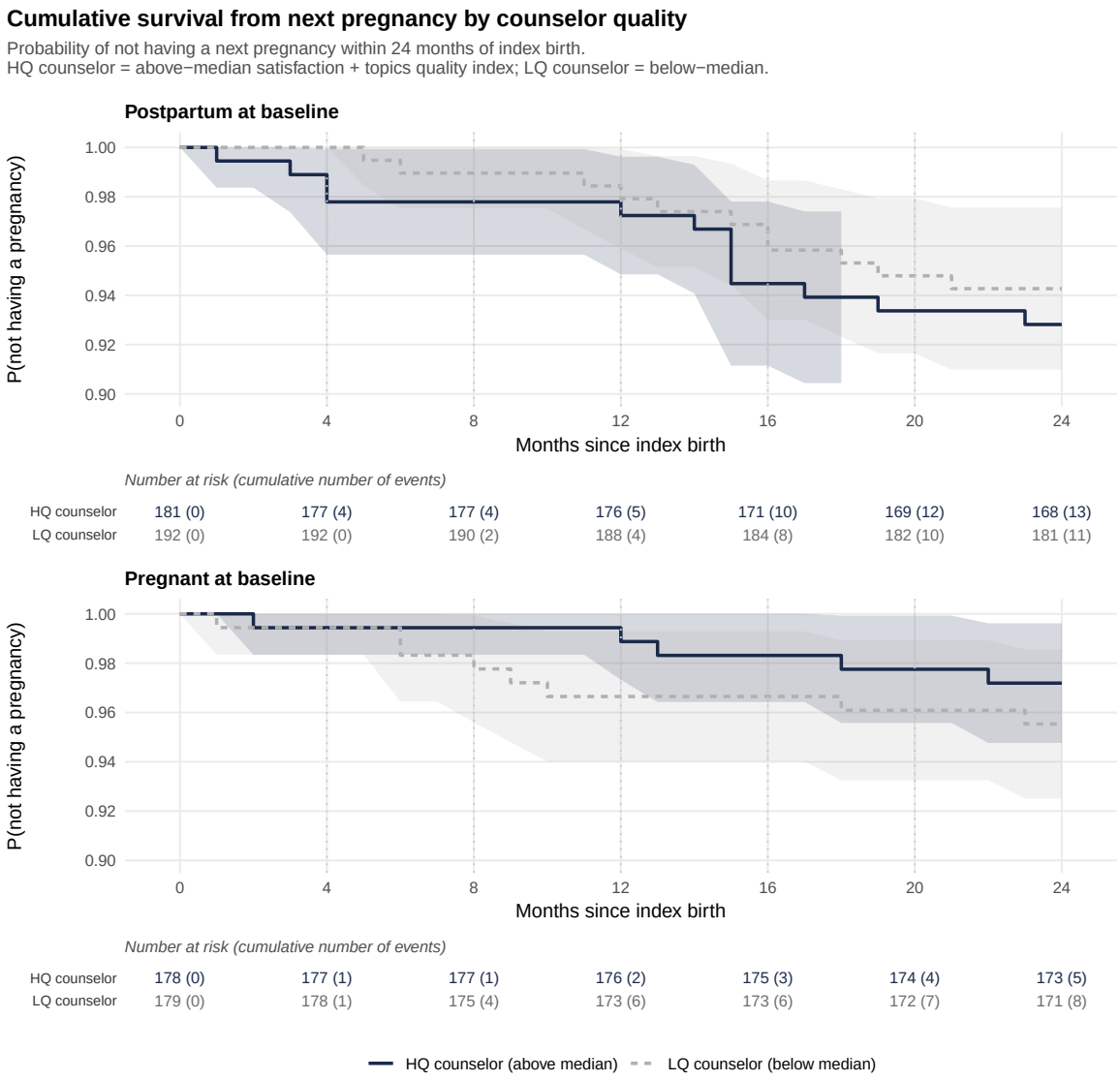
5.2 Survival Curves of Next Pregnancy by Counselor Quality

Figure 6 reports Kaplan-Meier-style cumulative survival curves of the probability of not having a next pregnancy (any subsequent pregnancy event, including those ending in miscarriage, abortion, or stillbirth) within 24 months of the index birth date, separately for postpartum (top panel) and pregnant-at-baseline (bottom panel) women. Within each panel, we plot two curves: women who received higher-quality counseling (above-median woman-level satisfaction-and-topics quality) and women who received lower-quality counseling (below-median). The split is on the woman's realized counseling quality, which is the endogenous treatment, not on counselor assignment, so the contrast is descriptive rather than an assignment-based reduced form.

In the postpartum panel, the two curves are virtually indistinguishable for the first four months, after which the higher-quality curve falls below the lower-quality curve from month 12 onward. By month 24, the higher-quality group has accumulated 13 pregnancies (out of 181 women) versus 11 in the lower-quality group (out of 192). In the pregnant-at-baseline panel, the pattern is reversed and emerges earlier: the lower-quality curve drops below the higher-quality curve around month 6

and stays below thereafter. By month 24, the higher-quality group has accumulated 5 pregnancies (out of 178 women) versus 8 in the lower-quality group (out of 179). The descriptive pattern in this figure aligns with the IV decomposition reported above and with the interval-hazard estimates in Appendix Table A4.

Figure 6. Kaplan-Meier Survival Curves of Next Pregnancy by Realized Counseling Quality



Notes: Probability of not having a next pregnancy within 24 months of index birth, by realized counseling-quality group, separately for postpartum (top) and pregnant-at-baseline (bottom) women. HQ = above-median woman-level satisfaction-and-topics quality index; LQ = below-median. Women are grouped by their own realized quality, which is the endogenous treatment, not by counselor assignment. Shaded bands show 95% pointwise confidence intervals. The number-at-risk row reports women still pregnancy-free at each four-month mark, with cumulative events in parentheses. This figure plots the unadjusted survival contrast; corresponding 2SLS interval-hazard IV magnitudes are reported in Appendix Table A4.

6 Discussion

Our findings refine three open questions in the family-planning counseling and judge-design literatures: how counseling quality shapes downstream behavior, what an examiner-design IV identifies when agents differ in skill as well as in preferences, and which features of the patient-provider interaction matter in settings where access is not the binding constraint. We discuss each in turn.

The behavioral story that emerges from Panels A through C is one of take-up rather than continuation. Higher-quality counseling produces sharply more FP-service receipt for both postpartum and pregnant women, but the gains do not consistently extend to longer-window outcomes such as LARC adoption or sustained method use. This pattern is consistent with the interpretation that quality affects the woman's decision to start a method without changing her decision to continue, and aligns with prior evidence that contraceptive continuation is governed by a different set of margins than initiation, including side-effect experience, partner dynamics, and provider follow-up (Ali et al., 2012; Castle and Askew, 2015). Dieye et al. (2026) report a parallel decoupling in a Kenyan postpartum digital-counseling RCT: gains in knowledge, intentions, and perceived counseling quality coexist with a precise null on modern contraceptive use, suggesting that the high-quality-counseling-to-sustained-use pathway is not automatic. Our findings refine this picture by isolating a specific channel through which counseling quality does shift behavior at the take-up margin: peri-partum injectable coverage, particularly among third-trimester pregnant women.

The suggestive visit-density pattern documented in Section 4.2 offers one interpretation of how higher-quality counselors allocate time: pregnant women receive directionally more visits per month while postpartum women receive directionally fewer, even though both subgroups receive fewer total sessions. The subgroup difference is only marginally significant ($p_{\text{PP}=\text{Preg}} = 0.071$), so we do not lean on it as an established finding; read cautiously, this asymmetric allocation would be consistent with a deliberate scheduling choice in which post-delivery FP initiation is planned during the antenatal window, when a joint decision with the partner is still being established, and a lighter touch suffices during the postpartum window, when a method has typically already been adopted. The same asymmetric allocation pattern shows up in the spousal-knowledge result: husbands of higher-quality-counseled pregnant women become more informed about FP while husbands of higher-quality-counseled postpartum women become less informed, with the subgroup difference significant at conventional levels. We read this as suggestive evidence that higher-quality counseling reorganizes the spousal information environment in opposite directions for the two subgroups: a joint-decision frame for pregnant women whose post-delivery FP plan is still being established, and a woman-centered frame for postpartum women whose method is already in place.

The reproductive heterogeneity in Panel D admits a coherent reading through the timing-misalignment hypothesis developed in Section 5: for third-trimester pregnant women, the peri-partum engagement window aligns with the post-delivery fertile horizon and produces a measurable -8.0 percentage-point reduction in 24-month pregnancy, while for postpartum women already weeks or months past delivery the same engagement window covers only a small fraction of the remaining

fertile horizon, and the injectable protection initiated near delivery runs out by the 12–18 month catch-up window in which we detect a +10.5 percentage-point increase in pregnancy occurrence. We offer this timing-misalignment hypothesis as one reading rather than as a formally tested mechanism, alongside the three alternative readings of the +12.0 percentage-point postpartum increase in mistimed pregnancy we discussed in Section 4.2: real behavioral failure of peri-partum counseling, selection in reported wantedness, and noise in a low-frequency outcome. These readings are not mutually exclusive, and we do not adjudicate among them.

A structural feature of our setting, and of any examiner design with a small number of decision-makers, is that the identifying contrast can be disproportionately driven by one decision-maker. With $G = 6$ counselors, our leave-one-out exercise documents that the FP-coverage mechanism is the most robust subset of our findings, while the reproductive panel does not survive excluding counselor 6, with its estimates collapsing toward zero or reversing sign. We treat the contrast itself as informative: Chan et al. (2022) show that examiner-IV estimates can be misleading when agents differ in skill as well as in preferences, which is why we report the full leave-one-counselor-out exercise and let it discipline which results we are willing to generalize. Our reading is that the mechanism we identify — compressed dose, expanded content, and peri-partum injectable take-up — is the part of our contribution that should bear policy weight, while any claim running through the reproductive panel alone is contingent on the contribution of one counselor.

Three limitations bear acknowledging. First, the setting is urban Lilongwe, Malawi, and external generalization to rural Malawi, other sub-Saharan settings, or non-RCT clinic conditions requires additional evidence; the methodological contribution of the paper is the identification strategy rather than the specific quantitative magnitudes. Second, with a 24-month follow-up window the reproductive panel is identified off a small number of complier-margin events, which both motivates and constrains our leave-one-out interpretation. Third, our quality index is patient-reported rather than based on direct observation of counselor-side behavior, which conflates objective counseling features (topic coverage, session structure) with patient-perceived relational quality.

With these limitations in mind, our results suggest that investing in counselor quality at the peri-partum window is a candidate policy lever for expanding family-planning take-up among both pregnant and postpartum women, with the strongest evidence base for the injectable-coverage margin in the third trimester. Downstream reproductive effects in our setting are contingent on the contribution of one counselor and should not be generalized without further evidence.

7 Conclusion

This paper provides causal evidence on family-planning counseling quality, distinct from access to counseling, by exploiting random first-counselor assignment in a Malawian RCT. We build a Counselor Quality Tendency instrument from the leave-one-out, scheduling-residualized average of patient-reported first-session quality and use it to identify the local average treatment effect of higher counseling quality on 19 outcomes organized into four panels. Higher counseling quality

compresses session dose while expanding per-session content for both postpartum and pregnant women and produces sharp gains in family-planning take-up. The take-up gains are robust to dropping any of the six counselors and concentrate sharply in peri-partum injectable coverage, with pregnant women in the third trimester gaining +10.7 percentage points in injectable coverage of the post-index fertile window and postpartum women gaining +19.4 percentage points in any family-planning service received. Reproductive outcomes diverge sharply by baseline pregnancy status but the heterogeneity is identified primarily from one counselor's contribution.

Two methodological points are worth emphasizing. First, our setting is one in which examiners (counselors) plausibly differ in skill as well as in preferences, and the standard convex-combination interpretation of an examiner-IV LATE (Kolesár, 2013; Chan et al., 2022) therefore captures the effect of being assigned a counselor with a higher first-session quality tendency, rather than a pure preference shift. Second, with $G = 6$ counselors the identifying contrast is structurally sensitive to the contribution of any one counselor; we therefore report a full leave-one-counselor-out diagnostic in the main text, in keeping with the recommendation that examiner designs demonstrate robustness to the set of examiners included (Chyn et al., 2025). The substantive payoff is that we identify the FP-coverage mechanism as the policy-relevant subset of our results: a counseling-quality intervention designed to expand peri-partum injectable take-up has a defensible empirical basis in our data, while a counseling-quality intervention designed to move downstream reproductive outcomes does not.

The broader empirical literature has documented that in many low- and middle-income settings access and price are not the binding constraints on contraceptive use (Dupas et al., 2025; Miller et al., 2025). If access is not the binding constraint, the question becomes which features of the patient-provider interaction are the binding constraint. Our findings suggest one answer: the breadth of topics covered in the first counseling session, combined with patient satisfaction, generates substantial between-counselor variation in downstream service take-up, and the take-up gains are concentrated precisely in the peri-partum injectable margin that has the largest scope for protection in this setting.

Three extensions follow naturally from these findings. First, a follow-up window longer than two years would help separate whether the take-up gains we document translate into sustained method continuation; the current 24-month panel is identified off a small number of reproductive events and cannot speak credibly to longer-run continuation outcomes. Second, validating the patient-reported quality index against video-coded session transcripts or independent observation of counselor behavior would clarify how much of the patient-reported variation reflects observable counseling features versus relational features that are harder to standardize and scale. Third, deploying the design in settings with larger counselor pools would enable formal monotonicity testing along the lines of Frandsen et al. (2023) and would reduce the small-G sensitivity that constrains our leave-one-out interpretation here.

CRediT authorship contribution statement

Marie C. Montás: Conceptualization, Methodology, Data Curation, Visualization, Formal analysis, Writing – original draft, Validation. **Sarah Vincent:** Methodology, Writing – original draft, Validation. **Mahesh Karra:** Conceptualization, Methodology, Writing – original draft, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A: Tables

Appendix Table A1. Quality Index Component Diagnostics

Component	Mean (SD)	Between-counselor η^2	ANOVA $F(5, 895)$	Role
Patient satisfaction (1–4)	3.31 (0.48)	0.599	267.5^{***}	**Selected (primary index)**
Topics requested (0–9)	3.15 (2.67)	0.900	1616.9^{***}	**Selected (primary index)**
Session duration (min)	13.35 (1.81)	0.004	0.69	Dropped from primary
Questions answered (0/1)	0.99 (0.12)	0.033	6.10^{***}	Dropped from primary
Cronbach's α (standardized):				
Satisfaction + topics (primary index)	$\boldsymbol{\alpha = 0.74}$			
All four components (robustness)	$\alpha = 0.33$			
Pairwise correlation between selected components:				
Satisfaction × topics	$r = 0.58$, $p < 0.001$			

Notes: Sample restricted to treatment-arm women with a first-counselor assignment ($N = 901$), measured at the first counseling session. Between-counselor η^2 is the share of total variance that lies between the six counselors, from a one-way analysis of variance with counselor identity as the grouping variable. Higher η^2 indicates the component captures more systematic counselor-level variation, which is the dimension the examiner-design instrument leverages. Patient satisfaction and topics requested have substantial between-counselor variation ($\eta^2 = 0.60$ and $\eta^2 = 0.90$) and load on a single factor with Cronbach's $\alpha = 0.74$, supporting their joint use as a scalar quality index. Session duration ($\eta^2 = 0.004$) and the questions-answered indicator ($\eta^2 = 0.033$) have essentially no between-counselor variation and are dropped from the primary index; the four-component bundle has $\alpha = 0.33$, indicating it is not a coherent unidimensional construct. $^{***}p < 0.01$.

Appendix Table A2. Robustness: Four-Component Endog — Mechanism Outcomes

	Panel A: Counseling continuity				*Panel B: Counseling topics (post-1st-session)*					
	Total sessions	Same counselor	Couns sessions per month	Husband present	FP information	Birth spacing	Pregnancy info	Partner engagement	Child health	Breastfeeding
Pooled										
Estimate	-0.529**	-0.188***	0.026	-0.111***	0.010	0.724***	1.159***	0.939***	0.254***	0.449***
	(0.234)	(0.026)	(0.027)	(0.040)	(0.047)	(0.059)	(0.062)	(0.063)	(0.053)	(0.059)
Postpartum (preg=0)										
Estimate	-0.722**	-0.216***	-0.024	-0.155***	-0.069	0.684***	1.225***	0.908***	0.129*	0.406***
	(0.344)	(0.041)	(0.025)	(0.055)	(0.068)	(0.084)	(0.089)	(0.092)	(0.074)	(0.089)
Pregnant (preg=1)										
Estimate	-0.357	-0.162***	0.072	-0.069	0.080	0.759***	1.100***	0.967***	0.366***	0.487***
	(0.321)	(0.034)	(0.048)	(0.061)	(0.064)	(0.082)	(0.088)	(0.085)	(0.075)	(0.078)
p(PP=Preg)	0.438	0.314	0.086	0.307	0.113	0.520	0.321	0.633	0.025	0.494
N	896	896	811	781	896	896	896	896	896	896
Mean DV	4.643	0.865	0.382	0.133	0.873	0.483	0.204	0.328	0.688	0.543

Notes: Robustness check using a four-component endogenous variable: an above-median binary indicator for the woman-level mean of an index combining patient satisfaction, breadth of topics covered, session duration, and a questions-answered indicator. Instrument: the corresponding leave-one-out, scheduling-residualized Counselor Quality Tendency built from the four-component index. Heteroskedasticity-robust standard errors in parentheses below each estimate. Stars: * p<0.10, ** p<0.05, *** p<0.01.

Appendix Table A3. Robustness: Four-Component Endog — FP Coverage and Reproductive Outcomes

	Panel C: FP coverage					*Panel D: Reproductive*				
	Any FP service	Any injectable	Number of FP services	Injectable coverage	Husband knows FP	Birth within 12 months	Birth within 24 months	Pregnancy within 12 months	Pregnancy within 24 months	Mistimed pregnancy
Pooled										
Estimate	0.136**	0.135**	0.719***	0.087***	-0.030	0.026	0.064*	-0.002	0.043	0.059
	(0.069)	(0.060)	(0.265)	(0.033)	(0.026)	(0.017)	(0.035)	(0.025)	(0.040)	(0.037)
Postpartum (preg=0)										
Estimate	0.203**	0.102	0.725*	0.060	-0.087*	0.048	0.160**	0.040	0.134*	0.139**
	(0.103)	(0.087)	(0.381)	(0.039)	(0.048)	(0.032)	(0.066)	(0.045)	(0.072)	(0.067)
Pregnant (preg=1)										
Estimate	0.070	0.168**	0.714**	0.116**	0.035*	0.002	-0.037**	-0.046**	-0.052	-0.025
	(0.093)	(0.085)	(0.360)	(0.051)	(0.018)	(0.003)	(0.017)	(0.019)	(0.037)	(0.033)
p(PP=Preg)	0.343	0.594	0.984	0.382	0.021	0.149	0.005	0.076	0.023	0.032
N	724	724	724	701	591	701	701	701	701	702
Mean DV	0.230	0.141	0.536	0.051	0.983	0.005	0.028	0.024	0.053	0.033

Notes: Robustness check using a four-component endogenous variable: an above-median binary indicator for the woman-level mean of an index combining patient satisfaction, breadth of topics covered, session duration, and a questions-answered indicator. Instrument: the corresponding leave-one-out, scheduling-residualized Counselor Quality Tendency built from the four-component index. Heteroskedasticity-robust standard errors in parentheses below each estimate. Stars: * p<0.10, ** p<0.05, *** p<0.01.

Appendix Table A4. Interval-Hazard IV: Effect of Higher Counseling Quality on Pregnancy Occurrence by Months since Index Birth

	Months since index birth (non-overlapping intervals)			
	0–6 months	6–12 months	12–18 months	18–24 months
Pooled				
Estimate	0.004	-0.007	0.057**	-0.018**
	(0.018)	(0.014)	(0.029)	(0.009)
Postpartum (preg=0)				
Estimate	0.022	0.014	0.105**	-0.023
	(0.032)	(0.024)	(0.052)	(0.014)
Pregnant (preg=1)				
Estimate	-0.013	-0.029**	0.008	-0.014
	(0.012)	(0.015)	(0.026)	(0.010)
p(PP=Preg)	0.295	0.120	0.100	0.579
N	701	701	701	701
Mean DV (pooled)	0.013	0.011	0.020	0.009
Mean DV (PP)	0.015	0.010	0.028	0.013
Mean DV (Preg)	0.011	0.011	0.011	0.006

Notes: Two-stage least squares estimates of higher counseling quality on the probability that a woman's next pregnancy event occurred within each non-overlapping interval from her index birth. Endogenous variable: an above-median indicator for the woman's average satisfaction-and-topics quality across all attended sessions. Instrument: the leave-one-out, scheduling-residualized Counselor Quality Tendency (CQT) of the woman's first counselor. The four intervals partition the 24-month follow-up window and together sum to the pregnancy-within-24-months outcome reported in Table 5. Each pregnancy outcome counts any subsequent pregnancy regardless of whether it ended in a live birth, miscarriage, abortion, or stillbirth, matching the Panel D at-risk population. Heteroskedasticity-robust standard errors in parentheses. The row *p(PP=Preg)* reports the p-value for testing equality of the two subgroup effects. Stars: * p<0.10, ** p<0.05, *** p<0.01.

Appendix Table A5. Counseling Topic Coverage by Counselor: Session-Level Descriptive Statistics

	Counselor						
Topic share	1	2	3	4	5	6	Overall
Panel A: Topic coverage (share of sessions)							
FP methods	0.438	0.369	0.231	0.373	0.493	0.947	0.442
Pregnancy	0.091	0.052	0.087	0.041	0.049	0.908	0.175
Birth spacing	0.324	0.115	0.242	0.062	0.284	0.967	0.299
Side effects of FP	0.317	0.218	0.504	0.065	0.309	0.773	0.351
FP services	0.189	0.258	0.049	0.000	0.331	0.557	0.192
STI prevention	0.201	0.069	0.101	0.181	0.325	0.615	0.229
Partner involvement	0.208	0.087	0.023	0.005	0.022	0.877	0.166
Child health	0.194	0.125	0.156	0.304	0.253	0.949	0.308
Breastfeeding	0.200	0.228	0.121	0.094	0.300	0.939	0.273
Panel B: Summary measures							
Mean topics/session	2.16	1.52	1.51	1.12	2.37	7.53	2.43
Topic breadth (0–9)	9	9	9	8	9	9	—
Mean minutes/session	12.87	13.39	12.94	13.17	13.04	13.13	13.08
Herfindahl index	0.13	0.15	0.19	0.22	0.14	0.11	—
N sessions	571	504	900	861	550	512	3,898

Notes: Each cell in Panel A reports the share of sessions where the topic was discussed, by the counselor who conducted the session (not restricted to first session), across all 3,898 counseling sessions with a valid recorded session length. Panel B reports summary measures. Topic breadth = number of distinct topics the counselor ever covers (out of 9). Herfindahl index $= \sum_{k=1}^{9} \tilde{s}_k^2$, where $\tilde{s}_k = s_k / \sum_j s_j$ is topic k's share of the counselor's total topic coverage, normalized so the shares sum to one; higher values indicate that coverage concentrates on fewer topics. Counselor 6 (highest CQT) has the lowest concentration, spreading attention nearly evenly across topics, whereas Counselor 4 (lower CQT) is the most concentrated, focusing on a narrower set of topics.

Appendix Table A6. Robustness: Leave-One-Counselor-Out IV Estimates — Pooled Sample

Outcome	Main	Drop C1	Drop C2	Drop C3	Drop C4	Drop C5	Drop C6
Total sessions	-0.422**	-0.480**	-0.462**	-0.452*	-0.279	-0.620***	0.493
Same counselor	-0.173***	-0.217***	-0.158***	-0.146***	-0.118***	-0.192***	-0.390***
Couns sessions per month	0.021	0.018	0.024	0.006	0.012	0.048**	-0.035
Husband present	-0.093**	-0.128***	-0.103***	-0.156***	-0.047	-0.081**	0.090
Discussed FP info	0.017	0.004	0.012	0.023	0.000	0.025	0.130**
Discussed birth spacing	0.614***	0.771***	0.647***	0.796***	0.509***	0.463***	0.248**
Discussed pregnancy	0.977***	1.074***	1.040***	1.132***	0.909***	0.933***	0.063
Discussed partner engagement	0.770***	1.101***	0.818***	0.803***	0.632***	0.664***	0.083
Discussed child health	0.183***	0.238***	0.238***	0.078	0.215***	0.265***	-0.519***
Discussed breastfeeding	0.344***	0.483***	0.403***	0.233***	0.269***	0.492***	-0.518***
Any FP service	0.132**	0.119*	0.131**	0.160**	0.119**	0.128**	0.167
Any injectable	0.129**	0.119*	0.132**	0.151**	0.115**	0.137***	0.129
Number of FP services	0.689***	0.651**	0.674***	0.815***	0.656***	0.657***	0.791**
Injectable coverage	0.083***	0.079**	0.084***	0.103***	0.078***	0.079***	0.069*
Birth within 12 months	0.025	0.028	0.023	0.029	0.021	0.021	0.040
Birth within 24 months	0.056*	0.065*	0.061*	0.080**	0.050*	0.045	-0.015
Pregnancy within 12 months	-0.003	-0.003	-0.001	0.010	-0.002	-0.010	-0.030
Pregnancy within 24 months	0.036	0.039	0.044	0.057	0.031	0.035	-0.074*
Mistimed pregnancy	0.048	0.053	0.057	0.054	0.053*	0.050	-0.083***

Notes: Two-stage least squares estimates of higher counseling quality on each outcome under the main specification (column 1) and after sequentially dropping each of the six counselors (columns 2–7). Panels A and B (continuity and topics) are estimated at the woman level (one observation per woman); Panels C and D (FP coverage and reproductive) at endline. Endogenous variable: an above-median indicator for the woman's average satisfaction-and-topics quality across all attended sessions. Instrument: the leave-one-out, scheduling-residualized Counselor Quality Tendency (CQT) of the woman's first counselor (recomputed within each leave-one-out subsample). Heteroskedasticity-robust standard errors. Stars: * p<0.10, ** p<0.05, *** p<0.01.

Appendix Table A7. Robustness: Leave-One-Counselor-Out IV Estimates — Postpartum Sample (preg=0)

Outcome	Main	Drop C1	Drop C2	Drop C3	Drop C4	Drop C5	Drop C6
Total sessions	-0.623**	-0.696**	-0.635**	-0.611*	-0.395	-0.934***	0.071
Same counselor	-0.204***	-0.240***	-0.188***	-0.179***	-0.144***	-0.230***	-0.498***
Couns sessions per month	-0.025	-0.019	-0.023	-0.049*	-0.034	0.005	-0.073
Husband present	-0.139***	-0.186***	-0.139***	-0.203***	-0.071	-0.126**	-0.099
Discussed FP info	-0.056	-0.086	-0.055	-0.042	-0.046	-0.061	-0.005
Discussed birth spacing	0.592***	0.715***	0.635***	0.824***	0.520***	0.390***	0.100
Discussed pregnancy	1.060***	1.162***	1.127***	1.250***	0.955***	1.017***	0.047
Discussed partner engagement	0.759***	1.110***	0.808***	0.791***	0.615***	0.627***	0.051
Discussed child health	0.066	0.126	0.131**	-0.050	0.116*	0.151**	-0.917***
Discussed breastfeeding	0.292***	0.495***	0.376***	0.153	0.179**	0.488***	-1.072***
Any FP service	0.194**	0.166	0.200**	0.252**	0.177**	0.199**	0.159
Any injectable	0.102	0.074	0.107	0.119	0.089	0.122	0.141
Number of FP services	0.712**	0.623	0.688*	0.825*	0.693**	0.730**	1.142
Injectable coverage	0.061*	0.049	0.057	0.070	0.061*	0.064*	0.114
Birth within 12 months	0.045	0.051	0.044	0.051	0.039	0.039	0.046
Birth within 24 months	0.144**	0.150**	0.153**	0.197***	0.126**	0.125**	-0.013
Pregnancy within 12 months	0.036	0.030	0.040	0.064	0.028	0.031	-0.032
Pregnancy within 24 months	0.118*	0.127*	0.130*	0.163**	0.103*	0.107*	-0.079
Mistimed pregnancy	0.120**	0.131*	0.132**	0.152**	0.120**	0.107*	-0.101**

Notes: Two-stage least squares estimates of higher counseling quality on each outcome under the main specification (column 1) and after sequentially dropping each of the six counselors (columns 2–7). Panels A and B (continuity and topics) are estimated at the woman level (one observation per woman); Panels C and D (FP coverage and reproductive) at endline. Endogenous variable: an above-median indicator for the woman's average satisfaction-and-topics quality across all attended sessions. Instrument: the leave-one-out, scheduling-residualized Counselor Quality Tendency (CQT) of the woman's first counselor (recomputed within each leave-one-out subsample). Heteroskedasticity-robust standard errors. Stars: * p<0.10, ** p<0.05, *** p<0.01.

Appendix Table A8. Robustness: Leave-One-Counselor-Out IV Estimates — Pregnant Sample (preg=1)

Outcome	Main	Drop C1	Drop C2	Drop C3	Drop C4	Drop C5	Drop C6
Total sessions	-0.253	-0.295	-0.312	-0.320	-0.175	-0.360	0.782**
Same counselor	-0.146***	-0.197***	-0.132***	-0.119***	-0.094***	-0.160***	-0.316***
Couns sessions per month	0.061	0.051	0.066	0.054	0.053	0.086**	-0.009
Husband present	-0.049	-0.072	-0.068	-0.111*	-0.024	-0.038	0.225*
Discussed FP info	0.078	0.081	0.070	0.077	0.042	0.097*	0.223***
Discussed birth spacing	0.633***	0.820***	0.658***	0.772***	0.499***	0.524***	0.350***
Discussed pregnancy	0.906***	0.999***	0.965***	1.034***	0.867***	0.864***	0.074
Discussed partner engagement	0.779***	1.093***	0.827***	0.813***	0.647***	0.695***	0.104
Discussed child health	0.282***	0.334***	0.330***	0.185***	0.304***	0.359***	-0.248*
Discussed breastfeeding	0.387***	0.473***	0.426***	0.299***	0.349***	0.495***	-0.139
Any FP service	0.072	0.075	0.062	0.072	0.062	0.058	0.172
Any injectable	0.155**	0.162*	0.157**	0.182**	0.140**	0.151**	0.121
Number of FP services	0.667**	0.678*	0.661**	0.805**	0.620**	0.585*	0.564
Injectable coverage	0.106**	0.109**	0.111**	0.137***	0.096**	0.095**	0.037
Birth within 12 months	0.005	0.005	0.002	0.005	0.003	0.003	0.035
Birth within 24 months	-0.033*	-0.024	-0.036**	-0.042*	-0.028*	-0.040**	-0.016
Pregnancy within 12 months	-0.043**	-0.037*	-0.044**	-0.047**	-0.032*	-0.054**	-0.029
Pregnancy within 24 months	-0.048	-0.052	-0.047	-0.053	-0.043	-0.041	-0.071
Mistimed pregnancy	-0.026	-0.028	-0.022	-0.049	-0.016	-0.010	-0.071**

Notes: Two-stage least squares estimates of higher counseling quality on each outcome under the main specification (column 1) and after sequentially dropping each of the six counselors (columns 2–7). Panels A and B (continuity and topics) are estimated at the woman level (one observation per woman); Panels C and D (FP coverage and reproductive) at endline. Endogenous variable: an above-median indicator for the woman's average satisfaction-and-topics quality across all attended sessions. Instrument: the leave-one-out, scheduling-residualized Counselor Quality Tendency (CQT) of the woman's first counselor (recomputed within each leave-one-out subsample). Heteroskedasticity-robust standard errors. Stars: * p<0.10, ** p<0.05, *** p<0.01.

Appendix Table A9. Balance: CQT Instrument and Ancillary-Component Take-Up

Ancillary component	CQT coef.	Robust SE	p-value	Mean	N
Transport voucher	0.090*	(0.048)	0.060	0.247	724
Clinic reimbursement	0.098**	(0.048)	0.041	0.243	724
Any ancillary component	0.098**	(0.048)	0.043	0.251	724

Notes: Each row reports an OLS regression of an ancillary-component take-up indicator on the range-normalized Counselor Quality Tendency (CQT) instrument (satisfaction-and-topics specification) and the full baseline covariate set X_i, estimated on the endline treatment sample (year 2018). The CQT coefficient is the change in take-up associated with moving from the lowest to the highest counselor-quality tendency. *Transport voucher* and *Clinic reimbursement* indicate any use of the free-taxi and clinic-reimbursement components respectively; *Any ancillary component* is their union. Take-up of these components is a *post-treatment* outcome that may itself be shifted by counseling quality, so it is not conditioned on in the main specification (Section 3.2.7); Appendix Tables A10 and A11 show the IV estimates are robust to controlling for it. Heteroskedasticity-robust (HC3) standard errors in parentheses. Stars: * p<0.10, ** p<0.05, *** p<0.01.

Appendix Table A10. Robustness: Take-Up Controls — Mechanism Outcomes (Continuity and Topics)

	Panel A: Counseling continuity				*Panel B: Counseling topics (post-1st-session)*					
	Total sessions	Same counselor	Couns sessions per month	Husband present	FP information	Birth spacing	Pregnancy info	Partner engagement	Child health	Breastfeeding
Pooled										
Estimate	-0.487**	-0.171***	0.020	-0.095***	0.005	0.609***	0.987***	0.776***	0.175***	0.344***
	(0.199)	(0.022)	(0.023)	(0.037)	(0.040)	(0.051)	(0.052)	(0.055)	(0.046)	(0.052)
Postpartum (preg=0)										
Estimate	-0.703**	-0.203***	-0.026	-0.141***	-0.070	0.586***	1.073***	0.767***	0.056	0.293***
	(0.294)	(0.035)	(0.023)	(0.051)	(0.059)	(0.075)	(0.080)	(0.085)	(0.067)	(0.082)
Pregnant (preg=1)										
Estimate	-0.303	-0.145***	0.060	-0.051	0.069	0.629***	0.914***	0.784***	0.276***	0.388***
	(0.269)	(0.028)	(0.040)	(0.054)	(0.053)	(0.067)	(0.067)	(0.071)	(0.063)	(0.065)
p(PP=Preg)	0.315	0.199	0.071	0.227	0.080	0.663	0.128	0.875	0.016	0.361
N	896	896	811	781	896	896	896	896	896	896
Mean DV	4.643	0.865	0.382	0.133	0.873	0.483	0.204	0.328	0.688	0.543

Notes: Robustness check that re-estimates every main IV coefficient with an additional binary control for whether the woman received any ancillary intervention component (a transport voucher or clinic-visit reimbursement). Endogenous variable: an above-median indicator for the woman's average satisfaction-and-topics quality across all attended sessions. Instrument: the leave-one-out, scheduling-residualized Counselor Quality Tendency (CQT) of the woman's first counselor. Heteroskedasticity-robust standard errors in parentheses. The row *p(PP=Preg)* reports the p-value for the test of equality of the postpartum and pregnant subgroup effects. Stars: * p<0.10, ** p<0.05, *** p<0.01.

Appendix Table A11. Robustness: Take-Up Controls — FP Coverage and Reproductive Outcomes

	Panel C: FP coverage					*Panel D: Reproductive*			
	Any FP service	Any injectable	Number of FP services	Injectable coverage	Husband knows FP	Birth within 24 months	Pregnancy within 12 months	Pregnancy within 24 months	Mistimed pregnancy
Pooled									
Estimate	0.004	0.052	0.398**	0.058**	-0.029	0.059*	-0.001	0.041	0.051
	(0.013)	(0.036)	(0.166)	(0.024)	(0.025)	(0.032)	(0.022)	(0.037)	(0.034)
Postpartum (preg=0)									
Estimate	0.005	-0.014	0.280	0.023	-0.084*	0.149**	0.038	0.125*	0.125**
	(0.025)	(0.058)	(0.258)	(0.030)	(0.046)	(0.059)	(0.039)	(0.065)	(0.061)
Pregnant (preg=1)									
Estimate	0.004	0.114***	0.511**	0.093***	0.030	-0.031*	-0.042**	-0.046	-0.024
	(0.004)	(0.042)	(0.206)	(0.035)	(0.018)	(0.017)	(0.019)	(0.034)	(0.031)
p(PP=Preg)	0.958	0.073	0.481	0.125	0.021	0.004	0.064	0.020	0.030
N	724	724	724	701	591	701	701	701	702
Mean DV	0.230	0.141	0.536	0.051	0.983	0.028	0.024	0.053	0.033

Notes: Robustness check that re-estimates every main IV coefficient with an additional binary control for whether the woman received any ancillary intervention component (a transport voucher or clinic-visit reimbursement). Endogenous variable: an above-median indicator for the woman's average satisfaction-and-topics quality across all attended sessions. Instrument: the leave-one-out, scheduling-residualized Counselor Quality Tendency (CQT) of the woman's first counselor. Heteroskedasticity-robust standard errors in parentheses. The row *p(PP=Preg)* reports the p-value for the test of equality of the postpartum and pregnant subgroup effects. Stars: * p<0.10, ** p<0.05, *** p<0.01.

Appendix Table A12. Per-Panel Summary-Index IV Estimates

	Panel A Continuity	Panel B Topics	Panel C FP coverage	Panel D Reproductive
Pooled				
Estimate	-0.321***	1.070***	0.312**	0.177
	(0.060)	(0.092)	(0.129)	(0.148)
Postpartum (preg=0)				
Estimate	-0.520***	0.998***	0.244	0.588**
	(0.076)	(0.139)	(0.187)	(0.266)
Pregnant (preg=1)				
Estimate	-0.153*	1.132***	0.377**	-0.245**
	(0.091)	(0.122)	(0.174)	(0.123)
p(PP=Preg)	0.002	0.469	0.597	0.005
N	896	896	724	702

Notes: Each column reports 2SLS estimates of higher counseling quality on a single summary index for the panel, constructed as the equal-weighted mean of the standardized (mean 0, SD 1) outcomes in that panel, following Kling et al. (2007) and Anderson (2008). Indices are in standard-deviation units. Aggregating each panel into one index controls the family-wise error rate across the nineteen outcomes and gains power by averaging out idiosyncratic noise. Endogenous variable, instrument, covariates, and samples are identical to Table 5 (Panels A and B at the woman level; Panels C and D at endline). Within Panel C the four FP-service outcomes are strongly positively correlated ($r = 0.55$ to 0.93), while husband's awareness of FP use is near-orthogonal; Panel A bundles less-coherent continuity components that are aligned by the direction of the treatment effect rather than by strong mutual correlation. Heteroskedasticity-robust standard errors in parentheses. Stars: * p<0.10, ** p<0.05, *** p<0.01.

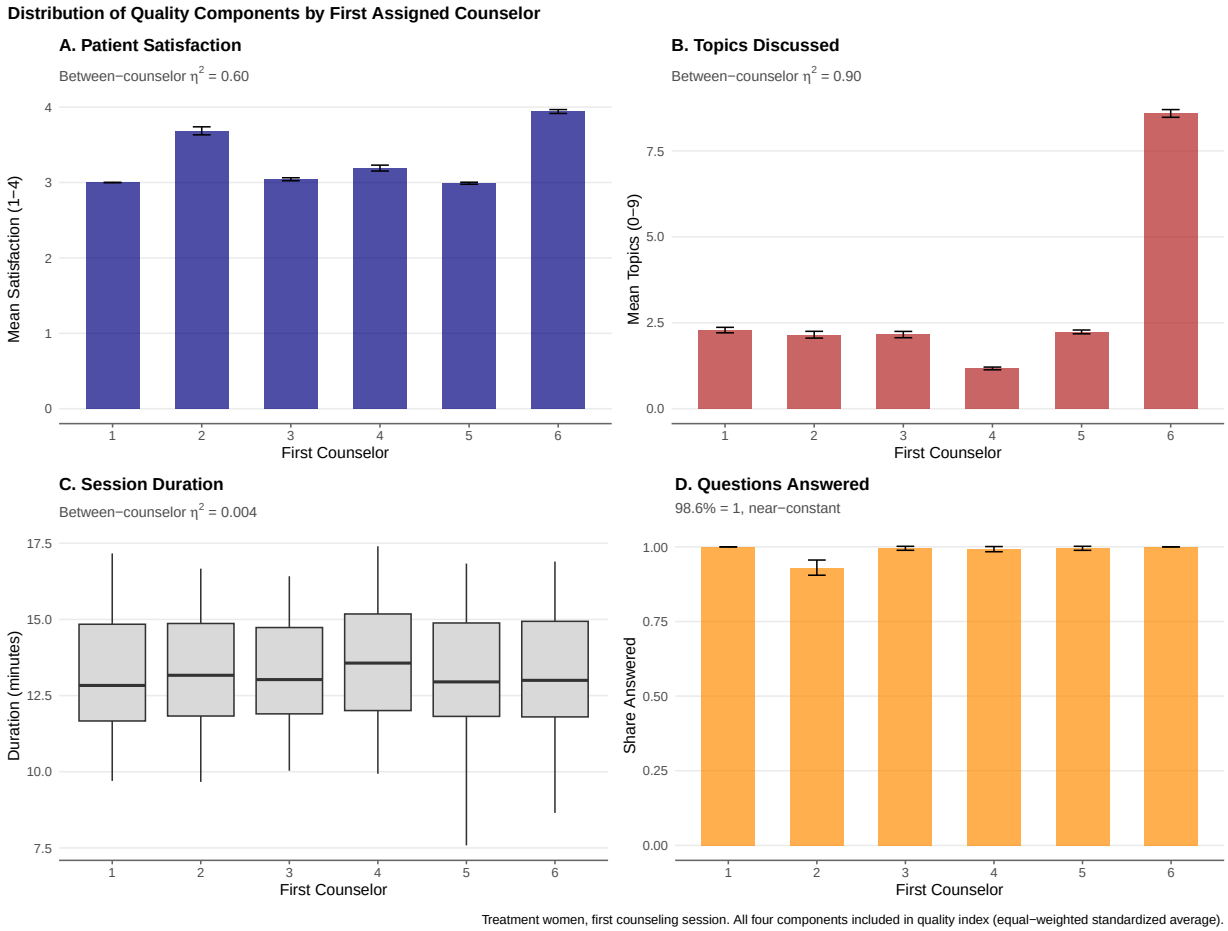
Appendix Table A13. Wild-Cluster-Bootstrap Inference (Clustering on the Six Counselors)

	Pooled	Postpartum	Pregnant
	Coefficient (WCB p-value)		
Panel A: Counseling continuity			
Same counselor	-0.173 (0.119)	-0.204 (0.099*)	-0.146 (0.160)
Total sessions	-0.422 (0.597)	-0.623 (0.431)	-0.253 (0.806)
Sessions per month	0.021 (0.588)	-0.025 (0.550)	0.061 (0.282)
Husband present	-0.093 (0.608)	-0.139 (0.453)	-0.049 (0.769)
Panel B: Counseling topics			
FP information	0.017 (0.483)	-0.056 (0.436)	0.078 (0.111)
Birth spacing	0.614 (0.491)	0.592 (0.579)	0.633 (0.456)
Pregnancy info	0.977 (0.575)	1.060 (0.222)	0.906 (0.263)
Partner engagement	0.770 (0.483)	0.759 (0.568)	0.779 (0.585)
Child health	0.183 (0.652)	0.066 (0.829)	0.282 (0.543)
Breastfeeding	0.344 (0.606)	0.292 (0.709)	0.387 (0.487)
Panel C: FP coverage			
Any FP service	0.132 (0.046**)	0.194 (0.142)	0.072 (0.022**)
Any injectable	0.129 (0.029**)	0.102 (0.140)	0.155 (0.037**)
Number of FP services	0.689 (0.031**)	0.712 (0.026**)	0.667 (0.058*)
Injectable coverage	0.083 (0.058*)	0.061 (0.061*)	0.106 (0.222)
Husband knows FP	-0.027 (0.260)	-0.082 (0.228)	0.031 (0.446)
Panel D: Reproductive			
Birth within 24 months	0.056 (0.630)	0.144 (0.607)	-0.033 (0.248)
Pregnancy within 12 months	-0.003 (0.767)	0.036 (0.581)	-0.043 (0.187)
Pregnancy within 24 months	0.036 (0.797)	0.118 (0.728)	-0.048 (0.034**)
Mistimed pregnancy	0.048 (0.681)	0.120 (0.683)	-0.026 (0.333)

Notes: Each cell reports the 2SLS coefficient (identical to Table 5) and, in parentheses, the wild-cluster-bootstrap p-value for the null that the coefficient is zero. Bootstrap clustering is on the six counselors using Webb six-point weights and 9,999 replications (Cameron et al., 2008; MacKinnon and Webb, 2017); subgroup effects are recovered as direct coefficients from a fully-interacted reparameterization. Because the Counselor Quality Tendency instrument varies essentially at the counselor level, with only six clusters the wild-cluster bootstrap is conservative and cannot separately distinguish the counselor-level mechanism channels (Panels A and B) from cluster-level noise, even where the point estimates are large; the family-planning take-up effects in Panel C nonetheless remain significant under this most conservative small-sample procedure. We therefore treat the wild-cluster bootstrap as a complement to, not a substitute for, the woman-level robust standard errors, the leave-one-counselor-out exercise, and the summary-index aggregation, on which our inference primarily rests. Stars: * p<0.10, ** p<0.05, *** p<0.01.

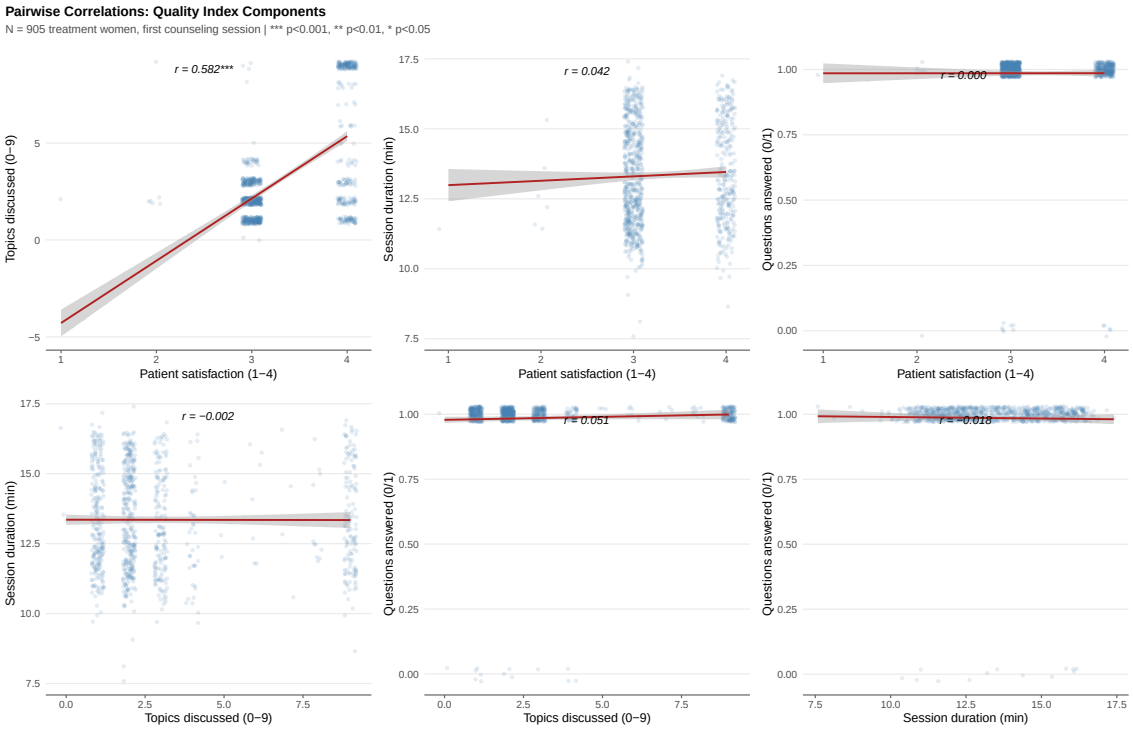
Appendix B: Figures

Figure B1. Distribution of Quality Components by First Assigned Counselor



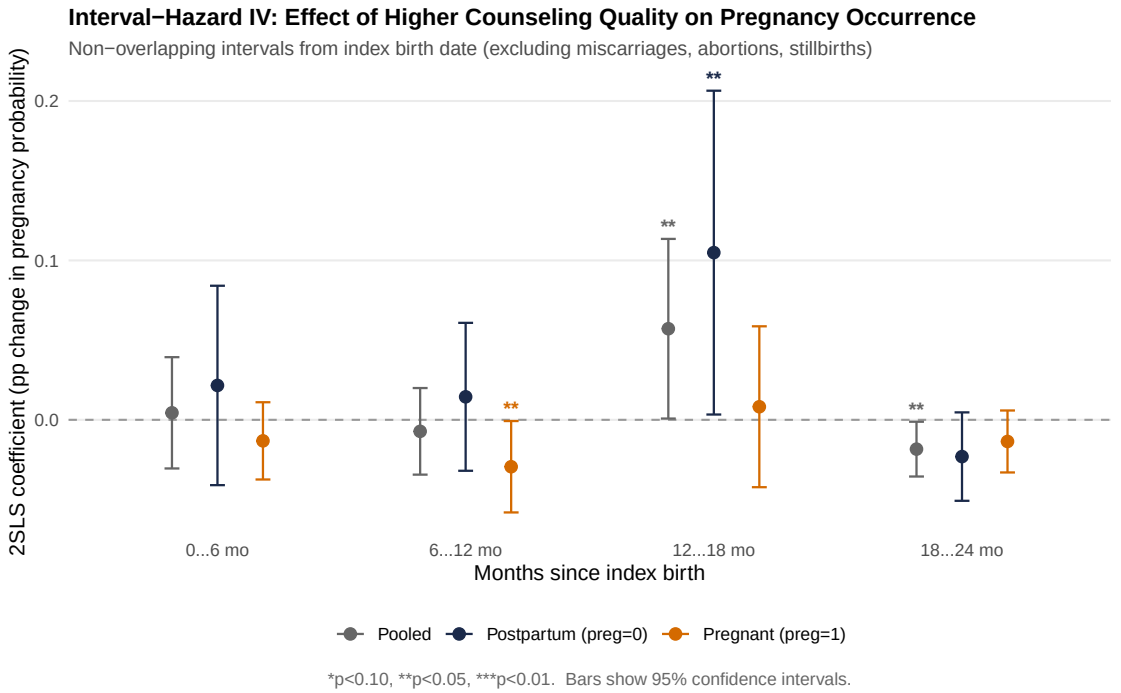
Notes: Distribution of the four candidate first-session quality components by first assigned counselor. Panel A: patient satisfaction (1–4 ordinal). Panel B: number of FP topics requested (0–9). Panel C: session duration in minutes (boxplot). Panel D: whether all questions were answered (binary). Between-counselor η^2 values reported in each subtitle.

Figure B2. Pairwise Correlations of Quality Components at the First Counseling Session



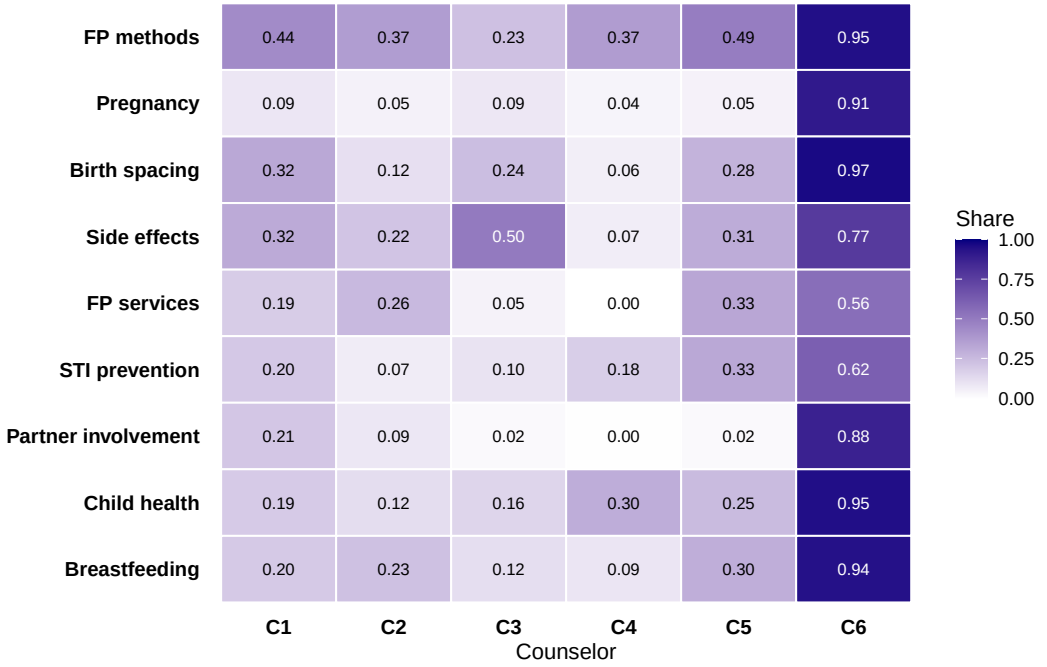
Notes: Pairwise Pearson correlations among the four candidate first-session quality components. Satisfaction and topics: $r = 0.58$, $p < 0.001$; all other pairwise correlations $|r| < 0.06$.

Figure B3. Interval-Hazard IV: Effect of Higher Counseling Quality on Pregnancy Occurrence by Months since Index Birth



Notes: 2SLS effect of higher counseling quality on the probability of a pregnancy event in each non-overlapping six-month interval from the index birth, for the pooled sample and both pregnancy-status subgroups (postpartum in navy, pregnant in orange). Bars show 95 percent confidence intervals; stars indicate two-sided significance at the 10, 5, and 1 percent levels.

Figure B4. Topic-Coverage Concentration by Counselor



Notes: Each cell reports the share of the counselor's sessions in which the topic was discussed, across all 3,898 counseling sessions; the cell values are the by-counselor shares in Panel A of Appendix Table A5. Counselors are labeled by their actual identifiers. Counselor 6, who has the highest counseling-quality tendency, covers nearly every topic in nearly every session, while the lower-quality counselors cover a sparser and more idiosyncratic set of topics. This concentration is the systematic between-counselor variation in content that the examiner-design instrument leverages.

References

Afulani, P. A., Nakphong, M. K., and Sudhinaraset, M. (2023). Person-centred sexual and reproductive health: A call for standardized measurement. *Health Expectations*, 26(4):1384–1390.

Ali, M. M., Cleland, J., and Shah, I. H. (2012). Causes and consequences of contraceptive discontinuation: Evidence from 60 demographic and health surveys. Technical report, World Health Organization, Geneva.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association*, 103(484):1481–1495.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Princeton, NJ.

Ashraf, N., Field, E., and Lee, J. (2014). Household bargaining and excess fertility: An experimental study in Zambia. *American Economic Review*, 104(7):2210–2237.

Athey, S., Bergstrom, K., Hadad, V., Jamison, J. C., Özler, B., Parisotto, L., and Sama, J. D. (2023). Can personalized digital counseling improve consumer search for modern contraceptive methods? *Science Advances*, 9:eadg4420.

Bau, N., Henning, D. J., Low, C., and Steinberg, B. (2024). Family planning, now and later: Infertility fears and contraceptive take-up. NBER Working Paper 32735, National Bureau of Economic Research.

Bruce, J. (1990). Fundamental elements of the quality of care: A simple framework. *Studies in Family Planning*, 21(2):61–91.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, 90(3):414–427.

Castle, S. and Askew, I. (2015). Contraceptive discontinuation: Reasons, challenges, and solutions. Technical report, Population Council and FP2020, New York.

Cavallaro, F. L., Benova, L., Owolabi, O. O., and Ali, M. (2020). A systematic review of the effectiveness of counselling strategies for modern contraceptive methods. *BMJ Sexual & Reproductive Health*, 46(4):254–269.

Chan, D. C., Gentzkow, M., and Yu, C. (2022). Selection with variation in diagnostic skill: Evidence from radiologists. *Quarterly Journal of Economics*, 137(2):729–783.

Chang, K. T., Mukanu, M., Bellows, B., Hameed, W., Kalamar, A. M., Grépin, K. A., Gul, X., and Chakraborty, N. M. (2019). Evaluating quality of contraceptive counseling: An analysis of the Method Information Index. *Studies in Family Planning*, 50(1):25–42.

Chyn, E., Frandsen, B., and Leslie, E. (2025). Examiner and judge designs in economics: A practitioner's guide. *Journal of Economic Literature*, 63(2):401–439.

Cleland, J. et al. (2012). Contraception and health. *The Lancet*, 380(9837):149–156.

Dahl, G. B., Kostol, A. R., and Mogstad, M. (2014). Family welfare cultures. *Quarterly Journal of Economics*, 129(4):1711–1752.

Dehlendorf, C., Henderson, J. T., Vittinghoff, E., Steinauer, J., and Hessler, D. (2018). Development of a patient-reported measure of the interpersonal quality of family planning care. *Contraception*, 97(1):34–40.

Delavande, A. (2008). Pill, patch, or shot? subjective expectations and birth control choice. *International Economic Review*, 49(3):999–1042.

Dieye, I., Wickramanayake, A., Akinyi, S., Cheruiyot, P., Ogola, T., Sitrin, D., Pfitzer, A., McConnell, M., Cohen, J., and Rokicki, S. (2026). Can remote personalized digital counseling improve postpartum contraceptive use? Working paper, preliminary.

Dobbie, W., Goldin, J., and Yang, C. S. (2018). The effects of pre-trial detention on conviction, future crime, and employment. *American Economic Review*, 108(2):201–240.

Doyle, J. J. (2007). Child protection and child outcomes: Measuring the effects of foster care. *American Economic Review*, 97(5):1583–1610.

Dupas, P. (2011). Do teenagers respond to HIV risk information? Evidence from a field experiment in Kenya. *American Economic Journal: Applied Economics*, 3(1):1–34.

Dupas, P., Jayachandran, S., Lleras-Muney, A., and Rossi, P. (2025). The negligible effect of free contraception on fertility: Experimental evidence from Burkina Faso. *American Economic Review*, 115(8):2659–2688.

Frandsen, B. R., Lefgren, L. J., and Leslie, E. C. (2023). Judging judge fixed effects. *American Economic Review*, 113(1):253–277.

Goldsmith-Pinkham, P., Hull, P., and Kolesár, M. (2025). Leniency designs: An operator's manual. NBER Working Paper 34473, National Bureau of Economic Research. R&R, Journal of Economic Perspectives.

Holt, K. et al. (2019). Measurement of perceived quality of care: Developing a new scale for family planning. *Studies in Family Planning*, 50(3):279–294.

Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475.

Jain, A. K. and Hardee, K. (2018). Revising the FP quality of care framework in the context of rights-based family planning. *Studies in Family Planning*, 49(2):171–179.

Karra, M. et al. (2022). The causal effect of a family planning intervention on women's contraceptive use and birth spacing. *Proceedings of the National Academy of Sciences*, 119(22):e2200279119.

Kling, J. R. (2006). Incarceration length, employment, and earnings. *American Economic Review*, 96(3):863–876.

Kling, J. R., Liebman, J. B., and Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica*, 75(1):83–119.

Kolesár, M. (2013). Estimation in an instrumental variables model with treatment effect heterogeneity. Working Paper 2013-2, Princeton University, Department of Economics.

Kruk, M. E., Gage, A. D., Arsenault, C., Jordan, K., Leslie, H. H., Roder-DeWan, S., Adeyi, O., Barker, P., Daelmans, B., Doubova, S. V., English, M., García-Elorrio, E., Guanais, F., Gureje, O., Hirschhorn, L. R., Jiang, L., Kelley, E., Lemango, E. T., Liljestrand, J., Malata, A., Marchant, T., Matsoso, M. P., Meara, J. G., Mohanan, M., Ndiaye, Y., Norheim, O. F., Reddy, K. S., Rowe, A. K., Salomon, J. A., Thapa, G., Twum-Danso, N. A. Y., and Pate, M. (2018). High-quality health systems in the Sustainable Development Goals era: time for a revolution. *The Lancet Global Health*, 6(11):e1196–e1252.

MacKinnon, J. G. and Webb, M. D. (2017). Wild bootstrap inference for wildly different cluster sizes. *Journal of Applied Econometrics*, 32(2):233–254.

Maggio, D., Karra, M., and Canning, D. (2024). Family planning and children's human capital: Experimental evidence from urban Malawi. *Demography*, 61(5):1667–1698.

Miller, G., de Paula, Á., and Valente, C. (2025). Subjective expectations and demand for contraception. *Journal of Econometrics*, 249:105997.

National Statistical Office [Malawi] and ICF (2017). Malawi demographic and health survey 2015–16.

Pasha, O., Goudar, S. S., Patel, A., Garces, A., Esamai, F., Chomba, E., Moore, J. L., Kodkany, B. S., Saleem, S., Derman, R. J., Liechty, E. A., Hibberd, P. L., Hambidge, K. M., Krebs, N. F., Carlo, W. A., McClure, E. M., Koso-Thomas, M., and Goldenberg, R. L. (2015). Postpartum contraceptive use and unmet need for family planning in five low-income countries. *Reproductive Health*, 12(Suppl 2):S11.

Sudhinaraset, M., Afulani, P. A., Diamond-Smith, N., Golub, G., and Srivastava, A. (2018). Development of a person-centered family planning scale in India and Kenya. *Studies in Family Planning*, 49(3):237–258.

Tully, K. P., Stuebe, A. M., and Verbiest, S. B. (2017). The fourth trimester: A critical transition period with unmet maternal health needs. *American Journal of Obstetrics and Gynecology*, 217(1):37–41.

Wagner, Z., Moucheraud, C., Shah, M., Wollum, A., Friedman, W., and Dow, W. H. (2025). Reducing bias among health care providers: Experimental evidence from Tanzania, Burkina Faso and Pakistan. *The Economic Journal*, 135(670):1891–1922.