NBER WORKING PAPER SERIES

REDUCING BIAS AMONG HEALTH CARE PROVIDERS: EXPERIMENTAL EVIDENCE FROM TANZANIA, BURKINA FASO, AND PAKISTAN

Zachary Wagner
Corrina Moucheraud
Manisha Shah
Alexandra Wollum
Willa H. Friedman
William H. Dow

Working Paper 31269
http://www.nber.org/papers/w31269

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2023

This work was funded by the Bill and Melinda Gates Foundation The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Reducing Bias Among Health Care Providers: Experimental Evidence From Tanzania, Burkina Faso, and Pakistan

Zachary Wagner, Corrina Moucheraud, Manisha Shah, Alexandra Wollum, Willa H. Friedman, and William H. Dow

NBER Working Paper No. 31269
May 2023
JEL No. D12,I11,I12,O12

ABSTRACT

Bias among health care providers can lead to poor-quality care and poor health outcomes, and it can exacerbate disparities. We use a randomized controlled trial to evaluate an intervention to reduce family planning provider bias towards young women in 227 clinics in Tanzania, Burkina Faso, and Pakistan. The intervention educated providers about bias towards young women, facilitated communication about bias with other providers, and offered non-financial public awards to clinics with the least biased care. After 12 months, the intervention led to less-biased attitudes and beliefs among providers and more comprehensive counseling. Clients also perceived better treatment at intervention clinics compared to control clinics. Despite reductions in reported bias, we find mixed evidence regarding changes in method dispensing

Zachary Wagner
RAND Corporation
1776 Main Street
Santa Monica, CA 90401
zwagner@rand.org

Corrina Moucheraud
University of California, Los Angeles
cmoucheraud@ucla.edu

Manisha Shah
Department of Public Policy
University of California, Los Angeles
Luskin School of Public Affairs 3250
Public Affairs Building
Los Angeles, CA 90095-1656
and NBER
manishashah@ucla.edu

Alexandra Wollum
University of California, Los Angeles
awollum@ucla.edu

Willa H. Friedman
Department of Economics
University of Houston
3623 Cullen Boulevard Room 204
Houston, TX 77204-5019
willa.friedman@gmail.com

William H. Dow
University of California, Berkeley
School of Public Health
2121 Berkeley Way #5324
Berkeley, CA 94720-7360
and NBER
wdow@berkeley.edu

A randomized controlled trials registry entry is available at
https://www.socialscienceregistry.org/trials/5354

1 Introduction

Bias among health care providers towards certain types of patients is well-documented in a variety of settings (Mishra et al., 2021; Kwan, 2020; Ikeda et al., 2019; Vela et al., 2022). Negative provider biases can lead to poor-quality care and poor health outcomes, and can exacerbate disparities (Vela et al., 2022). Provider bias stems from beliefs about certain groups often based on social norms and cultural beliefs (Greenwald and Krieger, 2006). As such, reducing provider bias is complicated and interventions that attempt to do so are nascent and with mixed results (Vela et al., 2022). Is provider bias in the health sector malleable, and if so, does reducing provider bias lead to better outcomes for groups that experience bias? We examine this issue in the context of bias in family planning (FP) service delivery in Tanzania, Burkina Faso, and Pakistan, three countries where over half of young women do not have modern contraception needs satisfied and unintended pregnancy rates are high (Kantorová et al., 2021).

Bias based on age in FP care in low- and middle income countries (LMICs) is well-documented (see Solo and Festin (2019) for a review). This bias can stem from paternalistic attitudes towards young women, such as wanting to protect youth from poor decisions, as well as judgment and animosity, such as negative feelings toward clients perceivesd to be promiscuous. Bias against providing FP services to young clients in particular may be driven by the fact that young women are more likely to be unmarried or nulliparous, two characteristics that are sometimes stigmatized in FP settings. For example, providers may believe that unmarried women should not have sex or that married women without children should focus on having children before using contraception. Provider surveys from India, Kenya, Senegal, and Nigeria reveal that providers sometimes refuse services or do not offer certain methods if a client is "too young," not married, or does not have children (Calhoun et al., 2013; Schwandt et al., 2017; Tumlinson et al., 2015; Sidze et al., 2014). A recent study from Nigeria used "mystery clients" to show that older, married women with children received far better services than younger, unmarried women without children (Sieverding et al., 2018). A discrete choice experiment (DCE) conducted as formative work for the current study finds that providers report more restrictive care to unmarried and/or nulliparous women, regardless of age (Dieci et al., 2021). Young women also have a higher unmet need for contraception than older women, which could be partly ex-

plained by provider bias at FP clinics (Kantorová et al., 2021). This evidence of unequal service provision suggests that interventions to reduce provider bias could improve services offered to young women and reduce age-based disparities in modern contraception use.

In this paper, we study the impact of The Beyond Bias project, which set out to reduce provider bias experienced by young women, aged 15-24. The year-long intervention was designed to be broadly appropriate for LMIC settings, but as effects may be context-dependent, we test across three disparate settings: Tanzania, Burkina Faso, and Pakistan. The objective was to change harmful beliefs held by FP providers about young FP clients, which in turn might increase the range of contraceptive methods offered and improve the quality of interactions from the perspective of young clients. The multi-pronged intervention disseminated information to providers about provider bias in FP services, facilitated peer-support and knowledge sharing amongst providers, and fostered peer-comparison through non-financial performance-based incentives which honored the best performing and most improved clinics at quarterly award ceremonies.

In September of 2020, we implemented a 12-month long randomized controlled trial (RCT) with 227 clinics across the three countries to evaluate the impact of the intervention. We use data from client exit surveys (continuously collected), mystery client visits (collected at endline), provider surveys (collected at endline), and administrative data on service delivery statistics (reported monthly) to assess outcomes.

Our first key finding is that the intervention reduces biases in attitudes and beliefs as measured in provider surveys – an important and novel finding given the widely perceived difficulty in counteracting deep cultural norms. An *unbiased index* made up of several measures of biased attitudes and beliefs increased by 0.2 standard deviations (SDs) in all three countries ($p<.01$) at endline compared to the control group. We observe improvements in provider attitudes and beliefs related to technical FP knowledge (e.g., appropriateness of long acting methods for young women) as well as attitudes and beliefs related to underlying social and cultural norms about sexual activity and FP use (e.g., acceptability of young women having sex before marriage).

Our second key finding is that the intervention improves some key provider behaviors, as reported by real and "mystery" clients. This indicates that changes in provider-reported

beliefs are not just reflecting social desirability bias, but that a bias-reduction intervention can improve FP care quality. Intervention providers are more likely to counsel young women on the full range of FP methods,[1] ask more key questions, and provide more information about the methods compared to control providers. The intervention also improved the way clients felt they were treated by the provider, and clients were more likely to say they would recommend the clinic to a friend. These improvements in counseling and perceived treatment are mostly driven by improvements in Tanzania and Pakistan; effect sizes for these outcomes in Burkina Faso are small and not statistically significant.

The third key finding is that despite improvements in the number of methods and information discussed during counseling, we find mixed evidence with respect to the types of methods received by young clients in the exit survey data. While young clients in Tanzania are 50% less likely to leave without a method from treatment clinics compared to control clinics, there is no change in this outcome in Burkina Faso or Pakistan. However, we also note that most FP clients received a method in all three countries at baseline so there was limited room for improvement.

Our mystery clients did not take a method by design due to ethical concerns, but they were more likely to report that they "could have taken a method" at intervention clinics; and they thought they could have taken a broader range of methods, including their (randomly-assigned) method of choice. It is not clear if this subjective belief about the potential for receiving a method reflects a change in method-dispensing practices, but the fact that their experience at treatment clinics made them think they had access to a broader range of methods is promising.

The intervention was designed to reduce disparities based on age, marital status, and parity and we show that the intervention reduces some key disparities while other disparities still remain. Disparities in counseling comprehensiveness experienced by nulliparous mystery clients are eliminated by the intervention, and disparities in perceived access to FP methods experienced by younger mystery clients decreased (latter not statistically significant). However, the intervention did not change disparities in perceived treatment experienced by unmarried

[1] We consider the full range to be IUDs, implants, injectables, and the pill

mystery clients.

This paper builds on existing literature in several ways. First, this is the first study we are aware of that tests an intervention designed to reduce provider bias in FP service delivery and one of only a few designed to reduce provider bias in health care service delivery (Vela et al., 2022). We show that FP providers' beliefs about the decisions of young women, unmarried women, and nulliparous women are malleable and amenable to intervention. This has implications for efforts to change biased beliefs in other types of healthcare service delivery. Discrimination in health care delivery in LMICs with respect to patient characteristics such as caste (Mishra et al., 2021), socioeconomic status (Kwan, 2020), and HIV status (Ikeda et al., 2019) are well-documented, but evidence on whether the biased beliefs that drive this discrimination can be addressed through intervention is not well understood. Our results demonstrate that even negative beliefs about patients rooted in cultural and social norms can be updated.

Second, we contribute to the literature on measuring bias in FP care. Most studies documenting provider bias are qualitative or from the provider's perspective (Solo and Festin, 2019; Calhoun et al., 2013; Schwandt et al., 2017; Tumlinson et al., 2015; Sidze et al., 2014). There are only a few papers that explore the manifestations of bias using quantitative data on FP client outcomes, and they show conflicting results. Moucheraud et al. (2022) finds that adolescent FP visits are more comprehensive and include more WHO-recommended care and counseling activities relative to adult women FP visits in the Democratic Republic of Congo, Malawi, Senegal, and Tanzania. Yet, analysis of exit surveys from India, Niger, Senegal, Nigeria, and Kenya find that age, marital status, and parity do not predict FP care quality (Speizer and Calhoun, 2020). Using mystery client data in Nigeria, Sieverding et al. (2018) show that young women who are unmarried and nulliparous get far worse care than older women who are married and parous, whereas a recent mystery client study from Malawi finds that adult mystery clients receive worse counseling than adolescent mystery clients (Hazel et al., 2021). Our study measures bias using several different data sources, allowing for triangulation; and we experimentally change biased beliefs, which allows us to document the extent to which these affect client outcomes. We show that reducing biased beliefs can improve some important client outcomes (counseling and interactions), but this might not affect the types of methods received by clients.

Third, we contribute to the literature on the high unmet need for contraception among adolescents. Over 40% of women aged 15-19 in LMICs do not have their contraception needs met compared to 25% of all women in LMICs (Kantorová et al., 2021). Provider bias is widely believed to be a key driver of this high unmet need (Sully et al., 2019). However, we show that disparities in care based on age, parity, and marital status are relatively small, which suggests that the difference in unmet need for young women may be driven primarily by disparities in utilization of services rather than differential care.

Fourth, it is rare for the same intervention to be tested with an RCT across three quite different countries. This approach addresses one of the main critiques of RCTs in economics (Deaton, 2020; Ravallion, 2020) and adds external validity to our findings, making our results more useful for global implementation and scale-up.

2 Setting and Sample

The study took place in Tanzania, Pakistan, and Burkina Faso. These countries represent three distinct regions of the world: East Africa, West Africa, and South Asia. An intervention that is successful in all three settings would have a strong argument for broad external validity and scale-up to other settings. Table A1 shows various indicators of development and contraception use for each country to illustrate how countries compare. In Tanzania, 45% of the population is below the international poverty line compared to 31% in Burkina Faso, and only 5% in Pakistan. All three countries have over two thirds of the population living in rural areas and about a quarter of the population completing secondary school. The total fertility rate is about 5 in both Tanzania and Burkina Faso compared to 3.4 in Pakistan. About 35% of women aged 15-49 years old in Tanzania were using contraception compared to 27% in Burkina Faso and 24% in Pakistan. Tanzania also had the highest rate of unintended pregnancy (105 per 1,000 women aged 15-49) compared to Burkina Faos (75) and Pakistan (71). All three countries had a substantial unmet need to for contraception ranging from 13% in Pakistan to 21% in Burkina Faso.

2.1 Family Planning Clinics

We enrolled 73 clinics in Tanzania (Dar es Salaam), 76 clinics in Pakistan (Karachi), and 78 clinics in Burkina Faso (Ouagadougou, Banfora, and Bobo). These clinics were the universe of clinics with which Pathfinder International, the implementing organization, had an existing relationship. In Burkina Faso, clinics were located in urban, peri-urban, and rural areas. Clinics in Tanzania and Pakistan were primarily in urban or peri-urban areas areas. Clinics in Tanzania and Burkina Faso were all public sector facilities while clinics in Pakistan were private, single-provider clinics. Clinics in Tanzania had about 4 providers per clinic that served about 227 clients per month (Panel A of Table 1). Over half of the clients surveyed were 25 or older. Burkina Faso had an average of about 11 providers serving 102 clients per month. All providers in Pakistan operated their own private clinic where they were usually the only provider and served an average of 21 clients per month.

2.2 Family Planning Providers

All providers at enrolled clinics were eligible for study participation. We randomly sampled 642 total providers (Burkina Faso n=313, Tanzania n=259, Pakistan n=70) to participate in an endline provider survey that we used to assess provider knowledge, attitudes, and beliefs that relate to bias (described more in Section 5.1). In Burkina Faso, the number of providers surveyed at each clinic was proportional to the total number of personnel at each clinic; however, all providers were invited to participate in the survey if the facility had fewer than 5 providers. In Tanzania and Pakistan, we attempted to survey all FP providers at enrolled clinics (average of 4 per clinic in Tanzania and 1 per clinic in Pakistan). In Pakistan, 6 enrolled providers did not complete the provider survey because they were on sabbatical; 3 from the treatment group and 3 from the control group. We had a response rate for the provider survey of 95% in Burkina Faso, 82% in Tanzania, and 92% in Pakistan. Panel B of Table 1 shows characteristics of the providers in our sample.

2.3 Family Planning Clients

We conducted client exit surveys continuously throughout the study period (described more in Section 5.1). Between September 2020 and August 2021, we surveyed 77,971 clients across

the three countries; 26,915 in Tanzania, 9,869 in Pakistan, and 41,187 in Burkina Faso. Panel C of Table 1 shows that 7% of clients were under 20 years old and 57% were 25 years old or more in Tanzania. Client ages were similar in Burkina Faso (10% and 61%, respectively) and Pakistan (8% and 52%, respectively). Sixty-five percent of clients were married in Tanzania, 78% in Burkina Faso, and 99.7% in Pakistan. It was rare for nulliparous women to come into the clinics in any of the three countries; 5% had no children in Tanzania, 9% in Burkina Faso, and 2% in Pakistan. One key takeaway from these numbers is that women under 20 and women without children — two target groups for the Beyond Bias intervention — rarely come into the clinic seeking FP care in any of the countries. The small share of clients in our target population visiting the clinic could be the result of expectations of provider bias; the clients most likely to experience bias might avoid FP clinics due to fear of negative treatment and poor care.

3 Conceptual Framework

Provider bias is a psycho-social phenomenon that materializes within a provider's psyche (Greenwald and Krieger, 2006). Provider bias stems from negative attitudes and beliefs (implicit or explicit) about certain clients or client characteristics. For example, providers' might think that unmarried adolescent women should not be having sex, or that women should have children soon after getting married. FP provider bias towards youth could be rooted in good intentions, such as wanting to protect young women from perceived bad decisions, but it could also be rooted in negative judgement and animosity, such as thinking that women who have sex before marriage are immoral and "bad". FP provider biases that stems from animosity might be more comparable to racial, ethnic, or class-based bias. FP provider biases could also be reinforced by inaccurate technical knowledge about FP care standards (e.g., the inaccurate belief that IUDs cause infertility for nulliparous women). We expect biased attitudes and beliefs to affect the quality of care provided to certain women, resulting in different methods received and interactions experienced across client types, leading to disparities. Thus, changing providers' attitudes and beliefs and correcting inaccurate knowledge could improve the quality of care offered to young, unmarried, and/or nulliparous women and thereby reduce care inequalities and disparities in the types of methods received.

However, updating biased attitudes and beliefs might not be sufficient to change practice if

providers are not motivated or empowered to change their actions, mirroring a broader literature on the "know-do" gap (Mohanan et al., 2015; Gage et al., 2018; Leonard and Masatu, 2010). Ensuring that providers consistently act on updated attitudes and beliefs requires that the new information remains salient and that they are motivated and have the tools to use the new information to improve practice. For example, Vela et al. (2022) show that training to reduce bias changes behavior in the short term, but then providers revert back to old habits.

With this framework in mind, the Beyond Bias intervention was designed with the goals of 1) changing biased attitudes and beliefs about the target groups, 2) updating inaccurate knowledge about appropriateness of care for the target groups, 3) keeping the new information salient over time, and 4) motivating and empowering providers to continue acting on the new information resulting in a sustained change in practice. While the intervention addresses both implicit and explicit biases, we do not attempt to measure implicit bias directly.

The Beyond Bias intervention bundle was developed using a human centered design (HCD) approach that involved in-depth qualitative work over about two years (for more info on HCD see IDEO (2011)). The HCD process aimed to identify key drivers of bias through surveys, qualitative interviews, a literature review, and focus groups. The project team used the results from this process to develop an intervention that addressed the key drivers of bias identified through this process. The team tested several potential interventions through rapid prototyping and pilot testing, and finalized a package of interventions that had the best potential for reducing provider bias and improving outcomes for the target groups, and that had potential for scale up. The design process is described in detail in Y-Labs (2019).

4 Study design and intervention

4.1 Description of intervention

The Beyond Bias intervention included three components designed to address provider bias: An intensive one-day training (Summit); a forum for continued learning and peer support (Connect); and a non-financial awards (NFAs) program based on peer comparison of care provided to young clients. Summit was designed to address #1 above (change negative attitudes and beliefs about the target groups); Connect was designed to address #1-3 above (change

negative attitudes and beliefs, update inaccuracies, keep information salient over time), and the NFAs were designed to motivate and empower providers to continue acting on the new information resulting in a sustained change in practice (#4).

Summit was a one-day training implemented by Pathfinder at the beginning of the intervention (see Figure A2 for complete study timeline). It included personal stories from youth who had experienced bias, stories from providers who have provided biased care, information that could help reduce bias (e.g., appropriateness of care standards), and an action plan for addressing bias. Summits were implemented by Pathfinder country offices in late September 2020/early October 2020. Approximately 98% of providers in the treatment group attended the Summit (see Table A14 for participation numbers in each intervention component).[2] Summit was designed by the HCD process to address provider bias by providing new information through personal stories to help providers update any biased attitudes and beliefs towards the target group.

At Summit, providers were also introduced to the additional interventions of Connect and the NFAs. The Connect forum started right after Summit and included all providers. The goal of Connect was to provide ongoing information to providers and allow them a space to problem solve together to apply unbiased practices in their daily work. Each Connect group was assigned a Pathfinder staff-person to moderate the conversations. Connect included realistic case studies of youth clients, practical tips shared by trusted technical experts to dispel medical misinformation, a safe space for providers to share struggles and successes with peers, and regular review of unbiased service-delivery goals to support providers in maintaining motivation and group commitment. Connect was implemented in two phases—an intensive phase lasting for the first 10 weeks of the 12 month intervention where Pathfinder staff engaged providers once a week followed by a continuous learning phase where implementing staff engaged participants once a month.

In Tanzania, Connect took place in-person at intervention health facilities and through ongoing WhatsApp conversations. In Burkina Faso, Connect sessions were held in-person only (no online WhatsApp forum). In Pakistan, Connect operated exclusively via a WhatsApp group

[2]In Tanzania, 162 FP providers attended the Summit (98%), 37 in Pakistan (91%), and 436 in Burkina Faso (98%).

for all intervention providers. Connect was designed to help cement the new information learned about provider bias and keep these concepts salient throughout the course of the intervention. The discussions and exchange of information with other providers (including clinic leadership) was also intended to empower providers to change behavior. All providers assigned to the treatment group across all countries participated in Connect to some extent. On average, in-person Connect sessions had 78% attendance in Tanzania and 80% in Burkian Faso. In Pakistan, 62% of providers had 2 or more quarters where they engaged with the WhatsApp forum at least every other week.

The NFAs were designed as performance-based social incentives.[3] The idea was to have clinics compete with one another to improve their services for clients under 25. Using data from client exit surveys, each clinic received an NFA score that was calculated every 3 months based on six principles of unbiased care developed by Pathfinder: (1) safe, welcoming space; (2) sensitive communication; (3) simple, comprehensive counseling; (4) seek understanding and agreement; (5) say yes to a safe method, and (6) security of information (see Appendix A13 for complete list of variables that we used to construct NFA scores).

We used the client exit survey data to construct scores every 3 months, at which point there was an NFA ceremony (3 ceremonies total).[4] At the first ceremony NFAs were given to the highest performing facility within each NFA cohort. Facilities were separated into Rewards cohorts based on geographic delineations (4 districts in Tanzania and Burkina Faso, and 3 neighborhoods in Pakistan). In the last two ceremonies, NFAs were also given to the most improved facilities. As a reward, each clinic received a rewards kit that could be taken back to the clinic for an additional ceremony with the rest of the staff. The kit included merit pins, stickers, and a certificate signed by the ministry of health. In addition, every facility received a report card with their performance and actionable suggestions from the past 3 months for improvement at each ceremony.

Since Summit, Connect and NFAs were implemented as a bundle to all treatment clinics and providers, our ability to disentangle the effects of any one of these three interventions relative

[3]The HCD process revealed that providers preferred social awards over financial rewards.

[4]Ceremonies were held in late January/Early February 2021, May 2021, and August 2021. In Tanzania, 100% of facilities sent at least one representative to all three ceremonies. In Pakistan, 44% of providers attended all three rewards ceremonies. In Burkina Faso, 92% had a representative at all three ceremonies.

to the other is limited.

4.2 Randomization

Random assignment was conducted at the clinic level prior to the start of the intervention in June 2019 with 75 clinics in Tanzania, 78 in Burkina Faso, and 80 in Pakistan. Clinics were not aware that they were being randomized. We stratified randomization using available baseline data, which varied by country. In Tanzania, we stratified on district, urban/non-urban, volume of FP clients under age 20, and number of providers in the clinic. In Burkina Faso, we stratified randomization on district, share of FP users under age 20, number of providers at the clinic, and whether the facility was recently added to the Pathfinder network. In Pakistan, we stratified on number of clients of different age groups, whether the provider was a midwife, and whether the facility had newly joined the Greenstar network.

5 Data Collection, Main Outcomes, and Balance

5.1 Data Collection

One strength of this study is our ability to evaluate outcomes using several different data sources. We collected data at the provider level using endline provider surveys and a discrete choice experiment (DCE); the client level using client exit surveys and mystery client visits; and the clinic level using administrative data (see Table 2 for a list of data sources and timeline of collection). For most outcomes experienced by clients we prefer the mystery client data because mystery clients were trained to observe key outcomes during their visits, they are not subject to social desirability bias, and the roles were standardized to reflect a client likely to be the target of negative provider bias. Real clients could also have sample selection bias if only the most courageous and determined clients choose to visit the clinic. However, mystery clients also have some weaknesses relative to other sources. For example, mystery clients do not actually take an FP method due to ethical reasons. Similarly, mystery clients might not present in the same way as real clients. Mystery client data was only collected at endline whereas exit surveys were conducted throughout the study. As such, we present outcomes that relate to the client-perspective for both mystery clients and exit surveys. Below, we provide more details on each data source.

Client Exit Surveys.

We collected client exit surveys in all study clinics at baseline and then continuously for the duration of the intervention. In each country, the client exit surveys were administered by youth enumerators (18-24 years old).[5] Enumerators attempted to approach all clients who visited the clinic to get consent for an interview and surveyed as many as they could each day. The survey collected information in the local languages on basic demographics about the client (age, parity, and marital status), information on the questions the provider asked and the services offered, and the subjective experience of the client. In our main analyses, we restricted the exit survey data to women under age 25 because this is the target population for the intervention. We show in the appendix that our results are consistent when we include all clients from the exit survey (Table A16). We include older clients when assessing disparities. We also dropped 347 surveys collected between September 2020 and August 2021 based on quality control measures that we implemented.[6]

Mystery Client Visits.

Each clinic in Tanzania and Burkina Faso received four anonymous visits from members of the research team who posed as real family planning clients. Providers in Pakistan received only two visits because these were smaller clinics with only one provider, and we feared more than two visits would raise suspicion. Mystery client visits took place between 10 and 12 months after the start of the intervention. The mystery client approach allowed us to record data on client-provider interactions without providers knowing they were being observed, while keeping the characteristics of clients consistent across providers. We worked with local survey firms to train young female enumerators to act as family planning clients. The mystery client visits

[5]We used youth enumerators to make young clients feel comfortable. Youth enumerators visited their assigned clinics for two to three days a week throughout the intervention. At the clinic's premises, the enumerators approached female clients upon their exit, and asked them for verbal consent to conduct a survey about their visit to the clinic. If the client provided consent and reported that either family planning services was the reason for their visit or they received family planning counseling during the visit, the enumerator administered the survey.

[6]Surveys from clients who received services but took less than 2 minutes to complete were dropped because this was too short a time frame to carefully complete the survey. Surveys that were submitted to the server over 31 days after they had been administered were dropped because this indicated they were incomplete. Surveys that reported receiving multiple hormonal methods were excluded because this is not possible and indicates the rest of the responses might also be poor quality.

were unannounced and anonymous. This is a validated methodology to measure quality of family planning care (Chandra-Mouli et al., 2018; Tumlinson et al., 2013, 2014; Fitzpatrick and Tumlinson, 2017; King et al., 2019; Sieverding et al., 2018; Das et al., 2021).

We worked with Pathfinder country offices and local partners to develop profiles that were realistic in the country context to avoid mystery clients being "discovered." To isolate the causal effect of age, marital status, and parity we created eight profiles that included every combination of marital status (married/unmarried), parity (one child or no children), and age (16/17 or 24 years old) and conducted an equal number of visits for each profile. We then randomly assigned four profiles to each clinic in Tanzania and Burkina Faso, and two profiles to each clinic in Pakistan. We stratified profile assignment such that half of visits in each clinic were unmarried/married, half were younger/older, and half were nulliparous/parous.[7] This design (shown in Table 3) allows us to estimate the effects of each attribute on outcomes of interest without confounding. We also assigned mystery clients to have a preference for either injectables or a long-acting method (implant in Tanzania and Burkina Faso and IUD in Pakistan), which we used in our "method of choice" analyses.

In Pakistan, unmarried women rarely seek family planning services, so it was not feasible to have very many unmarried profiles without raising suspicion. We also determined that it could be dangerous for an unmarried women to say she already had a child. Given this, we only included 3 profiles; all mystery clients in Pakistan were about the same age (19 or 20 years old) with some variation in marital status and parity (see Table 3).

Mystery clients received a seven-day training to learn and practice the roles, and then several days of piloting with real visits. Mystery clients went through the standard process to see the provider (check in at register, wait in line, etc.). They did not actually receive a family planning method; instead, they informed the provider at the end of the visits that they wanted to think about the information and talk to their mother (if unmarried) or husband (if married) before proceeding. There was little indication that mystery clients were discovered by the providers they interacted with.[8] Within an hour of leaving the facility, mystery clients completed a

[7]This ensured that there were an equal number of visits for each profile, that treatment and control clinics were balanced on the profile attributes of the visits they received and that profile attribute levels were balanced with other attribute levels (e.g., unmarried and married were equally likely to be paired with being young)

[8]Only one mystery client thought they might have been discovered and we were only aware of one provider

debriefing survey administered by their supervisor. The survey recorded information on the quality of services received, including the methods the provider counseled on, whether they were made to feel comfortable, and whether the provider asked about their preferences. They also recorded whether they thought they could have received each of a set of specific methods. All mystery clients were blinded to whether the facility was an intervention facility or a control facility.

Provider Surveys.

The study team collected data from providers at endline using two instruments: a provider survey and a DCE, which we implemented during a single data collection session. The provider survey recorded information on 1) demographic characteristics and background information; 2) general attitudes and beliefs about young, unmarried, and nulliparous family planning clients; and 3) details on the clinic environment where they practice, especially related to family planning services for youth.

The DCE presented a hypothetical client – only telling the provider her age, marital status, and parity – and asked the provider how they would provide family planning services to this client (e.g., what methods they would counsel on). Each provider was asked about four unique profiles with a different combination of age (15, 20, or 25), marital status (married or unmarried), and parity (no children or one child). We randomly assigned profile combinations to providers, ensuring treatment and control were balanced on attributes and that profile attributes were balanced with other profile attributes (e.g., unmarried and married profiles were equally paired with parity). This allows us to estimate the effect of each attribute on a provider's self-reported behavior.

Administrative Data on Service Delivery Statistics.

Service delivery statistics are routinely collected by clinics in Tanzania and Burkina Faso as part of the Ministry of Health monitoring system. Data are added to a national database each

who thought a client was our mystery client. The client they described was not part of our study. During piloting we followed up with providers to see if the provider suspected any of the mystery clients were not real clients, which none did.

month to which we were granted access. In Pakistan, these data are collected by Greenstar, the social franchising organization with which providers were affiliated. The service delivery data include the monthly number of new contraceptive users, returning contraceptive users, age categories of users, and method mix.

5.2 Main Outcomes

We categorize outcomes into four distinct domains, and pre-specified a primary outcome for each domain that we pre-registered. The domains follow a theory of change through which we expect provider bias to affect outcomes (see Figure A1). The first domain, *provider attitudes/beliefs*, is the level at which the Summit and Connect components directly intervened. We hypothesize that biased attitudes/beliefs affect the second domain, *patient-centered FP care*, which in turn should affect the third domain, *methods received* by the client. Finally, all of these domains could affect the fourth domain, the client's *perceived treatment* during the visit. Below, we describe the outcomes included in each of these domains in detail.

Provider attitudes and beliefs: Unbiased Index.

The pre-specified primary outcome for this domain was an index of unbiased beliefs. We use statements from the provider survey about how age, marital status, and parity affect family planning service provision with which providers could strongly agree, agree, disagree, or strongly disagree. These include statements such as "It is okay for young clients who are unmarried to use contraception" and "Married clients and unmarried clients should have the same FP options." We construct an index by combining responses to 47 of these statements. In addition, we construct four sub-indices, each made up of a subset of the 47 questions : 1) FP care specific attitudes and beliefs (28 items), 2) underlying attitudes and beliefs (if the statement is broader than just FP care and based more on social and cultural norms) (9 items), 3) beliefs about the professional environment (7 items), and 4) beliefs about the community (3 items). We code all items so that higher scores represent less biased attitudes or environments. In order to construct the indices, we implement the mean effects approach (Kling et al., 2007), which uses the mean of all 47 variables after they have been standardized relative to the control group. The resulting indices are in units of the (average) control group standard deviations. We conduct

this process separately for each country and for each sub-index.

Patient-centered FP care outcomes.

This domain captures key elements of FP care that relate to quality of care standards and patient-centeredness. We use three data sources to measure patient-centered FP care: client exit survey, mystery clients, and the DCE from the provider survey. The primary outcome for this domain is whether the client was counselled on all methods that fit with the client's preferences ("full range of methods"). We include IUD, implant, injectable, and pill as the maximum full range of methods, as they account for over 98% of all methods dispensed (see Figure A5). We define this outcome equal to 1 if the provider discussed all four methods during the counseling session. In the exit survey analysis, we removed some methods from the full range if clients reported specific preferences that made certain methods not applicable (e.g., do not want to take a pill) based on the WHO Decision-Making Tool for Family Planning Clients and Providers and recommendations made in the WHO Family Planning Global Handbook for providers (WHO, 2005, 2018) (See Appendix Table A2 for a list of methods that correspond to different preferences). For clients who were seeking a method switch, their current method was not included in the definition of the full range. Clients who came in to continue use of the same method were excluded from analyses using this outcome.

We also analyze several additional outcomes related to patient-centered FP care.

1. *Able to receive services*: This variable is equal to 1 if the client is able to see a provider and receive counseling.
2. *Essential questions asked*: This outcome is based on a set of key questions that the provider should ask during a counseling session 1) asked about method preferences, 2) asked about birth spacing and limiting preferences, and 3) asked if they have any questions. We created binary variables for each of these three questions and took the mean to create the outcome.
3. *Method Information Index*: The method information index is a widely used measure of the key information provided to the client during a provider-client interaction (Chang et al., 2019). The measure is set to 1 if the client answered "yes" to all of the following

questions (all binary variables):[9]

(a) Were you informed about other methods of family planning aside from the one you received?

(b) Were you informed about possible side effects or problems with the method you took?

(c) Were you told what to do if you experienced side effects or problems?

The resulting outcome is a binary variable, which is consistent with how it is used in the Family Planning literature. All other index based outcomes are in terms of standard deviations.

Method dispensing outcomes.

Our primary method dispensing outcome is whether a client received any modern contraceptive method during her visit. We set this binary outcome to 1 if the client received an IUD, implant, injectable, pill, condoms, emergency contraception, and/or permanent methods. The client exit survey is the only data source where clients receive a method. However, we trained the mystery clients to assess whether they thought they could take each type of method if they were a real client, which we also use to assess method dispensing outcomes. We also analyze whether the client received her method of choice, whether she received a long acting reversible contraception method (LARC), and whether she received an injectable. We chose to examine LARCs and injectables specifically because these were methods for which providers reported biases during formative work, especially for nulliparous women and for which there were disparities at baseline.

If a client did not receive services, each of these outcomes was coded as zero. In addition to the outcomes listed above, we examine method mix (share of all methods that are each type) and total number of methods dispensed using the client exit survey and administrative data.

[9] In the client exit survey, only those who received a method were asked this set of questions. We adapted this measure for mystery clients so that those who were counseled on more than one method and were informed of possible side effects or problems with the methods they were offered were assigned a value of 1.

Perceived treatment index.

The primary outcome for this domain is a perceived treatment index, which we assess using the client exit survey and the mystery clients. We use the mean effects approach (see description of approach in provider survey section) to construct an index based on 29 subjective questions from the client exit survey and 22 items in the mystery client debrief survey. We trained mystery clients extensively on what kinds of things to look for (e.g., overt judgment as well as subtle judgment though body language or facial expressions). We prioritize mystery clients for this outcome because real clients might give a socially desirable response during client exit surveys (Dunsch et al., 2018) unlike the mystery clients. In addition to the perceived treatment index, we analyze client satisfaction based on whether the client would recommend the clinic to a friend. We also measure whether the client felt judged or scolded by the provider and whether the client thought the provider made them feel uncomfortable because of her sex life.

5.3 Balance between treatment and control

We use baseline administrative data to check for balance against treatment and control status. In addition, we use endline provider surveys to assess balance on provider traits that are fixed over time. Baseline administrative data in Pakistan is not available and we assess balance only on fixed provider characteristics.

Table A3 shows that intervention and control clinics in Tanzania and Burkina Faso were well balanced on the number of providers, client volume, client age, and method mix. Providers were well balanced on qualifications in Tanzania and Burkina Faso. There is some imbalance in provider qualifications in Pakistan; intervention providers were more likely to be midwives or nurse midwives whereas control providers were more likely to be lady health visitors (LHVs) or doctors. However, it is not clear how to interpret the direction of the bias because the control group has more doctors, which have the most training, but also has the most LHVs, which have the least training. We examined the methods that providers were trained on and provided at their facility in Pakistan and found no substantive difference between the intervention and control providers. Ultimately the samples seem fairly well balanced at baseline.

6 Empirical Framework

We use a similar regression framework for all data sources and outcomes. We estimate intention to treat (ITT) regressions for each country separately and we also pool all three countries. Because we measure the same outcomes with several different data sources, we estimate regressions separately for each data source and compare the effects across data sources. Consistent results across data sources adds validity to the results. For all primary outcomes, we estimate an unadjusted analysis and an analysis that includes a set of covariates. The results are qualitatively the same whether or not we control for covariates. We cluster standard errors by clinic in all models because the intervention was randomized at the clinic level (Abadie et al., 2017). The regression models take the following form:
Separate for each country:

$$y_{if} = \beta_0 + \beta_1 \text{treat}_f + \epsilon_{if} \tag{1}$$

All countries pooled:

$$y_{ifc} = \beta_0 + \beta_1 \text{treat}_f + \alpha_c + \epsilon_{ifc} \tag{2}$$

Where y_{if} is an outcome for individual i (either a provider in the provider survey, a real client in the exit survey, a mystery client, or a profile in the DCE) from facility f. Treat indicates the facility's treatment assignment and β_1 is the ITT effect of the treatment. In equation 2 we pool all countries and include a set of country fixed effects (α_c). We also control for a set of covariates that vary by country and data source in some models. Information on which covariates are included in each model are listed in the notes of Table 4.

The client exit survey data is longitudinal and collected from the onset of the study. This allows us to estimate intervention impacts for each month of the study using the following regressions.

Separate for each country:

$$y_{ift} = \beta_0 + \sum_t \beta_t(\text{treat}_f \times \text{month}_t) + \sum_t \lambda_t \text{month}_t + \epsilon_{ift} \tag{3}$$

All countries pooled:

$$y_{iftc} = \beta_0 + \sum_t \beta_t(\text{treat}_f \times \text{month}_t) + \sum_t \lambda_t \text{month}_t + \lambda_c + \epsilon_{iftc} \tag{4}$$

Where the β_t represent the impact of the intervention in study month t and all other terms are the same as in equations 1 and 2. Our hypothesis is that outcomes might improve around the time of the rewards ceremony when the desired behavior is more salient, and thus treatment effects would be larger in the months immediately before and after the ceremony.

6.1 Multiple hypothesis testing

We adjust our p-values for multiple hypothesis testing using the sharpened q-value method to control the false discovery rate as outlined by Anderson (2008). We use this method to adjust p-values for our four primary outcomes and report sharpened q-values in brackets. When analyzing multiple outcomes within a domain (or family), we adjust for the number of outcomes in the respective domain.[10]

7 Results

7.1 Did treatment result in less biased provider attitudes and beliefs?

Figure 1 and Table 4 (columns 1 and 2) show that treatment resulted in less biased attitudes and beliefs with respect to age, marital status, and parity. Each point estimate in Figure 1 represents the (unadjusted) treatment effect estimated using equations 1 and 2 with 95% confidence intervals. The intervention improved the unbiased index by about 0.2 standard deviations in the pooled analysis, and this effect size is similar across all countries. Effects on the main index are driven by changes in FP care specific attitudes and beliefs (i.e. knowledge of

[10] We do this process separately within each country and when pooling countries

care standards) and underlying attitudes and beliefs (i.e. beliefs related to cultural and social norms). Attitudes and beliefs about the professional environment and community factors that could be drivers of bias improved slightly in Burkina Faso but not in the other countries (these types of beliefs were not explicitly targeted by the intervention). Column 2 of Table 4 shows that including covariates does not substantively change effect sizes. Table 4 also presents the sharpened q-values in brackets and results remain statistically significant after this adjustment.

To make these effect sizes more concrete, Tables A4 and A5 show the specific questions that went into the FP care specific sub-index and the underlying attitudes and beliefs sub-index, with treatment effects for each question. This shows significant improvements in 22 of 27 FP care specific statements in the pooled analysis and 7 of 9 of the underlying attitudes and beliefs statements.

Table A6 shows that providers in the intervention group are also more likely to report in the DCE that they would offer counseling (i.e., not deny services; 1.5 percentage point increase), provide a modern method (3.3 percentage point increase), and counsel on the full the range of methods (14.8 percentage point increase) compared to providers in the control arm. Figure 2 shows that interventions providers are significantly less likely to report imposing method restrictions because a client was "too young" (23 percentage point reduction), unmarried (7 percentage point reduction), or did not have children (15 percentage point reduction). Figure 2 shows pooled results but effect sizes are similar across countries.

The improvement we observe in provider attitudes and beliefs as well as the improvement in self-reported behaviors is encouraging. However, it is possible that providers in the treatment arm overstate positive attitudes and beliefs about the target population because the intervention trained them on which responses were "correct". Next, we assess whether the reduction in self-reported biased attitudes and beliefs translates to behavior change and better care for young women.

7.2 Did treatment improve patient-centered FP care for young women?

The primary patient-centered FP care outcome is counseling on a full range of methods (pill, injectable, IUD, and implant). Column 3 of Table 4 shows that in the mystery client data,

counseling on the full range of methods increased by 6.5 percentage points when pooling countries (off a base of 63.2%, p = 0.073). The effect is most pronounced in Pakistan (23 percentage point increase; p<0.01), modest in Tanzania (7 percentage point increase and not statistically significant), and negative in Burkina Faso. Interestingly, results using the client exit survey data are quite consistent. Counseling on the full range improved in all three countries; with a pooled effect of 9.3 percentage points off a base of 50.5% (p<0.05; column 3 of Table A7). The result in Burkina Faso is also positive and statistically significant on this outcome using client exit survey data.

Figure A3 plots treatment effects for all pre-specified FP care outcomes for all three data sources and regression results are reported in Table A7. In pooled analyses, we observe significant improvements in all outcomes across all data sources; clients are more likely to receive services,[11] be counseled on more methods, be given more information about the methods, and be asked more questions.

Positive effects in pooled analysis with the mystery client data are driven by Tanzania and Pakistan as effects are close to zero in Burkina Faso in the mystery client data. In the client exit survey data, all three countries had similar effects sizes for the four outcomes, and all pooled effects are statistically significant.

7.3 Did treatment impact methods received by young women?

Figure A4 and column 5 of Table 4 show that treatment increases the likelihood that clients who completed exit surveys received any method (pooled effect of 0.019; p = 0.085; q = 0.045). This pooled effect is driven by Tanzania as effect sizes in Pakistan and Burkina Faso are close to zero. In Tanzania, the intervention increased the share of clients that received a method by 4.7 percentage points (p<0.05). About 9% of clients in the control arm did not receive a method in Tanzania, so this effect size represents nearly a 50% reduction in the share of women not receiving a method.

A key limitation to this outcome is that over 90% of clients in the control group received a method, so there is very little scope for improvement. Reaching 100% of clients receiving

[11] Receiving services in the mystery client data is not significant but almost all mystery clients received services so there was no scope for improvement.

a method is an unrealistic target because some clients only want information or delay getting a method until they discuss it with family. Overall, clinics in our study setting appear to be poor locations for an intervention designed to improve method uptake because most women that come to the clinic receive a method in absence of the intervention.

Despite more comprehensive counseling and more patient-centered care, we find little evidence of a change in the types of methods clients received. The client exit survey data shows there was no change in the likelihood that young women received a LARC or an Injectable (top panel of Figure A4), methods for which we found evidence of bias in the formative work. We do find a modest improvement in clients receiving their method of choice, but this is entirely driven by women in Tanzania receiving a method rather being turned away, not a change in the types of methods dispensed.[12] Figure A5 shows that method types reported in the client exit survey were comparable between the intervention and control sites for all age groups with no statistical differences. Figures A6, A7, and A8 plot types of methods over time from the administrative data and show no substantive change in method mix following the start of the intervention.

Mystery clients' perceptions of method dispensing.

Although the intervention did not change the types of methods real clients received, mystery clients thought they would have been able to take a broader range of methods at intervention clinics. Mystery clients did not take a method but reported whether they "thought they could take" each method "if they were a real client" (middle panel in Figure A4 and "MC" columns of Table A8). All pooled effects for method dispensing outcomes in the mystery client data were significant and, similar to the client exit survey data, were mostly driven by improvements in Tanzania. Mystery clients visiting intervention clinics in Tanzania were far more likely to think they could take any method (13 percentage points) and their method of choice (17 percentage points). In Pakistan, the intervention demonstrated positive effects on method dispensing according to mystery client perceptions, but none of these effects were statistically significant. In Burkina Faso, the intervention had no effect on mystery client perceptions of whether they

[12] If a woman did not receive a method, we coded this as not receiving method of choice. When we restrict to clients who received a method, we do not observe a significant improvement in method of choice.

could take a method or which methods they could take. Overall, mystery clients in Tanzania and Pakistan seemed to believe they had access to an expanded range of methods in treatment facilities relative to control facilities.

Using mystery clients to measure method dispensing has both strengths and limitations relative to the client exit survey. On one hand, a key strength is that we designed the mystery client profiles such that they were representative of young women who we would expect to experience bias. Real clients might have private information that makes them less likely to experience bias and more likely to receive their preferred method; e.g., a young unmarried client without children might also be a student and know that providers are more sympathetic to students. Thus, our mystery client profiles could be more representative of clients who avoid the clinic out of fear of biased care (e.g., visits from nulliparous clients were very rare in the client exit surveys but half of mystery client visits were from nulliparous profiles). On the other hand, mystery clients do not actually take a method (they say they need to think about it and talk it over with their husband or mother) and therefore it is not clear how their subjective expectations of receiving a method correlate with actually receiving a method. It is certainly promising that the intervention significantly improved mystery clients' perceptions about method restrictions and this could have implications for the intervention encouraging women who traditionally avoid FP clinics out of fear of bias to start visiting clinics for care. But it is unclear whether this actually translates to better outcomes for real clients.

7.4 Did treatment impact client perceptions on how they were treated?

Mystery clients are the preferred data source for perceived treatment outcomes because they received training on what to look for during interactions and they do not have the incentive to provide a socially desirable response, which can bias responses about satisfaction (Dunsch et al., 2018). Mystery clients perceived significantly better treatment at intervention clinics relative to control clinics. Columns 7 and 8 of Table 4 shows the intervention increased the perceived treatment index by about 0.2 standard deviations in pooled analysis ($p<.01$), which was driven by improvements in Tanzania and Pakistan, with a much smaller effect in Burkina Faso. Table A9 shows the effect individually for each question that went into the perceived treatment index. Figure A9 shows that mystery clients in Tanzania and Pakistan were also

more likely to recommend the clinic to a friend, less likely to be judged or scolded, and less likely to be treated worse because of marital status or parity.

Table A10 shows regression results for all perceived treatment outcomes and compares treatment effects between the client exit survey and the mystery clients. In client exit survey data, while the treatment effects are mostly positive, they are smaller in magnitude than the mystery client results. In addition, except for Tanzania, they are not statistically significant at conventional levels.

7.5 Did treatment increase youth client volume?

As word spreads about improved care and better treatment of clients, we might expect more young clients who would would otherwise forgo FP care, to start visiting the clinic. Table A11 shows that the average number of clients per month over the 12 months of intervention did not appear to increase at intervention sites for any age group in Burkina Faso or Tanzania. In Pakistan, there appears to have been a modest but statistically significant increase in client flow, which was mostly driven by clients under 20 years old (about 2 additional youth clients per month per clinic). Figures A10, A11, and A12 plots the treatment effect on client volumes over time by age, which shows that youth client volumes were not increasing over time at intervention facilities. Taken together, this provides little evidence that the intervention increased client volume in Burkina Faso and Tanzania, but may have modestly increased youth client volume in Pakistan. This lack of change in client volume could be because there were no outreach efforts in conjunction with the intervention and not enough time passed for word to spread through social networks. It is possible that as more time passes, more youth clients will get word of this improved care and choose to attend these clinics.

7.6 Do treatment effects evolve over time?

Our client exit survey data was collected daily during the intervention period which allows us to assess 1) whether the treatment effects wain or get more pronounced with more intervention exposure, and 2) whether there are distinct changes in behavior that correspond with the timing of the different intervention components (e.g. do providers give particularly good services around the timing of the awards ceremony?). Figures A13 through A16 plot the monthly

treatment effects from equations 3 and 4 using the client exit survey data. In Tanzania and Pakistan, counseling on the full range of methods improved over time. In Tanzania in particular, the effect size started to increase directly after the first award ceremony suggesting that provider behavior might have responded to the ceremony. However, there is no clear pattern of treatment effects over time for the other outcomes and, for the most part, there does not appear to be a bump in the treatment effect around the timing of the ceremonies for other outcomes.

7.7 Did treatment reduce disparities by age, marital status, and parity?

A key premise of the Beyond Bias intervention was that young women receive particularly poor FP care and treatment from FP providers, especially when they are unmarried or nulliparous. As such, an objective of the intervention was to reduce the effect of age, marital status, and parity on FP counseling and interactions with providers. In the analyses that follow, we show three key findings. First, these different client characteristics lead to disparities along different outcome domains in the control group: nulliparous mystery clients are less likely to be counselled on the full range of methods than mystery clients with children, younger mystery clients thought they had less access to FP methods than older mystery clients, and unmarried mystery clients perceive worse treatment by the provider than married mystery clients. Second, the intervention reduced some key disparities while others were not affected: disparities in counseling experienced by nulliparous mystery clients were eliminated in all three countries and disparities in perceived access to FP methods for younger mystery clients were reduced in Tanzania, but disparities in perceived treatment for unmarried mystery clients were not affected by the intervention. Third, some of the improvements generated by the intervention were experienced by all types of clients, not just the clients who were targeted.

To estimate the effect of age, marital status, and parity on FP outcomes and how the intervention changed their effects, we use the following equation.

$$\begin{aligned} y_{if} = \alpha + \beta_1 \text{young}_{if} + \beta_2 \text{unmarr}_{if} + \beta_3 \text{nullip}_{if} + \delta \text{treat}_f + \\ \gamma_1(\text{young}_{if} \times \text{treat}_f) + \gamma_2(\text{unmarr}_{if} \times \text{treat}_f) + \gamma_3(\text{nullip}_{if} \times \text{treat}_f) + \epsilon_{if} \end{aligned} \quad (5)$$

In this setup, the β's represent the disparity for outcome y in the control group and the γ's represent the treatment effect on the disparity (or heterogeneity in the treatment effect). To interpret a result as showing that a *disparity* is reduced by the intervention, we would expect two estimates from this equation: 1) The disparity exists in the control group, which would show up as a negative β, and 2) The gap is smaller in the treatment group than in the control group, which would show up as a positive γ. The δ is the treatment effect for the group that was not targeted by the intervention (older, married women, with children). We focus on the mystery client data for this analysis because our design allows us to isolate the causal effect of each attribute on provider behavior; mystery client profiles were nearly identical aside from variation in age, marital status, and parity, and we randomized profiles to clinics (see section 5.1 for more details). Provider behavior towards real clients is less reliable for two reasons. First, real clients may have confounding factors that are correlated with age, marital status, and parity that we do not observe. Second, those who choose to visit an FP clinic may be differently selected by age, parity, and marital status. However, we have more statistical power in the client exit survey data because of a larger sample. We present results using those data in the appendix. We also did not vary age in Pakistani mystery clients, so estimates of age disparities are for all countries using the client exist survey data but only for Burkina Faso and Tanzania when using the mystery clients.

Counseling disparities.

Columns 1-4 show of Table 5 shows regression results from equation 5 for counseling on the full range of methods. Coefficients on *Nulliparous* are negative and of an important magnitude (>0.11) in all three countries, suggesting nulliparous women experience a disparity in counseling in the control group. The positive and sizable coefficients on *Treatment X Nulliparous* show that the intervention mostly eliminated the counseling disparity for nulliparous women (statistically significant coefficient of 0.125 in pooled analysis). This pattern was also present in the client exit survey data in Pakistan (Table A12). Coefficients on *Young* and *Unmarried* show that these characteristics are not driving counseling disparities. Moreover, small an insignificant coefficients on the *Treatment* variable and other interactions suggest the treatment effects for counseling are driven almost entirely by improvement for nulliparous women.

Method access disparities.

Columns 5-8 show that disparities in perceived method access are driven by the mystery client's age (negative coefficients on *Young*. Younger mystery clients were 13.4 percentage points less likely to think they could take any method compared to older mystery clients in the control group; the intervention reduced this disparity by more than half in Tanzania (coefficient on *Treatment X Young* is 0.106) but this was not significant, and there was no evidence for a reduction in this disparity in Burkina Faso. The small and insignificant coefficients on *Treatment* and mostly positive interaction terms show that treatments effects for this outcome were mostly driven by the target groups.

Perceived treatment disparities.

Columns 9-12 of Table 5 shows that disparities in perceived treatment were driven by marital status; coefficients on *Unmarried* are mostly negative and significant.[13] The intervention did not reduce this disparity. Moreover, treatments effects for the target groups were similar to treatment effects non-targeted clients (coefficient on *Treatment* is positive and significant and interaction terms are small an insignificant). This suggests that the intervention changed the way providers treated all clients, not just clients targeted by the intervention.

8 Robustness of results

8.1 Spillovers

Providers can switch between clinics in Tanzania and Burkina Faso. Ten percent of providers in Tanzania and 9 percent in Burkina Faso reported providing FP services at another clinic besides their current clinic in the previous 12 months. This could lead to spillover of the intervention if providers who participated in Beyond Bias switched to a control clinic during the study period or if control providers switched to a treatment clinic. This would lead us to underestimate the treatment effect. About 7.1% of control providers in Tanzania reported participating in some

[13]There were very few unmarried clients in Pakistan (0.03%; see Table 1) and our unmarried mystery client profiles were designed to be as socially acceptable as possible, so the effect of marital status in Pakistan should be interpreted with caution.

part of the intervention and 3.9% of control providers reported participating in some part of the intervention in Burkina Faso. If the providers in Tanzania who were exposed to the intervention had the same effect size for counseling on the full range of methods as we report in Table 4, 0.111, then the "true" effect size accounting for spillovers would be 0.117. Therefore while our effect sizes might be slightly underestimated, it is unlikely to impact our conclusions.

8.2 Non-participation in Pakistan

In Pakistan, an initial 82 providers identified by Pathfinder were invited to participate in the study. Prior to the start of the intervention, 23 providers were replaced because they were unable or unwilling to participate.[14] Pathfinder replaced these providers with providers who agreed to allow enumerators at their clinics and could attend the Summit event (if assigned to treatment). Non-participation was less common in the control group because the commitment in the treatment group was more substantial (19 clinics in the intervention group were replaced compared to 4 in the control group). As a result the treatment group was more likely to have replacement providers. Providers in the control group might no longer be a valid counterfactual for providers in the treatment group if availability/willingness to participate is correlated with outcomes of interest. To test this, soon after the real Summit event took place, we asked providers in the control group if they would be willing to participate in a Summit event similar to how we approached the intervention group to participate. We conduct a sensitivity analysis in which we restrict to 1) the 22 treatment group providers who were part of the original randomization and 2) the 21 control group providers who were part of the original randomization *and* agreed to participate in the intervention.[15] The top panel of Appendix Table A15 presents results with and without excluding these providers. Results are largely unchanged and qualitatively similar when we drop the replacements from the treatment group and non-participating providers from the control group. This suggests that non-participation and replacement of providers in Pakistan did not bias our treatment effects in important ways.

[14]Seven providers would not allow exit surveys, 8 providers were not able or not willing to partake in the intervention activities, 4 providers had personal issues that prohibited participation, and 4 providers were no longer affiliated with Pathfinder.

[15]Of the 43 providers that satisfy this criteria, 5 did not complete a provider survey and 1 did not receive mystery client visits.

We also test the sensitivity of our pooled analyses to excluding Pakistan and only including the countries in which randomization was not compromised (second panel of Table A15). The odd columns present the estimates from our full analysis in Table 4, and the even columns only include Tanzania and Burkina Faso. When excluding Pakistan, the estimated treatment effect on counseling on full range becomes close to zero but the other outcomes remain statistically significant and of a similar magnitude.

9 Discussion and Conclusion

This paper demonstrates that a relatively simple intervention can change biased provider attitudes and beliefs about contraception services for young women, and that this leads to better care and client interactions. The intervention caused providers to report less biased attitudes toward young women across three quite distinct countries, which is promising for the generalizability of our results. In Tanzania and Pakistan, the intervention led to improved counseling and better perceptions of how clients felt they were treated during FP visits. In Tanzania, the intervention also increased the likelihood that clients left the clinic with a method. The treatment effects were partly driven by reductions in disparities based on age and parity, suggesting that the intervention worked partly though the expected channels. Despite more comprehensive counseling, the intervention did not change the types of contraceptive methods clients received nor did it have an important effect on uptake outside of Tanzania. Most women in the control group received a method and reported that it was their method of choice, so there was also limited room for (measurable) improvement in method dispensing outcomes.

Although we do not find important improvements in method dispensing, this does not mean that the intervention does not have the potential to improve these outcomes. The improved attitudes and beliefs, counseling quality, and client experiences from Beyond Bias could cause more young women to come into the clinic in the longer-term, and receive better care when they do. The target populations account for a small portion of clients who attend the clinic (e.g., only 8% percent of clients were age 15-19, and only 6% did not have a child; less than 1 percent of clients were unmarried in Pakistan). They might avoid the clinic because they are worried about poor treatment or judgment from providers. A recent study from Senegal shows that women avoid clinics where there is greater provider bias (Speizer et al., 2021). Some women

might avoid clinics altogether if they do not know of a clinic where they could get unbiased care. In the United States, Black patients utilized more preventive health services when they were randomly assigned to a Black doctor compared to a White doctor, mostly because of better communication between patient and provider (Alsan et al., 2019). We do not find any evidence of increased client volume resulting from the Beyond Bias intervention, but this could be because the community was not aware of the improved treatment and counseling they would receive at intervention clinics. Over time, more women who avoid coming to FP clinics in absence of the intervention because of fear of provider bias could start visiting the clinic, thus reducing disparities in modern contraception use. Future work should explore coupling the Beyond Bias intervention with outreach efforts to inform the community about clinics that have less biased care.

The results in Tanzania were most promising. This is the only country where we see significant improvements in every outcome domain, and effects were mostly consistent across data sources. Tanzania was the only country where we observe an improvement in method uptake and also the country where we observed the largest effects on mystery client perceptions of taking a method. Engagement with the intervention was also very strong. In Pakistan, the intervention showed promise, but engagement was much lower than in Tanzania. Burkina Faso showed similar improvements in provider attitudes and beliefs as Tanzania and Pakistan, but we do not see improvements in care quality and perceived treatment.

The Beyond Bias intervention demonstrates that it is possible to change negatively biased beliefs that providers have about certain patients and that this can lead to improved patient-centered care and better client experiences (but not in Burkina Faso). Despite improvements in care, there is limited evidence that this intervention impacted the types of FP methods received among young women who visited the clinic. Future work should test whether coupling this intervention with community outreach efforts can increase modern contraception use for young women who would not otherwise come into the clinic.

References

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. (2017). When should you adjust standard errors for clustering? Technical report, National Bureau of Economic Research.

Alsan, M., Garrick, O., and Graziani, G. (2019). Does diversity matter for health? experimental evidence from oakland. *American Economic Review*, 109(12):4071–4111.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association*, 103(484):1481–1495.

Bearak, J. M., Popinchalk, A., Beavin, C., Ganatra, B., Moller, A.-B., Tunçalp, Ö., and Alkema, L. (2022). Country-specific estimates of unintended pregnancy and abortion incidence: a global comparative analysis of levels in 2015–2019. *BMJ global health*, 7(3):e007151.

Calhoun, L. M., Speizer, I. S., Rimal, R., Sripad, P., Chatterjee, N., Achyut, P., and Nanda, P. (2013). Provider imposed restrictions to clients' access to family planning in urban uttar pradesh, india: a mixed methods study. *BMC Health Services Research*, 13(1):1–13.

Chandra-Mouli, V., Lenz, C., Adebayo, E., Lang Lundgren, I., Gomez Garbero, L., and Chatteriee, S. (2018). A systematic review of the use of adolescent mystery clients in assessing the adolescent friendliness of health services in high, middle, and low-income countries. *Global health action*, 11(1):1536412.

Chang, K. T., Mukanu, M., Bellows, B., Hameed, W., Kalamar, A. M., Grépin, K. A., Gul, X., and Chakraborty, N. M. (2019). Evaluating quality of contraceptive counseling: an analysis of the method information index. *Studies in family planning*, 50(1):25–42.

Das, V., Daniels, B., Kwan, A., Saria, V., Das, R., Pai, M., and Das, J. (2021). Simulated patients and their reality: An inquiry into theory and method. *Social Science & Medicine*, page 114571.

Deaton, A. (2020). Randomization in the tropics revisited: a theme and eleven variations. Technical report, National Bureau of Economic Research.

Dieci, M., Wagner, Z., Friedman, W., Burgess, S., Vandermark, J., McCoy, S. I., Shah, M., and Dow, W. H. (2021). Measuring family planning provider bias: A discrete choice experiment among burkinabé, pakistani, and tanzanian providers. *Studies in Family Planning*, 52(3):299–320.

Dunsch, F., Evans, D. K., Macis, M., and Wang, Q. (2018). Bias in patient satisfaction surveys: a threat to measuring healthcare quality. *BMJ global health*, 3(2):e000694.

Fitzpatrick, A. and Tumlinson, K. (2017). Strategies for optimal implementation of simulated clients for measuring quality of care in low-and middle-income countries. *Global Health: Science and Practice*, 5(1):108–114.

Gage, A. D., Kruk, M. E., Girma, T., and Lemango, E. T. (2018). The know-do gap in sick child care in ethiopia. *PLoS One*, 13(12):e0208898.

Greenwald, A. G. and Krieger, L. H. (2006). Implicit bias: Scientific foundations. *California law review*, 94(4):945–967.

Haakenstad, A., Angelino, O., Irvine, C. M., Bhutta, Z. A., Bienhoff, K., Bintz, C., Causey, K., Dirac, M. A., Fullman, N., Gakidou, E., et al. (2022). Measuring contraceptive method mix, prevalence, and demand satisfied by age and marital status in 204 countries and territories, 1970–2019: a systematic analysis for the global burden of disease study 2019. *The Lancet*, 400(10348):295–327.

Hazel, E., Mohan, D., Chirwa, E., Phiri, M., Kachale, F., Msukwa, P., Katz, J., and Marx, M. A. (2021). Disrespectful care in family planning services among youth and adult simulated clients in public sector facilities in malawi. *BMC Health Services Research*, 21(1):1–13.

ICF (2022). The demographic and health survey program website. accessed 12/2022 from http://www.dhsprogram.com.

IDEO (2011). *Human Centered Design: Toolkit*. IDEO.

Ikeda, D. J., Nyblade, L., Srithanaviboonchai, K., and Agins, B. D. (2019). A quality improvement approach to the reduction of hiv-related stigma and discrimination in healthcare settings. *BMJ Global Health*, 4(3):e001587.

Kantorová, V., Wheldon, M. C., Dasgupta, A. N., Ueffing, P., and Castanheira, H. C. (2021). Contraceptive use and needs among adolescent women aged 15–19: Regional and global estimates and projections from 1990 to 2030 from a bayesian hierarchical modelling study. *PloS One*, 16(3):e0247479.

King, J. J., Das, J., Kwan, A., Daniels, B., Powell-Jackson, T., Makungu, C., and Goodman, C. (2019). How to do (or not to do)... using the standardized patient method to measure clinical quality of care in lmic health facilities. *Health policy and planning*, 34(8):625–634.

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

Kwan, A. T. T. (2020). *Can we improve quality of care in private health sectors? Evidence from a randomized field experiment in Kenya*. University of California, Berkeley.

Leonard, K. L. and Masatu, M. C. (2010). Professionalism and the know-do gap: Exploring intrinsic motivation among health workers in tanzania. *Health economics*, 19(12):1461–1477.

Mishra, P. S., Veerapandian, K., and Choudhary, P. K. (2021). Impact of socio-economic inequity in access to maternal health benefits in india: evidence from janani suraksha yojana using nfhs data. *PloS one*, 16(3):e0247935.

Mohanan, M., Vera-Hernández, M., Das, V., Giardili, S., Goldhaber-Fiebert, J. D., Rabin, T. L., Raj, S. S., Schwartz, J. I., and Seth, A. (2015). The know-do gap in quality of health care for childhood diarrhea and pneumonia in rural india. *JAMA pediatrics*, 169(4):349–357.

Moucheraud, C., McBride, K., Heuveline, P., and Shah, M. (2022). Preventing, but not caring for, adolescent pregnancies? disparities in the quality of reproductive health care in subsaharan africa. *Journal of Adolescent Health*.

Ravallion, M. (2020). Should the randomistas (continue to) rule? Technical report, National Bureau of Economic Research.

Schwandt, H. M., Speizer, I. S., and Corroon, M. (2017). Contraceptive service provider imposed restrictions to contraceptive access in urban nigeria. *BMC Health Services Research*, 17(1):1–9.

Sidze, E. M., Lardoux, S., Speizer, I. S., Faye, C. M., Mutua, M. M., and Badji, F. (2014). Young women access and use of contraception: the role of providers' restrictions in urban senegal. *International Perspectives on Sexual and Reproductive Health*, 40(4):176.

Sieverding, M., Schatzkin, E., Shen, J., and Liu, J. (2018). Bias in contraceptive provision to young women among private health care providers in south west nigeria. *International Perspectives on Sexual and Reproductive Health*, 44(1):19–29.

Solo, J. and Festin, M. (2019). Provider bias in family planning services: a review of its meaning and manifestations. *Global Health: Science and Practice*, 7(3):371–385.

Speizer, I. S. and Calhoun, L. M. (2020). Examination of bias across client exit interview data sources. full access full choice technical report.

Speizer, I. S., Guilkey, D. K., Winston, J., and Calhoun, L. M. (2021). Does provider bias affect choice of a facility for family planning services by women in urban senegal? *Studies in Family Planning*.

Sully, E. A., Biddlecom, A., Darroch, J. E., Riley, T., Ashford, L. S., Lince-Deroche, N., Firestein, L., and Murro, R. (2019). Adding it up: Investing in sexual and reproductive health 2019. Report, Guttmacher Institute.

Tumlinson, K., Okigbo, C. C., and Speizer, I. S. (2015). Provider barriers to family planning access in urban kenya. *Contraception*, 92(2):143–151.

Tumlinson, K., Speizer, I. S., Archer, L. H., and Behets, F. (2013). Simulated clients reveal factors that may limit contraceptive use in kisumu, kenya. *Global Health: Science and Practice*, 1(3):407–416.

Tumlinson, K., Speizer, I. S., Curtis, S. L., and Pence, B. W. (2014). Accuracy of standard measures of family planning service quality: findings from the simulated client method. *Studies in Family Planning*, 45(4):443–470.

Vela, M. B., Erondu, A. I., Smith, N. A., Peek, M. E., Woodruff, J. N., and Chin, M. H. (2022). Eliminating explicit and implicit biases in health care: Evidence and research needs. *Annual review of public health*, 43.

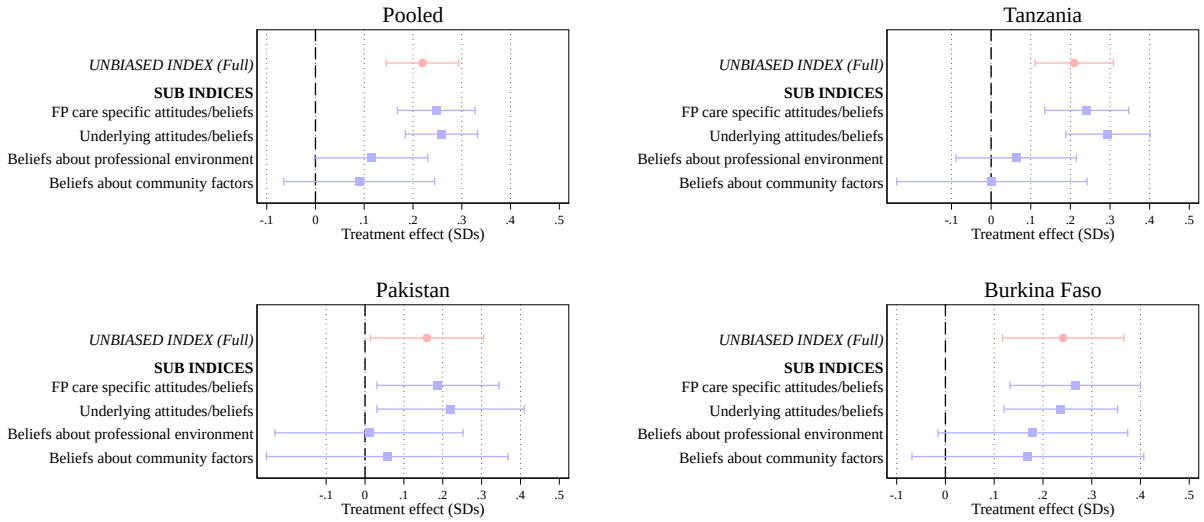
WHO (2005). Decision-making tool for family planning clients and providers.

WHO (2018). Family planning: a global handbook for providers. 2018 edition. *Family planning: a global handbook for providers.*

World Bank (2021). World bank development indicators.

Y-Labs (2019). Beyond bias: Design research report.

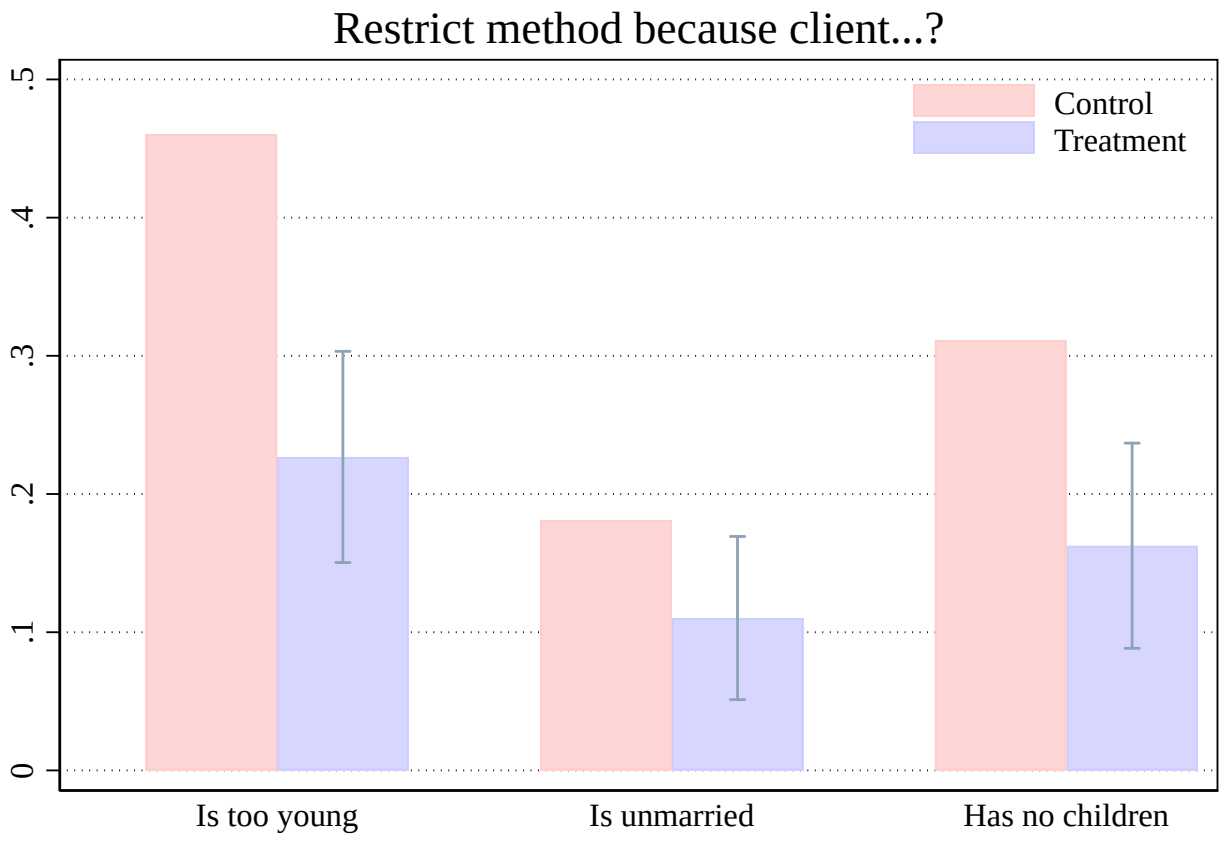
Figure 1: Impact of treatment on bias index from provider survey



Note: Positive means less bias

Notes: Treatment effects and 95% confidence intervals were estimated using equation 1 and 2. Pooled effects control for country fixed-effects. Bias indices were created using the endline provider survey. Indices were created by taking the average of a standardized set of variables by country (see Tables A7 to A11 for complete list of questions for each sub-index). Items that were negatively phrased were re-coded so that higher values represented more supportive family planning attitudes.
Source: Endline provider survey

Figure 2: Impact of treatment on provider reports of restricting methods (all countries pooled)



Notes: Data is based on provider survey responses at endline to three questions about whether provider would restrict certain methods because a client is too young, unmarried, or has no children. Estimates are pooled across countries based on equation 2. Error bars are 95% CIs based on on the treatment effect (β_1 from equation 2). Standard errors were clustered by clinic.
Source: Endline provider survey

Table 1: Description of Sample

	Tanzania	Pakistan	Burkina Faso
FP Clinics Enrolled			
Number of Clinics	73	76	78
Sector	Public	Private	Public
Prior Pathfinder Training	Yes	No	Yes
Providers per Clinic	4.288	1.000	11.090
FP Clients Per Month	226.548	21.099	101.707
FP Providers Surveyed			
Number of Providers	259	70	313
Provider Age	39.871	43.104	41.216
Doctor/Health officer	0.027	0.071	0.048
Nurse/Nurse-Midwife	0.842	0.100	0.032
Midwife	0.131	0.643	0.920
Lady Health Visitor	0.000	0.186	0.000
FP Clients Surveyed			
Number of clients	26915	9869	41187
Age 15-19	0.069	0.082	0.104
Age 20-24	0.357	0.394	0.287
Age 25+	0.574	0.524	0.608
Married	0.653	0.997	0.778
Nulliparous	0.047	0.024	0.086

Notes: Prior Pathfinder Training refers to clinic-wide training on youth friendly service provision that was conducted as part of pre-existing Pathfinder projects. The average FP clients per month in Tanzania is calculated between January 2021 to August 2021. In Pakistan, FP clients per month is estimated between September 2020 to June 2021. Only 70 of the 76 enrolled clinics in Pakistan completed an endline provider survey; 3 in the treatment group and 3 in the control group did not complete an endline survey because providers were on sabbatical.
Source: Data on FP clinics is from administrative data. Data on FP providers is from endline provider survey. Data on FP clients is from ongoing exit surveys and aggregated over the 12 months of the intervention.

Table 2: Data Sources

	Baseline	Ongoing	Endline
Client Exit Surveys	No	Yes	Yes
Provider Surveys	No	No	Yes
Discrete Choice Experiment	No	No	Yes
Mystery Client Visits	No	No	Yes
Administrative Data	Yes*	Yes	Yes

Notes: "Ongoing" means that data was continuously collected throughout the study.
*Baseline administrative data was not available in Pakistan.

Table 3: Mystery Client Design

	Tanzania		Pakistan		Burkina Faso		Total	
Mystery Client Profile	C	T	C	T	C	T	C	T
Younger, Unmarried, No Children	19	18	19	21	19	18	57	57
Younger, Unmarried, One Child	18	18	0	0	20	22	38	40
Younger, Married, No Children	18	18	24	31	21	20	63	69
Younger, Married, One Child	19	18	25	30	19	19	63	67
Older, Unmarried, No Children	18	18	0	0	20	20	38	38
Older, Unmarried, One Child	19	18	0	0	19	18	38	36
Older, Married, No Children	19	18	0	0	19	19	38	37
Older, Married, Parous	18	18	0	0	21	20	39	38
Total	148	144	68	82	158	156	374	382

Notes: The values in each cell represents the number of mystery client visits we conducted for the respective role/country/group. Some profiles were not feasible in Pakistan due to cultural sensitivities. Unequal numbers between treatment and control are due to an odd number off facilities.
T = Treatment group
C = Control group

Table 4: Effect of treatment on primary outcomes

	Unbiased Index		*Counselled on Full Range*		*Given Any Method*		*Perceived Treatment Index*	
	(Provider Survey)		(Mystery Clients)		(Exit Survey)		(Mystery Clients)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Pooled	0.220***	0.216***	0.065*	0.065*	0.019*	0.018	0.215***	0.214***
SE	(0.038)	(0.036)	(0.036)	(0.036)	(0.011)	(0.011)	(0.051)	(0.050)
q-value	[0.001]	[0.001]	[0.045]	[0.053]	[0.045]	[0.055]	[0.001]	[0.001]
Cntl Mean	-0.000	0.002	0.632	0.632	0.907	0.908	0.001	0.001
N	642	642	752	752	30886	30886	752	752
Tanzania	0.210***	0.208***	0.069	0.040	0.047**	0.029**	0.298***	0.289***
SE	(0.050)	(0.047)	(0.055)	(0.052)	(0.020)	(0.014)	(0.081)	(0.076)
q-value	[0.001]	[0.001]	[0.058]	[0.126]	[0.014]	[0.029]	[0.001]	[0.001]
Cntl Mean	0.000	0.001	0.723	0.737	0.911	0.920	-0.000	0.004
N	259	259	292	292	10934	10934	292	292
Pakistan	0.159**	0.160**	0.228***	0.227***	0.004	-0.008	0.334***	0.335***
SE	(0.074)	(0.077)	(0.084)	(0.084)	(0.052)	(0.052)	(0.087)	(0.088)
q-value	[0.025]	[0.029]	[0.013]	[0.014]	[0.305]	[0.282]	[0.002]	[0.002]
Cntl Mean	-0.000	-0.000	0.309	0.309	0.813	0.822	-0.000	-0.000
N	70	70	150	150	4659	4659	150	150
Burkina Faso	0.241***	0.219***	-0.017	0.000	0.002	-0.000	0.080	0.088
SE	(0.063)	(0.051)	(0.056)	(0.049)	(0.010)	(0.008)	(0.086)	(0.064)
q-value	[0.002]	[0.001]	[1.000]	[0.991]	[1.000]	[0.991]	[1.000]	[0.348]
Cntl Mean	0.000	0.012	0.699	0.690	0.934	0.936	-0.000	-0.004
N	313	313	310	310	15293	15293	310	310
Controls	No	Yes	No	Yes	No	Yes	No	Yes

Notes: Each estimate is the coefficient on treatment from equation 1 or 2. Odd columns are from unadjusted regressions (pooled unadjusted models include only country fixed-effects). Even columns include controls. Clinic-clustered standard errors are in parentheses. Sharpened q-values are in brackets. The data source is in parentheses at the top of each column. The provider survey and mystery client data were collected at endline and the exit survey data combines all data for the full 12 month study period. Controls in column 2 include randomization strata, provider age, provider marital status, provider parity, economic status of clients, provider qualifications, and numbers of days per week providing FP care. Controls in columns 4 and 8 include randomization strata and indicators for each mystery profile. Controls in column 6 include randomization strata, client age, marital status, parity, education level, and perceived socioeconomic status, and whether the client was a new FP user.
*** p<0.01, ** p<0.05, * p<0.1
Source: Mystery client data and provider surveys were collected at endline. Client exit survey data was collected continuously and is aggregated over the full 12 month intervention period.

Table 5: Effect of treatment on disparities by age, marital status, and parity (mystery clients)

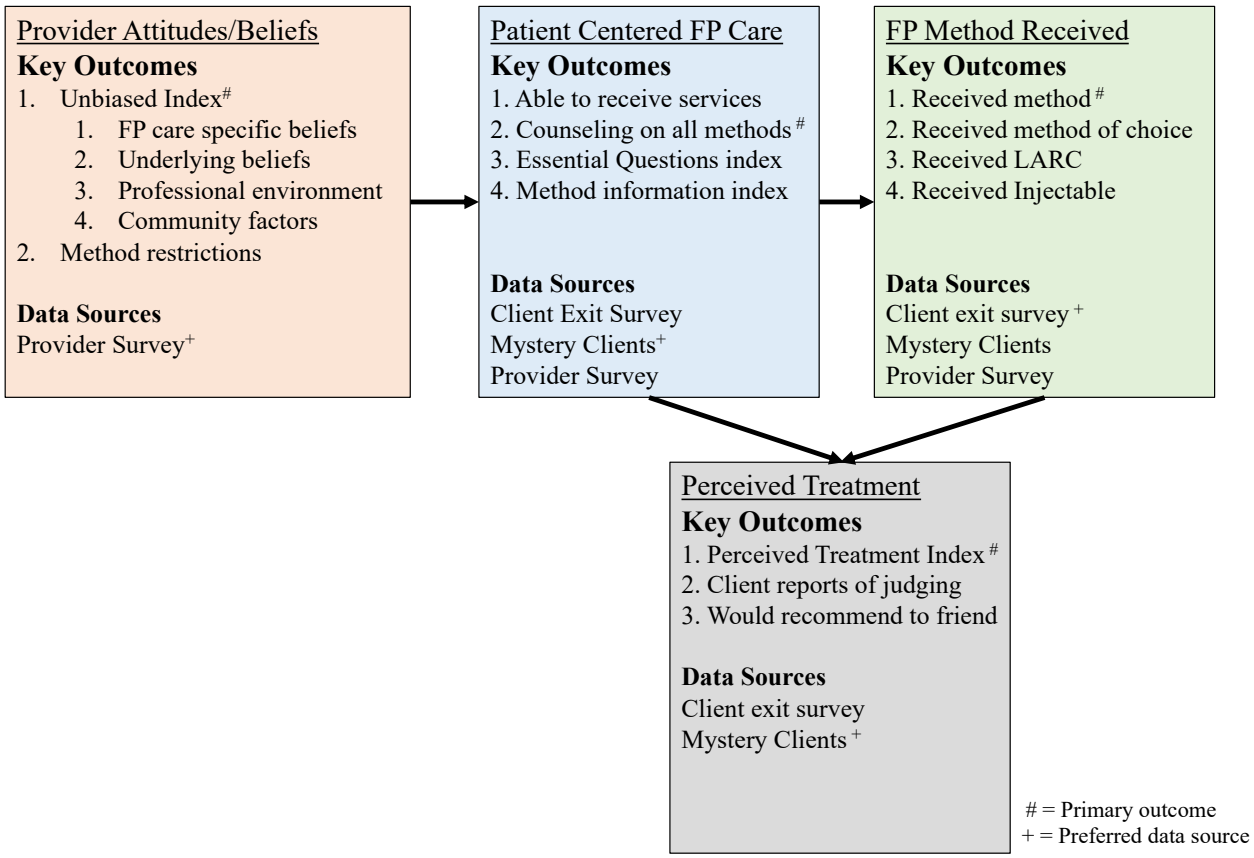
	Counseling on Full Range				*Could Take An FP Method*				*Perceived Treatment Index*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
VARIABLES	Pooled	TZ	PK	BF	Pooled	TZ	PK	BF	Pooled	TZ	PK	BF
Treatment	0.004	-0.021	0.207	-0.038	0.005	-0.018	0.060	0.009	0.235**	0.373**	0.360**	0.046
	(0.068)	(0.104)	(0.137)	(0.098)	(0.064)	(0.086)	(0.094)	(0.111)	(0.091)	(0.154)	(0.155)	(0.115)
Young	-0.007	0.041		0.013	-0.134***	-0.176**		-0.100	-0.061	0.082		-0.125*
	(0.045)	(0.064)		(0.068)	(0.046)	(0.069)		(0.073)	(0.059)	(0.096)		(0.067)
Unmarried	0.069	0.068	0.066	0.038	-0.029	-0.041	-0.116	-0.010	-0.187***	-0.292***	0.092	-0.197**
	(0.043)	(0.064)	(0.147)	(0.066)	(0.044)	(0.073)	(0.073)	(0.069)	(0.054)	(0.074)	(0.123)	(0.089)
Nulliparous	-0.135***	-0.122**	-0.110	-0.141*	-0.023	0.041	0.118	-0.125	-0.002	0.253***	-0.109	-0.232***
	(0.043)	(0.057)	(0.128)	(0.074)	(0.046)	(0.055)	(0.087)	(0.085)	(0.049)	(0.066)	(0.136)	(0.068)
Treatment X Young	0.020	-0.068		-0.026	0.003	0.106		-0.084	-0.084	-0.241*		-0.083
	(0.065)	(0.091)		(0.104)	(0.060)	(0.083)		(0.108)	(0.082)	(0.132)		(0.115)
Treatment X Unmarried	-0.038	0.099	-0.218	-0.034	0.057	0.110	0.133	-0.001	0.025	0.074	-0.113	0.091
	(0.062)	(0.098)	(0.196)	(0.092)	(0.057)	(0.091)	(0.099)	(0.099)	(0.074)	(0.109)	(0.154)	(0.119)
Treatment X Nulliparous	0.125**	0.149*	0.124	0.102	0.055	0.084	-0.083	0.072	0.034	0.017	0.008	0.059
	(0.062)	(0.080)	(0.182)	(0.104)	(0.059)	(0.071)	(0.114)	(0.115)	(0.071)	(0.100)	(0.178)	(0.110)
N	752	292	150	310	756	292	150	314	752	292	150	310
Controls	No	No	No	No	No	No	No	No	No	No	No	No
Country FE	Yes	No	No	No	Yes	No	No	No	Yes	No	No	No

Notes: Estimates are from equation 5. We did not vary mystery client ages in Pakistan which is why no coefficients on *young* were estimated in columns 3, 7, and 11. Pooled coefficients on *young* only include observations from Tanzania and Burkina Faso. *** p<0.01, ** p<0.05, * p<0.1.
Source: Mystery client data collected at endline.

FOR ONLINE PUBLICATION

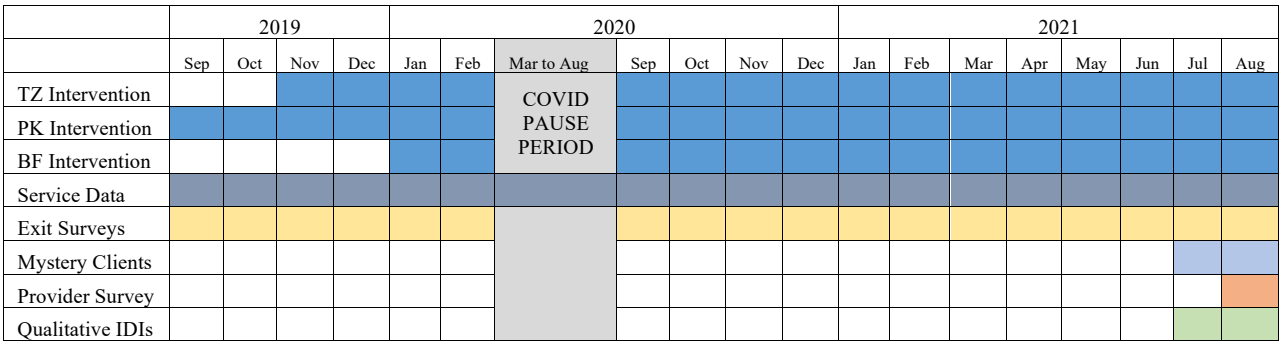
Appendix

Figure A1: Theory of change and key quantitative outcomes



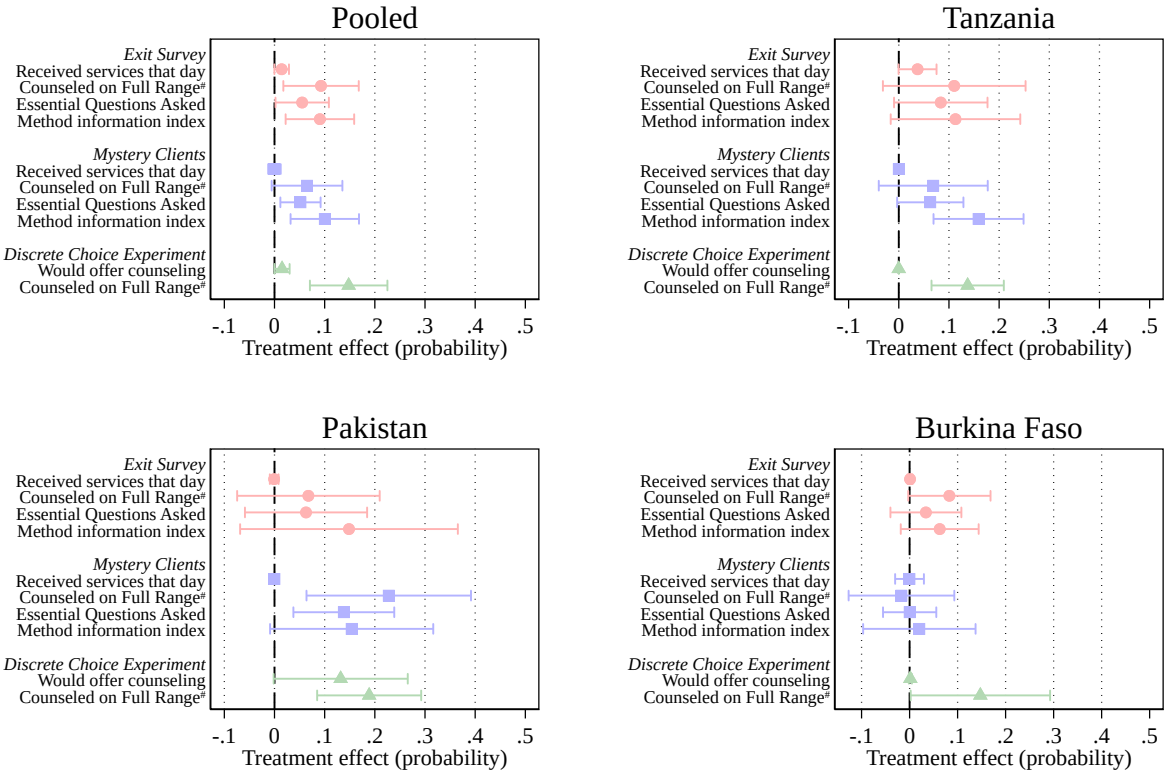
Notes: Each box represents the outcomes included in each domain. Each domain has a primary outcome that we pre-registered, which is indicated with a "#". Several domains had multiple data sources that measured the same thing. We indicate which data source is our preferred source with "+".

Figure A2: Study Timeline



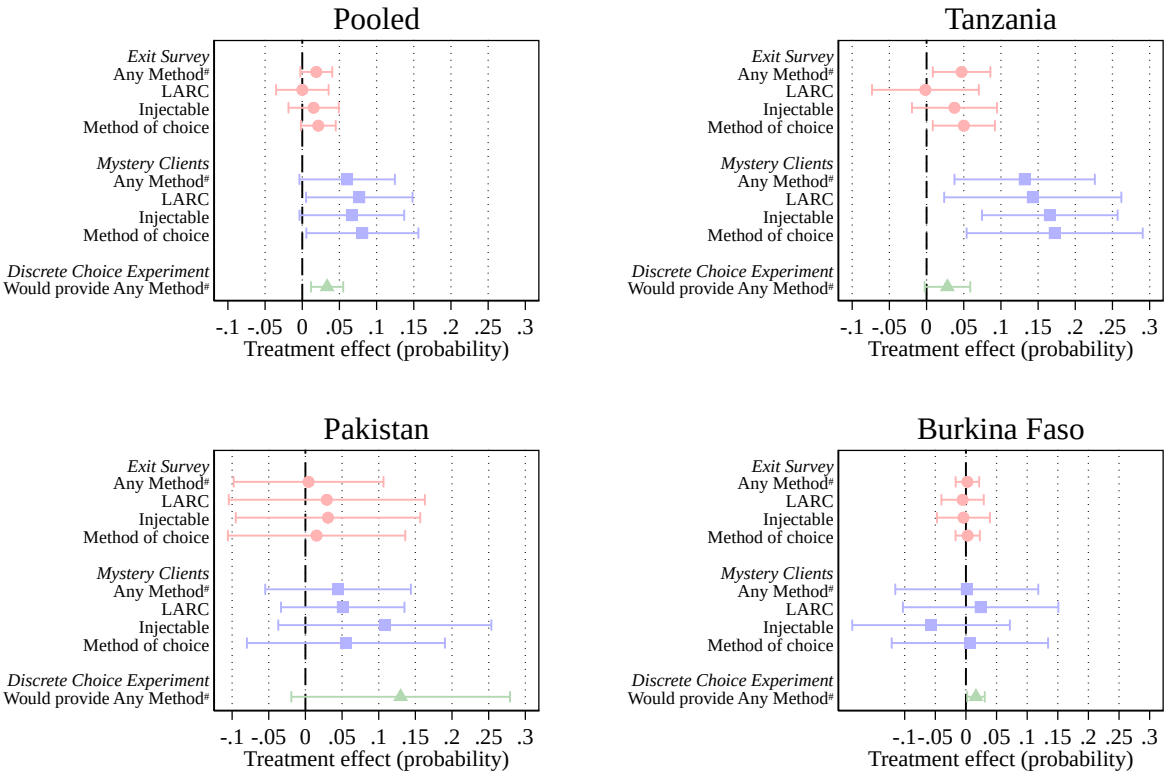
Notes: We focus on the last 12 months of the intervention after the COVID pause period.

Figure A3: Impact of treatment on patient-centered FP care outcomes



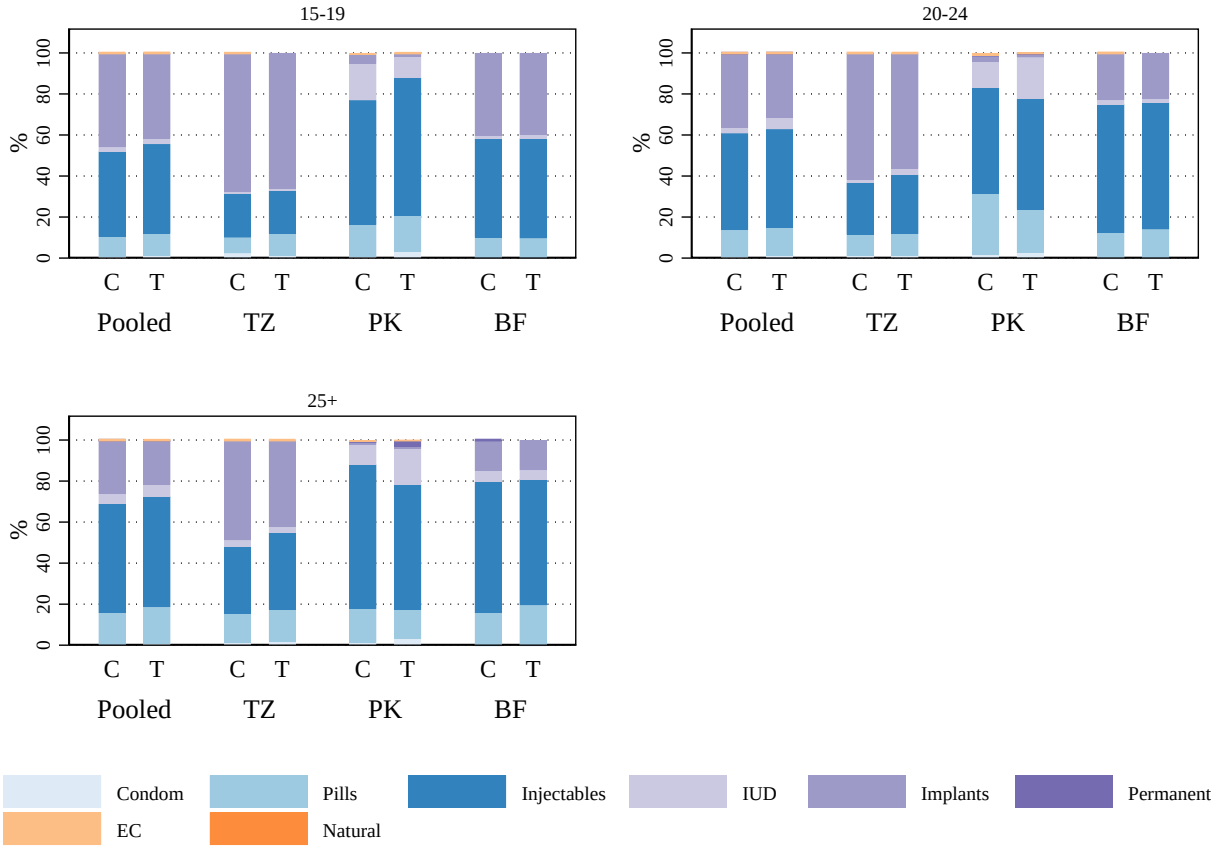
Notes: Each point represents the treatment effect estimated from equations 1 and 2. Error bars are 95% confidence intervals estimated with standard errors clustered by clinic. Pooled results control for country fixed effects. All outcomes are binary aside from "essential questions asked" and "method information index" which are each based on sets of three questions and take on a value of 1 if all three questions had a "yes" response. Counseling of the full range of methods is the pre-specified primary outcome for this domain.
Source: Mystery client data and discrete choice experiment data were collected at endline. Client exit survey data was collected continuously and is aggregated over the full 12 month intervention period.

Figure A4: Impact of treatment on method dispensing outcomes



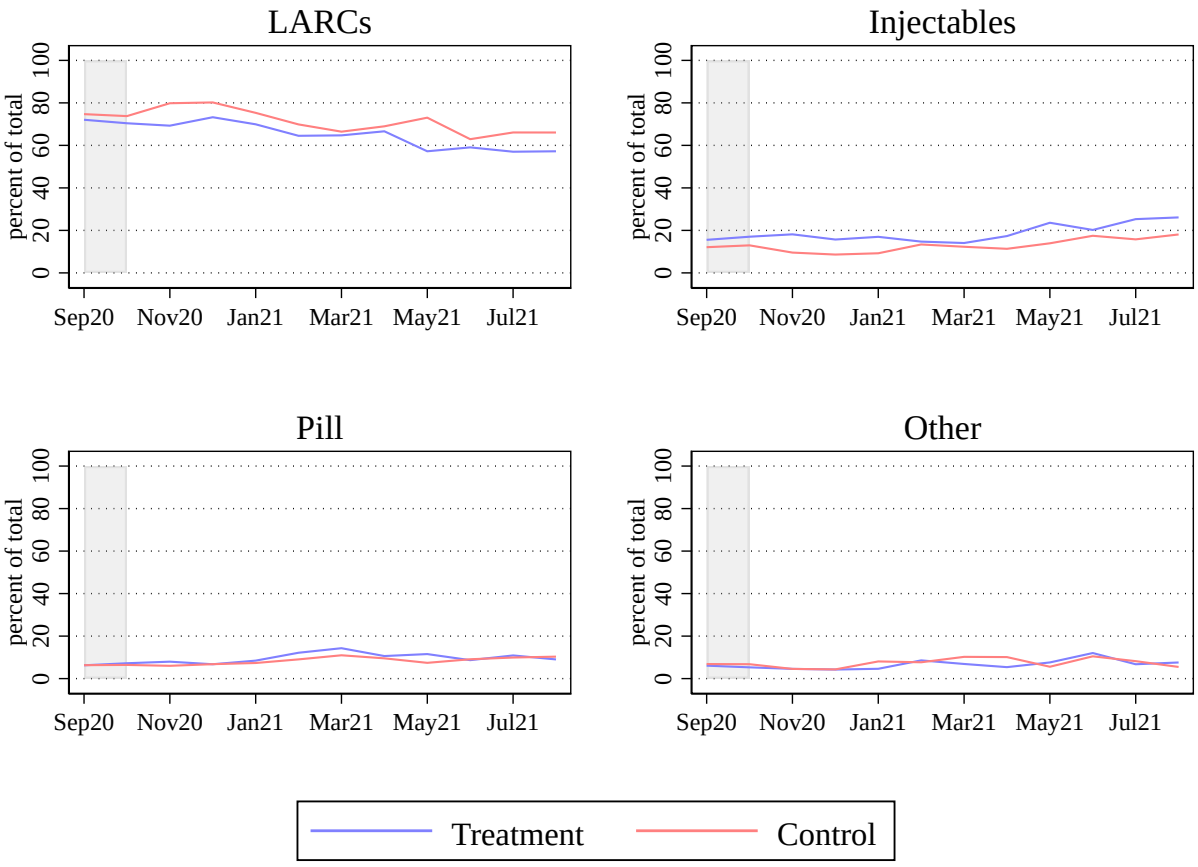
Notes: Each point represents the treatment effect estimated from equations 1 and 2. Error bars are 95% confidence intervals estimated with standard errors clustered by clinic. Pooled results control for country fixed effects. All outcomes are binary. Whether the client received any method is the pre-specified primary outcome for this domain. Mystery client data uses mystery clients perceptions of whether they thought they could take each method type. LARC is Long Active Reversible Contraception and includes IUDs and implants.
Source: Mystery client data and discrete choice experiment data were collected at endline. Client exit survey data was collected continuously and is aggregated over the full 12 month intervention period.

Figure A5: Method Mix (client exit survey)



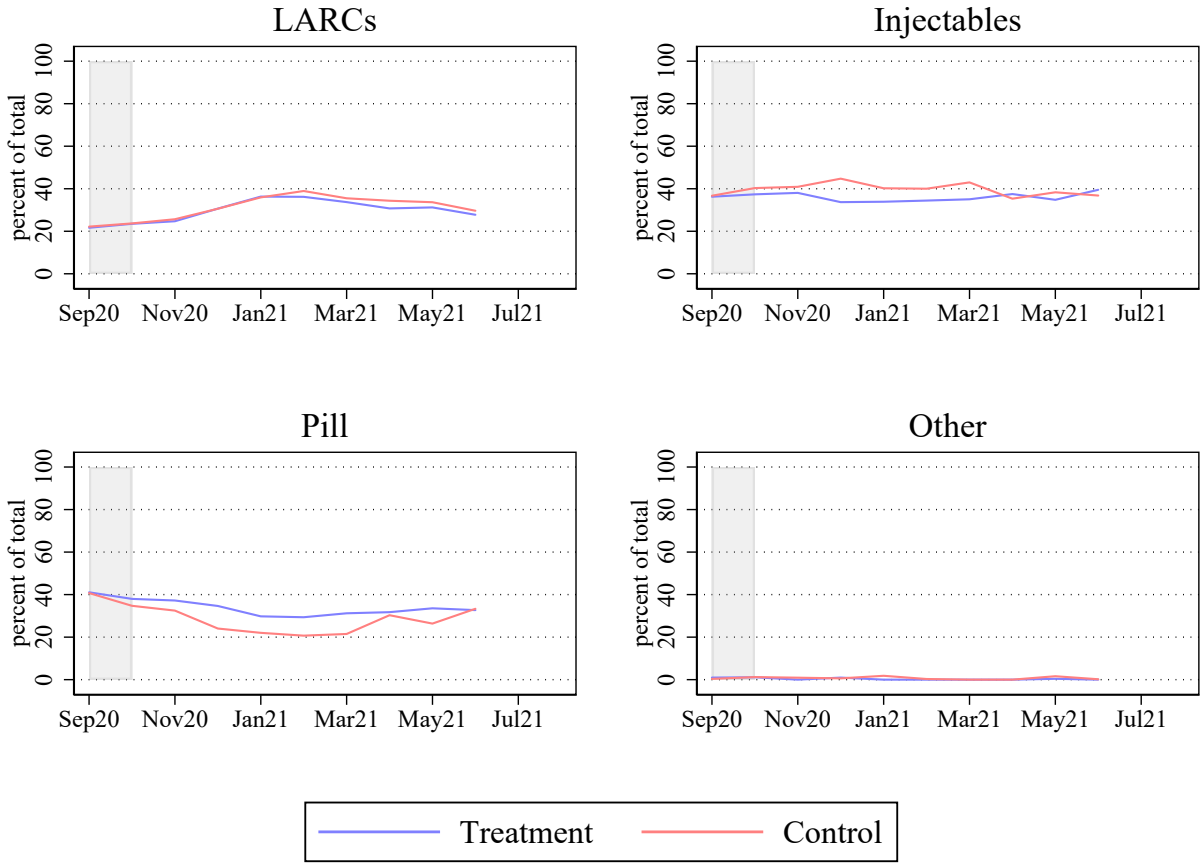
Notes: Bars portray the methods that clients reported receiving as share of the total number of methods reported. Less than a half of a percent reported receiving more than one method and were excluded from this figure. C = Control; T = Treatment; EC = Emergency Contraception.
Source: Exit survey data collected continuously and aggregated over the full 12 month intervention period.

Figure A6: Method mix over time (Tanzania)



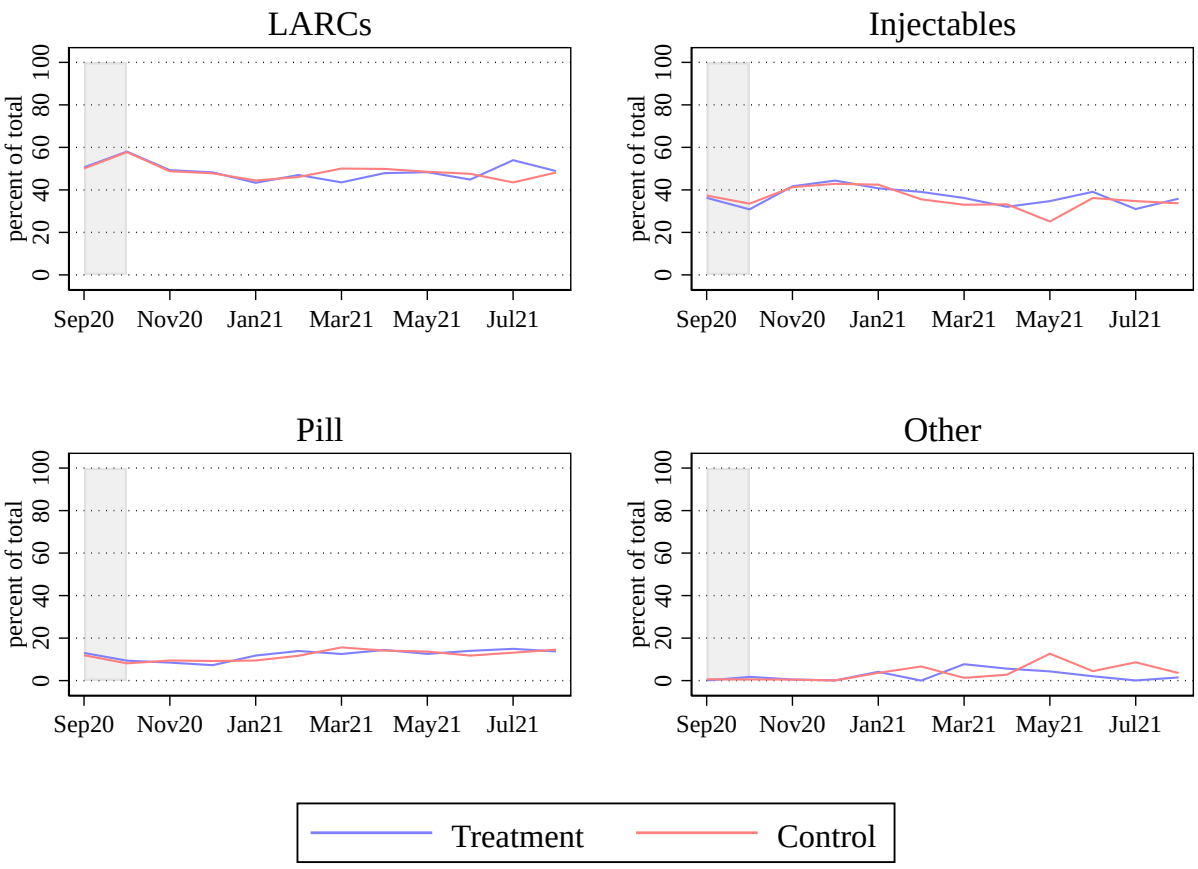
Notes: Data include all new users of family planning of any age. "Other" methods include female and male condoms, tubal ligations, and vasectomies.
Source: Administrative data routinely reported by each clinic to the Ministry of Health.

Figure A7: Method mix over time (Pakistan)



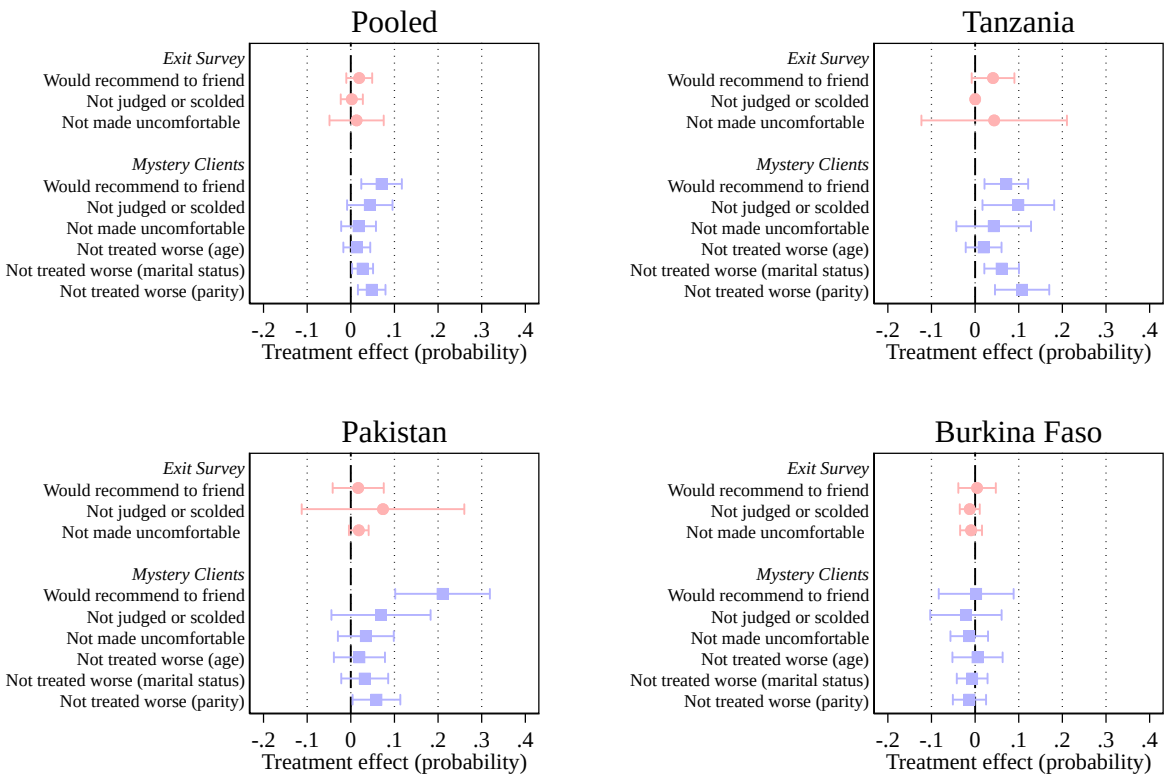
Notes: Data are for 75 of 76 facilities. Data from July and August were excluded because a large proportion of facilities were missing data. All types of clients (new users, returning users) are included in this figure. "Other" methods include sterilizations and condoms.
Source: Administrative data reported to Green Star, the social franchising organization with which each clinic was affiliated

Figure A8: Method mix over time (Burkina Faso)



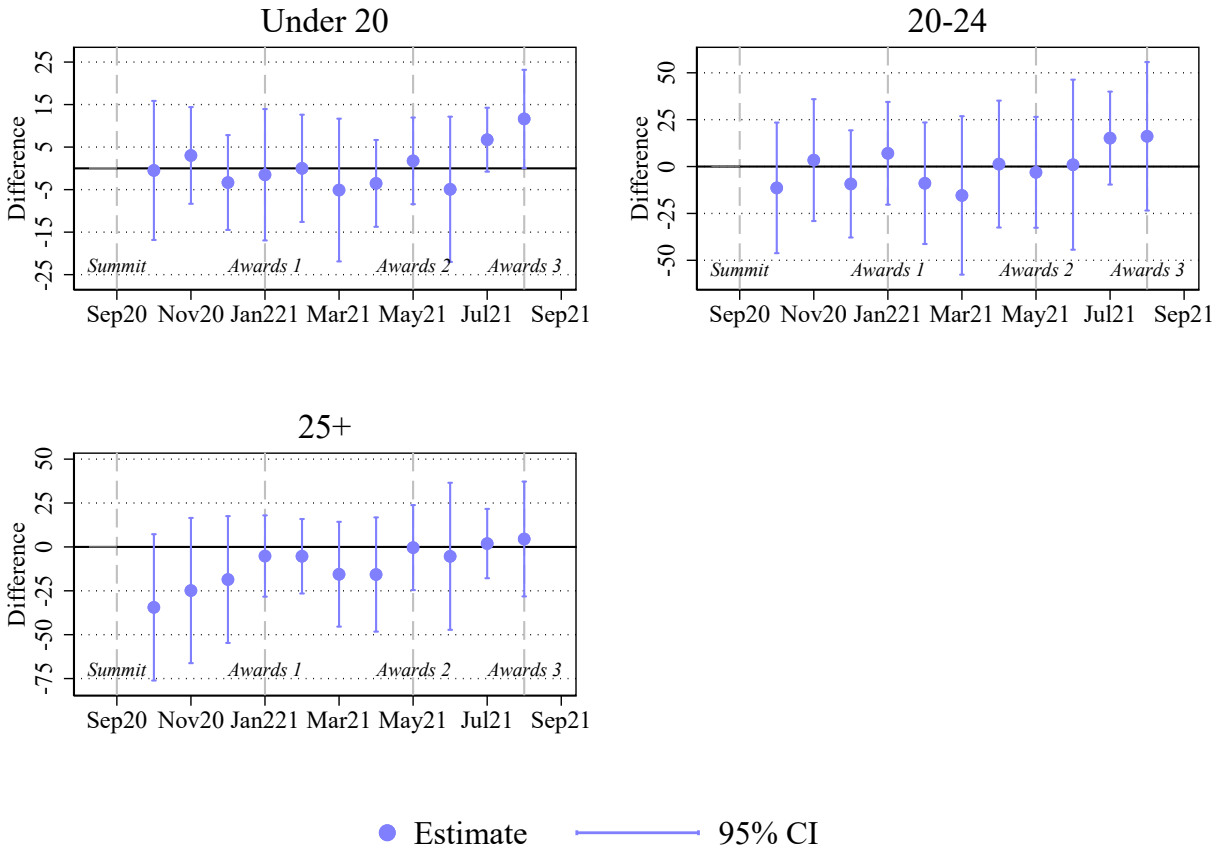
Notes: Data include all new users of family planning of any age. Other methods include female and male condoms, tubal ligations, and vasectomies.
Source: Administrative data routinely reported by each clinic to the Ministry of Health.

Figure A9: Impact of treatment on client perceptions of treatment (secondary outcomes)



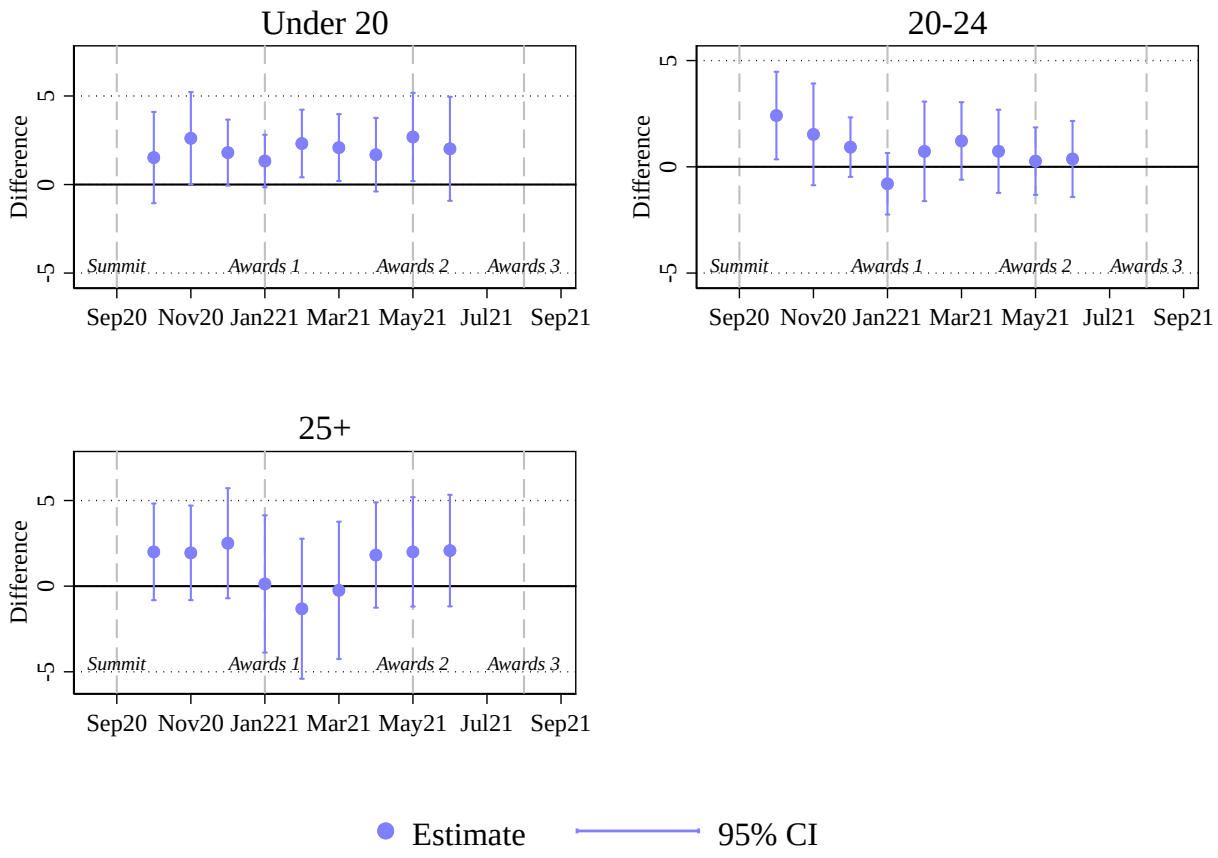
Notes: Each point represents the treatment effect estimated from equations 1 and 2. Error bars are 95% confidence intervals estimated with standard errors clustered by clinic. Pooled results control for country fixed effects. All outcomes are binary.
Source: Mystery client data was collected at endline. Client exit survey data was collected continuously and is aggregated over the full 12 month intervention period.

Figure A10: Client Volume By Age (Tanzania)



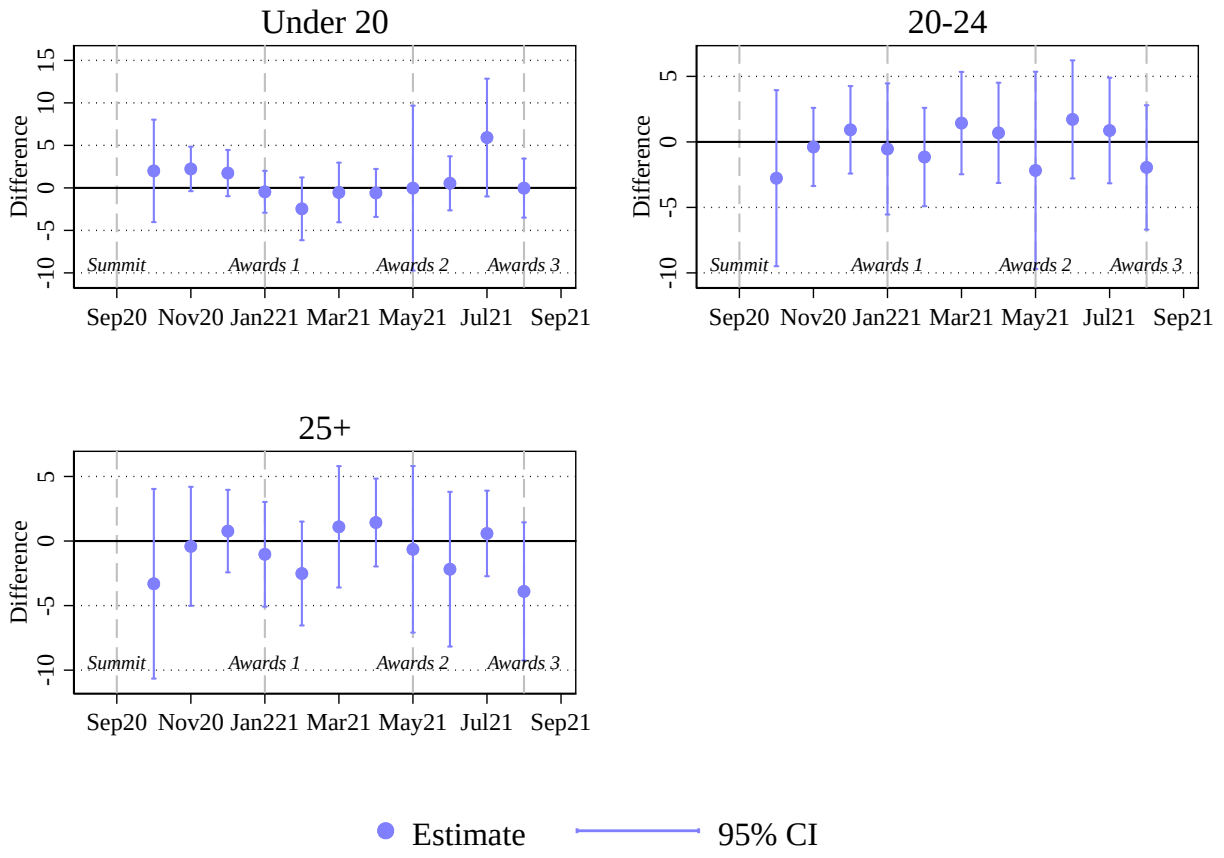
Notes: Plot shows treatment effects by month estimated using equation 3. Data include all new users of family planning.
Source: Administrative data routinely reported by each clinic to the Ministry of Health

Figure A11: Client Volume By Age (Pakistan)



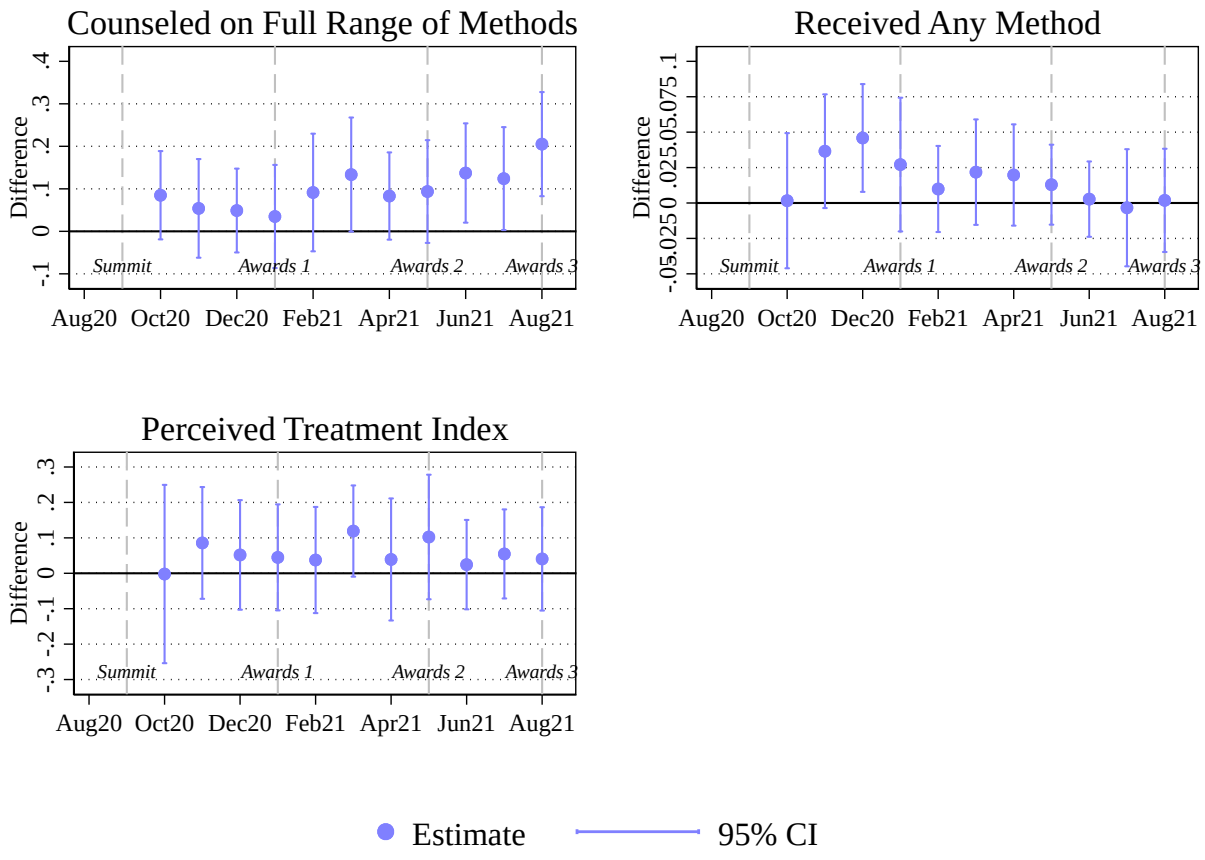
Notes: Plot shows treatment effects by month estimated using equation 3. Data are for 75 of 76 facilities. Data from July and August were excluded because a large proportion of facilities were missing data. All types of clients (new users, returning users) are included in this figure. "Other" methods include sterilizations and condoms.
Source: Administrative data reported to Green Star, the social franchising organization with which each clinic was affiliated. Gray shaded area represents data collected prior to the restart of the intervention.

Figure A12: Client Volume By Age (Burkina Faso)



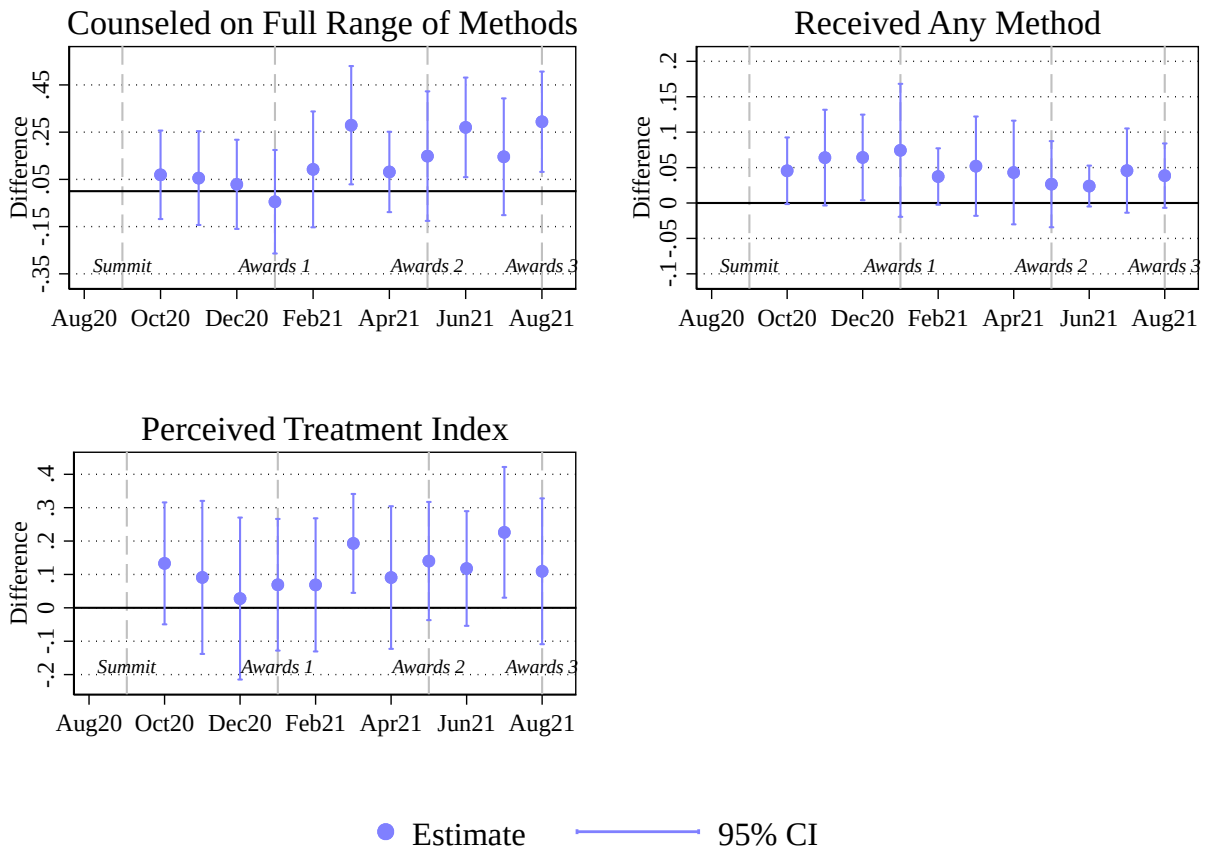
Notes: Data include all new users of family planning.
Source: Plot shows treatment effects by month estimated using equation 3. Administrative data routinely reported by each clinic to the Ministry of Health

Figure A13: Primary outcomes over time (Countries Pooled)



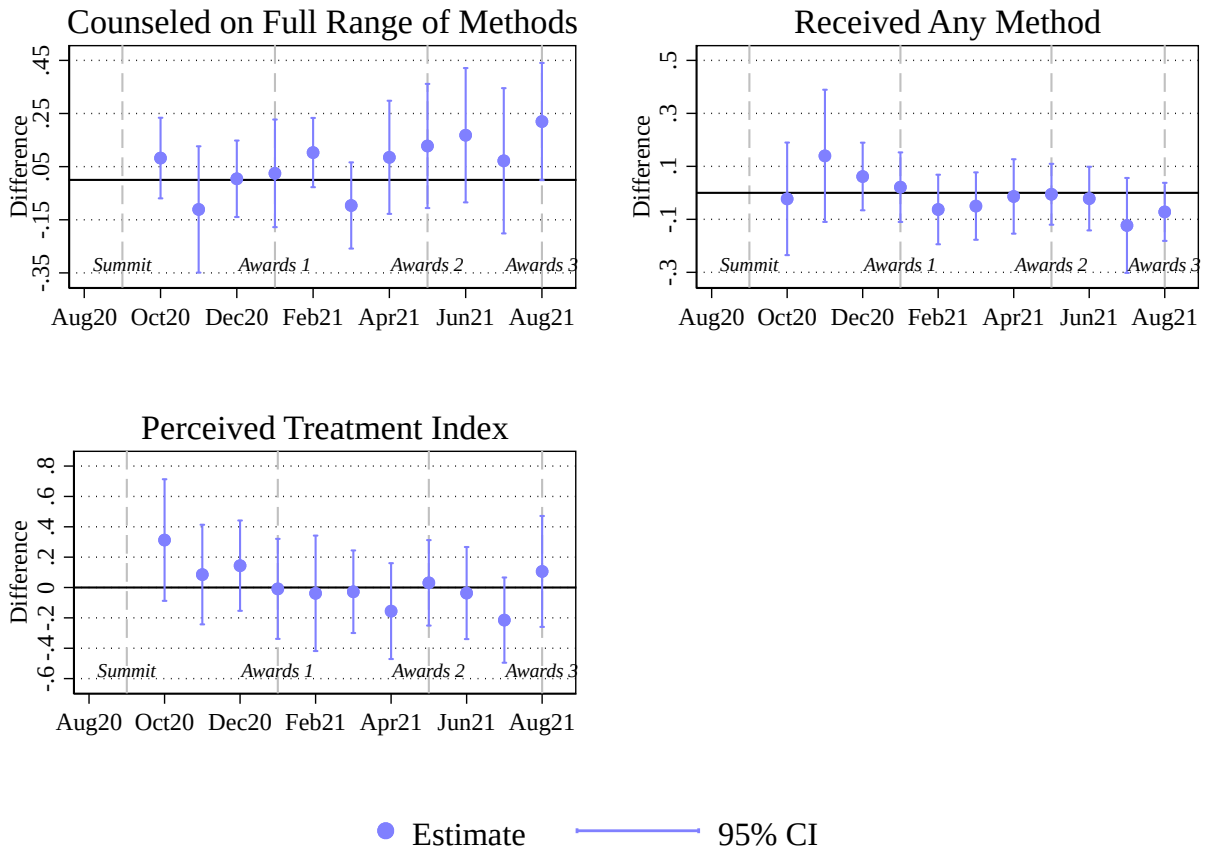
Notes: Plot shows treatment effects by month estimated using equation 3. Data includes clients under 25.
Source: Exit survey data collected continuously and aggregated for each study month.

Figure A14: Primary outcomes over time (Tanzania)



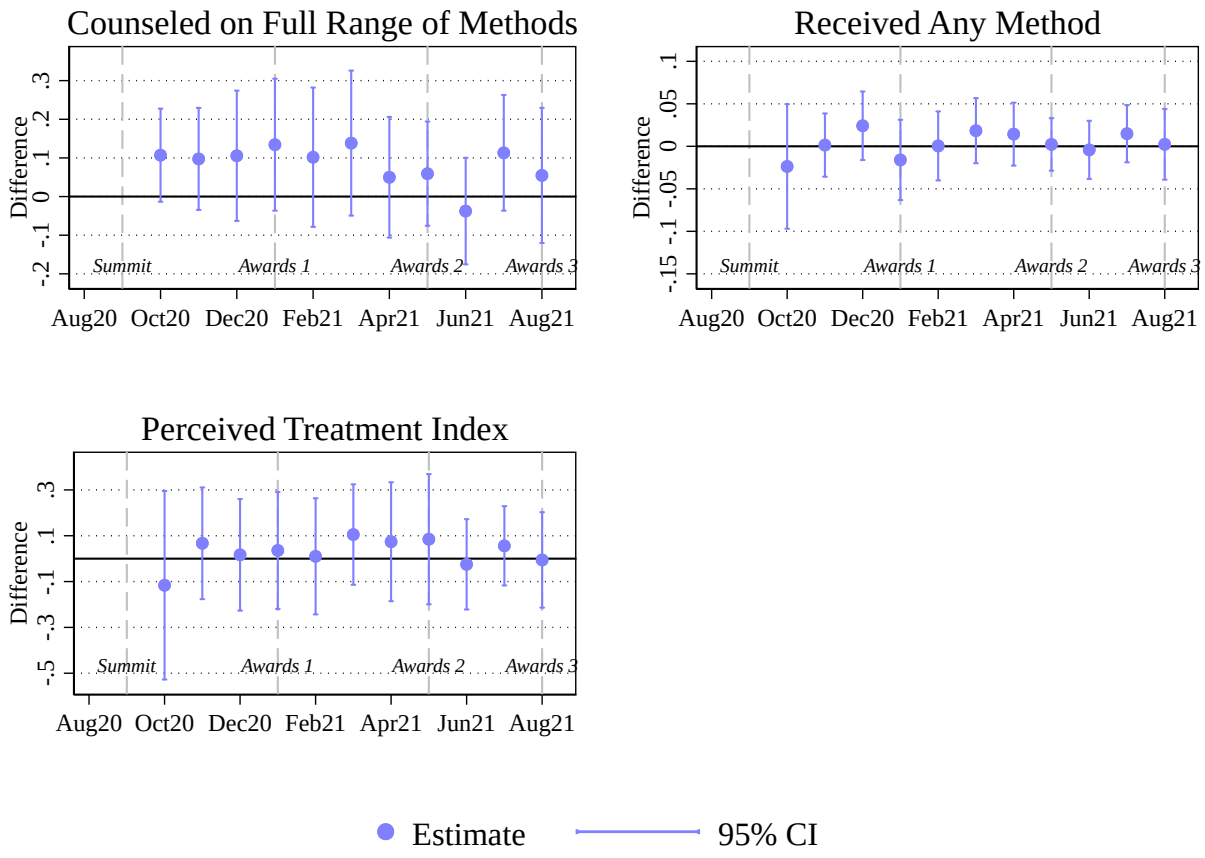
Notes: Plot shows treatment effects by month estimated using equation 3. Data includes clients under 25.
Source: Exit survey data collected continuously and aggregated for each study month.

Figure A15: Primary outcomes over time (Pakistan)



Notes: Plot shows treatment effects by month estimated using equation 3. Data is from the exit survey and includes clients under 25.
Source: Exit survey data collected continuously and aggregated for each study month.

Figure A16: Primary outcomes over time (Burkina Faso)



Notes: Plot shows treatment effects by month estimated using equation 3. Data includes clients under 25.
Source: Exit survey data collected continuously and aggregated for each study month.

Table A1: Development indicators and contraception use in study countries

	Tanzania	Pakistan	Burkina Faso
GDP per capita (USD, 2020)	1,099	1,505	893
Below international poverty line	45%	5%	31%
Rural	64%	63%	69%
At least secondary education	23%	25%	27%
Gender development index	0.94	0.81	0.90
Total fertility rate	4.8	3.4	5.0
Contraception prevalence (woman 15-49)	35%	24%	27%
Unintended pregnancy rate (per 1,000 women 15-49)	105	71	75
Unmet need for contraception	16%	13%	21%

Notes: GDP per capita, share below poverty line, rural residency, total fertility rate, and contraception prevalence are from the World Bank. Secondary education is from the Demographic and Health Surveys (ICF, 2022). Unintended pregnancy rate is from Bearak et al. (2022). Unmet need for contraception is from Haakenstad et al. (2022). The international poverty line is $2.15 per day.

Table A2: Full range of methods based on different client preferences (client exit data)

Method Preferences	Methods
No Preference	IUD, Implant, Injectable, Pill
Easy to hide	IUD, Implant, Injectable
Easy to stop using	Pill, Injectable
Last a long time	IUD, Implant
Does not require taking a pill	IUD, Implant, Injectable
Does not require a procedure	Injectable, Pill
Does not want another child	Permanent, IUD, Implant

Notes: Respondents reported method preferences during exit survey. The "Methods" column includes all methods considered to be the "full range" of methods.

Table A3: Balance between treatment and Control Clinics/Providers

	Pooled	*Tanzania*			*Pakistan*			*Burkina Faso*		
	(Diff)	(C)	(T)	(Diff)	(C)	(T)	(Diff)	(C)	(T)	(Diff)
Baseline Administrative Data										
Clinics		37	36					39	39	
Providers per facility	–0.236	4.432	4.139	0.293				10.769	11.41	–0.641
Clients per month										
Age <20	2.721	20.243	24.931	4.687				5.615	6.369	0.754
Age 20–24	–0.711	53.631	52.889	–0.742				7.287	7.277	–0.010
Age 25+	–7.083	79.045	67.847	–11.198				7.667	6.169	–1.497
Methods per month										
LARC	–7.555	125.383	112.032	–13.350				10.790	11.303	0.513
Injectable	3.562	14.275	21.153	6.878				6.836	6.831	–0.005
Pill	–0.029	6.122	6.537	0.415				3.051	2.574	–0.477
Other	–0.751	7.140	5.944	–1.195				0.256	0.097	–0.159
Total	–5.073	152.919	145.667	–7.252				20.569	19.815	–0.754
Endline Provider Survey										
Number of Providers		141	118		32	38		154	159	
Number of Clinics		37	35		32	38		39	39	
Provider age	0.815	39.273	40.581	1.308	43.233	43.000	–0.233	40.893	41.526	0.633
Provider married	–0.029	0.799	0.784	–0.014	0.875	0.895	0.020	0.765	0.713	–0.051
Provider has at least one child	0.015	0.907	0.907	0.000	0.759	0.895	0.136	0.909	0.911	0.002
Provider has ever used FP	0.002	0.772	0.795	0.023	0.480	0.500	0.020	0.903	0.885	–0.017
Provider qualifications										
Doctor/Health officer	0.028	0.007	0.051	0.044	0.125	0.026	–0.099	0.026	0.069	0.043
Nurse/Nurse-Midwife	–0.014	0.865	0.814	–0.052	0.063	0.132	0.069	0.032	0.031	–0.001
Midwife	0.005	0.128	0.136	0.008	0.531	0.737	0.206*	0.942	0.899	–0.042
Lady Health Visitor	–0.019*	N/A	N/A	N/A	0.281	0.105	–0.176*	N/A	N/A	N/A
Economic status of clients										
Well off or middle class	–0.038	0.128	0.059	–0.068*	0.281	0.105	–0.176*	0.065	0.082	0.017
Poor or vulnerable	0.006	0.348	0.271	–0.076	0.312	0.237	–0.076	0.305	0.396	0.091
A mix	0.031	0.496	0.644	0.148**	0.344	0.474	0.130	0.578	0.491	–0.087
Don't know	0.002	0.028	0.025	–0.003	0.063	0.184	0.122	0.052	0.031	–0.021
Frequency of FP care										
1–2 days per week	0.008	0.035	0.034	–0.002	0.031	0.026	–0.005	0.019	0.038	0.018
3–5 days per week	0.017	0.837	0.831	–0.006	0.250	0.368	0.118	0.143	0.157	0.014
6–7 days per week	–0.025	0.128	0.136	0.008	0.719	0.605	–0.113	0.838	0.805	–0.033

Notes: Pakistan administrative data was only available starting in September 2020 so is not included. "Other" methods in Tanzania include female and male condoms and female and male sterilization. In Burkina Faso, "other" methods include female and male condoms, natural methods, and other methods. "Clients per month" and "Methods per month" are for new clients only. Lady Health Visitors is a cadre of providers specific to Pakistan and therefore is missing in Tanzania and Burkina Faso. Difference between control and treatment group means and significance estimated using linear regression with standard errors clustered at the facility. In Pakistan, there were no doctors and all providers in the Doctor/Health officer/DHMS row had a DHMS qualification. *** p<0.01, ** p<0.05, * p<0.1
Source: Baseline administrative data and endline provider survey data

Table A4: Effect of treatment on individual provider attitudes related to family planning care in provider survey

	Pooled		*Tanzania*			*Pakistan*			*Burkina Faso*	
	(Diff)	(C)	(T)	(Diff)	(C)	(T)	(Diff)	(C)	(T)	(Diff)
Number of Providers		141	118		32	38		154	159	
Number of Clinics		37	35		32	38		39	39	
Agree or strongly agree...										
Okay for young clients who are unmarried to use contraception	0.062***	0.894	0.966	0.072**	0.375	0.526	0.151	0.955	0.987	0.033*
Married & unmarried clients should have the same FP options	0.043	0.674	0.712	0.038	0.500	0.711	0.211*	0.890	0.899	0.010
Young women under 20 have capacity to make health care decisions, including about contraception	0.058**	0.922	0.907	–0.015	0.469	0.816	0.347***	0.903	0.956	0.053*
You feel comfortable providing contraception to an unmarried client	0.063***	0.887	0.975	0.088**	0.594	0.789	0.196*	0.974	0.987	0.013
You feel comfortable providing contraception to a married client who has not yet had a child	0.092***	0.809	0.949	0.141***	0.813	0.816	0.003	0.890	0.962	0.073**
You are always confident you have the clinical skills to provide contraceptives safely to young, unmarried clients	0.037**	0.957	0.992	0.034*	0.750	0.868	0.118	0.948	0.969	0.021
Providing appropriate contraceptives to young, unmarried clients is completely in your control	0.063*	0.929	0.915	–0.014	0.813	0.816	0.003	0.747	0.887	0.140**
Hormonal methods are safe for youth's growing bodies (under 20)	0.189***	0.440	0.678	0.238***	0.406	0.316	–0.090	0.532	0.742	0.210***
You would provide FP services to a client even if you thought they were too young to use contraception	0.099***	0.638	0.856	0.218***	0.562	0.579	0.016	0.942	0.962	0.021
Disagree or strongly disagree...										
Unmarried youth clients require consent from parents before FP provided	0.079***	0.830	0.915	0.085**	0.219	0.447	0.229**	0.935	0.975	0.040
Young married clients require husband's consent before FP provided	0.155***	0.582	0.771	0.190***	0.156	0.368	0.212**	0.760	0.874	0.114**
Married women who do not have any children should not use FP	0.124***	0.660	0.814	0.154***	0.375	0.526	0.151	0.851	0.943	0.093***
Contraceptives are more appropriate for women at least 20 years old than for women ¡20	0.117***	0.695	0.839	0.144***	0.250	0.289	0.039	0.818	0.931	0.113***
Contraceptives are more appropriate for married women	0.040*	0.837	0.907	0.070*	0.219	0.289	0.071	0.955	0.962	0.008
It is important for women who have not been pregnant to prove their fertility before using some FP methods	0.147***	0.496	0.788	0.292***	0.406	0.342	–0.064	0.831	0.906	0.074*
For a family planning consult, it is not important to know if the client is married cli	0.163***	0.348	0.610	0.263***	0.188	0.421	0.234**	0.701	0.767	0.066
If the client hasn't yet had a child, she should avoid using injectables	0.232***	0.383	0.627	0.244***	0.281	0.526	0.245**	0.656	0.874	0.218***
If the client hasn't yet had a child, she should avoid using an IUD	0.097***	0.809	0.924	0.115***	0.438	0.632	0.194	0.883	0.943	0.060
If the client hasn't yet had a child, she should avoid using implants	0.067**	0.844	0.907	0.063	0.281	0.395	0.113	0.890	0.950	0.060*
You usually know what a young client under 20 needs for FP as soon as they come in	–0.029	0.319	0.246	–0.073	0.250	0.211	–0.039	0.662	0.673	0.011
You usually decide what family planning method young clients under 20 should use	0.028	0.589	0.653	0.064	0.250	0.263	0.013	0.968	0.969	0.001
You prefer not to provide FP methods to unmarried clients if they will not take HIV test	0.033	0.738	0.805	0.067	0.469	0.605	0.137	0.994	0.975	–0.019
Young women without children should not use products that might delay fertility once stopped	0.224***	0.426	0.602	0.176***	0.156	0.132	–0.025	0.448	0.767	0.319***
It's okay for another medical professional to come into the room when you're giving an FP consultation	0.113***	0.716	0.864	0.148**	0.719	0.474	–0.245**	0.604	0.767	0.163**
Young clients (under 20) are not capable of choosing the method that is best for them	0.067*	0.582	0.686	0.105	0.156	0.105	–0.051	0.812	0.874	0.063
A client with one daughter will have different FP needs than a client with one son	–0.029	0.950	0.890	–0.061	0.594	0.605	0.012	0.981	0.969	–0.012
You usually counsel unmarried clients under 20 to practice abstinence	0.222***	0.397	0.686	0.289***	0.312	0.237	–0.076	0.597	0.830	0.233***

Notes: Control and treatment group means and significance estimated using linear regression with standard errors clustered at the facility. *** p<0.01, ** p<0.05, * p<0.1

Source: Endline provider survey.

Table A5: Effect of treatment on individual provider attitudes and beliefs that could be drivers of bias

	Pooled			*Tanzania*			*Pakistan*			*Burkina Faso*		
	(Diff)	(C)	(T)	(Diff)	(C)	(T)	(Diff)	(C)	(T)	(Diff)		
Number of Providers		141	118		32	38		154	159			
Number of Clinics		37	35		32	38		39	39			
Agree or strongly agree...												
Sex is part of a healthy life for young women	0.135***	0.284	0.525	0.242***	0.781	0.789	0.008	0.773	0.849	0.076		
It is okay for young women to have sex before marriage	0.126***	0.121	0.339	0.218***	0.062	0.053	–0.010	0.468	0.547	0.080		
Disagree or strongly disagree...												
As a provider, you have a responsibility to teach young women how to behave	0.016	0.014	0.076	0.062**	0.125	0.079	–0.046	0.058	0.050	–0.008		
Young couples should have children as soon as possible after marriage	0.103***	0.553	0.610	0.057	0.344	0.526	0.183	0.494	0.616	0.123**		
Providing contraceptives for unmarried young women may make them more promiscuous	0.113***	0.631	0.771	0.140**	0.125	0.316	0.191*	0.838	0.912	0.074*		
Important to help youth clients understand when their actions are immoral/irresponsible	0.035	0.014	0.068	0.054**	0.031	0.053	0.021	0.065	0.088	0.023		
Young FP clients may need to be punished for bad behavior	0.075**	0.730	0.780	0.049	0.594	0.711	0.117	0.857	0.943	0.086**		
Irresponsible for young women to have sex before they are married	0.141***	0.574	0.737	0.163**	0.219	0.132	–0.087	0.695	0.868	0.173***		
Providing young unmarried clients with FP makes you worry about spread of HIV	0.126***	0.674	0.822	0.148***	0.156	0.342	0.186*	0.766	0.862	0.095*		

Notes: Control and treatment group means and significance estimated using linear regression with standard errors clustered at the facility. *** p<0.01, ** p<0.05, * p<0.1
Source: Endline provider survey.

Table A6: Effect of treatment on outcomes in DCE

	Would Offer Counseling	*Would Provide Modern Method*	*Would Counsel on Full Range of Methods*
Pooled	0.015**	0.033***	0.148***
SE	(0.008)	(0.011)	(0.039)
q-value	[0.017]	[0.003]	[0.001]
Cntl Mean	0.977	0.938	0.318
N	2568	2568	2568
Tanzania	0.000	0.028*	0.137***
SE	(0.000)	(0.016)	(0.037)
q-value	[.]	[0.040]	[0.001]
Cntl Mean	1.000	0.949	0.096
N	1036	1036	1036
Pakistan	0.132*	0.130*	0.189***
SE	(0.068)	(0.076)	(0.053)
q-value	[0.061]	[0.066]	[0.002]
Cntl Mean	0.789	0.719	0.055
N	280	280	280
Burkina Faso	0.002	0.016**	0.147*
SE	(0.002)	(0.007)	(0.074)
q-value	[0.119]	[0.082]	[0.082]
Cntl Mean	0.998	0.977	0.560
N	1252	1252	1252

Notes: Each estimate is the coefficient on treatment from equation 1 or 2. Clinic-clustered standard errors are in parentheses. Sharpened q-values are in brackets. All providers in Tanzania said they would offer counseling so the treatment effect was not estimable for this outcome. *** p<0.01, ** p<0.05, * p<0.1
Source: Endline discrete choice experiment

Table A7: Effect of treatment on person-centered care outcomes

	Received Services		*Counseled on Full Range*		*Method Information Index*		*Essential Questions Asked*	
	Exit Survey	Mystery Clients	Exit Survey	Mystery Clients	Exit Survey	Mystery Clients	Exit Survey	Mystery Clients
Pooled	0.014*	-0.000	0.093**	0.065*	0.055**	0.052**	0.091**	0.100***
SE	(0.007)	(0.006)	(0.038)	(0.036)	(0.027)	(0.020)	(0.035)	(0.035)
q-value	[0.034]	[0.330]	[0.034]	[0.052]	[0.034]	[0.019]	[0.034]	[0.018]
Cntl Mean	0.978	0.995	0.505	0.632	0.708	0.440	0.673	0.615
N	31603	756	15080	752	31871	752	28363	756
Tanzania	0.037*	0.000	0.111	0.069	0.084*	0.063*	0.113*	0.159***
SE	(0.019)	(0.000)	(0.073)	(0.055)	(0.048)	(0.034)	(0.066)	(0.046)
q-value	[0.138]	[.]	[0.138]	[0.112]	[0.138]	[0.072]	[0.138]	[0.003]
Cntl Mean	0.951	1.000	0.508	0.723	0.744	0.592	0.706	0.723
N	11452	292	6504	292	11157	292	10243	292
Pakistan	-0.001	0.000	0.068	0.228***	0.063	0.138***	0.149	0.154*
SE	(0.004)	(0.000)	(0.072)	(0.084)	(0.062)	(0.051)	(0.111)	(0.083)
q-value	[0.892]	[.]	[0.892]	[0.014]	[0.892]	[0.014]	[0.892]	[0.024]
Cntl Mean	0.994	1.000	0.177	0.309	0.728	0.333	0.632	0.456
N	4051	150	2934	150	4671	150	3807	150
Burkina Faso	0.001	-0.000	0.083*	-0.017	0.034	0.000	0.063	0.020
SE	(0.003)	(0.015)	(0.044)	(0.056)	(0.038)	(0.028)	(0.041)	(0.060)
q-value	[0.555]	[1.000]	[0.328]	[1.000]	[0.365]	[1.000]	[0.328]	[1.000]
Cntl Mean	0.994	0.987	0.677	0.699	0.676	0.346	0.655	0.589
N	16100	314	5642	310	16043	310	14313	314

Notes: Each estimate is the coefficient on treatment from equation 1 or 2. All models are unadjusted. Pooled unadjusted models include only country fixed-effects. Clinic-clustered standard errors are in parentheses. Sharpened q-values are in brackets. All mystery clients in Tanzania and Pakistan received services so the treatment effect was not estimable for this outcome.*** p<0.01, ** p<0.05, * p<0.1
Source: Mystery client data was collected at endline. Client exit survey data was collected continuously and is aggregated over the full 12 month intervention period.

Table A8: Effect of treatment on method dispensing outcomes

	Any Method		*LARC*		*Injectable*		*Method of Choice*	
	Exit Survey	Mystery Clients	Exit Survey	Mystery Clients	Exit Survey	Mystery Clients	Exit Survey	Mystery Clients
Pooled	0.019*	0.060*	0.000	0.077**	0.015	0.067*	0.022*	0.081**
SE	(0.011)	(0.033)	(0.018)	(0.037)	(0.017)	(0.036)	(0.012)	(0.038)
q-value	[0.205]	[0.072]	[0.986]	[0.072]	[0.335]	[0.072]	[0.205]	[0.072]
Cntl Mean	0.907	0.717	0.364	0.463	0.418	0.337	0.885	0.424
N	30886	756	30886	735	30886	726	30766	750
Tanzania	0.047**	0.132***	-0.002	0.143**	0.038	0.166***	0.050**	0.172***
SE	(0.020)	(0.048)	(0.037)	(0.061)	(0.029)	(0.047)	(0.021)	(0.060)
q-value	[0.045]	[0.008]	[0.369]	[0.011]	[0.156]	[0.003]	[0.045]	[0.008]
Cntl Mean	0.911	0.750	0.582	0.642	0.225	0.216	0.888	0.453
N	10934	292	10934	292	10934	292	10915	292
Pakistan	0.004	0.044	0.029	0.051	0.031	0.108	0.015	0.055
SE	(0.052)	(0.051)	(0.068)	(0.043)	(0.064)	(0.074)	(0.062)	(0.069)
q-value	[1.000]	[0.741]	[1.000]	[0.741]	[1.000]	[0.741]	[1.000]	[0.741]
Cntl Mean	0.813	0.882	0.134	0.059	0.431	0.294	0.755	0.176
N	4659	150	4659	150	4659	150	4574	150
Burkina Faso	0.002	0.001	-0.005	0.024	-0.004	-0.057	0.003	0.006
SE	(0.010)	(0.060)	(0.018)	(0.065)	(0.022)	(0.066)	(0.010)	(0.065)
q-value	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
Cntl Mean	0.934	0.608	0.275	0.493	0.550	0.483	0.922	0.519
N	15293	314	15293	293	15293	284	15277	308

Notes: Each estimate is the coefficient on treatment from equation 1 or 2. All models are unadjusted. Pooled models include only country fixed-effects. Clinic-clustered standard errors are in parentheses. Sharpened q-values are in brackets. *** p<0.01, ** p<0.05, * p<0.1

Source: Mystery client data was collected at endline. Client exit survey data was collected continuously and is aggregated over the full 12 month intervention period.

Table A9: Effect of treatment on perceived treatment index components

	Pooled	*Tanzania*			*Pakistan*			*Burkina Faso*		
	(Diff)	(C)	(T)	(Diff)	(C)	(T)	(Diff)	(C)	(T)	(Diff)
Number of Profiles		148	144		68	82		158	156	
Positive statements										
Treated with respect by Provider	0.142***	0.216	0.389	0.173***	0.721	0.854	0.133*	0.615	0.734	0.118
Treated in a friendly manner	0.147***	0.270	0.465	0.195***	0.559	0.732	0.173**	0.436	0.526	0.090
Provider paid attention to you	0.081**	0.223	0.347	0.124***	0.824	0.890	0.067	0.596	0.643	0.047
Felt Provider cares for you as person	0.184***	0.358	0.556	0.197***	0.397	0.732	0.335***	0.660	0.760	0.099
Felt she could trust the Provider	0.173***	0.176	0.354	0.178***	0.588	0.854	0.265***	0.487	0.610	0.123*
Felt safe in the health facility	0.141***	0.378	0.556	0.177***	0.912	0.976	0.064	0.532	0.675	0.143*
Provider clearly explained things	0.192***	0.169	0.319	0.151***	0.353	0.707	0.354***	0.301	0.455	0.153**
Provider talked about how you were feeling	0.016	0.000	0.028	0.028**	0.015	0.024	0.010	0.026	0.032	0.007
Felt listened to by the Provider	0.133***	0.135	0.306	0.170***	0.824	0.915	0.091	0.526	0.643	0.117
Provider considered personal situation methods	0.077***	0.095	0.125	0.030	0.441	0.780	0.339***	0.103	0.097	–0.005
Felt involved by in FP decisions	0.110***	0.101	0.243	0.142***	0.368	0.634	0.266***	0.327	0.331	0.004
Provider allowed you give opinion	0.071**	0.027	0.104	0.077***	0.353	0.585	0.232***	0.141	0.130	–0.011
Provider gave enough information	0.153***	0.135	0.229	0.094**	0.309	0.646	0.338***	0.321	0.442	0.121*
Provider let you say what mattered about method	0.050**	0.027	0.090	0.063**	0.221	0.439	0.218***	0.141	0.097	–0.044
Provider made an effort to ensure privacy	0.119***	0.318	0.486	0.169***	0.059	0.049	–0.010	0.391	0.526	0.135*
Feel personal info will be kept confidential	0.100**	0.365	0.521	0.156***	0.794	0.951	0.157***	0.577	0.597	0.020
Negative statements										
Felt judged by Provider	0.027	0.845	0.938	0.093**	0.838	0.890	0.052	0.910	0.864	–0.047
Felt scolded by Provider	0.022	0.939	0.979	0.040*	0.956	0.976	0.020	0.929	0.935	0.006
Made you uncomfortable because of sex life	0.018	0.784	0.826	0.043	0.941	0.976	0.034	0.962	0.948	–0.013
Provider pressured you to use their preferred method	0.043	0.493	0.618	0.125**	0.647	0.634	–0.013	0.904	0.896	–0.008
Could be seen by others	0.036	0.014	0.000	–0.014	0.456	0.537	0.081	0.750	0.812	0.062
Could be heard by others	0.068**	0.831	0.868	0.037	0.368	0.549	0.181*	0.737	0.779	0.042

Notes: Control and treatment group means and significance estimated using linear regression with standard errors clustered at the facility. *** p<0.01, ** p<0.05, * p<0.1
Source: Endline mystery client data.

Table A10: Effect of treatment on perceived treatment outcomes

	Perceived Treatment Index		*Would Recommend*		*Not Judged or Scolded*		*Did Not Feel Uncomfortable b/c of Sex Life*	
	Exit Survey	Mystery Clients	Exit Survey	Mystery Clients	Exit Survey	Mystery Clients	Exit Survey	Mystery Clients
Pooled	0.111	0.215***	0.019	0.071***	0.002	0.044	0.013	0.018
SE	(0.094)	(0.051)	(0.015)	(0.024)	(0.013)	(0.027)	(0.032)	(0.020)
q-value	[0.912]	[0.001]	[0.912]	[0.005]	[0.912]	[0.074]	[0.912]	[0.158]
Cntl Mean	-0.002	0.001	0.846	0.663	0.969	0.841	0.854	0.889
N	32303	752	32255	752	32282	752	32267	752
Tanzania	0.280**	0.298***	0.041	0.072***	0.000	0.099**	0.044	0.043
SE	(0.140)	(0.081)	(0.025)	(0.026)	(0.003)	(0.042)	(0.085)	(0.044)
q-value	[0.250]	[0.002]	[0.250]	[0.010]	[0.844]	[0.015]	[0.683]	[0.091]
Cntl Mean	0.006	-0.000	0.816	0.639	0.989	0.818	0.650	0.784
N	11477	292	11469	292	11476	292	11472	292
Pakistan	0.059	0.334***	0.017	0.210***	0.074	0.069	0.018	0.034
SE	(0.234)	(0.087)	(0.030)	(0.055)	(0.095)	(0.058)	(0.012)	(0.033)
q-value	[1.000]	[0.001]	[1.000]	[0.001]	[1.000]	[0.174]	[0.865]	[0.174]
Cntl Mean	0.131	-0.000	0.895	0.643	0.858	0.809	0.966	0.941
N	4695	150	4675	150	4680	150	4682	150
Burkina Faso	0.004	0.080	0.004	0.002	-0.012	-0.021	-0.009	-0.013
SE	(0.139)	(0.086)	(0.022)	(0.044)	(0.012)	(0.042)	(0.013)	(0.022)
q-value	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
Cntl Mean	-0.045	-0.000	0.852	0.691	0.980	0.878	0.966	0.962
N	16131	310	16111	310	16126	310	16113	310

Notes: Each estimate is the coefficient on treatment from equation 1 or 2. All models are unadjusted. Pooled unadjusted models include only country fixed-effects. Clinic-clustered standard errors are in parentheses. Sharpened q-values are in brackets. *** p<0.01, ** p<0.05, * p<0.1
Source: Mystery client data was collected at endline. Client exit survey data was collected continuously and is aggregated over the full 12 month intervention period.

Table A11: Effect of treatment on family planning client volume

	Pooled		*Tanzania*			*Pakistan*			*Burkina Faso*	
	(Diff)	(C)	(T)	(Diff)	(C)	(T)	(Diff)	(C)	(T)	(Diff)
Number of Clinics		37	36		37	38		39	39	
Total monthly family planning clients	6.872	226.742	226.346	–0.395	18.885	23.204	4.319**	95.079	108.335	13.256
Total monthly new users	–4.834	161.704	148.423	–13.281	8.604	9.382	0.778	35.577	34.577	–1.000
Number of clients by age (monthly average)										
Age <20	1.614	23.693	22.988	–0.704	2.665	4.856	2.192***	14.923	18.288	3.365
Age 20–24	0.619	65.886	63.870	–2.016	4.867	5.756	0.889	28.639	31.528	2.889
Age 25+	–1.193	71.298	59.365	–11.933	11.353	12.592	1.238	51.068	58.160	7.092
Number of methods by type (monthly average)										
LARC	–3.314	140.560	124.997	–15.564	5.746	6.730	0.984	36.626	37.835	1.209
Injectable	7.806*	43.805	57.260	13.456	7.508	8.385	0.878	39.201	46.744	7.543
Pill	2.245	26.507	28.677	2.170	5.495	7.997	2.502***	14.767	16.897	2.130
Other	0.611	17.003	15.549	–1.454	0.136	0.092	–0.044	4.485	6.859	2.374

Notes: Family planning clients in Burkina Faso do not include those renewing the same method. Other methods in Tanzania include female and male condoms and female and male sterilization. In Burkina Faso, other methods include female and male condoms, natural methods, and other methods. Data on returning users of LARC methods and other methods was not collected until 2021 in Tanzania and therefore the number of LARC and pill users in Tanzania only include January 2021 to August 2021. Age of family planning users was only available for new users so in Tanzania, disaggregation by age only includes new users while in Pakistan and Burkina Faso, both new and returning users are captured. The total number of family planning clients in Tanzania is included from January 2021 to August 2021. Data from July and August 2021 in Pakistan are excluded because a large proportion of facilities were missing data. *** $p<0.01$, ** $p<0.05$, * $p<0.1$
Source: Administrative data aggregated during months when intervention was active. In Pakistan this excludes the last two months of intervention because data was not available.

Table A12: Effect of treatment on disparities by age, marital status, and parity (exit survey)

	Counseling on Full Range				*Received Any Method*				*Perceived Treatment Index*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
VARIABLES	Pooled	TZ	PK	BF	Pooled	TZ	PK	BF	Pooled	TZ	PK	BF
Treatment	0.053**	0.083*	0.020	0.025	0.020	0.047**	0.003	0.005	0.041	0.063	0.149	0.010
	(0.027)	(0.050)	(0.037)	(0.040)	(0.013)	(0.018)	(0.059)	(0.009)	(0.052)	(0.069)	(0.116)	(0.078)
Young	0.041***	0.055***	0.024***	0.022**	0.011***	0.016***	0.013	0.004	0.002	0.010	0.054**	-0.012
	(0.007)	(0.011)	(0.009)	(0.010)	(0.003)	(0.004)	(0.014)	(0.004)	(0.007)	(0.012)	(0.025)	(0.009)
Unmarried	0.028	0.027	0.093	0.052*	0.002	0.002	-0.197	0.015**	-0.069**	-0.084**	-0.174	-0.035
	(0.023)	(0.032)	(0.109)	(0.026)	(0.011)	(0.013)	(0.127)	(0.006)	(0.032)	(0.035)	(0.234)	(0.051)
Nulliparous	0.038	0.085**	-0.081**	-0.002	-0.018	-0.022	-0.169	-0.021**	0.012	0.027	0.035	-0.024
	(0.025)	(0.040)	(0.032)	(0.024)	(0.012)	(0.019)	(0.124)	(0.009)	(0.032)	(0.034)	(0.105)	(0.035)
Treatment X Young	0.005	-0.004	0.013	0.022	0.001	-0.002	0.013	-0.002	-0.001	0.021	-0.066*	-0.007
	(0.010)	(0.015)	(0.019)	(0.014)	(0.004)	(0.005)	(0.020)	(0.004)	(0.012)	(0.019)	(0.036)	(0.014)
Treatment X Unmarried	0.036	0.033	-0.258**	0.011	0.011	-0.002	0.225	0.002	0.057	0.059	0.280	0.028
	(0.033)	(0.042)	(0.126)	(0.033)	(0.015)	(0.015)	(0.171)	(0.008)	(0.047)	(0.050)	(0.308)	(0.061)
Treatment X Nulliparous	0.009	0.010	0.210***	-0.001	-0.021	0.044**	-0.019	-0.023	-0.012	0.078*	-0.010	-0.008
	(0.034)	(0.056)	(0.051)	(0.033)	(0.019)	(0.022)	(0.145)	(0.019)	(0.047)	(0.046)	(0.154)	(0.048)
N	48,016	23,241	10,456	14,319	106,420	41,926	15,978	48,516	77,920	26,894	9,848	41,178
Controls	No	No	No	No	No	No	No	No	No	No	No	No
Country FE	Yes	No	No	No	Yes	No	No	No	Yes	No	No	No

Notes: Estimates are from equation 5. *** p<0.01, ** p<0.05, * p<0.1.
Source: Client exit survey data was collected continuously and is aggregated over the full 12 month intervention period.

Table A13: Variables used to calculated rewards scores

Safe welcoming space
% clients to whom service was not refused (unless for medical reasons)
% clients who felt that staff did not treat them with disrespect
% clients who felt that staff did not treat them in an unfriendly manner
% clients who felt that staff showed that they cared about them as a person
% clients who felt that staff paid attention to them
% clients who felt they could completely trust the FP provider
Sensitive communication
% clients who were invited to ask questions
% clients whose questions were answered to their satisfaction
% clients who felt they could ask staff any question
Seek understanding and agreement
% clients not asked about permission
% clients felt included in the decision
% clients asked about family planning preferences
Security of information
% of clients who were given privacy
% of clients who felt provider made an effort to give them privacy
Say yes to a safe method
% clients who were discouraged from using a method
% clients who receive the method of their choice
Simple, comprehensive counseling
% clients with whom two modern methods are discussed
% clients informed about at least two methods

Notes: Each of the six domains indicated in bold were developed by Pathfinder and represent the six principles of unbiased care. Each item represents the variables that went into the calculation of the clinic's score which was used to identify the clinics that received a non-financial incentive. The domain specific scores were calculated by taking the average for each domain and the total score was calculated by taking the average across all domains.
Source: Pathfinder International

Table A14: Participation in Intervention

	Tanzania	Pakistan	Burkina Faso
Summit			
Share of providers attended	0.983	N/A	0.975
Share of facilities with full participation	0.944	0.902	0.795
Connect			
Average in-person participation	0.783	N/A	0.798
Low participation	0.361	0.390	0.282
Medium participation	0.306	0.317	0.385
High participation	0.333	0.293	0.333
Non-Financial Awards			
Attended 0-1 ceremonies	0.00	0.146	0.00
Attended 2 ceremonies	0.00	0.415	0.077
Attended 3 ceremonies	1.00	0.439	0.923

Notes: Connect participation categories in Tanzania and Burkina Faso using the average number of providers that attended in-person Connect sessions quarterly as a percentage of the total number of FP providers employed at the facility. Low represents $\leq$70% of providers participating, medium participation represents 71-84% of providers participating, and high represents $\geq$80% of providers participating. In Pakistan, low participation represents 2+ quarters of passive activity on WhatsApp Connect. Medium participation represents 2+ quarters of inconsistent activity (i.e., providers participating every other week or twice a month). High participation represents 2+ quarters of consistent activity (i.e., participating in Connect activities every week).
Source: Pathfinder International program monitoring data

Table A15: Sensitivity analyses to non-participation in Pakistan

	Unbiased Index (Provider Survey)		*Counselled on Full Range* (Mystery Clients)		*Given Any Method* (Exit Survey)		*Perceived Treatment Index* (Mystery Clients)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Pakistan	0.159**	0.120	0.228***	0.266**	0.004	0.119	0.334***	0.423***
SE	(0.074)	(0.109)	(0.084)	(0.113)	(0.052)	(0.075)	(0.087)	(0.125)
q-value	[0.025]	[0.164]	[0.013]	[0.038]	[0.305]	[0.087]	[0.002]	[0.007]
Cntl Mean	-0.000	0.056	0.309	0.325	0.813	0.763	-0.000	-0.097
N	70	38	150	84	4659	2705	150	84
Pooled	0.220***	0.227***	0.065*	0.025	0.019*	0.021**	0.215***	0.186***
SE	(0.038)	(0.041)	(0.036)	(0.040)	(0.011)	(0.010)	(0.051)	(0.060)
q-value	[0.001]	[0.001]	[0.045]	[0.155]	[0.045]	[0.029]	[0.001]	[0.004]
Cntl Mean	-0.000	0.000	0.632	0.710	0.907	0.925	0.001	-0.000
N	642	572	752	602	30886	26227	752	602

Notes: Each estimate is the coefficient on treatment from equation 1 (Pakistan only) or 2 (Countries pooled excluding Pakistan). Odd columns are identical to table 4 and include all clinics. Even columns in Pakistan exclude treatment providers who were added after randomization and control providers who said they would not participate in a Summit. Even columns in the pooled analyses exclude Pakistan. Clinic-clustered standard errors are in parentheses. Sharpened q-values are in brackets. The data source is in parentheses at the top of each column.
*** p<0.01, ** p<0.05, * p<0.1
Source: Mystery client data and provider surveys were collected at endline. Client exit survey data was collected continuously and is aggregated over the full 12 month intervention period.

Table A16: Effect of treatment on client exit survey outcomes including clients of all ages

	Person-Centered FP Care Outcomes				*Method Dispensing Outcomes*			
	Received Services	Full Range	EQI	MII	Any Method	LARC	Injectable	Method of choice
Pooled	0.015	0.082**	0.056*	0.080**	0.020*	-0.002	0.004	0.024*
SE	(0.009)	(0.037)	(0.029)	(0.036)	(0.012)	(0.015)	(0.018)	(0.013)
Cntl Mean	0.976	0.460	0.674	0.650	0.899	0.307	0.458	0.876
N	76246	26609	76618	67185	73784	73784	73784	73548
Tanzania	0.040*	0.118*	0.082	0.112	0.053**	-0.011	0.046	0.055**
SE	(0.024)	(0.071)	(0.050)	(0.071)	(0.024)	(0.034)	(0.032)	(0.025)
Cntl Mean	0.944	0.446	0.698	0.658	0.895	0.509	0.264	0.870
N	26859	11444	25910	23197	25159	25159	25159	25122
Pakistan	-0.000	0.040	0.122**	0.121	-0.018	0.054	-0.047	0.001
SE	(0.004)	(0.063)	(0.060)	(0.082)	(0.061)	(0.045)	(0.075)	(0.068)
Cntl Mean	0.994	0.177	0.655	0.687	0.803	0.108	0.505	0.736
N	8272	5516	9749	7706	9731	9731	9731	9568
Burkina Faso	0.000	0.058	0.027	0.052	0.006	-0.008	-0.013	0.008
SE	(0.003)	(0.043)	(0.040)	(0.044)	(0.008)	(0.015)	(0.020)	(0.009)
Cntl Mean	0.995	0.644	0.660	0.635	0.929	0.222	0.576	0.915
N	41115	9649	40959	36282	38894	38894	38894	38858

Notes: Each estimate is the coefficient on treatment from equation 1 or 2. All models are unadjusted. Pooled unadjusted models include only country fixed-effects. Clinic-clustered standard errors are in parentheses. Full range means "counselled on full range of methods" and was only assessed for new FP clients or clients switching a method. *** p<0.01, ** p<0.05, * p<0.1

Source: Client exit survey data was collected continuously and is aggregated over the full 12 month intervention period.