Experimentally Testing New Measures of Preference-Behavior Alignment in Family Planning: A Mixed-Methods Study

Mahesh Karra[a] Ben Bellows[b] Pavani Chopra[c] Livia Montana[d]
Ilene Speizer[e] Beth Sully[f] Kelsey Holt[g] Jewel Gausman[h] Claire Rothschild[i]
Kerry MacQuarrie[j] Kristen Burke[k] Carolina Cardona[l] Celia Karp[m]
S Anukriti[n] Catalina Herrera Almanza[o] Sarah Vincent[p]

July 31, 2026

Abstract

Existing indicators of family planning demand and unmet need are defined in terms of current behavior alone, capturing neither contraceptive preferences nor method preference-behavior alignment. In a mixed methods study, we fielded a randomized survey experiment with 1,350 women of reproductive age in India and Kenya to evaluate proposed indicators of alignment. Our main analysis compares four indicators that classify women by their current method preference: preference-aligned fertility management (PFM), preference-use concordance (PUC), and conditional and unconditional contraceptive concordance (CCC and CCU). Unmet need was measured across all arms, and cognitive interviews were conducted to validate question phrasing and respondent comprehension. Adjusting for country, survey language, age, education, marital status, and parity, the predicted share of women classified as aligned was 53 percent for PFM, 70 percent for PUC, 34 percent for CCC, and 70 percent for CCU, differing significantly across most pairwise comparisons even after correcting for multiple hypothesis testing. This range reflects two sources of variation: a construct difference, with PFM and PUC defining

[a]**Corresponding Author**. Mailman School of Public Health, Columbia University. mvk2109@cumc.columbia.edu
[b]Nivi, Washington, DC, USA. ben@nivi.io
[c]Nivi, Washington, DC, USA. pavani@nivi.io
[d]Center for Social and Economic Research, University of Southern California, Washington, DC, USA. lmontana@usc.edu
[e]Gillings School of Public Health, University of North Carolina Chapel Hill, Chapel Hill, NC, USA. ilene_speizer@unc.edu
[f]Guttmacher Institute, New York, USA. esully@guttmacher.edu
[g]University of California San Francisco, San Francisco, CA, USA. kelsey.holt@ucsf.edu
[h]Guttmacher Institute, New York, USA. jgausman@guttmacher.edu
[i]Population Services International, Seattle, WA. crothschild@psi.org
[j]Washington, DC, USA. kmacqua@gmail.com
[k]University of Central Florida. kristen.burke@ucf.edu
[l]Bloomberg School of Public Health, Johns Hopkins University, Baltimore, USA. ccardon4@jhu.edu
[m]Bloomberg School of Public Health, Johns Hopkins University, Baltimore, USA. ckarp2@jhu.edu
[n]World Bank, Washington DC, USA. sanukriti@worldbank.org
[o]University of Illinois-Urbana Champaign, Urbana, IL, USA. cataher@illinois.edu
[p]Mailman School of Public Health, Columbia University, New York, USA. sv2890@cumc.columbia.edu

alignment at the method-specific level (treating a user who would prefer a different method as discordant) and CCC and CCU at the use level (treating that same user as aligned, since she is in fact using a method); and an elicitation-strategy difference, where hypothetical questions ("If you could switch, would you") elicit more reported switching than direct questions ("Do you want this method?"). When CCC and CCU were reconstructed to treat a user who would prefer a different method as discordant, matching the definition already used by PFM and PUC, the two reconstructed hypothetical measures (modified CCC and modified CCU) cluster at around 20 percent, while the two direct measures (PFM and PUC) ranged from 53 to 70 percent. Elicitation strategy therefore affects measured alignment more than construct choice alone.

Keywords: family planning; measurement; preference-aligned fertility management; contraceptive concordance; intention to use; unmet need; indicators; experiment; India; Kenya

1 Introduction

Reproductive agency is a cornerstone of women's health, embodying the principle that individuals have the right to make decisions over their own reproductive health care that reflect their preferences (Bhan et al., 2022; Upadhyay et al., 2014; Purdy, 2006). A high-quality family planning and reproductive health (FP/RH) program is therefore not only determined by the achievement of good reproductive health outcomes but also prioritizes helping individuals, as the primary decision-makers, to maximize a complex and evolving set of preferences around their future fertility, health, and well-being (Karra, 2022; Diamond-Smith et al., 2018; Dehlendorf et al., 2016). For this reason, an individual's demand for contraception and FP services differs from most other interventions in health; while one can presume that individuals consistently have a demand for health interventions that reduce their risk of morbidity and mortality, the same cannot be said for the demand for contraception since women and couples may seek to become pregnant over their lifetimes (Cleland et al., 2014). As a result, it has become incumbent on FP/RH programs to: 1) demonstrate that a demand for FP/RH exists; and 2) measure the extent to which this demand is met.

There has been a growing call from researchers, policymakers, and practitioners to develop new indicators that effectively promote women's reproductive health and agency (Bingenheimer et al., 2023; Speizer et al., 2025). By reorienting measurement around person-centered, preference-based indicators, FP providers and programs can better identify gaps in access and care that are articulated by women themselves (Senderowicz, 2020; Holt et al., 2017, 2024); responding to these gaps, in turn, would positively contribute to women's agency, health, and well-being (Speizer et al., 2025). To date, public health systems rely on indicators such as contraceptive use and unmet need to guide funding allocations, program design, and accountability. These indicators shape programmatic decisions affecting all populations, including how need is defined, which populations are prioritized, and how success is evaluated; however, they have been critiqued for assuming a demand for contraception among women who are not using contraception but who report a desire to delay or avoid pregnancy (Karra, 2022; Cleland et al., 2014; Senderowicz et al., 2023). As a result, these indicators risk misclassifying women's preferences and incentivizing uptake-oriented policies that do not reflect women's reproductive goals, especially in contexts where contraceptive decision-making is shaped by complex social, relational, and structural factors.

To address these gaps, efforts have been made to introduce new, person-centered indicators that capture the alignment between preferences and behavior in FP decision-making to both estimate and meet the demand for FP/RH services. While calls for new measures have been enthusiastic, the development and implementation of new indicators have been

conspicuously slow. Recent advances, which were highlighted in the 2023 and 2025 special issues in *Studies in Family Planning* and at conference sessions at the Population Association of America (PAA) annual meeting, have focused on conceptualizing new approaches to person-centered, rights- and justice-based measurement in FP/RH (Bingenheimer et al., 2023; Speizer et al., 2025). Several new indicators have been proposed in response to these calls (Senderowicz, 2020; Holt et al., 2023; Vincent et al., 2025; Bullington et al., 2025); however, their implementation and validation have been limited to a few settings, and they have not been systematically compared to one another or to existing standards. As a result, policymakers lack causal evidence on how different indicator choices shape estimates of demand, subgroup disparities, and program performance. This mixed-methods study addresses that exact gap by treating measurement itself as the object of evaluation to generate evidence on how alternative indicators influence inference and decision-making in FP policy and programs.

The objective of this study is to generate causal evidence on the performance and policy relevance of recently developed, person-centered FP indicators by means of: 1) a randomized survey experiment with women aged 18-49 in India and Kenya; and 2) cognitive interviews with women to understand how indicators were interpreted and whether survey questions were effectively phrased and posed to elicit the intended information. Nivi's WhatsApp-based digital infrastructure provides the platform for recruiting and efficiently implementing the survey experiment with respondents in both settings. As part of the experiment, alternative measures of preference-behavior alignment are experimentally tested against commonly used indicators such as unmet need and intention to use contraception. By randomizing respondents to different indicator variants, the study estimates how measurement choice affects prevalence estimates, subgroup classification, and agreement across measures. The broader purpose of this study is to contribute to an emerging international agenda aimed at centering women's experiences, choices, and preferences within FP measurement systems. By refining these indicators, the project aims to improve the conceptual and empirical tools that can inform program design, policy-making, and global monitoring efforts.

2 Data and Experimental Design

2.1 Study Setting

The data come from a randomized survey experiment conducted among women of reproductive age in India and Kenya. Respondents were recruited and surveyed through Nivi, a digital, WhatsApp-based, self-care platform that delivers reproductive health information

and services to women in both countries (CITE NIVI). The platform setting allows the measurement experiment to be embedded directly within an existing survey workflow, so that randomization to an indicator condition occurs unobtrusively and respondents are able to complete the assigned module as part of a single, self-administered interview. Conducting the experiment in two distinct contexts allows us to assess whether differences across indicators are a stable feature of the measurement instruments themselves or are sensitive to the contraceptive method mix, fertility norms, and FP program environment of a particular setting.

India and Kenya are particularly suitable for evaluating these new measures. Both countries have diverse reproductive health landscapes, high levels of engagement with digital health platforms, and a growing reliance on self-care and self-guided decision-making tools. In India, contraceptive prevalence is characterized by high rates of female sterilization, with method mix varying significantly by geography, age, and urban-rural residence (Government of India, 2022). In Kenya, the method mix is more diverse, with a growing uptake of long-acting reversible contraceptives alongside traditional and short-acting methods (Kenya National Bureau of Statistics and ICF International, 2023). These distinct contexts provide an opportunity to test how well the new indicators perform across varying reproductive health environments.

Nivi's WhatsApp-based digital platform provides a unique infrastructure for reaching women confidentially and efficiently. This study leverages Nivi's infrastructure to implement a randomized survey experiment and to recruit women for cognitive interviews that probe how they interpret key measurement items. The rationale behind using a mixed-methods approach is to complement robust quantitative testing with qualitative inquiry to understand how women make sense of complex constructs and to determine whether survey questions are appropriately phrased to elicit the intended information.

The broader purpose of this study is to contribute to an emerging international agenda aimed at centering women's experiences, choices, and preferences within family planning measurement systems. By refining these indicators, the project aims to improve the conceptual and empirical tools that can inform reproductive health program design, policy-making, and global monitoring efforts.

2.2 Sample

Eligible respondents were women aged 18 to 49 who were not currently pregnant at the time of the survey. After screening, eligible women who consented were randomized to one of four measurement arms and completed the corresponding module. All respondents,

regardless of assigned arm, answered a common block of screening and background questions that included age, education, marital status, parity, and fertility preferences (whether the respondent wants a/another child and the desired timing of the next birth). This common block provides the source of the covariates that are used for adjustment and of the traditional comparison measure of unmet need and unwanted use described in Section 2.5. As these items precede and are independent of the randomized module, they are measured identically across all four arms.

2.3 Experimental Design and Randomization

Figure A1 presents the experimental design. The experiment manipulates a single factor: the measurement instrument used to assess preference-behavior alignment. Each respondent was randomly assigned to one of five arms. Four operationalize competing indicators of current preference-behavior alignment and constitute the main analysis: 1) preference-aligned fertility management (PFM), as defined by Holt et al. (2023); 2) preference-use concordance (PUC), as proposed by CITE ILENE; 3) conditional contraceptive concordance, as originally defined by Vincent et al. (2025) (CCC); and 4) unconditional contraceptive concordance (CCU), a modified version of CCC. The fifth arm fielded a modified measure of intention to use contraception (ITU).[1] Because ITU is anchored on *future* intention rather than current method preference, we treat it as a conceptually distinct construct and report it separately in Appendix B. Because the only element that varies across arms is the set of questions used to classify a woman as aligned or misaligned, the design isolates the causal effect of *indicator choice* on measured alignment. Randomization was embedded in the survey workflow with equal target allocation across arms.

The four indicators differ along two conceptual dimensions that organize the analysis. The first is the *construct* each indicator measures. PFM and PUC operationalize alignment at the *method-specific* level: a current user who would prefer a different method than the one she is using is classified as *misaligned*, because her behavior does not match her stated method preference. CCC and CCU instead operationalize alignment at the *contraceptive use* level: the same switcher is classified as *aligned*, because she wants to be using a method and is in fact using one. The second dimension is the *elicitation strategy*. PFM and PUC use *direct* elicitation, asking the respondent whether her current behavior is what she wants ("Do you want to be using this method?"). CCC and CCU use *hypothetical* or counterfactual elicitation, inviting the respondent to consider deviating from her current behavior ("If you could switch, would you?"). These two dimensions, construct and elicitation strategy, are

[1] Unlike the conventional ITU indicator, which is defined only among non-users, our measure classifies both users and non-users so that it is comparable to the other four arms.

the lens through which the indicators are compared throughout the study.

Figure 1 maps the indicators onto these two dimensions, with example wording and the adjusted share of women classified as aligned under each indicator. The figure also highlights the two re-constructed ("modified") versions of CCC and CCU, defined in Section 2.5: holding the elicitation strategy fixed, but treating a user who would prefer a different method as discordant, moves CCC and CCU from the use-level row into the method-specific row, where they converge with one another (and, as a supplementary result, with ITU).

Figure 1: Alignment indicators by construct and elicitation strategy

	Direct elicitation ("Do you want to be using this method right now?")	**Hypothetical elicitation** ("If you could switch, would you?")
Method-specific a user who would prefer a different method is treated as *discordant*	**PFM** — 53% **PUC** — 70%	**ITU** — 23.5% (future-intention framing) *Modified CCC* — 20.3% *Modified CCU* — 20.5%
Use-level a user who would prefer a different method is treated as *aligned*	*(no indicator)*	**CCC** — 34% **CCU** — 70%

Note: Each cell reports the indicator's adjusted predicted alignment (Table A8; modified CCC and CCU from Table A13). PFM and PUC ask respondents directly whether their current method is what they want; CCC, CCU, and ITU invite respondents to consider switching to, stopping, or starting a method. Re-constructing CCC and CCU under the method-specific definition (*modified*) moves them into the upper-right cell, where all three hypothetical elicitation indicators converge at around 20%. The empty lower-left cell reflects that no indicator we test pairs a use-level construct with direct elicitation. Cells report the adjusted share of women classified as aligned (predicted probabilities from Table A8 and Table A13; ITU from Appendix B). Modified CCC and CCU re-classify a user who would prefer a different method as discordant, matching the PFM/PUC method-specific definition, which moves them into the method-specific row.

2.4 Measurement of Preference-Behavior Alignment Indicators

Each arm classifies a woman as aligned (= 1) or misaligned (= 0) using a short sequence of questions specific to that arm. In every arm, current contraceptive use is established first and serves as the behavioral anchor against which stated preferences are compared. We describe the construction and coding of each indicator in turn. 'Don't know' and 'refused' responses on the preference-determining questions are treated as missing and excluded from the corresponding indicator.

2.4.1 Arm 1: Preference-Aligned Fertility Management (PFM)

PFM is a direct elicitation, method-specific indicator constructed in two steps. In following Holt et al. (2023), *Step 1* assesses use-level alignment by comparing a woman's stated preference for using a method - whether she wants to be using a method to avoid pregnancy[2] with whether she is currently using one (current users include those who report using "sometimes"). A woman is use-level aligned if her preference for use equals her behavior (a user who wants to use, or a non-user who does not want to use). *Step 2* adds a method-specific test applied to current users only: the user is asked whether she wants to be using *this* method right now.[3] A user who answers "no" (i.e., would prefer a different method) is reclassified as misaligned even though her use-level preference matched her behavior. Non-users carry forward their Step 1 classification, since the method-specific question does not apply to them. A woman is coded PFM-aligned (= 1) only if she is aligned at both steps.

2.4.2 Arm 2: Preference-Use Concordance (PUC)

PUC is also a direct elicitation, method-specific indicator, constructed analogously to PFM but eliciting the use-level preference separately for users and non-users. As proposed by CITE, current use is established first. Among users, the use-level preference is taken from whether the woman reports that her current use is what she wants;[4] among non-users, it is taken from whether she reports that not using is *not* what she wants (i.e., a non-user who says her non-use is not what she wants is coded as wanting a method).[5] As in PFM, Step 1 codes a woman as use-level aligned when this preference matches her behavior, and Step 2 applies a method-specific test to current users (whether she wants to be using her current method now).[6] A woman is coded PUC-aligned (= 1) only if she is aligned at both steps; non-users carry forward their Step 1 classification.

2.4.3 Arm 3: Conditional Contraceptive Concordance (CCC)

CCC is a hypothetical elicitation indicator operationalized at the contraceptive use level (Vincent et al., 2025). Current use is established first. A woman is classified as *wanting FP* if: i) she is a non-user who reports she would use a method;[7] or ii) she is a user who reports she would *not* stop using a method;[8] Users who report they would stop are routed past

[2] "Do you currently want to be using any method to avoid pregnancy?"
[3] "Do you want to be using this method right now?"
[4] "Is using a method what you want?"
[5] "Is not using a method what you want?" - coded as wanting a method when the answer is "no".
[6] "Do you want to be using [METHOD] right now?"
[7] "Would you use a method to avoid pregnancy?" (asked of non-users).
[8] "Would you stop using your method?" (asked of users).

the switch question by the conditional skip and are classified as not wanting FP, regardless of whether they would in fact switch to a different method. The stop-and-switch group is therefore not observable under CCC, which is a structural limitation of the conditional design that the Arm 4 (CCU) instrument is built to correct.

A woman is coded CCC-aligned (= 1) when her preference for FP equals her current use (a user who wants to be using a method, or a non-user who does not). Because alignment is defined at the use level, a user who would prefer to switch to a different method is still counted as aligned. The defining feature of CCC is its *conditional skip pattern*: the switch question is asked conditionally, so that users who report they would stop their method are routed away from the subsequent switch question. As a result, the switcher population is only partially observed under CCC, a limitation that the Arm 4 design is constructed to correct.

2.4.4 Arm 4: Unconditional Contraceptive Concordance (CCU)

CCU is identical in construct to CCC, a hypothetical elicitation, use-level indicator, but removes the conditional skip pattern. The switch question is asked of *all* current users rather than only of those who would not stop, so that the full population of women who would prefer to switch methods is observed. A woman is classified as wanting family planning if she is a non-user who would use a method, a user who would not stop, or a user who would switch; she is coded CCU-aligned (= 1) when her preference for family planning equals her current use. As with CCC, a switcher is counted as aligned under the use-level definition. The contrast between CCC and CCU therefore isolates the effect of the conditional skip pattern, holding the construct and the elicitation strategy fixed.

2.5 Key Outcomes and Coding

The analysis uses one primary outcome, a binary indicator of preference-behavior alignment, together with a set of comparison measures and a categorical decomposition that allow us to interpret *why* the indicators diverge.

2.5.1 Primary outcome: binary alignment

The primary outcome, *align*, is the respondent's classification under her assigned arm's indicator, coded 1 for aligned and 0 for misaligned, as defined in Section 2.4. Operationally, *align* takes the value of PFM in Arm 1, PUC in Arm 2, CCC in Arm 3, and CCU in Arm 4. Differences in the adjusted prevalence of *align* across arms are the central object of the

analysis, since under randomization any systematic difference is attributable to the indicator rather than to the women being measured.

2.5.2 Current contraceptive use

Current use is the common behavioral anchor underlying every indicator and is also reported on its own. It is coded 1 for any woman who reports currently using (or sometimes using) a method, drawn from the use question in her assigned arm, and 0 otherwise. Because the use question is asked comparably across arms, current use provides a behavioral benchmark across the experiment.[9]

2.5.3 Traditional comparison measures: unmet need and unwanted use

To benchmark the new indicators against established measures, we construct two traditional measures from the common screening block, so that they are available for *all* respondents regardless of assigned arm. *Unmet need* is coded 1 for a woman who wants to space or limit childbearing - that is, who wants no (more) children or wants to wait two or more years for her next birth - but is not currently using a method, and 0 otherwise. *Unwanted use* (sometimes "unwanted family planning use") is coded 1 for a woman who wants a child within two years yet is currently using a method, and 0 otherwise (Canning and Karra, 2023). These follow the conventional logic that ties contraceptive demand to fertility preferences, in contrast to the preference-based logic of the new indicators.[10]

2.5.4 Control concordance and demand satisfied

From the two traditional measures, we derive two reference concordance measures. *Control concordance* is the complement of any traditional discordance: a woman is discordant if she has either unmet need or unwanted use, and concordant (= 1) otherwise. *Demand satisfied* is defined more narrowly as the complement of unmet need ($1 - \textit{unmet need}$). Both serve as external benchmarks against which each new indicator's agreement is assessed (via Cohen's κ and McNemar's test) in the Results, Section 4.

[9] Unmet need and unwanted FP use for each respondent are calculated using the respective current use questions from each assigned arm.

[10] This is a simplified operationalization of unmet need built from fertility timing preferences and current use; it does not incorporate the postpartum amenorrhea, fecundity, or pregnancy intention refinements of the full DHS algorithm (Bradley et al., 2012; Cleland et al., 2014).

2.5.5 Method-specific (modified) concordance

To compare the indicators on a common conceptual footing, we re-construct CCC and CCU under the method-specific definition used by PFM and PUC, in which a switcher is treated as *misaligned*. In the modified measures (*ccc_mod*, *ccu_mod*), a current user is coded aligned (= 1) only if she would *neither* stop *nor* switch her method; a user who would stop, or who would switch, is coded misaligned; non-users are coded as in the original (aligned if they do not want a method). The modified measures are combined with the unchanged PFM and PUC into a four-arm method-specific alignment variable used in the corresponding tables.

2.5.6 Decomposition into alignment categories

Finally, to show that indicators producing similar binary alignment rates may classify *different* women as misaligned, we decompose the binary outcome into five mutually exclusive categories defined by the cross-classification of preference and current use, with a separate switcher subcategory among current users:

1. **Aligned — wanted user**: currently using, wants a method, and satisfied with the current method (not a switcher);
2. **Aligned — wanted non-user**: not using and does not want a method;
3. **Misaligned — unmet need**: not using but wants a method;
4. **Misaligned — unwanted use**: using but does not want any method;
5. **Misaligned — switcher**: using and wants a method, but would prefer a different one.

Each indicator assigns respondents to these categories using its own preference and switching questions, so the category distributions differ across arms both because the arms survey different (randomized) respondents and because they ask different questions. The four cross-classification categories are also tabulated for traditional unmet need as a comparison column, with unwanted use defined as in subsection 2.5 (currently using and wanting a child within two years); current users who express no spacing or limiting preference form a residual category, and the switcher category does not apply.

2.5.7 Covariates

All adjusted models control for the background characteristics measured in the common block: country (India vs. Kenya), survey language, age, education (completed university or

vocational / adult education vs. below), marital status (married vs. not), and parity (any children vs. none). Country is included in all adjusted specifications because, as the balance tests show, the share of respondents from each country is the principal dimension on which the arms are imbalanced.

2.6 Cognitive Interviews

To complement the randomized experiment, we conducted cognitive interviews to examine the response processes underlying the alignment-determining questions and understand whether women interpret these questions as the indicators' designers intended. Whereas the experiment measures how indicator choice affects the *prevalence* of measured alignment, the cognitive interviews seek to examine *why* a given question elicits the responses it does, and in particular whether the preference and switching items that distinguish the indicators carry their intended meaning for respondents. We followed standard cognitive interviewing methodology (CITE), combining a paraphrase task (respondents restated each question in their own words) with concurrent and retrospective verbal probes targeting the specific terms on which the indicators turn.

The interview guide covered the items that operationalize each construct: 1) the use-level preference item used by PFM and PUC ("Do you currently want to be using any method to avoid pregnancy?") and the method-specific item that distinguishes their Step 2 ("Do you want to be using this method right now?"); and 2) the hypothetical elicitation items used by CCC and CCU ("If you had the choice and ability to stop using your family planning method, would you choose to stop?" and the parallel switching item)[11] It also probed the behavioral anchor common to all arms ("Are you currently using any method…" / "Are you currently doing something or using any method…") and the key terms embedded in these items: *currently / right now*, *want*, *choice* and *ability*, *method* versus *family planning method*, *delay* versus *avoid*, and *future*. Respondents were also asked to compare paired short and long phrasings of selected items and to state which they preferred and why.

Participants were a purposive subsample of reproductive-age women (18-49, not currently pregnant) recruited through the same Nivi platform used for the experiment in India with interviews conducted in English or Hindi. Due to resource constraints, no cognitive interviews could be completed in Kenya. Interviews were conducted by phone in a private setting, audio-recorded with consent, and translated and transcribed for analysis; respondents were not asked to disclose their own contraceptive use or reproductive history beyond what they volunteered. The design was iterative and proceeded in two rounds: a first round

[11] For ITU, future intention items ("…at any time in the future") were added to the hypothetical eliciation items.

administered the original question wording, and findings from that round informed targeted revisions that were re-tested in a second round. Sampling continued toward thematic saturation, defined as the point at which additional interviews surfaced no new patterns of misunderstanding, ambiguity, or unintended interpretation (CITE ON SATURATION).

The analysis reported here draws on the India cognitive interviews: a first round with 11 English-speaking and 7 Hindi-speaking women, and a second round with 3 English-speaking women who re-tested the revised wording. Transcripts were analyzed thematically, organized first by construct and then by term. For each probed term we classified respondents' comprehension as consistent with, partially consistent with, or inconsistent with the intended meaning, and we coded in particular whether the preference items were processed as questions about *preference* or were instead answered as questions about current *behavior*, social *norms*, or future *plans*. Because the qualitative sample is small and not designed for statistical inference, we report comprehension patterns as counts of respondents rather than as population estimates and use illustrative verbatim responses (identified by participant number, location, and education) to characterize each pattern.

3 Empirical Strategy

3.1 Identification

Because respondents are randomly assigned to one of the four measurement arms, the distribution of background characteristics is, in expectation, identical across arms. Any systematic difference in measured alignment across arms is therefore attributable to the indicator itself rather than to differences in the women being measured. The estimand of interest is the causal effect of *indicator choice*, the randomized factor, on the probability that a woman is classified as preference-aligned. We take preference-aligned fertility management (PFM, Arm 1) as the reference category throughout, so that each contrast is interpreted relative to PFM. All models adjust for the small number of pre-randomization covariates measured in the common screening block - country (Kenya vs. India), survey language (English), age, education (completed university or vocational / adult education), marital status (married), and parity (any children) - both to absorb residual chance imbalance (most notably the cross-arm imbalance in country composition documented in the balance tests) and to improve precision. Heteroskedasticity-robust standard errors are used in all specifications.

3.2 Main Specification

We denote y_i as the binary alignment outcome for woman i that is equal to one if she is classified as aligned under her assigned arm's indicator and zero otherwise, and we define $Arm_i \in \{1, \ldots, 5\}$ as her randomized arm. We estimate a logit model of the form

$$\Pr(y_i = 1 \mid Arm_i, \mathbf{X}_i) = \Lambda\left(\alpha + \sum_{k=2}^{4} \beta_k \, \mathbf{1}[Arm_i = k] + \mathbf{X}_i'\boldsymbol{\gamma}\right), \tag{1}$$

where $\Lambda(\cdot)$ is the logistic CDF, $\mathbf{1}[\cdot]$ is the indicator function, the omitted category is PFM (so β_k captures the shift in the log-odds of alignment for arm k relative to PFM), and $\mathbf{X}_i$ is the vector of covariates described above. We report the model both unadjusted (arm dummies only) and adjusted (including $\mathbf{X}_i$). The omnibus null hypothesis that the four indicators yield the same alignment rate,

$$H_0 \colon \ \beta_2 = \beta_3 = \beta_4 = 0,$$

is assessed with a Wald test, which is distributed χ^2 with three degrees of freedom under the null.

Because logit coefficients are not directly interpretable on the probability scale, we summarize the arm effects as adjusted average marginal effects (AMEs). For each non-reference arm k, the AME is the average over the estimation sample of the difference in predicted probability of alignment between assignment to arm k and assignment to PFM, holding each woman's covariates at their observed values:

$$\mathrm{AME}_k = \frac{1}{N} \sum_{i=1}^{N} \Big[\widehat{\Pr}(y_i = 1 \mid Arm = k, \mathbf{X}_i) - \widehat{\Pr}(y_i = 1 \mid Arm = 1, \mathbf{X}_i)\Big], \tag{2}$$

expressed in percentage points. We also report, for each arm, the model-implied predicted probability of alignment, obtained as the predictive margin from the adjusted logit in Equation (1).

3.3 Pairwise Comparisons and Multiple-Comparison Adjustment

The omnibus test and the PFM-referenced contrasts do not exhaust the comparisons of interest, since the substantive question is whether *any* pair of indicators yields different alignment rates. We therefore estimate all six pairwise differences in adjusted alignment probability. For transparency on the probability scale and to make the contrasts directly

comparable, these are obtained from a linear probability model (LPM),

$$y_i = \alpha + \sum_{k=2}^{4} \beta_k \, \mathbf{1}[Arm_i = k] + \mathbf{X}_i'\boldsymbol{\gamma} + \varepsilon_i, \tag{3}$$

from which all $\binom{4}{2} = 6$ pairwise contrasts are computed. Because evaluating six contrasts inflates the family-wise error rate, we report each contrast's p-value three ways: unadjusted, Bonferroni-adjusted, and Sidak-adjusted across the six comparisons. As a companion, we also report the pairwise differences with Sidak-adjusted 95% confidence intervals, which are correspondingly widened to control the experiment-wise error rate.

3.4 Heterogeneity

We test whether the arm effects differ across pre-specified subgroups, country (Kenya vs. India), parity (any children vs. none), and education (university or vocational / adult education vs. less). For each subgroup variable g, we fit a logit model that fully interacts arm assignment with the subgroup,

$$\Pr(y_i = 1 \mid \cdot) = \Lambda\left(\alpha + \sum_{k=2}^{4} \beta_k \, \mathbf{1}[Arm_i = k] + \delta \, g_i + \sum_{k=2}^{4} \theta_k \left(\mathbf{1}[Arm_i = k] \times g_i \right) + \mathbf{Z}_i'\boldsymbol{\gamma} \right), \tag{4}$$

and conduct a joint Wald test of the null that all interaction terms are zero ($H_0: \ \theta_2 = \theta_3 = \theta_4 = 0$), distributed χ^2 with three degrees of freedom. Here, $\mathbf{Z}_i$ represents the vector of covariates excluding the subgroup variable. Because three subgroup tests are conducted, we interpret them against a Bonferroni-corrected threshold across the three pre-specified tests. To visualize the pattern, we report subgroup-specific predicted probabilities of alignment by arm, obtained from logit models stratified by subgroup level (equivalent in expectation to the interaction model used for the test).

3.5 Agreement with Traditional Measures

The cross-arm comparisons above are between-subjects, since each woman is observed under only one indicator. A separate question is how each new indicator agrees with the established measures *within* an arm, where the new indicator and the traditional measures are observed on the same respondents. This within-arm comparison is a paired (within-subject) one, so we summarize agreement with Cohen's κ and test the equality of marginal prevalence with McNemar's test. For each arm, we compute these against two traditional benchmarks constructed from the common screening block: i) *control concordance*, defined

as the complement of any traditional discordance (unmet need or unwanted use); and ii) the *complement of unmet need* ("demand satisfied"). For a given arm, with b and c denoting the two off-diagonal cell counts of the 2×2 cross-tabulation of the alignment indicator against the benchmark, McNemar's statistic is computed as $(b-c)^2/(b+c)$ and referred to a χ^2 distribution with one degree of freedom. Low κ values indicate that the new indicator captures information distinct from the traditional measure; a significant McNemar test indicates that the two measures classify different overall proportions of women as aligned.

3.6 Placing Indicators on a Common Method-Specific Footing

The five indicators differ on two dimensions simultaneously, construct (method-specific vs. use-level) and elicitation strategy (direct vs. hypothetical), which confounds a simple comparison of their prevalences. To separate these dimensions, we re-estimate the analysis after re-constructing the two use-level indicators, CCC and CCU, under the method-specific definition used by PFM and PUC, in which a woman who would prefer to switch methods is treated as misaligned (the modified indicators *ccc_mod* and *ccu_mod* defined in Section 2.5). We combine the unchanged PFM and PUC with the modified CCC and CCU into a single four-arm method-specific alignment outcome and re-run the logit, predicted-probability, and pairwise-contrast specifications of Equations (1)-(3), again with Bonferroni- and Sidak-adjusted p-values across the resulting contrasts. With construct now held constant across the four arms, residual differences between the direct elicitation indicators (PFM, PUC) and the hypothetical elicitation indicators (modified CCC, modified CCU) are interpreted as an elicitation strategy effect.

To quantify how much the method-specific re-construction changes each use-level indicator, we compare the original and modified versions within their own arms, a paired comparison, since both are computed for the same Arm 3 (CCC) and Arm 4 (CCU) respondents. We report the prevalence under each definition together with Cohen's κ and a McNemar test of the change in marginal prevalence, computed as above.

3.7 Decomposition of Misalignment

The binary alignment outcome conceals *which kind* of misalignment an indicator identifies. To make this visible, we classify every respondent into one of five mutually exclusive categories defined by the cross-classification of preference and current use, with a separate switcher subcategory among current users: aligned wanted user, aligned wanted non-user, unmet need, unwanted use, and switcher (defined in Section 2.5). Each indicator assigns respondents to these categories using its own preference and switching questions. We present

the resulting category distributions as within-arm percentages, first pooled across both countries and then separately for India and Kenya, alongside traditional unmet need as a comparison column, for which unwanted use follows the subsection 2.5 definition (currently using and wanting a child within two years) and where current users with no stated spacing or limiting preference form a residual category. This decomposition is descriptive: it shows that indicators yielding similar binary alignment rates can nonetheless distribute the misaligned across substantively different categories.

4 Results

4.1 Descriptive Statistics and Randomization

Table A1 presents baseline characteristics across the four experimental arms. The analytic sample comprises 1,350 women across the four experimental arms ($n_{PFM} = 371$, $n_{PUC} = 334$, $n_{CCC} = 310$, $n_{CCU} = 335$).[12] The sample is predominantly university or vocational / adult-educated (65-69 percent), and the majority of women are married (59-62 percent) and have had at least one child (77-81 percent). Current contraceptive use is high, ranging from 69 percent in the PFM arm to 78 percent in CCC.[13]

The four arms are well-balanced across most baseline covariates, including marital status, education, parity, and fertility intentions.[14] Some variation in country composition remains, with the share of Kenyan respondents ranging from 71 percent in the PFM arm to 65 percent in the PUC arm. The pairwise balance tests in Table A2, however, show that this variation does not translate into broad covariate imbalance across arms: none of the six pairwise comparisons reaches individual or joint significance, indicating that the contrasts are statistically balanced at the 5 percent level (PFMPUC, $p = 0.203$; PFM-CCC, $p = 0.451$; PFM-CCU, $p = 0.206$; PUC-CCC, $p = 0.616$; PUC-CCU, $p = 0.163$; CCC-CCU, $p = 0.485$). Because measured alignment differs between Kenya and India (Table A10 reports the alignment outcome by country), even this modest variation in country composition could bias raw cross-arm mean comparisons, but not the adjusted models. All subsequent analyses therefore adjust for country, in addition to survey language, age, education, marital

[12] The analytic sample for the alignment-outcome regressions is smaller (N = 1,214) due to item-level missingness on the alignment-determining survey questions.

[13] Because current use is drawn from each arm's own randomized module rather than a common pre-randomization item, the cross-arm range may partly reflect differences in question wording, ordering, and priming across arms.

[14] Current contraceptive use, which is measured within each arm rather than as a common baseline item, is also similar across arms (69 to 78 percent), though this comparison should be interpreted descriptively given the randomized arm-specific measurement.

status, and parity; we also report unadjusted descriptive results (Table A3 and Table A4) for transparency.

4.2 Main Results

Table A3 reports unadjusted means of the alignment outcome by arm. The four indicators produce a wide range of alignment rates: the unadjusted prevalence of alignment is 53.7 percent for PFM, 69.2 percent for PUC, 34.3 percent for CCC, and 70.5 percent for CCU. The indicators split into two clear groups: those that produce unadjusted alignment prevalences above 50 percent (PFM, PUC, and CCU) and those that produce prevalences below 40 percent (CCC). The pairwise t-tests in Table A4 confirm that comparisons across these groups are large in magnitude and statistically significant, while contrasts within the upper group (PFM, PUC, CCU) are smaller.

Table A5 reports the unadjusted and adjusted logit estimates of the probability of alignment, with PFM as the reference category. The omnibus Wald test of arm equality is strongly rejected in both specifications (χ^2 between 97 and 99, df = 3, $p < 0.001$), confirming significant differences in alignment prevalences across arms. The adjusted average marginal effects (AMEs) for the three non-PFM arms are all significant at $p < 0.001$: PUC is 16.2 percentage points (p.p.) higher than PFM, CCC is 18.9 p.p. lower, and CCU is 17.0 p.p. higher. Table A6 shows that all three PFM contrasts survive both Bonferroni and Sidak multiple-comparison adjustment at $p < 0.001$. The only pairwise contrast that fails to reach significance under adjustment is CCU vs PUC (Sidak $p = 0.77$), which captures the close proximity of the two upper-group indicators. Table A8 reports the adjusted predicted probabilities of alignment evaluated at sample covariate means: the adjusted prevalence of alignment under PFM is estimated at 53.3 percent, under PUC at 69.5 percent, under CCC at 34.4 percent, and under CCU at 70.3 percent, with non-overlapping confidence intervals between all groups except PUC and CCU.

Table A9 explores heterogeneity by presenting the joint Wald tests for arm-by-subgroup interactions. The country interaction is significant at the conventional 0.05 level ($\chi^2(3) = 9.55$, $p = 0.023$) but does not survive Bonferroni correction across the three pre-specified tests (threshold $p = 0.0167$). As Table A10 shows, the country effect is concentrated in CCU and CCC, both of which show higher adjusted prevalence of alignment in Kenya than in India. The adjusted prevalence of CCU alignment is 62.6 percent in India versus 74.3 percent in Kenya (an approximately 12 p.p. gap), and the adjusted prevalence of CCC alignment is 23.2 percent in India versus 40.1 percent in Kenya (an approximately 17 p.p. gap). PFM and PUC are relatively stable across countries. The country heterogeneity is

thus a meaningful feature of the data, concentrated in the hypothetical elicitation indicators (CCU and CCC), which register consistently higher alignment in Kenya, while the direct elicitation indicators (PFM and PUC) show little cross-country variation. Because survey language could itself drive part of this cross-country difference, all adjusted models control for the language in which the module was administered; the country effect therefore reflects residual between-country variation net of survey language, though language and country are necessarily correlated and we cannot fully separate translation effects from other contextual differences.

Table A11 and Table A12 report agreement statistics between each alignment indicator and two traditional measures of preference-behavior concordance: a control concordance measure that is defined as the complement of any discordance from unmet need or unwanted family planning (UFP) use (Table A11), and a measure that is the complement of unmet need, typically interpreted as "demand satisfied" (Table A12). The Cohen's κ statistics of indicator agreement are small for every indicator. Against the control concordance measure, κ is 0.12 for PFM, 0.28 for PUC, 0.10 for CCC, and 0.23 for CCU. The κ statistics against the complement of unmet need in Table A12 follow a similar pattern: 0.13 for PFM, 0.21 for PUC, 0.05 for CCC, and 0.33 for CCU.[15]

To assess whether the indicators agree once they are placed under a common conceptual framework, we re-construct the CCC and CCU measures using the PFM/PUC method-specific definition, where any user who would prefer to switch methods is classified as discordant regardless of her use status. Placing the measures on a common method-specific footing produces results that reframe the original analysis. Table A13 reports the adjusted predicted probabilities of method-specific alignment by arm. The small shifts that do appear (PFM 52.9 to 53.0 percent; PUC 67.8 to 67.9 percent) arise not because the PFM and PUC variables change (they do not) but because the four-arm method-specific model is estimated on a slightly different analytic sample. Specifically, the modified CCC and CCU coding alters item-level missingness, changing the pooled estimation sample (N falls from 1,255 to 1,214), and the predictive margins for PFM and PUC are evaluated at that sample's covariate means. The modified CCC indicator, however, falls to 20.3 percent from 34.3 percent, and the modified CCU indicator falls substantially to 20.5 percent from 70.5 percent. The CCU drop of 50.0 p.p. reflects the fact that the unconditioned design captures the full population of users who would prefer to switch a substantial group that the original use-level definition counted as aligned. The CCC drop is smaller because the conditional skip pattern

[15] In Appendix B, we show that ITU's near-zero κ values indicate that the new indicators capture substantively different information from the traditional measures, with McNemar tests rejecting equality of marginal prevalence at $p < 0.001$ across all indicators in Table A12.

already excluded much of the switcher population from the originally-aligned pool. Notably, the modified CCC and CCU values are now in close proximity to the ITU adjusted estimate of 17.5 percent in Appendix B, suggesting that all three hypothetical elicitation indicators converge on similar alignment estimates when held to a strict method-specific standard.

The pairwise comparisons in Table A14 highlight the two-family structure with more clarity. Within the direct elicitation family (PFM and PUC), the 16.0 p.p. PUC-PFM gap is statistically significant under every multiple-comparison procedure ($p < 0.001$). Within the hypothetical elicitation family, the modified CCC and modified CCU produce essentially identical estimates a 0.3 p.p. statistically insignificant gap ($p = 0.93$). All four cross-family contrasts are large (33 to 49 p.p.) and remain significant at $p < 0.001$ under every multiple comparison procedure. Conditional on a common method-specific definition, the hypothetical elicitation indicators classify substantially more women as discordant than the direct elicitation indicators do.

Country is the principal dimension on which the arms are imbalanced, and Kenya and India differ on the alignment outcome (Table A10), so we re-estimate the four-arm method-specific analysis separately within each country to confirm that the direct-versus-hypothetical pattern is not an artifact of pooling two heterogeneous settings. Table A15 and Table A16 report the adjusted predicted probabilities, and Table A17 and Table A18 the corresponding pairwise contrasts. The two-family structure is preserved in both settings: the direct elicitation indicators remain far higher than the hypothetical elicitation indicators (PFM and PUC at 57.1 and 68.9 percent in India, 51.8 and 69.9 percent in Kenya; modified CCC and modified CCU at 16.3 and 16.6 percent in India, 22.6 and 22.2 percent in Kenya), and within the hypothetical family the modified CCC and modified CCU estimates remain statistically indistinguishable in each country (a 0.2 p.p. gap in India, $p = 0.97$; a 0.5 p.p. gap in Kenya, $p = 0.91$). All four cross-family contrasts are large (29 to 52 p.p.) and significant at $p < 0.001$ under Bonferroni correction in both countries. The one contrast that is not stable across settings is the within-direct family PUC-PFM gap, which is 18.1 p.p. and significant in Kenya ($p_{Bonf} = 0.001$) but only 11.4 p.p. and not significant after correction in India ($p_{unadj} = 0.105$, $p_{Bonf} = 0.629$); the modified hypothetical measures also sit several points higher in Kenya than in India, echoing the country heterogeneity documented in Table A10 for the use-level indicators. The finding that holding the construct fixed, hypothetical elicitation yields substantially lower alignment than direct elicitation continues to hold within each country and is not therefore an artifact of pooling India and Kenya, though per-country cell sizes are modest (N = 389 in India, 800 in Kenya) and the within-country comparisons are, in turn, less precise.[16]

[16]Stratifying by country removes only the between-country component of the variation in current contra-

Table A19 quantifies the reclassification within each arm using paired agreement statistics. The CCC modification reclassifies a moderate share of women: prevalence falls from 34.3 percent to 20.3 percent, with a Cohen's κ of 0.67 (indicating substantial agreement), and the McNemar test rejects equality of marginal prevalence at $p < 0.001$. The CCU modification is far more consequential: prevalence falls from 70.5 percent to 20.5 percent, with a κ of 0.19 (only slight agreement), and the McNemar test rejects at $p < 0.001$. The asymmetry between Arms 3 and 4 has a substantive interpretation. CCU's unconditioned design reveals that approximately half of all Arm 4 respondents are current users who, when given the opportunity, express a preference to switch methods. CCC's conditional design misses most of this group because the survey skip pattern routes women who say they would stop their method away from the switch question entirely, which is confirmed by the earlier findings.

4.3 Decomposition of Misalignment Categories

Table A20 through Table A22 decompose the binary alignment outcome into five mutually exclusive categories defined by the 2-by-2 classification of preference and current use alignment, with a separate switcher subcategory for current users who want to be using a method but would prefer to use a different one.[17] This decomposition reveals that the indicators differ not only in how many women they classify as misaligned but, more substantively, in which categories of misalignment they identify, a distinction that the binary outcome obscures.

Table A20 presents the pooled distribution across both countries. Among PFM respondents, misaligned women are more evenly distributed across unmet need (21.8 percent), unwanted use (10.3 percent), and switchers (13.6 percent), for a total of 45.7 percent misaligned. PUC produces a similar distribution but with a smaller misaligned share overall (9.3 percent unmet need, 9.3 percent unwanted use, and 7.8 percent switchers, yielding 26.4 percent total misaligned). The contraceptive concordance indicators produce more lopsided distributions. CCC categorizes slightly less than half of its sample (48.3 percent) as unwanted use (substantially more than any other indicator), which is consistent with the

ceptive use; because current use is mechanically tied to the alignment outcome, any within-country differences in current use across arms remain unadjusted, and the stratified estimates do not address that residual. Randomization with equal target allocation is the basis for treating those within-country differences as chance imbalance rather than a confounder.

[17] Because switchers are isolated as their own subcategory rather than absorbed into the aligned categories, the misaligned totals reported here correspond to the method-specific definition of alignment (the modified CCC and CCU of Section 3.6), in which a switcher is counted as misaligned. They therefore differ from the original use-level CCC and CCU prevalences (34.3 and 70.5 percent aligned) in Table A19, under which switchers are counted as aligned.

survey's conditional skip pattern that routes women who say they would stop their method into a category that the CCC indicator must, by construction, code as unwanted use regardless of their underlying preference. CCU produces the inverse finding, whereby 51.4 percent of its sample is classified as switchers, with only 8.2 percent as unwanted use. CCC and CCU therefore identify roughly the same total proportion of misaligned women (about 80 percent in each) but distribute them across fundamentally different misalignment categories; for CCU, the switcher row (51.4 percent) is exactly the group that the original use-level definition counts as aligned. Traditional unmet need, now included as a comparison column, identifies 16.5 percent of the full sample as having unmet need and 11.1 percent as having unwanted use, with no switcher category by construction.

Table A21 and Table A22 stratify the analysis by country and reveal that the substantive findings are preserved in both contexts but with meaningful shifts in magnitude. Misalignment shares are systematically higher in India than in Kenya for most indicators, but the patterns that distinguish the indicators from one another remain stable: the CCC arm continues to assign approximately half of its sample to unwanted use in both countries (50.6 percent in India, 47.3 percent in Kenya), while the CCU arm continues to assign approximately half to switching (45.7 percent and 54.5 percent, respectively). The country split therefore confirms that the CCC-CCU divergence is structural to the indicator design rather than driven by contextual factors that are specific to either country.

4.4 Cognitive Interview Findings

The cognitive interviews do not change the experimental estimates, but they supply the response-process evidence needed to interpret them.

First, they provide a candidate mechanism for the study's central empirical finding: that elicitation strategy moves measured alignment more than construct choice. The direct "want" items anchor respondents to their current situation and are frequently answered as behavioral questions, so they rarely surface a preference to deviate from the status quo. The hypothetical "if you could stop / switch" items instead license respondents to consider deviating, and were treated as genuine decisions. The roughly 30 p.p. excess of women classified as wanting to switch under the hypothetical indicators is therefore consistent not with those indicators over-detecting, but with the direct indicators systematically under-detecting a switching preference that respondents hold but do not volunteer when asked directly.

Second, and as a corollary, the interviews caution against reading the direct elicitation indicators' lower measured misalignment as the "true" alignment rate. Because the use-level

"want" question is often not processed as a preference question at all, the high alignment recorded under PFM and PUC partly reflects a measurement property of the question, its tendency to be answered behaviorally, rather than a substantive finding that women's preferences are in fact aligned with their behavior.

Third, the interviews clarify the construct comparison and support the manuscript's modified indicator analysis. The method-specific distinction that separates PFM/PUC from CCC/CCU, wanting *this* method versus wanting *a* method, is one respondents struggle to draw, which is consistent with the experiment's finding that this distinction is sensitive to how it is elicited. And because the "choice and ability" conditioning that defines CCC is largely not comprehended, the conditional skip is doing less of the work the design intends; this strengthens the interpretability case for the unconditioned measure (CCU) and for the reconstructed method-specific versions, and is consistent with the convergence of modified CCC and modified CCU once a common method-specific standard is imposed.

Fourth, the interviews reinforce the decision to treat ITU separately and to locate it in the appendix. The future orientation that the manuscript flags as introducing ambiguity is borne out directly: respondents anchor "future" to horizons ranging from days to a lifetime, so the numerical proximity between ITU and the modified hypothetical measures reflects the shared elicitation framing rather than a stable underlying construct.

Finally, the interviews surface measurement caveats that cut across all arms and that the quantitative estimates cannot reveal on their own: the situational reading of "current," the unreliable capture of traditional and self-managed methods, and the spacing/permanence connotations of "family planning method." Each is a potential source of systematic miscoding, the first two affecting the behavioral anchor and the switcher pool in every arm, and the third affecting the hypothetical elicitation items specifically. Together they indicate that the choice among indicators should weigh not only the prevalence each produces but also the comprehension burden each imposes, and they motivate the need for wording refinements in future iterations of these indicators.

5 Discussion

Three key findings emerge from the empirical analysis. First, the four indicators do not converge on a common alignment rate. Adjusted alignment prevalence ranges from 34 percent (CCC) to 70 percent (CCU), with PFM (53 percent) and PUC (70 percent) spread between them. Statistical interchangeability holds for only one pair: the two hypothetical elicitation indicators, modified CCC and modified CCU, converge once both are placed on a common method-specific footing (a 0.3 p.p. gap, $p = 0.93$; Table A14). It does not extend to the other

within-family pairs, where the two direct, method-specific indicators differ significantly (the PUC-PFM gap is 16.0 p.p., $p < 0.001$; Table A14), and the original use-level CCC and CCU differ sharply (34.3 versus 70.5 percent unadjusted; Table A3). Across conceptual families, by contrast, the differences are large and robust to multiple-comparison adjustment. Second, the construct distinction between the families is consequential: PFM and PUC operationalize alignment at the method-specific level, while CCC and CCU operationalize alignment at the use level. The indicators therefore measure conceptually distinct constructs, and the difference in alignment prevalences reflects, to some degree, this construct difference rather than disagreement about which women are preference-aligned. The agreement statistics in Table A11 and Table A12 support this interpretation, with CCU (the indicator closest in operational logic to traditional concordance) producing the highest agreement with the unmet need benchmark.

The construct distinction is consequential for how these results should be interpreted. A central question related to the use of these indicators going forward is how each indicator family treats *switchers*, those women who would prefer to be using a method and are using a method, but want to be using a different method than the one they currently are. PFM and PUC take a conservative approach by operationalizing alignment at the *method-specific* level, where a user who would prefer to switch methods is classified as discordant because her current behavior is not what she states she would prefer. In contrast, CCC and CCU take a more liberal approach, operationalizing alignment at the *contraceptive use* level; a user who prefers to switch is still classified as aligned because she wants to be using a method and is, in fact, using one (i.e., her use status is consistent with her preference). The wide range of alignment prevalences in Table A8 (37 to 72 percent) reflects this conceptual divide alongside differences in elicitation strategy. The CCC arm, while operationally in the use-level family, produces a substantially lower prevalence than CCU because the survey's conditional skip pattern channels users who report a preference to stop their method away from the subsequent switch question, which effectively forces the assumption that "would stop" implies "would not switch" an assumption that CCU's unconditioned design corrects. A consensus has not yet been reached on which interpretation is more appropriate, and the choice has substantial empirical consequences. In our sample, applying the conservative method-specific approach to the CCC and CCU indicators (Table A13) reduces alignment prevalence from 34.3 to 20.3 percent for CCC and from 70.5 to 20.5 percent for CCU - a reduction of 50.0 p.p. in the latter case.

This reframing yields a substantive finding that the binary alignment outcome obscures: when the four indicators with directly comparable method-level questions are placed on a common method-specific footing, CCC and CCU identify a substantially larger misaligned

population than PFM does. PFM classifies approximately 47 percent of its sample as misaligned (Table A8), while the modified CCC and modified CCU classify approximately 80 percent of their respective samples as misaligned (Table A13). A likely reason is that the hypothetical elicitation strategy used by CCC and CCU surfaces switching preferences that the direct elicitation strategy used by PFM does not. Specifically, PFM asks "Do you want to be using this method?" and accepts the status quo as the implicit reference point; a respondent must actively reject her current behavior to register discordance. CCC and CCU ask "If you could switch, would you?" and invite the respondent to consider alternatives. Because the modified CCC and modified CCU apply the same method-specific stop/switch rule, the two binary measures are near-identical classifiers, so their convergence (a 0.3 p.p. gap, $p = 0.93$; Table A14) is partly built in by construction. What the convergence does establish is that the two estimates, which are produced on separate randomized arms, agree on roughly 20 percent alignment, which supports the assertion that the figure reflects the elicitation strategy rather than the particular sample of one arm.

A more granular comparison reinforces this conclusion. PFM is itself constructed as a two-step indicator: step 1 captures use-level alignment (whether a woman's preference for use matches her current behavior), and step 2 conditions on method-specific alignment via the direct question "Do you want to be using this method right now?". Comparing the step 1 to step 2 drop within PFM to the analogous modified drops in CCC and CCU clarifies where the elicitation effect operates. PFM's step 1 alignment prevalence is 68.2 percent and its step 2 alignment prevalence (the final PFM measure) is 53.7 percent, a 14.5 p.p. drop that reflects the share of switchers PFM identifies. Because PFM's step-2 question is posed only to current users (non-users retain their step-1 classification unchanged), this reclassification occurs entirely within the user subset, so the arm-level 14.5 p.p. drop understates method-level discordance among users specifically, where alignment falls from 83.1 percent to 63.8 percent. The analogous drop for CCC is 14.0p.p. (34.3 to 20.3 percent), broadly comparable to PFM, while the CCU drop is 50.0 p.p. (70.5 to 20.5 percent), approximately three to four times larger than either PFM or CCC. Restricting attention to current users within each arm yields the same picture: PFM identifies 20.3 percent of its users as switchers, CCC at 17.5 percent (likely an underestimate given the conditional skip pattern), and CCU at 70.9 percent. PFM's method-specific question is therefore not categorically less sensitive than CCC's conditional switching question, but neither approaches the sensitivity of CCU's unconditioned hypothetical framing. The elicitation effect described above is therefore not about whether the indicator asks any method-specific question at all (PFM does) but about how that question is framed, either directly ("Do you want this method?") or counterfactually ("If you could switch, would you?").

The apparent convergence between ITU and the modified CCC and CCU estimates should not be taken to imply that the three indicators measure the same underlying construct. ITU's future orientation introduces a fundamental ambiguity; a woman who expresses an intention to discontinue or switch at some point in the next 12 months may or may not hold that preference today, and stated future intentions are prone to "cheap talk" when they carry no immediate behavioral cost. The broader literature on preference-behavior alignment and person-centeredness in FP has, for this reason, focused on whether women's current preferences are aligned with their current behavior, on the view that inferring present alignment is both more interpretable and more actionable than future intent. From this perspective, the numerical agreement between ITU and the modified CCC and CCU indicators reflects the elicitation effect we describe above rather than a substantive equivalence of constructs, and the methodological case for the modified CCC and CCU as preferred measures is stronger than the case for ITU.

This finding has direct implications for programs and policy. A central programmatic goal of preference-behavior alignment measurement is to identify women whose contraceptive behavior does not match their stated preferences, so that programs can support them in moving on either margin, either toward use if they currently want a method, or toward switching if their current method is not what they prefer. From this perspective, the most programmatically useful indicator is the one that is most *sensitive* to misalignment, in the sense that it identifies the largest pool of women who may benefit from intervention. On this criterion, the hypothetical elicitation indicators perform better than direct elicitation indicators. Because each woman is observed under only one indicator, this is a comparison of equivalent randomized groups rather than of the same individuals: the hypothetical elicitation arms classify a larger share of women as misaligned (about 30 p.p. more under a common method-specific definition), and the additional misalignment they capture is concentrated among current users who would prefer to switch or discontinue, a category that direct elicitation largely leaves in the aligned group. Whether the additional identified women are realistically receptive to intervention is an open empirical question (some may be expressing hypothetical preferences that they would not act on), but they are, at minimum, the population that a switcher-targeted intervention would seek to reach. Moreover, method dissatisfaction is itself a programmatically actionable signal: a woman who is using a method but would prefer a different one can be reached through counseling, method switching, or resupply. This reinforces the case for a method-specific measure that surfaces this group rather than a use-level measure that absorbs it into the aligned category.

An additional question is whether the FP measurement field should adopt the method-specific definition of alignment, more generally. There are three reasons for doing so. First,

across the broader family of person-centered FP indicators, treating switchers as aligned versus misaligned is the single most substantively important classification choice, and operating under multiple definitions in the same measurement space will create a source of confusion in interpreting prevalence estimates across studies. Second, the method-specific definition identifies a larger pool of program-relevant misalignment. Third, a woman who explicitly states that she would prefer to be using a different method is, by her own stated preferences, not in a state of preference-behavior alignment, even if she remains a method user. Treating her as aligned would subordinate her own stated preference to her observed behavior, which runs counter to the principle of reproductive autonomy that motivates person-centered measurement in the first place (Bhan et al., 2022; Upadhyay et al., 2014; Purdy, 2006): respecting a woman's articulated method preference is part of what it means to measure whether her family planning use reflects her own choices. We acknowledge that the use-level definition has its own justifications, most notably that switching may not always be feasible in low-capacity settings, where "would prefer a different method" may reflect an aspirational preference rather than an actionable one. The consensus choice should therefore be made deliberately rather than by default.

The decomposition results in Table A20 through Table A22 provide further evidence that this is the right level of distinction to make. Indicators that produce similar binary alignment rates can categorize substantively different populations as misaligned. CCC and CCU produce similar method-specific alignment rates (20.5 versus 20.3 percent) but disagree on the composition of the misaligned: CCC places half of its sample in the unwanted use category and only 13.3 percent as switchers, while CCU places half in the switcher category and only 8 percent as unwanted users. Because the arms are randomized, the two groups are exchangeable, so this difference in the composition of misalignment is attributable to the one feature that differs by design, where CCC's conditional skip routes users who report a preference to stop their method away from the switch question and into the unwanted use cell, where they can no longer be identified as switchers, whereas CCU asks the switch question of all current users. From a measurement standpoint, this is a strong argument for the unconditioned CCU design over the conditioned CCC design, independent of the broader switcher treatment question. From a programmatic standpoint, the CCC indicator may give a misleading signal about what kind of misalignment is most prevalent in a given population.

Finally, the elicitation strategy result deserves emphasis in its own right. The gap between direct and hypothetical elicitation, even after holding the construct constant, suggests that survey question wording in FP measurement is itself a substantive determinant, and not merely a methodological detail. Direct questions accept the status quo as the implicit reference point and are likely to under-identify women who hold latent preferences for change

but who have not actively considered acting on them. Hypothetical questions invite respondents to engage with counterfactual scenarios and surface preferences that direct questions may miss. The convergence of three independent hypothetical elicitation indicators (ITU, modified CCC, and modified CCU) on a common alignment rate of approximately 20 percent provides evidence that this elicitation effect may be more than idiosyncratic. For population-level monitoring, this distinction matters considerably: estimates of contraceptive demand and program-relevant misalignment depend substantially on how the relevant questions are asked. As preference-behavior alignment indicators move toward wider use in population surveys, the choice of elicitation strategy deserves the same deliberate attention as the construct choice.

6 Conclusion

This study treated family planning measurement as the object of experimental evaluation, randomizing women in India and Kenya across four indicators of preference-behavior alignment to isolate how indicator choice shapes measured alignment. Two conclusions follow for the measurement field.

First, the field should converge on a single approach to measuring alignment, chosen deliberately rather than by default. Because the treatment of switchers is the most consequential classification decision, and because elicitation strategy affects measured alignment at least as much as construct choice, continuing to operate under multiple definitions will make prevalence estimates incomparable across studies. The case for a single standard does not by itself determine which measure should fill it; that is where our evidence speaks, and where our study's constraints bind. On that question, our evidence favors a measure with two properties, one speaking to its construct and the other to its elicitation.

The first property is one of construct: the measure should define alignment at the method-specific level, treating a woman who would prefer a different method than the one she uses as misaligned. Such a woman is, by her own stated preference, not in a state of preference-behavior alignment, and a use-level definition that counts her as aligned subordinates her articulated preference to her observed behavior. The second property concerns how that preference is elicited and has two features that need to be taken into account. The preference should be elicited counterfactually, by inviting every current user to consider whether she would switch, rather than through a direct question about whether her current method is what she wants. Direct "want" questions accept the status quo as their implicit reference point, so a respondent's latent preference to switch goes unrecorded unless she volunteers it. On the other hand, a counterfactual question surfaces those preferences, which is why

method-specific indicators that elicit directly classify far fewer women as misaligned than those that elicit alignment counterfactually, holding the construct fixed. Moreover, that switch question should be posed to every current user rather than through a conditional skip: holding construct and elicitation strategy fixed, a conditional design channels women who would stop their method away from the switch question and into a category from which they can no longer be identified as switchers, so that a measure built on it systematically understates the switching population it is meant to capture. A measure that is both counterfactual and unconditioned avoids these two limitations and is correspondingly more sensitive to the margin on which we observe the most variation.

This programmatic framing should not, however, be turned into a reason to discount the preferences themselves. A stated preference to switch is indicative of a woman's preference at the moment it is elicited, and under the person-centered framework that motivates these indicators, that is precisely the quantity that alignment is meant to capture; whether she would ultimately act on it is a separate matter. To treat a surfaced preference as less real because it was elicited hypothetically and carried no immediate behavioral cost would substitute our external judgment for her stated choice the same subordination of articulated preference to observed behavior that person-centered measurement exists to avoid. On that criterion, the more sensitive measure is not over-detecting misalignment; rather, it is recording preferences that less sensitive designs fail to elicit at all. What remains genuinely open is narrower and program-facing: how a stated willingness to switch relates to subsequent contraceptive behavior, which bears on how programs forecast demand and allocate resources rather than on whether a woman's preference counts. Establishing that relationship is a validation step that this study does not undertake. Our results establish that the measure is the most sensitive, that its sensitivity comes from identifiable design choices rather than from chance, and that the preferences it surfaces are women's own; whether those elicited preferences also predict subsequent behavior, and how programs should weigh them if they do not, is the empirical question that remains.

This tension also points beyond the indicators themselves. The use-level definitions count a switcher as aligned because the underlying 2×2 framework asks only whether preference and use agree on the *extensive* margin, i.e., whether a woman who wants to use is in fact using. But the margin on which we observe the most variation is the *intensive* one: not whether a woman uses, but whether she is using the method she would choose. An alignment framework organized solely around the extensive margin cannot represent this variation, which suggests that the field needs to move one step beyond the 2×2 framework toward a conception of alignment that treats method choice among users as a dimension of alignment in its own right.

These conclusions require further testing before any single measure is adopted at scale. Our estimates come from two settings and a single digital platform, and the within-country analyses, while preserving the central pattern, rest on modest per-country cell sizes. Establishing which approach best serves person-centered measurement will require replication across a wider range of settings, populations, and survey modes, and, most importantly, validation of elicited switching preferences against women's subsequent contraceptive behavior. As preference-based indicators move from conceptual proposals toward use in population monitoring, evidence of this kind is a prerequisite for measures that genuinely center women's own contraceptive and reproductive preferences.

7 Acknowledgments

We thank the International Union for the Scientific Study of Population (IUSSP) Panel on Rethinking Family Planning Measurement with a Rights- and Justice-Based Lens...Need to put a whole lot of acknowledgments here.

7.1 Ethical Considerations and Trial Registration

Ethical approval to conduct the study was received from the Columbia University Institutional Review Board (protocol number ACYY0849) and from the Population Council Institute (protocol number PCIIRB\2026-2027\003).

7.2 Other Disclosures

The findings, interpretations, and conclusions expressed in this study are entirely those of the authors. They do not necessarily represent the views of the World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

References

Bhan, N., A. Raj, E. E. Thomas, and P. Nanda (2022, December). Measuring womens agency in family planning: the conceptual and structural factors in the way. *Sexual and Reproductive Health Matters 30*(1), 2062161. _eprint: https://doi.org/10.1080/26410397.2022.2062161. 1, 25

Bingenheimer, J. B., K. Hardee, M. Hindin, A. Jain, J. Mumah, and J. v. Dam (2023). Introduction to the Special Issue: Indicators in Sexual and Reproductive Health and Rights. *Studies in Family Planning 54*(1), 9–16. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/sifp.12239. 1, 2

Bradley, S. E., T. Croft, J. Fishel, and C. Westoff (2012). Revising Unmet Need for Family Planning. DHS Analytical Studies No. 25, ICF International, Calverton, MD. 8

Bullington, B. W., S. Chung, C. W. Rothschild, D. Onyango, L. Senderowicz, E. Goland, A. Mwanyiro, B. Wekesa, B. Frizzelle, G. Golub, and K. Tumlinson (2025). Measuring misaligned contraceptive use among reproductive-aged women in Kisumu, Kenya: a cross-sectional, population-based study. *BMJ Public Health 3*, e001671. 2

Canning, D. and M. Karra (2023). Unwanted Family Planning: Prevalence Estimates for 56 Countries. *Studies in Family Planning 54*(1), 75–93. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/sifp.12230. 8

Cleland, J., S. Harbison, and I. H. Shah (2014, June). Unmet Need for Contraception: Issues and Challenges. *Studies in Family Planning 45*(2), 105–122. 1, 8

Dehlendorf, C., E. Fox, L. Sobel, and S. Borrero (2016, March). Patient-Centered Contraceptive Counseling: Evidence to Inform Practice. *Current Obstetrics and Gynecology Reports 5*(1), 55–63. 1

Diamond-Smith, N., R. Warnock, and M. Sudhinaraset (2018, August). Interventions to improve the person-centered quality of family planning services: a narrative review. *Reproductive Health 15*(1), 144. 1

Government of India (2022). National Family Health Survey, 2019-2021. Technical report, Ministry of Health and Family Welfare, New Delhi, India. 3

Holt, K., J. M. Caglia, E. Peca, J. M. Sherry, and A. Langer (2017, February). A call for collaboration on respectful, person-centered health care in family planning and maternal health. *Reproductive Health 14*(1), 20. 1

Holt, K., S. Challa, P. Alitubeera, L. Atuyambe, C. Dehlendorf, C. Galavotti, I. Idiodi, A. Jegede, E. Omoluabi, P. Waiswa, and U. Upadhyay (2024, February). Conceptualizing Contraceptive Agency: A Critical Step to Enable Human Rights-Based Family Planning Programs and Measurement. *Global Health: Science and Practice 12*(1). Publisher: Global

Health: Science and Practice. 1

Holt, K., C. Galavotti, E. Omoluabi, S. Challa, P. Waiswa, and J. Liu (2023, March). Preference-Aligned Fertility Management as a Person-Centered Alternative to Contraceptive Use-Focused Measures. *Studies in Family Planning 54*(1), 301–308. 2, 4, 6

Karra, M. (2022). Measurement of Unmet Need for Contraception: A Counterfactual Approach. *Studies in Family Planning 53*(4), 657–680. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/sifp.12216. 1

Kenya National Bureau of Statistics and ICF International (2023). Kenya Demographic and Health Survey 2022. Technical report, ICF International, Nairobi, Kenya. 3

Purdy, L. (2006, May). Women's reproductive autonomy: medicalisation and beyond. *Journal of Medical Ethics 32*(5), 287–291. 1, 25

Senderowicz, L. (2020). Contraceptive Autonomy: Conceptions and Measurement of a Novel Family Planning Indicator. *Studies in Family Planning 51*(2), 161–176. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/sifp.12114. 1, 2

Senderowicz, L., B. W. Bullington, N. Sawadogo, K. Tumlinson, A. Langer, A. Soura, P. Zabr, and A. Si (2023). Assessing the Suitability of Unmet Need as a Proxy for Access to Contraception and Desire to Use It. *Studies in Family Planning 54*(1), 231–250. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/sifp.12233. 1

Speizer, I. S., E. A. Sully, G. Binstock, and N. Saggurti (2025). Advancing New Directions for Family Planning Measurement. *Studies in Family Planning 56*(3), 361–371. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/sifp.70036. 1, 2

Speizer, I. S., E. A. Sully, Y. M. Hashem, and M. A. A. Oduro (2025). The State of Person-Centered Measurement for Family Planning Need and Use: A Scoping Review. *Studies in Family Planning 56*(3), 403–436. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/sifp.70019. 1

Upadhyay, U. D., S. L. Dworkin, T. A. Weitz, and D. G. Foster (2014). Development and Validation of a Reproductive Autonomy Scale. *Studies in Family Planning 45*(1), 19–41. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1728-4465.2014.00374.x. 1, 25

Vincent, S., C. Herrera-Almanza, S. Anukriti, and M. Karra (2025, September). Contraceptive Concordance. *Studies in Family Planning 56*(3), 532–549. 2, 4, 6

A Appendix

Figure A1: Experimental Design

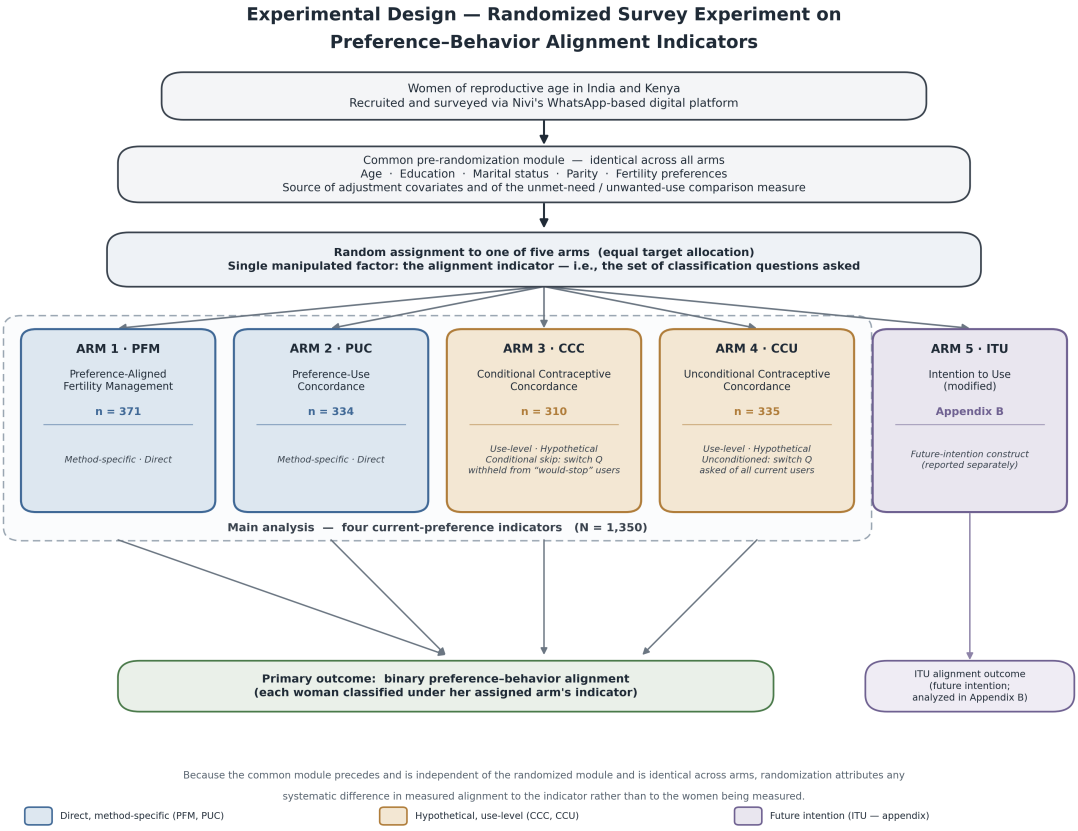

Table A1: Sample characteristics by arm

	PFM		PUC		CCC		CCU	
Country (Kenya = 1)	0.714	(0.452)	0.647	(0.479)	0.668	(0.472)	0.651	(0.477)
Language (English = 1)	0.757	(0.429)	0.751	(0.433)	0.739	(0.440)	0.728	(0.446)
Age (years)	26.181	(5.103)	26.841	(5.180)	26.542	(4.992)	26.078	(4.788)
University or Vocational / Adult Education (=1)	0.668	(0.471)	0.647	(0.479)	0.687	(0.464)	0.665	(0.473)
Marital Status (Married = 1)	0.605	(0.489)	0.610	(0.489)	0.594	(0.492)	0.620	(0.486)
Parity - Binary	0.771	(0.421)	0.793	(0.406)	0.810	(0.393)	0.805	(0.396)
Current use of FP (=1)	0.690	(0.463)	0.769	(0.422)	0.775	(0.418)	0.741	(0.439)
Wants none / no more	0.345	(0.476)	0.360	(0.481)	0.405	(0.492)	0.326	(0.469)
Wants to wait 2+ years	0.206	(0.405)	0.190	(0.393)	0.197	(0.399)	0.281	(0.450)
Wants to space 2+ / limit	0.552	(0.498)	0.550	(0.498)	0.602	(0.490)	0.607	(0.489)
Unmet Need (=1)	0.149	(0.356)	0.129	(0.335)	0.123	(0.329)	0.153	(0.360)
Unwanted FP (=1)	0.103	(0.304)	0.093	(0.291)	0.104	(0.305)	0.117	(0.322)
Concordance, control (1 - discordance) (=1)	0.749	(0.434)	0.778	(0.416)	0.773	(0.419)	0.731	(0.444)
Observations	371		334		310		335	

Note: Standard deviations in parentheses.

Table A2: Pairwise balance tests across arms

	PFM–PUC		PFM–CCC		PFM–CCU		PUC–CCC		PUC–CCU		CCC–CCU	
Country (Kenya = 1)	0.068	(0.055)	0.047	(0.192)	0.064	(0.071)	-0.021	(0.575)	-0.004	(0.913)	0.017	(0.650)
Language (English = 1)	0.006	(0.856)	0.019	(0.577)	0.030	(0.366)	0.013	(0.710)	0.024	(0.481)	0.011	(0.749)
Age (years)	-0.661	(0.089)	-0.361	(0.352)	0.103	(0.783)	0.299	(0.455)	0.763*	(0.048)	0.464	(0.230)
University or Vocational / Adult Education (=1)	0.022	(0.544)	-0.019	(0.605)	0.004	(0.915)	-0.040	(0.278)	-0.018	(0.626)	0.022	(0.544)
Marital Status (Married = 1)	-0.004	(0.909)	0.012	(0.754)	-0.014	(0.697)	0.016	(0.678)	-0.010	(0.788)	-0.026	(0.497)
Parity - Binary	-0.022	(0.483)	-0.039	(0.215)	-0.034	(0.263)	-0.017	(0.592)	-0.013	(0.685)	0.004	(0.890)
Current use of FP (=1)	-0.079*	(0.020)	-0.085*	(0.014)	-0.051	(0.142)	-0.006	(0.868)	0.028	(0.400)	0.034	(0.320)
Wants none / no more	-0.015	(0.695)	-0.060	(0.126)	0.019	(0.603)	-0.045	(0.265)	0.034	(0.375)	0.079*	(0.045)
Wants to wait 2+ years	0.016	(0.602)	0.009	(0.773)	-0.075*	(0.027)	-0.007	(0.827)	-0.091**	(0.008)	-0.084*	(0.016)
Wants to space 2+ / limit	0.002	(0.967)	-0.050	(0.204)	-0.055	(0.153)	-0.052	(0.203)	-0.057	(0.153)	-0.005	(0.903)
Unmet Need (=1)	0.020	(0.445)	0.026	(0.330)	-0.004	(0.881)	0.006	(0.826)	-0.024	(0.374)	-0.030	(0.275)
Unwanted FP (=1)	0.010	(0.659)	-0.001	(0.971)	-0.014	(0.552)	-0.011	(0.648)	-0.024	(0.313)	-0.013	(0.593)
Concordance, control (1 - discordance) (=1)	-0.030	(0.353)	-0.025	(0.450)	0.018	(0.585)	0.005	(0.880)	0.048	(0.151)	0.043	(0.208)
Observations	705		681		706		644		669		645	
Joint F-test p-value	0.203		0.451		0.206		0.616		0.163		0.485	
N	614		600		630		568		598		584	

Note: p-values in parentheses. Joint F-test from regression of arm assignment on all balance covariates within the pair. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A3: Means and standard deviations by arm

	PFM		PUC		CCC		CCU	
Aligned (=1)	0.537	(0.499)	0.692	(0.462)	0.343	(0.475)	0.705	(0.457)
Concordance, control (1 - discordance) (=1)	0.749	(0.434)	0.778	(0.416)	0.773	(0.419)	0.731	(0.444)
Unmet Need (=1)	0.149	(0.356)	0.129	(0.335)	0.123	(0.329)	0.153	(0.360)
Unwanted FP (=1)	0.103	(0.304)	0.093	(0.291)	0.104	(0.305)	0.117	(0.322)
Current use of FP (=1)	0.690	(0.463)	0.769	(0.422)	0.775	(0.418)	0.741	(0.439)
Wants none / no more	0.345	(0.476)	0.360	(0.481)	0.405	(0.492)	0.326	(0.469)
Wants to space 2+ / limit	0.552	(0.498)	0.550	(0.498)	0.602	(0.490)	0.607	(0.489)
Wants to wait 2+ years	0.206	(0.405)	0.190	(0.393)	0.197	(0.399)	0.281	(0.450)
Observations	371		334		309		334	

Note: Standard deviations in parentheses.

Table A4: Pairwise differences in means with t-test p-values

	PFM–PUC		PFM–CCC		PFM–CCU		PUC–CCC		PUC–CCU		CCC–CCU	
Aligned (=1)	-0.155***	(0.000)	0.195***	(0.000)	-0.167***	(0.000)	0.350***	(0.000)	-0.012	(0.746)	-0.362***	(0.000)
Concordance, control (1 - discordance) (=1)	-0.030	(0.353)	-0.025	(0.450)	0.018	(0.585)	0.005	(0.880)	0.048	(0.151)	0.043	(0.208)
Unmet Need (=1)	0.020	(0.445)	0.026	(0.330)	-0.004	(0.881)	0.006	(0.826)	-0.024	(0.374)	-0.030	(0.275)
Unwanted FP (=1)	0.010	(0.659)	-0.001	(0.971)	-0.014	(0.552)	-0.011	(0.648)	-0.024	(0.313)	-0.013	(0.593)
Current use of FP (=1)	-0.079*	(0.020)	-0.085*	(0.014)	-0.051	(0.142)	-0.006	(0.868)	0.028	(0.400)	0.034	(0.320)
Wants none / no more	-0.015	(0.695)	-0.060	(0.126)	0.019	(0.603)	-0.045	(0.265)	0.034	(0.375)	0.079*	(0.045)
Wants to space 2+ / limit	0.002	(0.967)	-0.050	(0.204)	-0.055	(0.153)	-0.052	(0.203)	-0.057	(0.153)	-0.005	(0.903)
Wants to wait 2+ years	0.016	(0.602)	0.009	(0.773)	-0.075*	(0.027)	-0.007	(0.827)	-0.091**	(0.008)	-0.084*	(0.016)
Observations	705		680		705		643		668		643	
Joint F-test p-value	0.164		0.000		0.014		0.000		0.207		0.000	
N	557		562		588		519		545		550	

Note: p-values in parentheses. Joint F-test from regression of arm assignment on all outcomes within the pair. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A5: Logit and AME estimates: probability of alignment by indicator (PFM = reference)

	Unadj. logit	Adj. logit	Adj. AME (pp)
main			
PUC vs PFM	0.661***	0.698***	0.162***
	(0.169)	(0.171)	(0.039)
CCC vs PFM	-0.801***	-0.784***	-0.189***
	(0.166)	(0.166)	(0.039)
CCU vs PFM	0.720***	0.738***	0.170***
	(0.166)	(0.168)	(0.038)
Kenya (vs India)		0.184	
		(0.197)	
English language		0.147	
		(0.203)	
Age (years)		-0.024*	
		(0.013)	
University or Vocational / Adult Education		-0.121	
		(0.135)	
Married		0.057	
		(0.155)	
Has 1+ children		0.336*	
		(0.174)	
Constant	0.150	0.307	
	(0.110)	(0.373)	
Joint test of arm equality: χ^2	96.65	98.51	115.56
degrees of freedom	3	3	3
p-value	0.000	0.000	0.000
Observations	1215	1214	1214

Note: Robust standard errors in parentheses. Cols. 1-2 report logit coefficients; Col. 3 reports adjusted average marginal effects (percentage points). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A6: Pairwise differences in alignment probability across indicators

		Diff.		
	Diff	p_unadj	p_bonf	p_sidak
PUC vs PFM	0.162	0.000	0.000	0.000
CCC vs PFM	-0.189	0.000	0.000	0.000
CCU vs PFM	0.170	0.000	0.000	0.000
CCC vs PUC	-0.351	0.000	0.000	0.000
CCU vs PUC	0.008	0.830	1.000	1.000
CCU vs CCC	0.359	0.000	0.000	0.000

Note: Each row is the row arm minus the column arm. Estimates from a linear probability model adjusting for country, language, age, education, marital status, and parity. p-values: unadjusted, Bonferroni-adjusted, and Sidak-adjusted (across the 6 pairwise contrasts).

Table A7: Pairwise differences with Sidak-adjusted 95% confidence intervals

	Diff	Diff. ci_lower	ci_upper
PUC vs PFM	0.162	0.059	0.264
CCC vs PFM	-0.189	-0.292	-0.086
CCU vs PFM	0.170	0.070	0.270
CCC vs PUC	-0.351	-0.453	-0.248
CCU vs PUC	0.008	-0.092	0.108
CCU vs CCC	0.359	0.258	0.459

Note: Confidence intervals are widened to control the experimentwise error rate via Sidak's correction across the 6 pairwise contrasts.

Table A8: Adjusted probability of alignment by indicator

	Adj. P(aligned)
PFM (Arm 1)	0.533*** [0.479,0.587]
PUC (Arm 2)	0.695*** [0.641,0.748]
CCC (Arm 3)	0.344*** [0.290,0.399]
CCU (Arm 4)	0.703*** [0.652,0.754]
Observations	1214

Note: 95% confidence intervals in brackets. Predicted probabilities are average predictive margins (average marginal standardization): for each arm, the model-predicted probability is computed for every woman as if she were assigned to that arm and then averaged over the observed covariate distribution, as described in section 3. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A9: Joint Wald tests of heterogeneity in arm effects by subgroup

	χ^2 chi2	df	p
Country (Kenya vs India)	9.55	3	0.023
Parity (1+ children vs none)	4.49	3	0.213
Education (university vs less)	5.91	3	0.116

Note: Each row is a joint Wald test of the arm × subgroup interaction terms in a logit regression of alignment on arm, the subgroup, their interaction, and other covariates.

Table A10: Predicted probability of alignment by arm and subgroup

	India					
	India	Kenya	NoKids	OneKids	LessUni	Uni
PFM	0.574	0.518	0.465	0.550	0.635	0.481
PUC	0.689	0.699	0.643	0.709	0.677	0.705
CCC	0.232	0.401	0.275	0.364	0.379	0.328
CCU	0.626	0.743	0.773	0.687	0.687	0.712

Note: Predicted probabilities from logit models stratified by subgroup, adjusting for other covariates.

Table A11: Alignment indicator vs control concordance within each arm

	P(aligned)			
	p_align	p_concord	kappa	mcnemar_p
PFM (Arm 1)	0.537	0.749	0.117	0.000
PUC (Arm 2)	0.692	0.778	0.275	0.037
CCC (Arm 3)	0.343	0.773	0.103	0.000
CCU (Arm 4)	0.705	0.731	0.226	0.686

Note: Both measures are observed on the same respondents within each arm. Cohen's κ measures agreement between the two binary classifications; low values indicate the measures capture different information. McNemar's test compares the marginal prevalence of the two measures (paired binary).

Table A12: Alignment indicator vs concordance by unmet need within each arm

	P(aligned)			
	p_align	demand_satisfied	kappa	mcnemar_p
PFM (Arm 1)	0.537	0.851	0.128	0.000
PUC (Arm 2)	0.692	0.871	0.206	0.000
CCC (Arm 3)	0.343	0.877	0.053	0.000
CCU (Arm 4)	0.705	0.847	0.334	0.000

Note: Both measures are observed on the same respondents within each arm. Cohen's κ measures agreement between the two binary classifications; low values indicate the measures capture different information. McNemar's test compares the marginal prevalence of the two measures (paired binary).

Table A13: Method-specific alignment: predicted probability by indicator

	Adj. P(method-specific aligned)
PFM (Arm 1)	0.534***
	[0.481,0.588]
PUC (Arm 2)	0.695***
	[0.641,0.748]
CCC mod. (Arm 3)	0.205***
	[0.157,0.253]
CCU mod. (Arm 4)	0.203***
	[0.157,0.248]
Observations	1189

Note: 95% confidence intervals in brackets. CCC and CCU are modified to treat switchers as discordant, matching the PFM/PUC method-specific definition. Predicted probabilities from logit model adjusting for country, language, age, education, marital status, and parity. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A14: Pairwise differences in method-specific alignment probability

	Diff.			
	Diff	p_unadj	p_bonf	p_sidak
PUC vs PFM	0.160	0.000	0.000	0.000
CCC (mod.) vs PFM	-0.330	0.000	0.000	0.000
CCU (mod.) vs PFM	-0.333	0.000	0.000	0.000
CCC (mod.) vs PUC	-0.489	0.000	0.000	0.000
CCU (mod.) vs PUC	-0.492	0.000	0.000	0.000
CCU (mod.) vs CCC (mod.)	-0.003	0.929	1.000	1.000

Note: CCC and CCU are modified to treat switchers as discordant. Linear probability model adjusting for country, language, age, education, marital status, and parity.

Table A15: Method-specific alignment by indicator: India only

	Adj. P(method-specific aligned): India
PFM (Arm 1)	0.571***
	[0.470,0.672]
PUC (Arm 2)	0.689***
	[0.598,0.779]
CCC mod. (Arm 3)	0.163***
	[0.085,0.241]
CCU mod. (Arm 4)	0.166***
	[0.097,0.236]
Observations	389

Note: 95% confidence intervals in brackets. Sample restricted to India. Logit adjusting for language, age, education, marital status, and parity (country omitted, constant within stratum). Per-country cell sizes are modest; interpret with caution. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A16: Method-specific alignment by indicator: Kenya only

	Adj. P(method-specific aligned): Kenya
PFM (Arm 1)	0.518***
	[0.456,0.581]
PUC (Arm 2)	0.699***
	[0.633,0.765]
CCC mod. (Arm 3)	0.226***
	[0.166,0.286]
CCU mod. (Arm 4)	0.222***
	[0.163,0.281]
Observations	800

Note: 95% confidence intervals in brackets. Sample restricted to Kenya. Logit adjusting for language, age, education, marital status, and parity (country omitted, constant within stratum). Per-country cell sizes are modest; interpret with caution. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A17: Pairwise differences in method-specific alignment: India only

	Diff	Diff. p_unadj	p_bonf
PUC vs PFM	0.114	0.105	0.629
CCC (mod.) vs PFM	-0.410	0.000	0.000
CCU (mod.) vs PFM	-0.408	0.000	0.000
CCC (mod.) vs PUC	-0.524	0.000	0.000
CCU (mod.) vs PUC	-0.522	0.000	0.000
CCU (mod.) vs CCC (mod.)	0.002	0.971	1.000

Note: Sample restricted to India. Linear probability model; country omitted (constant within stratum). Bonferroni adjustment across the six within-country contrasts.

Table A18: Pairwise differences in method-specific alignment: Kenya only

	Diff	Diff. p_unadj	p_bonf
PUC vs PFM	0.181	0.000	0.001
CCC (mod.) vs PFM	-0.292	0.000	0.000
CCU (mod.) vs PFM	-0.297	0.000	0.000
CCC (mod.) vs PUC	-0.473	0.000	0.000
CCU (mod.) vs PUC	-0.478	0.000	0.000
CCU (mod.) vs CCC (mod.)	-0.005	0.908	1.000

Note: Sample restricted to Kenya. Linear probability model; country omitted (constant within stratum). Bonferroni adjustment across the six within-country contrasts.

Table A19: Agreement between original and modified concordance measures

	P(original)			
	p_orig	p_mod	kappa	mcnemar_p
CCC (Arm 3)	0.343	0.203	0.670	0.000
CCU (Arm 4)	0.705	0.205	0.189	0.000

Note: Original counts switchers as preference-aligned (use-level definition). Modified counts switchers as discordant (method-specific definition).

Table A20: Distribution of respondents by alignment category and indicator (percentages

	PFM				
	PFM	PUC	CCC	CCU	UnmetNeed
Aligned: wanted user (satisfied)	43.2	58.7	14.4	12.9	41.7
Aligned: wanted non-user	11.2	14.9	5.9	7.8	11.3
Misaligned: unmet need	21.8	9.3	18.1	19.7	16.5
Misaligned: unwanted use	10.3	9.3	48.3	8.2	11.1
Misaligned: switcher	13.6	7.8	13.3	51.4	0.0
Using, no stated fertility-timing demand	0.0	0.0	0.0	0.0	19.4
Observations (N)	331	269	271	294	1396

Note: Each column reports the percentage of respondents in the relevant arm (or full sample for Unmet Need) classified into each category. The Unmet Need column applies the unwanted use definition from subsection 2.5 (currently using and wanting a child within two years), which is consistent with Table A1 and Table A3, and current users with no stated spacing or limiting preference are reported in the residual category.

Table A21: Distribution of respondents by alignment category and indicator: India (percentages)

	PFM				
	PFM	PUC	CCC	CCU	UnmetNeed
Aligned: wanted user (satisfied)	34.1	43.9	4.5	1.9	43.0
Aligned: wanted non-user	25.3	26.5	11.2	15.2	11.4
Misaligned: unmet need	16.5	13.3	27.0	30.5	31.4
Misaligned: unwanted use	15.4	10.2	50.6	6.7	6.2
Misaligned: switcher	8.8	6.1	6.7	45.7	0.0
Using, no stated fertility-timing demand	0.0	0.0	0.0	0.0	8.1
Observations (N)	91	98	89	105	484

Note: Each column reports the percentage of respondents in the relevant arm (or full sample for Unmet Need) classified into each category. The Unmet Need column applies the unwanted use definition from subsection 2.5 (currently using and wanting a child within two years), which is consistent with Table A1 and Table A3, and current users with no stated spacing or limiting preference are reported in the residual category.

Table A22: Distribution of respondents by alignment category and indicator: Kenya (percentages)

		PFM			
	PFM	PUC	CCC	CCU	UnmetNeed
Aligned: wanted user (satisfied)	46.7	67.3	19.2	19.0	41.0
Aligned: wanted non-user	5.8	8.2	3.3	3.7	11.3
Misaligned: unmet need	23.8	7.0	13.7	13.8	8.6
Misaligned: unwanted use	8.3	8.8	47.3	9.0	13.7
Misaligned: switcher	15.4	8.8	16.5	54.5	0.0
Using, no stated fertility-timing demand	0.0	0.0	0.0	0.0	25.4
Observations (N)	240	171	182	189	912

Note: Each column reports the percentage of respondents in the relevant arm (or full sample for Unmet Need) classified into each category. The Unmet Need column applies the unwanted use definition from subsection 2.5 (currently using and wanting a child within two years), which is consistent with Table A1 and Table A3, and current users with no stated spacing or limiting preference are reported in the residual category.

B Intention to Use: The Fifth Indicator

The main analysis compares the four indicators that classify women by their *current* method preference. A fifth arm fielded a modified measure of intention to use contraception (ITU). Unlike the conventional ITU indicator, which is defined only among non-users, our measure classifies both users and non-users, so that an ITU-based alignment outcome can be placed on the same footing as the other four arms and can also be entered into the same regression framework. We report ITU separately here, rather than alongside PFM, PUC, CCC, and CCU, because it is anchored on a woman's *future* intention to use rather than on whether her current behavior matches her current preference. This makes ITU a conceptually distinct construct: it asks whether a woman intends to use contraception going forward, not whether she is presently in a state of preference-behavior alignment. The results below should therefore be read as a supplementary comparison rather than as a fifth entry in the main ranking of indicators.

Table B.1 reports unadjusted means by arm with ITU included. ITU produces the lowest unadjusted alignment of any arm, at 23.4 percent, well below even CCC (34.3 percent). ITU's low alignment is a property of the indicator: because it is anchored on future intention, it classifies as misaligned a larger share of women whose present behavior the other indicators would treat as concordant. The ITU arm also shows the lowest current use (63.7 percent) and the highest unmet need (19.6 percent) of the five arms, but these are descriptive quantities measured from the common pre-randomization block rather than products of the ITU classification, so their arm-level differences are better read as chance imbalance, complete-case denominator differences, or possible measurement differences across arms than as consequences of ITU's future-oriented logic.

Table B.2 reports the logit and average marginal effect (AME) estimates with ITU added as a fifth arm and PFM as the reference. The omnibus Wald test of arm equality is rejected at $p < 0.001$ (now on four degrees of freedom to reflect the five-arm design). ITU is the most discordant arm relative to PFM, where the adjusted AME is -30.0 p.p., roughly twice the magnitude of the CCC-PFM gap. Table B.4 expresses the same result as adjusted predicted probabilities evaluated at sample covariate means: ITU alignment is 23.5 percent (95% CI 16.3-30.7), the lowest of the five indicators and well separated from the direct elicitation measures. Table B.3 reports all ten pairwise contrasts; every contrast involving ITU is large and remains significant at $p < 0.001$ under both Bonferroni and Sidak correction across the ten comparisons. The only contrast that does not reach significance under adjustment remains CCU vs PUC, as in the four-arm analysis.

Table B.5 assesses how ITU agrees with the two traditional benchmarks within the Arm 5 sample. Agreement is effectively absent: Cohen's κ is -0.016 against control concordance and 0.004 against the complement of unmet need, both around zero, and the McNemar test rejects equality of marginal prevalence at $p < 0.001$ in each case. ITU thus captures information that is not merely distinct from the traditional measures but essentially uncorrelated with them at the individual level.

Finally, Table B.6 decomposes the ITU arm into the five alignment categories. ITU distributes its misaligned women differently from the contraceptive-concordance indicators: 33.6 percent are classified as unmet need and the same as unwanted use, while only 9.5 percent are identified as switchers (aligned categories account for the remaining 23.4 percent:

8.8 percent satisfied users and 14.6 percent wanted non-users). Where CCU's unconditioned design loads misalignment onto the switcher category, ITU's future-intention framing instead loads it onto the extensive-margin categories of unmet need and unwanted use, reflecting that the question turns on whether a woman intends to use at all rather than on whether her current method is the one she would choose.

Taken together, these results situate ITU relative to the main findings without displacing them. When the use-level indicators are re-constructed on the method-specific footing used in the main analysis (Table A13), the hypothetical elicitation measures converge near 20 percent; ITU's adjusted alignment of 23.5 percent sits just above that band. This proximity is consistent with the elicitation-effect interpretation developed in the main text—hypothetical and future-oriented framings surface more reported discordance than direct questions about the current method—but it should not be read as evidence that ITU measures the same construct. As discussed in Section 5, ITU's future orientation introduces an ambiguity that the present-focused indicators avoid, since an intention stated at no immediate behavioral cost need not reflect a woman's current preference. We therefore retain ITU as an informative point of comparison while basing the study's substantive conclusions on the four current-preference indicators.

Table B.1: Appendix: Means and standard deviations by arm, including ITU

	PFM		PUC		CCC		CCU		ITU	
Aligned (=1)	0.537	(0.499)	0.692	(0.462)	0.343	(0.475)	0.705	(0.457)	0.234	(0.425)
Concordance, control (1 - discordance) (=1)	0.749	(0.434)	0.778	(0.416)	0.773	(0.419)	0.731	(0.444)	0.735	(0.442)
Unmet Need (=1)	0.149	(0.356)	0.129	(0.335)	0.123	(0.329)	0.153	(0.360)	0.196	(0.398)
Unwanted FP (=1)	0.103	(0.304)	0.093	(0.291)	0.104	(0.305)	0.117	(0.322)	0.068	(0.253)
Current use of FP (=1)	0.690	(0.463)	0.769	(0.422)	0.775	(0.418)	0.741	(0.439)	0.637	(0.482)
Wants none / no more	0.345	(0.476)	0.360	(0.481)	0.405	(0.492)	0.326	(0.469)	0.402	(0.492)
Wants to space 2+ / limit	0.552	(0.498)	0.550	(0.498)	0.602	(0.490)	0.607	(0.489)	0.613	(0.488)
Wants to wait 2+ years	0.206	(0.405)	0.190	(0.393)	0.197	(0.399)	0.281	(0.450)	0.211	(0.409)
Observations	371		334		309		334		219	

Note: Standard deviations in parentheses. ITU (Arm 5) captures future intention and is reported separately from the four method-specific indicators.

Table B.2: Appendix: Logit and AME estimates including ITU (PFM = reference)

	Unadj. logit	Adj. logit	Adj. AME (pp)
main			
PUC vs PFM	0.661***	0.683***	0.159***
	(0.169)	(0.170)	(0.039)
CCC vs PFM	-0.801***	-0.792***	-0.191***
	(0.166)	(0.166)	(0.039)
CCU vs PFM	0.720***	0.727***	0.169***
	(0.166)	(0.167)	(0.038)
ITU vs PFM	-1.338***	-1.326***	-0.300***
	(0.230)	(0.233)	(0.046)
Kenya (vs India)		0.088	
		(0.187)	
English language		0.045	
		(0.196)	
Age (years)		-0.016	
		(0.013)	
University or Vocational / Adult Education		-0.093	
		(0.128)	
Married		-0.032	
		(0.146)	
Has 1+ children		0.281*	
		(0.165)	
Constant	0.150	0.333	
	(0.110)	(0.357)	
Joint test of arm equality: χ^2	144.94	143.73	190.12
degrees of freedom	4	4	4
p-value	0.000	0.000	0.000
Observations	1352	1351	1351

Note: Standard errors in parentheses. Robust standard errors in parentheses. Columns 1-2 report logit coefficients; Column 3 reports adjusted average marginal effects (percentage points). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.3: Appendix: Pairwise differences in alignment probability, including ITU

	Diff.			
	Diff	p_unadj	p_bonf	p_sidak
PUC vs PFM	0.159	0.000	0.000	0.000
CCC vs PFM	-0.192	0.000	0.000	0.000
CCU vs PFM	0.168	0.000	0.000	0.000
ITU vs PFM	-0.300	0.000	0.000	0.000
CCC vs PUC	-0.351	0.000	0.000	0.000
CCU vs PUC	0.009	0.807	1.000	1.000
ITU vs PUC	-0.459	0.000	0.000	0.000
CCU vs CCC	0.360	0.000	0.000	0.000
ITU vs CCC	-0.108	0.020	0.200	0.183
ITU vs CCU	-0.468	0.000	0.000	0.000

Note: Each row is the row arm minus the column arm. Linear probability model adjusting for country, language, age, education, marital status, and parity. p-values: unadjusted, Bonferroni-adjusted, and Sidak-adjusted across the 10 pairwise contrasts.

Table B.4: Appendix: Adjusted probability of alignment by indicator, including ITU

	Adj. P(aligned)
PFM (Arm 1)	0.535***
	[0.481,0.588]
PUC (Arm 2)	0.694***
	[0.640,0.747]
CCC (Arm 3)	0.343***
	[0.289,0.398]
CCU (Arm 4)	0.703***
	[0.652,0.754]
ITU (Arm 5)	0.235***
	[0.163,0.307]
Observations	1351

Note: 95% confidence intervals in brackets. Predicted probabilities from logit model adjusting for country, language, age, education, marital status, and parity, evaluated at sample means of covariates. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.5: Appendix: ITU alignment vs control concordance and unmet need (Arm 5)

	P(aligned)			
	p_align	p_compare	kappa	mcnemar_p
ITU vs control concordance	0.234	0.735	-0.016	0.000
ITU vs unmet-need concordance	0.234	0.804	0.004	0.000

Note: Both measures are observed on the same Arm 5 respondents. Cohen's κ measures agreement between the two binary classifications.

Table B.6: Appendix: Distribution of ITU (Arm 5) respondents by alignment category (percentages)

	ITU ITU
Aligned: wanted user (satisfied)	8.8
Aligned: wanted non-user	14.6
Misaligned: unmet need	33.6
Misaligned: unwanted use	33.6
Misaligned: switcher	9.5
Observations (N)	137

Note: Column reports the percentage of Arm 5 respondents in each category. Denominator in the bottom row.

C Survey Instrument

Testing New Measures of Preference-Behavior Alignment in Family Planning and Reproductive Health Decision-Making

Version Date: **26 March 2026**

Respondent ID Generation

Record the following place and date information BEFORE administering the RECRUITMENT SCRIPT, page 2.

1. Respondent Country Code (Enter 01-02): ☐☐
   1) India
   2) Kenya

2. Interview number – autogenerated 4 digits describing the sequence when the respondent is completing the survey. For example, if the respondent is the 15th person being surveyed, the interview number is '0015'.

☐☐☐☐

The **respondent ID number** is a 6-digit number that is created by joining the 2 numerical responses recorded above in order. An example of a complete respondent ID number is given below.

Example: For the 280th respondent who is interviewed in Kenya, the **respondent ID number** will be created as:

0	2	0	2	8	0

RESPONDENT ID NUMBER: ☐☐☐☐☐☐

Record assigned number above before proceeding to RECRUITMENT SCRIPT, page 2.

Recruitment Script

INTRODUCTION [ON WHATSAPP PLATFORM]

Nivi is conducting a survey and is asking questions about your family planning and reproductive health. You may participate if:

1. You are a woman between the ages of 18-49.
2. You are not currently pregnant.

This survey takes only about 5 minutes, and the information provided will be used to better understand this important issue. You will not be asked for your name, address, or other personal information that can identify you. The survey will not include your IP address. You don't have to answer any question you don't want to, and you can end the survey at any time. Any information you provide will be confidential.

DETERMINING PARTICIPATION

No.	Question	Response	Skip Rules
1	Are you willing to participate in this survey?	YES NO NOT NOW	IF NO, GO TO **END** IF YES, GO TO **THANKS**
END	Thank you for your time. Have a good day.		**END VISIT.**

THANKS

Thank you for agreeing to participate. You will now begin the survey.

Testing New Measures of Preference-Behavior Alignment in Family Planning and Reproductive Health Decision-Making

Respondent ID Number: | | | | | | |

PART 1
Background

We will ask you for some information about yourself.			
101	How old were you on your last birthday? RECORD AGE IN COMPLETED YEARS. IF DON'T KNOW, RECORD 88. IF REFUSED, RECORD 99.	AGE:	
102	What is the highest standard that you have completed?	Not Attended School……..……..…0 Not Completed Primary School……..…...…1 Completed Primary School……2 Completed Secondary School……3 Completed University……...……4 Completed Adult Education / Vocational Training…….5 Don't Know……..……88 Refused……..…..99	

PART 2
Reproduction

We would like to ask about all the births you have had during your life.

201	How many children have you given birth to in your lifetime? IF NONE, RECORD, RECORD 0. IF DON'T KNOW, RECORD 88. IF REFUSED, RECORD 99.	☐☐	**IF 0, GO TO 202A**
202A	Would you like to have a child, or would you prefer not to have any children?	Have a child……...1 None…..……...…….…….2 Says they can't get pregnant……..…...…….…...3 Up to God / fatalistic…….....4 Don't know…….….88 Refused…….....…….99	**IF 1, GO TO 203** **ELSE, GO TO S.3**
202B	Would you like to have another child, or would you prefer not to have any more children?	Have another child……..……....1 No more……...…….…….2 Says they can't get pregnant……..…...…….…...3 Up to God / fatalistic…….....4 Don't know…….….88 Refused…….....…….99	**IF 1, GO TO 203** **ELSE, GO TO S.3**
203	How long would you like to wait from now before the birth of (a/another) child?	Soon/Now……..……....1 Not now, but within next 1 year.…….……....2 More than 1 year but within next 2 years…..…....3 2 years or more……...4 Don't know…….….88 Refused…….....…….99	

PART 3
Contraception

[RANDOMIZE PARTICIPANT INTO ARM A, ARM B, ARM C, ARM D, OR ARM E]

ARM A: Preference Aligned Fertility Management – ORIGINAL

A301	Do you currently want to be using any method to avoid pregnancy – that is, to do something to keep it from happening? That is, I am asking whether you want to use a method to keep you from becoming pregnant, not whether you are using one.	Yes……...……...……..…1 No……...……...…...…0 Don't Know…….…….88 Refused…..…….…....99	
A302	Are you currently using any method to avoid pregnancy?	Yes……...……...……..…1 No……...……...…...…0 Sometimes..……...……..2 Refused…..…….…....99	**IF 0 OR 99, GO TO END**
A303	Which methods are you currently using? SELECT ALL MENTIONED.	Female Sterilization……...……....1 Male Sterilization……...…...2 IUD/PPIUD/Multiload for 5 years….....….3 Injectables…….……....4 Implants…….…….....5 Pill…….....6 Condom……...7 Female Condom……...8 Emergency Contraception……9 Diaphragm…….....10 Foam/Jelly…….11 Standard Days Method……...12 Lactational Amen. Method……...…..13 Rhythm Method……..……...14 Withdrawal…….15 Other Modern Method…….…..16 Other Traditional Method……..17 Don't know ……..……88 Refused…….……99	
A304	[REPEAT FOR EACH METHOD REPORTED IN A303] Do you want to be using [METHOD] right now?	Yes……...……...……..…1 No……...……...…...…0 Don't Know…….…….88 Refused…..…….…....99	**GO TO END**

ARM B: Preference-Use Concordance for Contraception

B301	Are you currently using any method to avoid pregnancy – that is, to do something to keep it from happening?	Yes……1 No……0 Sometimes……2 Refused……99	**IF 0 OR 99, GO TO B305**
B302	Which methods are you currently using? SELECT ALL MENTIONED.	Female Sterilization……1 Male Sterilization……2 IUD/PPIUD/Multiload for 5 years……3 Injectables……4 Implants……5 Pill……6 Condom……7 Female Condom……8 Emergency Contraception……9 Diaphragm……10 Foam/Jelly……11 Standard Days Method……12 Lactational Amen. Method……13 Rhythm Method……14 Withdrawal……15 Other Modern Method……16 Other Traditional Method……17 Don't know ……88 Refused……99	
B303	You said you are using a method right now to prevent pregnancy. Is that what you want – to be doing or using something to prevent pregnancy right now?	Yes……1 No……0 Don't Know……88 Refused……99	
B304	[CHECK B302. REPEAT FOR EACH METHOD REPORTED IN B302] You said you are using [METHOD]. Do you want to be using [METHOD] to prevent a pregnancy right now?	Yes……1 No……0 Don't Know……88 Refused……99	**GO TO END**
B305	You just said you are not using anything to prevent pregnancy right now. Is that what you want – to not be doing or using anything to prevent pregnancy right now?	Yes……1 No……0 Don't Know……88 Refused……99	**GO TO END**

ARM C: Contraceptive Concordance – CONDITIONED

C301	Are you currently doing something or using any method to delay or avoid getting pregnant?	Yes……..……..……..…….1 No.……..……..……..……2 Don't Know……..…….……88 Refused…….…….99	**IF 2, 88, OR 99, GO TO C306**
C302	Which method(s) are you currently using? SELECT ALL MENTIONED.	Female Sterilization……...……....1 Male Sterilization……...…...2 IUD/PPIUD/Multiload for 5 years….....….3 Injectables…….……...4 Implants…….…….....5 Pill…….....6 Condom……...7 Female Condom……...8 Emergency Contraception……9 Diaphragm…….....10 Foam/Jelly…….11 Standard Days Method……...12 Lactational Amen. Method……...…..13 Rhythm Method……..……...14 Withdrawal…….15 Other Modern Method…….…..16 Other Traditional Method……..17 Don't know ……..……88 Refused…….……99	
C303	If you could stop using your method, would you choose to stop?	Yes……..……..……..…….1 No.……..……..……..……2 Don't Know……..…….……88 Refused…….…….99	**IF 1, GO TO END**
C304	If you could switch to another method, would you choose to switch?	Yes……..……..……..…….1 No.……..……..……..……2 Don't Know……..…….……88 Refused…….…….99	**IF 2, 88, OR 99, GO TO END**
C305	Which method(s) would you want to switch to? SELECT ALL MENTIONED.	Female Sterilization……...……....1 Male Sterilization……...…...2 IUD/PPIUD/Multiload for 5 years….....….3 Injectables…….……...4 Implants…….…….....5 Pill…….....6 Condom……...7 Female Condom……...8 Emergency Contraception……9 Diaphragm…….....10 Foam/Jelly…….11 Standard Days Method……...12 Lactational Amen. Method……...…..13 Rhythm Method……..……...14 Withdrawal…….15 Other Modern Method…….…..16 Other Traditional Method……..17 Don't know ……..……88 Refused…….……99	**GO TO END**

C306	If you could use a method, would you choose to use a method?	Yes……..……...……..…….1 No.……..……...……..……2 Don't Know……..…….……88 Refused…….……99	**IF 2, 88, OR 99, GO TO END**
C307	Which method(s) would you want to use? SELECT ALL MENTIONED.	Female Sterilization……...……....1 Male Sterilization……...…...2 IUD/PPIUD/Multiload for 5 years….....….3 Injectables…….……...4 Implants…….…….....5 Pill…….....6 Condom……...7 Female Condom……...8 Emergency Contraception……9 Diaphragm…….....10 Foam/Jelly…….11 Standard Days Method……...12 Lactational Amen. Method……...…..13 Rhythm Method……..……...14 Withdrawal…….15 Other Modern Method…….…..16 Other Traditional Method……..17 Don't know ……..……88 Refused……..……99	**GO TO END**

ARM D: Contraceptive Concordance – UNCONDITIONED

D301	Are you currently doing something or using any method to delay or avoid getting pregnant?	Yes……..……..……..…….1 No.……..……..……..……2 Don't Know……..…….……88 Refused…….…….99	**IF 2, 88, OR 99, GO TO D306**
D302	Which method(s) are you currently using? SELECT ALL MENTIONED.	Female Sterilization……...……....1 Male Sterilization……...…...2 IUD/PPIUD/Multiload for 5 years….....….3 Injectables…….……...4 Implants…….…….....5 Pill…….....6 Condom……...7 Female Condom……...8 Emergency Contraception……9 Diaphragm…….....10 Foam/Jelly…….11 Standard Days Method……...12 Lactational Amen. Method……...…..13 Rhythm Method……..……...14 Withdrawal…….15 Other Modern Method…….…..16 Other Traditional Method……..17 Don't know ……..……88 Refused…….……99	
D303	If you could stop using your method, would you choose to stop?	Yes……..……..……..…….1 No.……..……..……..……2 Don't Know……..…….……88 Refused…….…….99	
D304	If you could switch to another method, would you choose to switch?	Yes……..……..……..…….1 No.……..……..……..……2 Don't Know……..…….……88 Refused…….…….99	**IF 2, 88, OR 99, GO TO END**
D305	Which method(s) would you want to switch to? SELECT ALL MENTIONED.	Female Sterilization……...……....1 Male Sterilization……...…...2 IUD/PPIUD/Multiload for 5 years….....….3 Injectables…….……...4 Implants…….…….....5 Pill…….....6 Condom……...7 Female Condom……...8 Emergency Contraception……9 Diaphragm…….....10 Foam/Jelly…….11 Standard Days Method……...12 Lactational Amen. Method……...…..13 Rhythm Method……..……...14 Withdrawal…….15 Other Modern Method…….…..16 Other Traditional Method……..17 Don't know ……..……88 Refused…….……99	**GO TO END**

D306	If you could use a method, would you choose to use a method?	Yes……..……...……..…….1 No.……..……...……..……2 Don't Know……..…….……88 Refused…….…….99	**IF 2, 88, OR 99, GO TO END**
D307	Which method(s) would you want to use? SELECT ALL MENTIONED.	Female Sterilization……...……....1 Male Sterilization……...…...2 IUD/PPIUD/Multiload for 5 years….....….3 Injectables…….……...4 Implants…….…….....5 Pill…….....6 Condom……...7 Female Condom……...8 Emergency Contraception……9 Diaphragm…….....10 Foam/Jelly…….11 Standard Days Method……...12 Lactational Amen. Method……...…..13 Rhythm Method……..……...14 Withdrawal…….15 Other Modern Method…….…..16 Other Traditional Method……..17 Don't know ……..…….88 Refused……..…….99	**GO TO END**

ARM E: Intention to Use

E301	Are you currently doing something or using any method to delay or avoid getting pregnant?	Yes……..……...……..…….1 No.……..……...……..……2 Don't Know……..…….……88 Refused…….…….99	**IF 2, 88, OR 99, GO TO E307**
E302	Which methods are you currently using? SELECT ALL MENTIONED.	Female Sterilization……...……....1 Male Sterilization……...…...2 IUD/PPIUD/Multiload for 5 years….....….3 Injectables…….……...4 Implants…….…….....5 Pill…….....6 Condom……...7 Female Condom……...8 Emergency Contraception……9 Diaphragm…….....10 Foam/Jelly…….11 Standard Days Method……...12 Lactational Amen. Method……...…..13 Rhythm Method……..……...14 Withdrawal…….15 Other Modern Method…….…..16 Other Traditional Method……..17 Don't know ……..……88 Refused…….……99	
E303	How likely is it that you will continue to use this method in the next 12 months?	Highly likely…….……...…...……..…….1 Likely…..……...……..……2 Unlikely...……...……..……3 Highly unlikely…….…….....……..…….4 Don't Know……..…….……88 Refused…….…….99	
E304	How likely is it that you will discontinue family planning completely in the next 12 months?	Highly likely…….……...…...……..…….1 Likely…..……...……..……2 Unlikely...……...……..……3 Highly unlikely…….…….....……..…….4 Don't Know……..…….……88 Refused…….…….99	
E305	How likely is it that you will switch to a different FP method?	Highly likely…….……...…...……..…….1 Likely…..……...……..……2 Unlikely...……...……..……3 Highly unlikely…….…….....……..…….4 Don't Know……..…….……88 Refused…….…….99	**IF 3, 4, 88 OR 99, GO TO END**
E306	What method do you think you will switch to? SELECT ALL MENTIONED.	Female Sterilization……...……....1 Male Sterilization……...…...2 IUD/PPIUD/Multiload for 5 years….....….3 Injectables…….……...4 Implants…….…….....5 Pill…….....6 Condom……...7 Female Condom……...8 Emergency Contraception……9 Diaphragm…….....10	**GO TO END**

		Foam/Jelly……..11 Standard Days Method……...12 Lactational Amen. Method……...…..13 Rhythm Method……..……...14 Withdrawal…….15 Other Modern Method…….…..16 Other Traditional Method……..17 Don't know ……..…….88 Refused……..…….99	
E307	Do you intend or plan to use a method to avoid or delay pregnancy at any point in the future, that is to do something to keep it from happening?	Yes……..……...……..…….1 No.……..……...……..……2 Don't Know……..…….……88 Refused……..…….99	**IF 2, 88, OR 99, GO TO END**
E308	When do you think you will start to use a family planning method?	Months…… ☐☐ Years…… ☐☐ Soon/Now……..……....1 After marriage……...…….……....2 After next pregnancy……..…..…....3 After reaching ideal family size……..……...4 Don't know…….….88 Refused……...…….99	
E309	What method do you think you will use? SELECT ALL MENTIONED.	Female Sterilization……...…….....1 Male Sterilization……...…...2 IUD/PPIUD/Multiload for 5 years….....….3 Injectables…….……....4 Implants…….…….....5 Pill…….....6 Condom……...7 Female Condom……...8 Emergency Contraception……9 Diaphragm…….....10 Foam/Jelly…….11 Standard Days Method……...12 Lactational Amen. Method……...…..13 Rhythm Method……..……...14 Withdrawal…….15 Other Modern Method…….…..16 Other Traditional Method……..17 Don't know ……..…….88 Refused……..…….99	

END

We have now reached the end of this survey. Thank you for your time and participation. Have a good day.

SAVE AND END SURVEY