Family Planning and Unintended Pregnancy: Experimental Evidence from Urban Malawi

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

September 18, 2026

PRELIMINARY - DO NOT CITE

Abstract

Family planning (FP) programs are widely promoted on the premise that they reduce unintended pregnancy, yet direct experimental evidence on this link is scarce, and the field has not converged on how unintended pregnancy should be measured. We use longitudinal data from a randomized FP trial in urban Malawi that collected both a prospective (ex-ante) and a retrospective (ex-post) measure of pregnancy intention in the same sample, along with detailed birth timing preferences. We show that a null treatment effect on a binary ex-ante measure of *unwanted* pregnancy - which flags only women who wanted no more children - reflects that narrower construct rather than an absence of intervention impact. Extending the ex-ante measure to also capture mistimed pregnancies over a 24-month horizon yields a treatment effect of 3.8 percentage points (p.p.), or a 44% relative reduction, that closely reproduces an ex-post retrospective estimate of unintended pregnancy (3.7 p.p.), and concordance between the two measures rises accordingly. An otherwise identical timing-inclusive measure that is built on a 12-month horizon reproduces the null result, which we attribute to low fecundability at 12 months among our immediate postpartum and pregnant-at-baseline sample rather than to a flaw in the timing-inclusive approach itself. These results offer causal estimates of FP's effect on unintended pregnancy specifically and indicate that reconciling ex-ante and ex-post approaches to measuring it depends both on incorporating a timing dimension and on allowing a follow-up horizon long enough to observe pregnancy risk in the population of interest.

Keywords: family planning; experiment; unintended pregnancy; unwanted birth; Malawi

[a]**Corresponding Author**. Mailman School of Public Health, Columbia University. mvk2109@cumc.columbia.edu
[b]Harvard T.H. Chan School of Public Health. mariemontas@fas.harvard.edu
[c]Mailman School of Public Health, Columbia University, New York, USA. sv2890@cumc.columbia.edu

1 Introduction

Family planning (FP) programs are widely promoted on the premise that expanding access to contraception and other FP services reduces unintended pregnancies and unwanted births (Starbird et al., 2016; UNFPA, 2022, 2014), a premise that has motivated global development agendas such as Sustainable Development Goal 3.7 (United Nations, 2018). The link between FP access and contraceptive use is well supported empirically.[1] The link between FP access and unintended pregnancy itself, however, has received far less direct testing. A meta analysis review of 38 FP counseling and service provision trials found that most studies documented effects on contraceptive use, but relatively few evaluated unintended pregnancy directly, and no meta analysis of that outcome was possible because trials defined pregnancy intention inconsistently (Nelson et al., 2022; Heil et al., 2021; Shlay et al., 2003; Stevens et al., 2017). Reviews of peri-abortion counseling (Gonzales-Huaman et al., 2021) and of contraceptive use more broadly (Tsui et al., 2010) reach similarly inconclusive verdicts further along the causal pathway. The chain from FP access to contraceptive use to reduced unintended pregnancy that motivates most policy discussion therefore remains considerably less tested at its later links than at its first.

A second source of ambiguity concerns measurement. *Ex-ante* measures elicit fertility preferences prospectively and classify a subsequent pregnancy as unintended if it violates those preferences; *ex-post* measures instead ask retrospectively whether an already-experienced pregnancy was wanted "at that time." The two are not interchangeable. Ex-ante measures are vulnerable to preference instability and, in practice, may omit the timing or spacing dimension of intention unless explicitly asked (Mueller et al., 2022), while ex-post measures are vulnerable to rationalization bias, recall error, and social desirability bias once the pregnancy's outcome is already known (Yeatman et al., 2013; Casterline and El-Zeini, 2007; Trinitapoli and Yeatman, 2018). The literature has not converged on which approach is preferable,[2] in part because the two measures are rarely fielded to the same women in the same study, so disagreements across studies may reflect differences in the population of interest or design as much as a genuine difference between measurement approaches.

This study addresses both gaps using experimental data from an RCT in urban Malawi. First, it provides causal estimates of an FP intervention's effect on unintended pregnancy itself, rather than on contraceptive use or pregnancy occurrence alone. Second, because the

[1] See Mwaikambo et al. (2011); Debpuur et al. (2002); Desai and Tarozzi (2011); Karra et al. (2022), among others.

[2] Rackin and Morgan (2018) make a modest, practical case for retrospective over prospective measurement, while acknowledging that both face real limitations, particularly when intentions are held with uncertainty (Bachrach and Morgan, 2013) or disrupted by events such as a child's death (Smith-Greenaway and Sennott, 2016).

endline survey collected both ex-ante and ex-post items on pregnancy intention in the same sample, along with birth timing preferences, it offers a within-study opportunity to compare the two approaches directly.

2 Data and Experimental Design

We use longitudinal data from the Malawi Family Planning Study (MFPS), a randomized controlled trial conducted in Lilongwe, Malawi, between September 2016 and February 2019.[3] The MFPS evaluated a multicomponent FP intervention addressing key barriers to contraceptive access, including knowledge gaps and constraints related to distance, service availability, and cost, among a cohort of married women who were pregnant or immediate postpartum at baseline. Women were interviewed annually over three survey waves.

Following baseline, women were individually randomized to treatment or control arms. The treatment arm received: 1) up to six private, home-based FP counseling visits with trained counselors; 2) free, private transportation to a high-quality clinic with reliable supply and short wait times; and 3) full coverage of costs for FP services received at that clinic or elsewhere, including methods, consultations, tests, and treatment for contraceptive-related side effects and contraindications. The control arm received publicly available information on FP methods and nearby clinics and was not otherwise recontacted until follow-up.

The MFPS has been the basis for several analyses of the intervention's first-stage effects (Karra et al., 2022; Maggio et al., 2024): two years after exposure, treated women were 5.9 percentage points (p.p.) more likely to use any contraceptive method, which was driven by increased adoption of long-acting reversible methods (particularly the implant), and 43 percent less likely to have a subsequent pregnancy within 24 months of their index birth. Pechar et al. (2026) further show that the intervention reduced within-woman variability in fertility preferences over time (without changing desired family size), particularly among women who were pregnant at baseline, and this stability was associated with higher contraceptive use. This combination of improved access and more stable, actionable fertility goals makes the MFPS well suited to examining how relaxing FP constraints shapes pregnancy intentions.

2.1 Eligibility, Sample Size and Randomization

Eligible women were married; currently pregnant or within six months postpartum at baseline; aged 18 to 35; permanent residents of Lilongwe; and had not been sterilized or undergone

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

a hysterectomy. Where multiple eligible women were present in a household, the youngest was selected. Enrolled women resided at least five households apart to limit spillovers. A total of 2,143 women were enrolled in 2016 and were randomized using stratified covariate-balanced randomization on parity, current contraceptive use, age at sexual debut, education, work status, and neighborhood, yielding 1,027 women assigned to the treatment arm and 1,113 women assigned to the control arm.[4] Table A1 shows the two arms are balanced on observable baseline characteristics, with only a minor difference in prior contraceptive use; a joint significance test similarly supports no systematic baseline differences across arms. Additional detail on the randomization protocol and covariate balance is presented in Karra and Canning (2020).

Of the 2,143 women enrolled, 1,672 (78 percent) were re-interviewed at endline. Maggio et al. (2024) and Pechar et al. (2026) find attrited and retained women to be statistically indistinguishable by treatment arm, supporting the internal validity of the experimental comparison, although attrition is related to some baseline characteristics. Specifically, women lost to follow-up were, on average, younger, less educated, of lower parity, and somewhat more likely to want additional children. We therefore interpret our estimates as representative of the retained sample.

2.2 Key Outcomes

The 2018 endline survey supports three binary outcome constructions, together yielding the four outcome measures of interest. Two of these constructions, the ex-post measure and the two timing-inclusive ex-ante measures, capture unintended pregnancy as conventionally defined: a pregnancy that was either unwanted (no children, or no more children, desired) or mistimed (wanted, but sooner than desired) (Santelli et al., 2003). The third construction captures only the unwanted component and is better understood as a measure of ex-ante *unwanted* pregnancy specifically. All four measures are defined only over pregnancies occurring after baseline, excluding the index pregnancy that qualified a woman for enrollment.

Ex-post unintended pregnancy. This measure asks, after the fact, whether a pregnancy or birth experienced during follow-up was wanted "at that time," combining a completed-birth item ("When you got pregnant with [last child], did you want to get pregnant at that time?") with a parallel item for an ongoing pregnancy. A woman is coded 1 if she reports a birth or ongoing pregnancy she did not want at the time, and 0 otherwise (including no follow-up pregnancy). Women with a follow-up pregnancy but no clear wantedness response are coded 0.

[4]Three women withdrew from the study following completion of the baseline and prior to randomization, resulting in a total of 2,140 women who were randomized into intervention and control arms.

Ex-ante unwanted pregnancy. This prospective measure flags women who reported at baseline that they wanted no more children, and codes them 1 if they had any pregnancy during follow-up. Because it only ever flags women who wanted no more children *at all*, it cannot register a pregnancy that violated a spacing rather than a limiting preference, operationalizing only half of the standard unintended pregnancy definition (Santelli et al., 2003).

Ex-ante unintended pregnancy, timing-inclusive (12- and 24-month cutoffs). These two measures extend the above unwanted pregnancy measure to also capture mistiming. Among women who wanted another child at baseline, a companion item asked desired waiting time in months, yielding a three-way baseline preference: 1) wants no more children; 2) wants another child, but not soon; or 3) wants another child soon (within a stated cutoff). A woman is "at risk" of unwanted or mistimed pregnancy under a cutoff if she falls in either of the first two categories, and is coded 1 under that cutoff if she also became pregnant within the corresponding follow-up window. Two cutoffs are reported, each paired with a follow-up window of matching length so that preference horizon and observation window align: 12 months (to the 2017 midline) and 24 months (to the 2018 endline, matching the intervention duration and the WHO birth spacing benchmark upon which the MFPS intervention was designed (Karra et al., 2020)). Women who wanted another child but gave no timing answer are left missing rather than defaulted to either category.

2.3 Covariates

The covariate-adjusted specification (model 2 in Table A3 and Table A4) conditions on nine baseline characteristics: neighborhood, age group, age at sexual debut, number of children alive, ever use of FP, education (primary or lower vs. secondary or higher), work status, religion (Christian vs. other), and ethnicity (Chewa vs. other). These are the same characteristics reported in Table A1, with one difference: the balance table reports age continuously (mean 24.9 years) for balance checking, while the regressions use a categorical age-group version of the same information.

3 Empirical Strategy

Our main results present an Intent-to-Treat (ITT) analysis, identifying the average treatment effect of assignment to treatment relative to control:

$$Y_{ia} = \alpha + \beta_T T_{ia} + \mathbf{X}_{ia}^{0}\gamma + \delta_a + \varepsilon_{ia} \tag{1}$$

Y_{ia} is the outcome of interest for woman i in enumeration area a (the ex-ante unwanted, ex-ante unintended, and ex-post pregnancy measures); T_{ia} indicates treatment assignment, with β_T our parameter of interest; $\mathbf{X}_{ia}^0$ is a vector of baseline covariates, including randomization strata; δ_a is a neighborhood fixed effect; and ε_{ia} is the error term. We report β_T from a naïve model (treatment only) and from the fully adjusted model above, with heteroskedasticity-robust standard errors throughout.

We test whether the ITT effect is modified by three baseline characteristics - pregnancy status, desire for another child, and current contraceptive use - via stratified subgroup regressions (rather than a pooled interaction model), for each moderator and each outcome definition, with and without the full covariate set. We test equality of subgroup-specific treatment coefficients using seemingly unrelated estimation.

4 Results

4.1 Sample and Baseline Balance

The analytic sample comprises the 1,672 women who were re-interviewed at endline: 890 control, 782 treatment. Table A1 compares 11 baseline characteristics across arms. Ten of 11 normalized differences fall below the conventional 0.10 imbalance threshold (Austin, 2009); the exception is baseline ever use of FP, with a normalized difference of -0.130 (74.6% control vs. 80.1% treatment), which we address by conditioning on ever use in the adjusted specification given its plausible relationship to future unintended pregnancy risk. The table's joint F-test ($p = 0.236$) indicates overall balance, so we do not treat this single imbalance as a threat to randomization, but we retain the covariate adjustment as a precaution.

4.2 Unadjusted Comparisons

Table A2 reports unadjusted differences in means for the four outcome measures. The timing-inclusive rows use a slightly smaller sample (1,657: 879 control, 778 treatment) than the ex-ante unwanted-pregnancy row (1,672), reflecting 15 women who wanted another child but gave no definite timing answer. The ex-post measure shows a large gap (7.6% control vs. 4.0% treatment; normalized difference of 0.156); the ex-ante unwanted pregnancy measure shows a much smaller one (2.9% control vs. 2.6% treatment; normalized difference of 0.022), about one-seventh the size. The two timing-inclusive measures diverge once each is paired with its matching follow-up window: the 24-month cutoff sits close to the ex-post gap (8.6% control vs. 4.9% treatment; normalized difference of 0.149), while the 12-month cutoff sits

close to the ex-ante unwanted-pregnancy gap instead (3.4% control vs. 2.4% treatment; normalized difference of 0.057).

4.3 Intention-to-Treat Effects

Table A3 regresses each outcome on treatment, without (model 1) and with (model 2) baseline covariates. In Panel A (unwanted pregnancy), the treatment effect is small and not significant (-0.004) against a control mean of 0.029. Under the ex-post definition (Panel B), the estimated effect of -0.037 against a control mean of 0.076 implies that the intervention reduced reported ex-post unintended pregnancy by 48.7% (47.4% with covariates) in the treatment arm relative to the control arm. At the 24-month cutoff (Panel D), the estimate closely reproduces the ex-post result (-0.038) against a control mean of 0.086, a 44.2% relative reduction in ex-ante unintended pregnancy (41.9% with covariates). At the 12-month cutoff (Panel C), the estimate instead resembles the ex-ante unwanted pregnancy null against a control mean of 0.034. Covariate adjustment shifts every panel only marginally, so this pattern is not likely attributable to confounding.

Panels A and C are close to indistinguishable from one another, as are Panels B and D, but the two pairs are not close to each other (Figure A4). Notably, Panel D's larger standard error (0.012, versus 0.008 in A and C) reflects higher outcome prevalence, not lower precision; what separates it from Panels A and C is a point estimate roughly four to nine times larger in magnitude. This distinction is central to how the null results in Panels A and C should be interpreted (Section 5).

4.4 Concordance between the Ex-Ante and Ex-Post Measures

Figure A1 through Figure A3 cross-tabulate the ex-post measure against each ex-ante measure by arm (restricted to women with both measures non-missing); Table A5 summarizes all three. At the 24-month cutoff, concordance with the ex-post measure is higher, and discordance is about half as common, than under unwanted pregnancy, in both arms (94.0% unwanted vs. 96.6% at 24 months in the control arm; 96.3% unwanted vs. 97.6% at 24 months in the treatment arm). At 12 months, concordance is essentially unchanged from the unwanted pregnancy baseline (93.0% at 12 months vs. 94.0% under unwanted pregnancy in the control arm; 96.4% at 12 months vs. 96.3% under unwanted pregnancy in the treatment arm). The composition of discordance shows the same split: under ex-ante unwanted pregnancy, discordance is "regret" dominated, i.e., a pregnancy the ex-ante measure did not flag as at-risk, but the woman retrospectively reported as unwanted, and asymmetric by arm (88.2% of control discordance vs. 67.9% of treatment discordance). Under the 24-month

measure, however, discordance is instead "wanted after all" dominated, i.e., a pregnancy that the ex-ante measure flagged as at risk of mistiming, but which the woman retrospectively reported as wanted, and the treatment arms look more alike (69.0% control vs. 72.2% treatment); under the 12-month measure, discordance remains regret-dominated and the cross-arm asymmetry persists (79.7% vs. 70.4%, though on only 59 and 27 discordant cases, respectively). In short, the 24-month cutoff resembles the ex-post measure on both concordance level and discordance composition, while the 12-month cutoff resembles ex-ante unwanted pregnancy on both – a pattern discussed further in Section 5.

4.5 Heterogeneity

Table A4 reports subgroup-specific treatment effects for each outcome across the three moderators from Section 3.

Baseline desire for another child produces the most coherent pattern, and it appears specifically under the 24-month measure. There, the effect concentrates among women who wanted another child at baseline (-0.059) and is small and insignificant among those who did not (-0.010); the equality-of-coefficients test clears conventional significance with and without controls. Under the 12-month measure, this split does not hold (did not want another: 0.002; wanted another: -0.019, with the equality test being far from significant). Substantively, the 24-month pattern is what a construct-mismatch account would predict, where an effect on mistiming is a coherent outcome only for women who intended another child, since a timing violation is undefined otherwise. That the 12-month cutoff does not reproduce this split is a limitation of that account taken alone (Section 5). Neither the ex-post measure nor the ex-ante unwanted pregnancy measure (Panels A and B) shows meaningful heterogeneity along this moderator, likely reflecting the small number of subsequent pregnancies overall (46 women).

The other two moderators show suggestive but statistically indistinguishable patterns, and here too, the 12-month cutoff does not parallel the 24-month cutoff. For baseline pregnancy status, point estimates run consistently larger among women not pregnant at baseline in Panels B and D (e.g., ex-post: -0.049 vs. -0.025; 24-month: -0.046 vs. -0.030), but the 12-month split is negligible (-0.011 vs. -0.009); none of the eight equality tests are significant. For baseline contraceptive use, the direction of the split is inconsistent across panels and no equality test approaches significance (range: 0.293-0.996). Given 24 equality tests in total, we treat these heterogeneity results as hypothesis-generating rather than confirmatory; that the desire-for-another-child split appears only at 24 months, and not at 12, indicates it is specific to that cutoff rather than a general feature of adding timing to the

ex-ante measure.

5 Discussion

This study's central finding is that measures targeting the same construct do not all agree empirically. The ex-post measure and the 24-month timing-inclusive ex-ante measure converge on treatment effects of similar magnitude (42% to 49% relative reductions); the 12-month timing-inclusive measure, despite targeting the same construct, produces a null estimate much like the ex-ante unwanted pregnancy measure, but for a different reason. Two explanations, one about which construct a measure captures and one about how much follow-up time it is given, together account for this pattern.

The first is construct mismatch, and for unwanted pregnancy it is definitional rather than merely empirical. The ex-post measure captures both the unwanted and mistimed components of unintended pregnancy (Santelli et al., 2003); however, ex-ante unwanted pregnancy captures only the former, so a woman whose spacing preference (rather than her limiting preference) was violated can never be coded as having an unintended pregnancy under that measure. Its null result is therefore not evidence against intervention impact - rather, it is likely a correctly estimated null on a narrower, non-overlapping construct. Because spacing is a documented channel through which this intervention operates (Karra et al., 2022), a measure confined to the unwanted component is structurally unable to register that mechanism. This is directly testable: once timing is added at the 24-month cutoff, the estimated effect moves from -0.004 to -0.038, closing the gap with the ex-post estimate, and the Panel D heterogeneity result reinforces the mechanism, where the effect concentrates among women who wanted another child at baseline, exactly those for whom a spacing violation is coherent.

The second explanation, and the reason the same construct fix does not also appear at 12 months, is exposure time. Study eligibility required a woman to be currently pregnant or within six months postpartum at baseline, so a large share of the sample remains in a period of low fecundability at 12 months, a period shaped by postpartum amenorrhea and lactational suppression of ovulation, and, for roughly half the sample, by a baseline that itself precedes any live birth. Accordingly, the 12-month control arm event rate (3.4%) runs at less than half the 24-month rate (8.6%), likely leaving too little variation for the measure to detect an effect regardless of arm. Unlike ex-ante unwanted pregnancy, whose null reflects a construct that can never register mistiming, the 12-month null plausibly reflects what has not yet happened by 12 months, even though the measure correctly targets unintended pregnancy.

A third possibility, differential rationalization by arm, is less likely to be the leading explanation. Retrospective wantedness items are generally vulnerable to ex-post rationalization (Rackin and Morgan, 2018; Mueller et al., 2022), and Yeatman et al. (2013) document non-trivial, distinct sources of error in prospective versus retrospective measurement in this same population. The cross-arm asymmetry in discordance composition we document under ex-ante unwanted pregnancy (Subsection 4.4) largely disappears at the 24-month cutoff, consistent with that asymmetry being a symptom of construct mismatch rather than of differential rationalization. However, its nominal persistence at 12 months rests on too few discordant cases (59 control, 27 treatment) to interpret and is consistent with, rather than competing against, the exposure time explanation.

Taken together, these results suggest that FP access can meaningfully reduce unintended pregnancy where it is measured over a horizon long enough for the relevant pregnancies to occur, and that the field's difficulty reconciling ex-ante and ex-post estimates may owe less to a fundamental incompatibility between the two approaches than to how narrowly ex-ante measures are typically constructed. Adding a timing dimension is a necessary but not sufficient fix: the chosen horizon must also be long enough given the study population's baseline fecundability, particularly where eligibility criteria concentrate women in a low-fecundability period. For measurement practice, this implies that studies aiming to compare prospective and retrospective approaches should field both within the same sample where possible, and should set timing cutoffs with reference to expected exposure time rather than defaulting to a single convention.

6 Conclusion

Whether this randomized FP intervention in urban Malawi reduced unintended pregnancy depends on how the outcome is measured. A narrow, prospective measure that only captures women who wanted no more children at all - properly understood as measuring unwanted, not unintended, pregnancy - shows no detectable effect. A retrospective measure, and a prospective measure that also accounts for desired birth spacing over a 24-month horizon, show reductions of roughly 42% to 49% relative to the control mean. The identical spacing information over a 12-month horizon shows no detectable effect, which we attribute to this sample's baseline fecundability rather than to a flaw in timing-inclusive measurement itself.

Beyond this trial, the analysis contributes one of the relatively few causal estimates of FP's effect on unintended pregnancy specifically, and offers a within-study comparison of prospective and retrospective measurement approaches. It suggests that some of the field's long-standing disagreement over which approach to trust may reflect how ex-ante

measures are typically constructed, both in terms of whether they incorporate timing and whether the chosen horizon suits the population's exposure time, rather than an inherent incompatibility between the two approaches. This argues for the embedding of more precise spacing preferences into ex-ante measures from the outset, setting the follow-up horizon with explicit reference to expected exposure time, and, where feasible, fielding both ex-ante and ex-post items in the same sample.

7 Acknowledgments

This research uses original data collected by Canning and Karra with support from Innovations for Poverty Action (IPA) in Malawi. Karra would like to acknowledge the dedication and support of Carly Farver, Patrick Baxter, Bagrey Ngwira, Reginald Chunda, Viola Nyirongo, Violet Chitsulo, Macdonald Salamu, and the entire MFPS team. The findings, interpretations, and conclusions expressed in this study are entirely those of the authors.

7.1 Funding Sources

This project was supported by two grants from the Hewlett Foundation and the Program for Women's Empowerment Research (POWER) (Grant No: 2018-8083, 2020-1162).

7.2 Ethical Considerations and Trial Registration

The MFPS trial was registered at the American Economics Association Registry for randomized controlled trials on May 7, 2015 (AEARCTR-0000697) and at the Registry for International Development Impact Evaluations (RIDIE) on May 28, 2015 (RIDIE-STUDY-ID-556784ed86956). Ethical approval to conduct the study was received from the Harvard University IRB (protocol number IRB16-0421) and the Malawi National Health Sciences Research Committee (protocol number 16/7/1628).

References

Austin, P. C. (2009, November). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine 28*(25), 3083–3107. 5

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

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

Debpuur, C., J. F. Phillips, E. F. Jackson, A. Nazzar, P. Ngom, and F. N. Binka (2002, June). The Impact of the Navrongo Project on Contraceptive Knowledge and Use, Reproductive Preferences, and Fertility. *Studies in Family Planning 33*(2), 141–164. 1

Desai, J. and A. Tarozzi (2011). Microcredit, Family Planning Programs, and Contraceptive Behavior: Evidence From a Field Experiment in Ethiopia. *Demography 48*(2), 749–782. 1

Gonzales-Huaman, P., J. E. Fernandez-Chinguel, and A. Taype-Rondan (2021, December). Peri-abortion contraceptive counseling: A systematic review of randomized controlled trials. *PLOS ONE 16*(12), e0260794. 1

Heil, S. H., H. S. Melbostad, A. K. Matusiewicz, C. N. Rey, G. J. Badger, D. S. Shepard, S. C. Sigmon, L. K. MacAfee, and S. T. Higgins (2021, October). Efficacy and Cost-Benefit of Onsite Contraceptive Services With and Without Incentives Among Women With Opioid Use Disorder at High Risk for Unintended Pregnancy. *JAMA Psychiatry 78*(10), 1–9. 1

Karra, M. and D. Canning (2020). The Effect of Improved Access to Family Planning on Postpartum Women: Protocol for a Randomized Controlled Trial. *JMIR Research Protocols 9*(8), e16697. 2, 3

Karra, M., D. Maggio, M. Guo, B. Ngwira, and D. Canning (2022). The causal effect of a family planning intervention on womens contraceptive use and birth spacing. *Proceedings of the National Academy of Sciences 119*(22), e2200279119. 1, 2, 8

Karra, M., D. Maggio, M. Guo, B. Ngwire, and D. Canning (2020). The Causal Effect of Improved Access to Postpartum Family Planning: Evidence from a Single-Blind, Randomized Controlled Trial in Urban Malawi. *Working Paper*. 4

Maggio, D., M. Karra, and D. Canning (2024, October). Family Planning and Children's Human Capital: Experimental Evidence From Urban Malawi. *Demography 61*(5), 1667–1698. 2, 3

Mueller, M. W., J. Hamory, J. Johnson-Hanks, and E. Miguel (2022). The illusion of stable fertility preferences. *Population Studies 76*(2), 169–189. 1, 9

Mwaikambo, L., I. S. Speizer, A. Schurmann, G. Morgan, and F. Fikree (2011). What works in family planning interventions: A systematic review. *Studies in Family Planning 42*(2), 67–82. 1

Nelson, H. D., A. Cantor, R. M. Jungbauer, K. B. Eden, B. Darney, K. Ahrens, A. Burgess, C. Atchison, R. Goueth, and R. Fu (2022, July). Effectiveness and Harms of Contraceptive Counseling and Provision Interventions for Women. *Annals of Internal Medicine 175*(7), 980–993. 1

Pechar, S., D. Maggio, and M. Karra (2026). Family Planning and the Dynamics of Fertility Preferences: Experimental Evidence from Urban Malawi. 2, 3

Rackin, H. M. and S. P. Morgan (2018). Prospective versus retrospective measurement of unwanted fertility: Strengths, weaknesses, and inconsistencies assessed for a cohort of US women. *Demographic research 39*, 61–94. 1, 9

Santelli, J., R. Rochat, K. Hatfield-Timajchy, B. C. Gilbert, K. Curtis, R. Cabral, J. S. Hirsch, and L. Schieve (2003). The Measurement and Meaning of Unintended Pregnancy. *Perspectives on Sexual and Reproductive Health 35*(2), 94–101. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1363/3509403. 3, 4, 8

Shlay, J. C., B. Mayhugh, M. Foster, M. E. Maravi, A. E. Barn, and J. M. Douglas (2003, August). Initiating contraception in sexually transmitted disease clinic setting: a randomized trial. *American Journal of Obstetrics & Gynecology 189*(2), 473–481. 1

Smith-Greenaway, E. and C. Sennott (2016, June). Death and Desirability: Retrospective Reporting of Unintended Pregnancy after a Childs Death. *Demography 53*(3), 805–834. 1

Starbird, E., M. Norton, and R. Marcus (2016, June). Investing in Family Planning: Key to Achieving the Sustainable Development Goals. *Global Health: Science and Practice 4*(2), 191–210. 1

Stevens, J., R. Lutz, N. Osuagwu, D. Rotz, and B. Goesling (2017, October). A randomized trial of motivational interviewing and facilitated contraceptive access to prevent rapid repeat pregnancy among adolescent mothers. *American Journal of Obstetrics & Gynecology 217*(4), 423.e1–423.e9. 1

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

Tsui, A. O., R. McDonald-Mosley, and A. E. Burke (2010, April). Family Planning and the Burden of Unintended Pregnancies. *Epidemiologic Reviews 32*(1), 152–174. 1

UNFPA (2014). Programme of Action, adopted at the International Conference on Population and Development. Technical report, United Nations Population Fund, Cairo, Egypt. 1

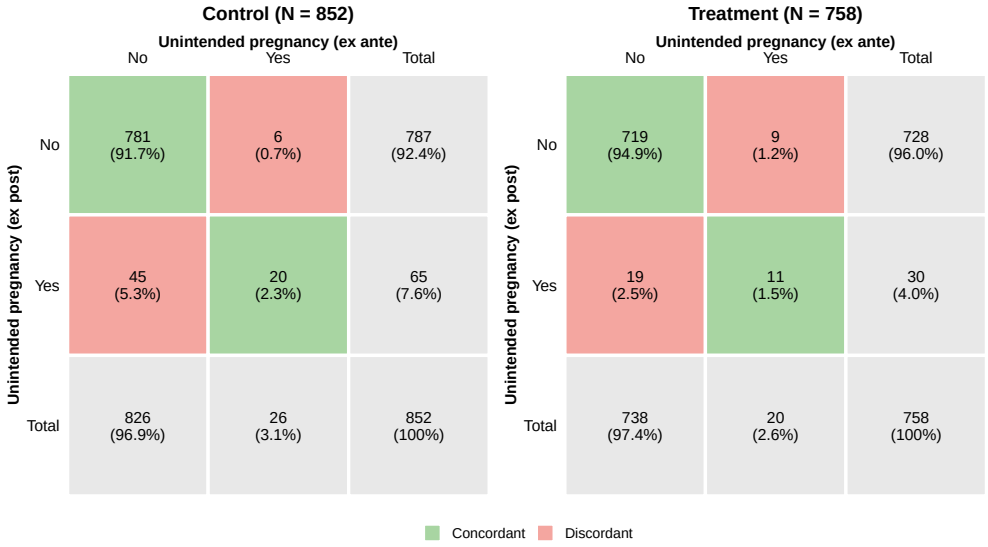
UNFPA (2022). Strategy for Family Planning, 2022-2030: Expanding Choices, Ensuring Rights in a Diverse and Changing World. Technical report, United Nations Population Fund, New York, NY. 1

United Nations (2018). Sustainable Development Goals. Technical report, United Nations, New York, NY. 1

Yeatman, S., C. Sennott, and S. Culpepper (2013). Young Womens Dynamic Family Size Preferences in the Context of Transitioning Fertility. *Demography 50*(5), 1715–1737. 1, 9

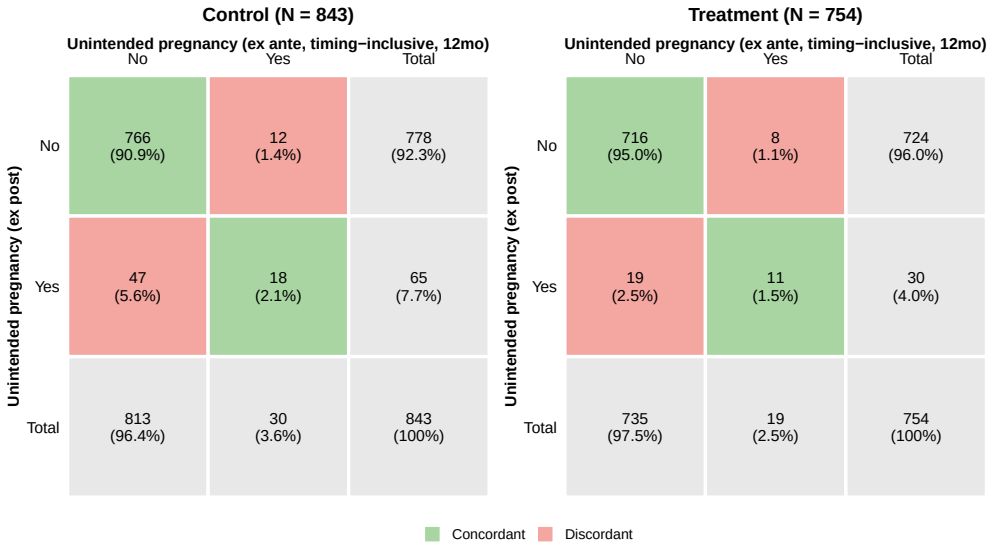
A Appendix

Figure A1: Concordance between Ex-Post Unintended Pregnancy and Original Ex-Ante Unwanted Pregnancy, by Arm



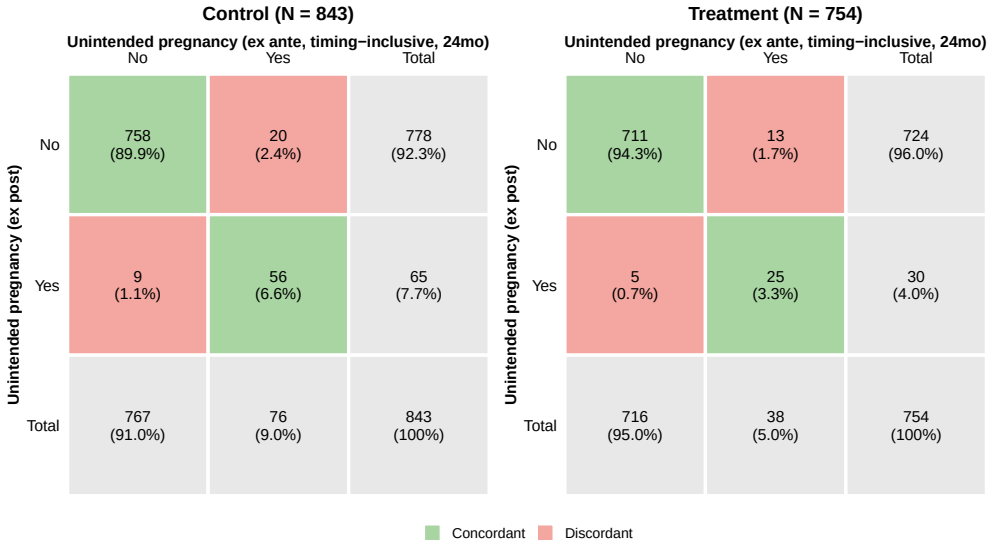
Notes: Endline sample (year = 2018), MFPS treatment + control. Cells are colored green for concordant women (both Yes or both No) and red for discordant women. Percentages are over the arm−specific total of women with non−missing values for both measures.

Figure A2: Concordance between Ex-Post and Ex-Ante (12-month window) Unintended Pregnancy, by Arm



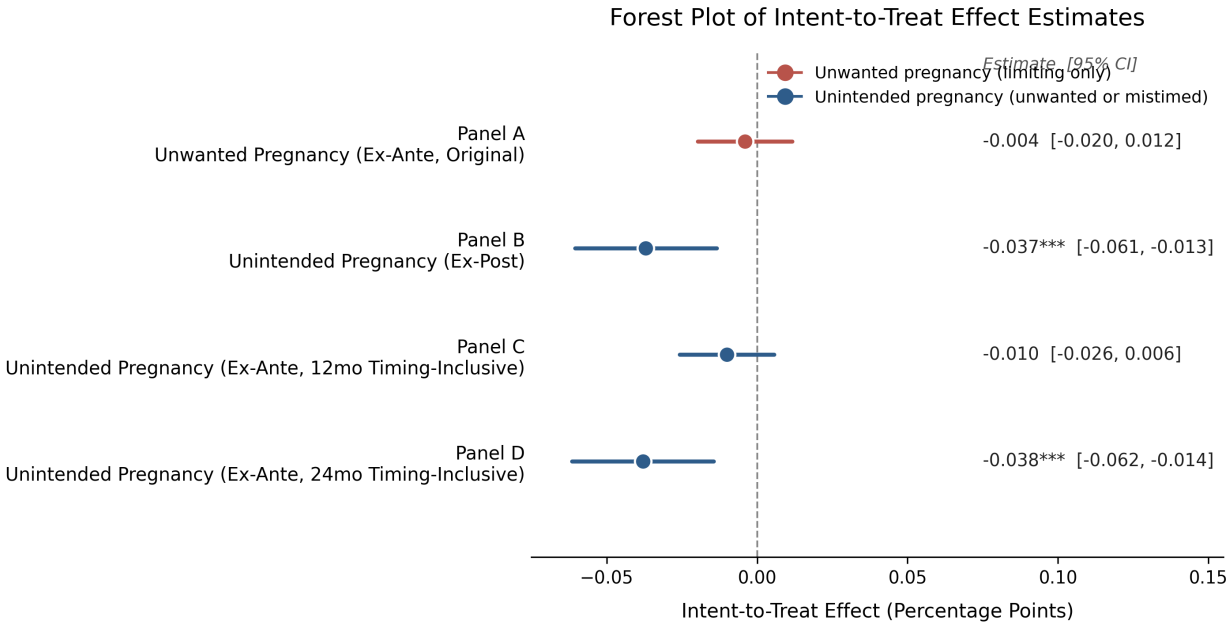
Notes: Endline sample (year = 2018), MFPS treatment + control. Cells are colored green for concordant women (both Yes or both No) and red for discordant women. Percentages are over the arm−specific total of women with non−missing values for both measures.

Figure A3: Concordance between Ex-Post and Ex-Ante (24-month window) Unintended Pregnancy, by Arm



Notes: Endline sample (year = 2018), MFPS treatment + control. Cells are colored green for concordant women (both Yes or both No) and red for discordant women. Percentages are over the arm−specific total of women with non−missing values for both measures.

Figure A4: Forest Plot of Intent-to-Treat Effect Estimates



Note: *** p<0.01. Model 1 (no covariates). Bars show 95% confidence intervals (coefficient ± 1.96×SE).

Table A1: Balance Table, Baseline Covariates

	Estimation Sample		**Control Group**		**Treatment Group**		
	N (1)	Mean (2)	N (3)	Mean (4)	N (5)	Mean (6)	Normalized Difference (7)
age_bl	1672	24.864	890	24.804	782	24.931	-0.027
Age at FSR	1667	18.941	886	18.823	781	19.074	-0.090
Number of children alive	1671	2.297	889	2.279	782	2.318	-0.031
preg_bl	1672	0.504	890	0.513	782	0.492	0.042
Injectable use BL	1672	0.245	890	0.238	782	0.253	-0.035
FP ever use - BL	1672	0.772	890	0.746	782	0.801	-0.130
Education level	1672	0.554	890	0.563	782	0.543	0.039
Work status	1672	0.100	890	0.100	782	0.100	0.001
Religion	1672	0.834	890	0.830	782	0.838	-0.019
Ethnicity	1672	0.425	890	0.426	782	0.425	0.003
Baseline: did not want another child	1672	0.436	890	0.429	782	0.444	-0.029
Observations for F-test							1667
F-test of joint significance: p-value							0.236

Note: Columns (1)-(2) pool both arms; Columns (3)-(4) are the control arm; Columns (5)-(6) are the treatment arm; Column (7) is the normalized difference $(\bar{x}_T - \bar{x}_C)/\sqrt{(s_T^2 + s_C^2)/2}$. The reported F-test is a test of joint significance.

Table A2: Balance Table, Outcomes

	Estimation Sample		**Control Group**		**Treatment Group**		
	N (1)	Mean (2)	N (3)	Mean (4)	N (5)	Mean (6)	Normalized Difference (7)
Unintended pregnancy (ex post)	1610	0.059	852	0.076	758	0.040	0.156
Unwanted pregnancy (ex ante)	1672	0.028	890	0.029	782	0.026	0.022
Unintended pregnancy (ex ante, timing-inclusive, 12mo cutoff, 12mo pregnancy win	1657	0.030	879	0.034	778	0.024	0.057
Unintended pregnancy (ex ante, timing-inclusive, 24mo cutoff, 24mo pregnancy win	1657	0.069	879	0.086	778	0.049	0.149
Observations for F-test							1597
F-test of joint significance: p-value							0.001

Note: Columns (1)-(2) pool both arms; Columns (3)-(4) are the control arm; Columns (5)-(6) are the treatment arm; Column (7) is the normalized difference $(\bar{x}_T - \bar{x}_C)/\sqrt{(s_T^2 + s_C^2)/2}$. The reported F-test is a test of joint significance.

Table A3: Main ITT Results: Unwanted (Panel A) and Unintended (Panels B-D) Pregnancy

Panel: A	**Unwanted Pregnancy: Ex-Ante**	
	(1)	(2)
treatment	-0.004	-0.004
	(0.008)	(0.008)
Observations	1672	1667
R^2	0.000	0.024
Control Mean	0.029	0.029
Panel: B	**Unintended Pregnancy: Ex-Post**	
	(1)	(2)
treatment	-0.037***	-0.036***
	(0.012)	(0.012)
Observations	1610	1605
R^2	0.006	0.020
Control Mean	0.076	0.076
Panel: C	**Unintended Pregnancy: Ex-Ante, Timing-Inclusive (12mo)**	
	(1)	(2)
treatment	-0.010	-0.010
	(0.008)	(0.008)
Observations	1657	1655
R^2	0.001	0.010
Control Mean	0.034	0.034
Panel: D	**Unintended Pregnancy: Ex-Ante, Timing-Inclusive (24mo)**	
	(1)	(2)
treatment	-0.038***	-0.036***
	(0.012)	(0.012)
Observations	1657	1655
R^2	0.006	0.014
Control Mean	0.086	0.086
Controls		x

Note: Robust standard errors in parentheses. *** $p < 0.01$. Controls include area, age group, age at sexual debut, children alive, ever use of FP, education, work, religion, and ethnicity.

Table A4: Heterogeneity (Subgroup) Analyses

Subgroup	N	(1) No controls Coef. (SE)	(2) With controls Coef. (SE)
Panel A: Unintended Pregnancy: Ex-Post			
Baseline pregnancy status			
Not pregnant at baseline	830	-0.049*** (0.018)	-0.049** (0.019)
Pregnant at baseline	780	-0.025* (0.013)	-0.019 (0.013)
p-value: equal coefficients (suest)		0.301	0.192
Baseline desire for another child			
Did not want another child at baseline	706	-0.025 (0.016)	-0.020 (0.016)
Wanted another child at baseline	891	-0.047*** (0.017)	-0.044*** (0.017)
p-value: equal coefficients (suest)		0.341	0.292
Baseline contraceptive use			
Not using FP at baseline	1200	-0.045*** (0.013)	-0.040*** (0.013)
Using FP at baseline	410	-0.015 (0.026)	-0.016 (0.026)
p-value: equal coefficients (suest)		0.293	0.405
Panel B: Unwanted Pregnancy: Ex-Ante			
Baseline pregnancy status			
Not pregnant at baseline	830	-0.009 (0.013)	-0.008 (0.013)
Pregnant at baseline	842	0.001 (0.010)	0.001 (0.010)
p-value: equal coefficients (suest)		0.527	0.579
Baseline desire for another child			
Did not want another child at baseline	729	-0.010 (0.018)	-0.007 (0.018)
Wanted another child at baseline	928	0.000 (0.000)	0.000 (0.000)
p-value: equal coefficients (suest)		–	–
Baseline contraceptive use			
Not using FP at baseline	1262	-0.002 (0.009)	-0.002 (0.009)
Using FP at baseline	410	-0.008 (0.017)	-0.005 (0.017)
p-value: equal coefficients (suest)		0.774	0.863
Panel C: Unintended Pregnancy: Ex-Ante, Timing-Inclusive (12mo)			
Baseline pregnancy status			
Not pregnant at baseline	822	-0.011 (0.014)	-0.010 (0.014)
Pregnant at baseline	835	-0.009 (0.009)	-0.004 (0.009)
p-value: equal coefficients (suest)		0.902	0.743
Baseline desire for another child			
Did not want another child at baseline	729	0.002 (0.012)	0.002 (0.012)
Wanted another child at baseline	928	-0.019 (0.012)	-0.020 (0.012)
p-value: equal coefficients (suest)		0.190	0.184
Baseline contraceptive use			
Not using FP at baseline	1248	-0.005 (0.008)	-0.002 (0.009)
Using FP at baseline	409	-0.026 (0.021)	-0.023 (0.021)
p-value: equal coefficients (suest)		0.357	0.358
Panel D: Unintended Pregnancy: Ex-Ante, Timing-Inclusive (24mo)			
Baseline pregnancy status			
Not pregnant at baseline	822	-0.046** (0.019)	-0.046** (0.020)
Pregnant at baseline	835	-0.030** (0.015)	-0.022 (0.015)
p-value: equal coefficients (suest)		0.515	0.341
Baseline desire for another child			
Did not want another child at baseline	729	-0.010 (0.018)	-0.007 (0.018)
Wanted another child at baseline	928	-0.059*** (0.017)	-0.056*** (0.017)
p-value: equal coefficients (suest)		0.048	0.043
Baseline contraceptive use			
Not using FP at baseline	1248	-0.038*** (0.014)	-0.032** (0.013)
Using FP at baseline	409	-0.038 (0.027)	-0.038 (0.029)
p-value: equal coefficients (suest)		0.996	0.855

Note: *** $p < 0.01$. Controls include area, age group, age at sexual debut, children alive, ever use of FP, education, work, religion, and ethnicity.

Table A5: Concordance Between Ex-Post Unintended Pregnancy and Ex-Ante Unwanted/Unintended Pregnancy, by Definition and Arm

			Discordant		Composition of Discordance	
Ex-Ante Definition	**Arm**	*N*	Concordant, %	Discordant, %	Regret, %	Wanted after all, %
Original (unwanted)	Control	852	94.0	6.0	88.2	11.8
	Treatment	758	96.3	3.7	67.9	32.1
Timing-inclusive (12mo, unintended)	Control	843	96.7	3.3	28.6	71.4
	Treatment	754	97.6	2.4	22.2	77.8
Timing-inclusive (24mo, unintended)	Control	843	96.4	3.6	33.3	66.7
	Treatment	754	97.6	2.4	27.8	72.2

Note: Endline sample (year = 2018), MFPS treatment + control, restricted in each row to women with non-missing values for both the ex-post and the indicated ex-ante measure (this restriction makes N here slightly smaller than the corresponding row of Table 2, which requires only the ex-ante measure to be non-missing). "Regret" cases are pregnancies the ex-ante measure did not flag as at-risk but which the woman retrospectively called unwanted (ex-ante = 0, ex-post = 1). "Wanted after all" cases are pregnancies the ex-ante measure flagged as at risk of mistiming but which the woman retrospectively called wanted (ex-ante = 1, ex-post = 0). Regret % and Wanted-after-all % are shares of that row's discordant cases and sum to 100.