

**DELIVERING REPRODUCTIVE HEALTH
RESULTS (DRHR) THROUGH NON-STATE
PROVIDERS IN PAKISTAN
MSI/PSI Impact Evaluation Report**

Sophie Witter, Rashid Zaman, Molly Scott and Rozina Mistry

March 2016

Preface

This work was carried out as a part of the evaluation of the UK Department for International Development-funded programme titled ‘Delivering Reproductive Health Results (DRHR) through Non-State Providers’.

The team leader of this evaluation was Professor Sophie Witter and the project manager was Dr Rashid Zaman. The remaining key team members of this evaluation include Dr Michele Binci, Dr Rozina Mistry, Dr Adrian Gheorghe, Shafique Arif, Molly Scott, Martina Garcia, Iftikhar Cheema and Abdur Rashid Bhatti. Megan Titchell was the project administrator. Oxford Policy Management’s Pakistan office undertook the primary data collection activities for this evaluation.

We are grateful to Sherwan Asif, Jane Rashad, Sajil Liaqat and Laura Norris for important suggestions and feedback on this analysis. We would also like to thank all the key stakeholders at Marie Stopes International (MSI), Population Services International (PSI) and DKT International for providing important information on and insights into the implementation of the programme. We are also thankful to all the respondents of the primary data collection activities in the households, health facilities and outlets in Punjab, Sindh and Khyber Pakhtunkhwa.

This project was carried out by HEART (Health & Education Advice & Resource Team). Please contact the project manager Dr Rashid Zaman for further details at rashid.zaman@opml.co.uk. The contact point for the client is Sherwan Asif (S-Asif@dfid.gov.uk). The client reference number for the project is A1054.

Disclaimer

The Health & Education Advice & Resource Team (HEART) provides technical assistance and knowledge services to the British Government’s Department for International Development (DFID) and its partners in support of pro-poor programmes in education, health and nutrition. The HEART services are provided by a consortium of leading organisations in international development, health and education: Oxford Policy Management, CfBT, FHI360, HERA, the Institute of Development Studies, IPACT, the Liverpool School of Tropical Medicine and the Nuffield Centre for International Health and Development at the University of Leeds. HEART cannot be held responsible for errors or any consequences arising from the use of information contained in this report. Any views and opinions expressed do not necessarily reflect those of DFID, HEART or any other contributing organisation.

HEART

Level 3, Clarendon House
52 Cornmarket Street
Oxford, OX1 3HJ
United Kingdom

Tel +44 (0) 1865 207 333
Email info@heart-resources.org
consultancy@heart-resources.org
Web www.heart-resources.org

Executive summary

Background on the Delivering Reproductive Health Results programme

The UK Department for International Development (DFID) Delivering Reproductive Health Results (DRHR) through Non-State Providers programme has supported Marie Stopes International (MSI) and Population Services International (PSI) over four years (2012–2016) in relation to expanding the delivery and utilisation of high-quality reproductive health (RH) services and commodities in under-served urban and rural areas in southern Punjab, northern Sindh, Khyber Pakhtunkhwa (KP) and the Federally Administered Tribal Areas (FATA). MSI works with Marie Stopes Society Pakistan (MSS) and uses a social franchising (SF) approach. PSI works with its local counterpart Greenstar Social Marketing (GSM) and uses a social marketing (SM) approach.

Study objectives and design

This report presents the findings of an evaluation of the DRHR programme's impact on access to and utilisation of RH services and commodities, equity and quality of care. A statistically representative 7,888 households were surveyed in 400 clusters in mid-2013 at baseline and the panel followed up at endline (late 2015). 6,336 households and 5,514 women were successfully followed up at endline. Outcomes were compared between married women of reproductive age (MWRAs) exposed to DRHR programme activities (PSI only and a combined MSI and PSI group) and those who were not (control). Propensity score matching was used to eliminate unobserved biases in selection into these groups and to provide robust quasi-experimental estimates of effect.

In addition to the household survey, facility and RH outlet provider surveys were conducted in treatment and control clusters at baseline and endline to assess changes to the quality and price of RH services and products (219 facilities at baseline and 188 at endline; 735 outlets at baseline and 620 at endline).

Preliminary findings were presented to and discussed with DFID and implementing partners in Pakistan in February 2016. The impact evaluation is accompanied by a report on the value for money of the DRHR programme (Oxford Policy Management (OPM), 2016).

Key results

Population characteristics

The demographic characteristics of the sample at baseline are described in the baseline report of 2013, which reveals a youthful population, with high rates of fertility, relatively low levels of education and an incidence of poverty that is higher than the national average, according to the National Income Support Programme (NISP) poverty scorecard. The samples are quite balanced according to age, with the PSI-only group containing a slightly younger sample of MWRA than the control group. The figures indicate a modest degree of targeting toward households falling into the poorest poverty quintile, especially so for the PSI-only treatment group.

At endline, the PSI-only group had significantly larger household sizes (with a mean of eight persons), compared to the two other groups, as well as more MWRAs and a lower mean poverty score, while the PSI+MSI group had a significantly higher proportion of female-headed households (over 9%). In relation to MWRA, a significantly lower proportion were literate in the PSI-only group (at 24%, this was less than half that of the control, which had 50% literacy in MWRA), which is matched by significantly fewer years of school in the PSI-only group (fewer than two years on

average). Correspondingly, the PSI-only group reported significantly earlier age at first marriage and age at first pregnancy, compared to the two other groups. These differences of characteristics are allowed for in the matched impact analyses.

Access to RH services

There has been little change descriptively in the time and cost required to reach the nearest RH provider across the three arms. In terms of changes to knowledge of the nearest RH provider, which may be one barrier to access, across all groups we found a decrease in the mean number of RH providers known to respondents in the community, driven by a decrease in awareness of government providers. This was accompanied by a small and marginally significant increase in the percentage of MWRAs who are aware that their nearest RH provider is non-state.

Utilisation of RH services

The proportion of MWRAs who had ever used contraception and any modern method showed significant increases in all groups. As this is an indicator that increases over time, this was anticipated. In relation to current use and unmet need, there was no significant change, with rates remaining high across all three groups (38% for the PSI-only group at endline, compared to 28% for PSI+MSI and control).

We also investigated knowledge of RH methods, as this is one important factor influencing uptake. Knowledge of any method was almost universal and has not changed. Awareness of modern methods appears to be increasing, and possibly displacing knowledge of traditional methods. All groups showed improved overall understanding of the efficacy and preventive effects of contraception, although in the PSI+MSI group a small but significant reduction was found in those knowing that contraception prevented unwanted pregnancies.

Another important driver is, of course, the coverage of the programme. The proportion of MWRAs who reported visiting an RH provider in the last three months dropped significantly for the PSI+MSI group (from 12% to 9%), with no significant change for the two other groups. However, uptake of subsidised family planning (FP) did increase for all three groups from low starting values, and significantly by 12% points for the PSI group (albeit from a very low starting value of 3%). In relation to exposure to behaviour change communication (BCC) activities, there were significant increases reported across all groups, but especially for the PSI+MSI group and the control.

In relation to household expenditure on contraception, there was a significant reduction in the proportion paying for their last method in the control group, but a significant increase in the PSI-only group. Costs increased hugely across all groups. However, sample sizes for this set of indicators are relatively small, so these figures should be treated with some caution.

Indicators of service acceptability were also examined. Many women prefer to see a female provider and the gap between these preferences and their actual experience is one indicator of the extent to which services are acceptable to MWRAs. For the PSI-only group, the gap narrowed over the project period, though largely through a reduction in preferences for a female provider. For the control group, the gap widened, largely through a decrease in those who saw a female provider at their last visit. Again, sample sizes for this class of indicators are too small to make conclusive statements about what has changed.

Among MWRAs not using contraception, the reasons for non-use were found to be relatively stable over time. There were significant increases in women reporting that they cannot fall pregnant or are menopausal, and a decrease in women reporting that they want more children. These effects may be due to the time interval between baseline and endline.

Equity

Disaggregating changes to utilisation by poor/non-poor households, a pro-poor pattern of change was found for the PSI-only group in relation to ever using contraception (perhaps partly linked to lower starting levels for the poorer households – 26%, as against 45% for non-poor). However, the control group also showed a pro-poor increase in ever using contraception (i.e. a bigger increase in reportedly ever using contraception for poor than for non-poor households), even though the starting level here was higher (56%, as against 61% for the non-poor).

It appears that increases in ever using modern contraceptive methods over time accrue to non-poor households faster than they do to poor households. Amongst non-poor households, the increase in ever using modern methods was large and always significant, while the corresponding point estimate for poor households was smaller and only significant in the PSI-only treatment group. Conversely, poor households appear to be more inclined to use traditional methods. In the PSI-only group, the increase in ever using traditional methods over time was 14.4% for poor households, compared with 9.5% for non-poor households.

We also investigated whether changing underlying awareness was equitable. Gains to awareness of modern methods accrue quite evenly to poor and non-poor MWRA in the PSI+MSI group. However, in the control group the increases are seen in the non-poor group only, which may indicate some success on the part of PSI and MSI in combination in regard to raising awareness among the poor.

In relation to access, we found a substantial and significant reduction in mean spend in transport to reach the nearest RH provider for the poor in the MSI+PSI group, and for the non-poor in the control. There was also a significant reduction in time taken to reach this provider in the PSI-only group for the non-poor.

Quality of RH services

Comparing the functionality of basic infrastructure in facilities in the combined intervention group with those in the control, most facilities scored highly for power, running water and latrines. There was a significant improvement in functioning fridges in the control group. For the PSI+MSI areas, only 60% were found with functioning fridges. Both groups saw deteriorating functioning of telephones.

There was a significant reduction in the proportion of facilities displaying standards of practice between baseline and endline for the control group, while this increased significantly in the intervention. However, intervention groups showed a significant reduction in the proportion cleaned daily with a disinfectant. There were significant reductions in stock outs for all contraceptive types in control facilities and for condoms and injectables (but not oral contraceptives) in intervention area facilities. There were significant improvements in all types of staffing across the intervention facilities.

In relation to fees charged by facilities, sample sizes restrict our analysis, but there does appear to be a significant reduction in general FP consultation costs charged by intervention area facilities.

For the retail outlets, those in PSI-only areas showed a small but significant reduction in the range of FP products supplied over the period, including for injectables. In PSI+MSI areas, there was no significant overall change but there was an increase in oral contraceptives supplied (from 29% to 39%) and a small decrease in condoms. In control areas, there were significant decreases overall

and for condoms. Prices at outlets varied in their levels and trends, but the most striking change was an increase across all groups for injectables.

From the client perspective, no significant change was registered in waiting times for MWRAs in the intervention arms, while a significant increase did occur in the control areas (though the level remained lower than in PSI only sites, for example). There was no significant change in any arm for the proportion told about possible side effects of contraceptives, which remains relatively low (27% PSI-only, 38% MSI+PSI, 34% control). There was a substantial increase in those told about alternative methods in MSI+PSI areas (from 27% to 46%), which corresponded to a similar increase in the proportion of MWRAs rating the service as good in these areas. There was also a significant but less large increase in 'good' ratings in PSI-only areas. Control areas saw significant increases in 'good' ratings and highly significant increases in 'excellent'.

Impact

As a result of the largely convergent trends noted in the descriptive analysis above, the impact estimates, which used propensity score matching to remove hidden selection bias, found no significant impact for the two intervention groups, compared to the control, for contraceptive access and uptake.

There was some evidence of increased exposure to communication materials in both intervention groups. Impact estimates also found a significant increase in the mean number of modern methods named by MWRAs in the PSI+MSI group, compared to the control.

Conclusions

The overall finding of a lack of significant effect of the interventions on access to and utilisation of contraception is disappointing. There is evidence that the intervention has improved awareness, compared to control areas, and in the descriptive analysis we see a relatively higher performance in some of the domains that should have a positive effect on utilisation. However, on most indicators, performance (change since baseline) was either comparable to or worse than in the control areas. Current use of FP across all groups and unmet need were relatively constant, as was access. Various possible explanations for this are discussed.

Table of contents

Preface	i
Executive summary	2
Background on the Delivering Reproductive Health Results programme	2
Study objectives and design	2
Key results	2
Population characteristics	2
Access to RH services	3
Utilisation of RH services	3
Equity	4
Quality of RH services	4
Impact	5
Conclusions	5
List of tables and figures	8
List of abbreviations	10
1 Introduction	1
1.1 Reproductive health context in Pakistan	1
1.2 Background on family planning service provision context in the study sites	1
1.3 The DRHR programme	3
1.3.1 SF model	3
1.3.2 SM model	4
1.4 DRHR programme impact evaluation	4
2 Impact evaluation methodology	5
2.1 What the evaluation is measuring	5
2.2 How the impact was measured	8
2.3 Survey instruments	9
2.3.1 Household panel survey	10
2.3.2 Facility and retailer questionnaires	10
2.4 Sampling strategy	10
2.5 Sample sizes	11
2.5.1 The location of the sample	13
2.5.2 The facility and outlet samples	14
2.6 Survey implementation	15
2.7 Data management and analysis	15
2.8 Ethical consideration	16
2.9 Limitations of the analysis	16
2.9.1 Attrition	16
2.9.2 Contamination	16
2.10 How to read tables in this report	17
3 Demographic characteristics	18
3.1 Age distribution of sampled MWRAAs	18
3.2 Poverty index quintiles	19
3.3 Characteristics of households and MWRA respondents at endline	20
4 Access to RH services	22

5	Utilisation of RH services	27
5.1	Current use of contraceptive methods	27
5.2	Knowledge	32
5.3	Coverage	37
5.4	Household costs	40
5.5	Acceptability and reduction of barriers	43
6	Equity	46
7	Quality of RH provision	58
7.1	Health facilities	58
7.1.1	Infrastructure	58
7.1.2	Standards, hygiene and stock outs	61
7.1.3	Fees	65
7.2	Outlets	67
7.2.1	Range of FP products	67
7.2.2	Costs of FP products	69
7.3	Quality of service at the source of last contraceptives	71
8	Impact estimates	74
9	Discussion and conclusion	79
Annex A	Econometric methods	81
A.1	Selecting the PSM model	81
A.2	Additional diagnostic test results	81
Annex B	Robustness check	89
Annex C	Sensitivity analysis	93
C.1	Purpose of the sensitivity analysis	93
C.2	Sensitivity analysis considering attrition at the woman level	93
C.3	Sensitivity analysis considering attrition at the woman level	95

List of tables and figures

Table 1:	Key impact indicators.....	7
Table 2:	Overall sample sizes by survey round and treatment group.....	12
Table 3:	Sample size used for analysis of the household data.....	13
Table 4:	Types of sampled health facilities	14
Table 5:	Types of sampled outlets.....	15
Table 6:	Household characteristics at endline	21
Table 7:	Characteristics of MWRA at endline	21
Table 8:	Average travel time and cost to reach nearest RH provider	23
Table 9:	Access to RH services in the community and nearby	25
Table 10:	Ever used contraceptives.....	28
Table 11:	Currently using contraceptives.....	30
Table 12:	Trends in knowledge of MWRAs regarding contraceptive methods.....	33
Table 13:	Trends in knowledge of MWRAs regarding efficacy and preventative effects of contraceptive methods.....	35
Table 14:	Coverage of FP and BCC services	38
Table 15:	Household FP costs.....	42
Table 16:	Acceptability	44
Table 17:	Descriptive evidence of trends in equity of utilisation among treatment groups.....	49
Table 18:	Descriptive evidence of trends in equity of utilisation among the control group.....	50
Table 19:	Descriptive evidence of trends in awareness of contraceptive methods, by poverty status, among the PSI-only and MSI+PSI groups	53
Table 20:	Descriptive evidence of trends in awareness of contraceptive methods, by poverty status, among the control group.....	54
Table 21:	Average travel time and cost to reach nearest RH provider, by poverty status.....	56
Table 22:	Health facility infrastructure.....	60
Table 23:	Facility standards of care and stock out	63
Table 24:	Health facility staffing.....	64
Table 25:	Health facility fees.....	66
Table 26:	FP products sold by outlets.....	68
Table 27:	Prices of FP products.....	70
Table 28:	Quality of service at the source of last contraceptives obtained	72
Table 29:	Impact estimates for contraceptive access.....	74
Table 30:	Impact estimates for contraceptive utilisation.....	75
Table 31:	Impact estimates for awareness of communication campaigns on FP	77
Table 32:	Impact estimates for contraception awareness	77
Table 33:	Main matching model specification and results: PSI-only vs control.....	83
Table 34:	Main matching model specification and results: PSI+MSI vs control	86
Table 35:	Nearest neighbour matching model specification and results: PSI-only vs control ..	90
Table 36:	Nearest neighbour matching model specification and results: PSI+MSI vs control.	91
Table 37:	Sample sizes and attrition.....	93
Table 38:	Sensitivity analysis: Differences at endline in the households where a replacement woman was interviewed compared with the households where the same woman was re-interviewed	94
Table 39:	Sensitivity analysis: Differences at endline in women that were interviewed for the first time at endline compared with women who were re-interviewed.	95
Table 40:	Sensitivity analysis: Differences at baseline between households that were successfully re-interviewed at endline and households that were not re-interviewed.....	96
Table 41:	Sensitivity analysis: Differences at baseline between women from households that were re-interviewed at endline and women from households that were not interviewed at endline	97
Figure 1:	Sampled districts according to treatment status	14
Figure 2:	Districts where DKT International operates	17
Figure 3:	Ages of sampled MWRA in PSI-only and control groups.....	18

Figure 4:	Ages of sampled women in PSI+MSI, and control groups.....	19
Figure 5:	Poverty quintiles	20
Figure 6:	Covariate balance for the main model before and after matching	78
Figure 7:	Covariate balance for before and after matching.....	85
Figure 8:	The distribution of propensity scores in the matched and unmatched sample: PSI- only vs control	85
Figure 9:	Covariate balance for the main model before and after matching: PSI+MSI vs control	88
Figure 10:	The distribution of propensity scores in the matched and unmatched sample: PSI+MSI vs control	88

List of abbreviations

ATT	Average treatment effect on the treated
BCC	Behaviour change communication
BHU	Basic Health Unit
CAMP	Community Appraisal and Motivation Program
CAPI	Computer-assisted personal interviewing
CBOs	Community-based organisations
DFID	Department for International Development (UK)
DHS	Demographic and Health Survey
DRHR	Delivering Reproductive Health Results
ERC	Ethical Review Committee
FATA	Federally Administered Tribal Areas
FP	Family planning
FWC	Family Welfare Centre
FWM	Field worker marketing agent
GSM	Greenstar Social Marketing
HH	Household
KP	Khyber Pakhtunkhwa
IEC	Information, Education and Communication
INGO	International NGO
IUD	Intrauterine device
LHW	Lady health worker
MCH	Maternal and child health
MDGs	Millennium Development Goals
MSI	Marie Stopes International
MSS	Marie Stopes Society Pakistan
MWRA	Married woman of reproductive age
NWRAs	Married women of reproductive age
NATPOW	The National Trust for Population Welfare

NGO	Non-governmental organisation
NISP	National Income Support Programme
OPM	Oxford Policy Management
PAP	Population Association of Pakistan
PAVHNA	Pakistan Voluntary Health & Nutrition Association
PKR	Pakistani rupee
PSI	Population Services International
PSM	Propensity score matching
PLSM	Pakistan Living Standards Measurement
PMT	Proxy Means Test
RH	Reproductive health
RHC	Reproductive Health Centre
RHS	Reproductive Health Services
RHPP	Reproductive health private provider
SF	Social franchising
SM	Social marketing
UNICEF	United Nations Children's Fund
UNFPA	United Nations Population Fund
YAN	Youth Advocacy Network

1 Introduction

1.1 Reproductive health context in Pakistan

High mortality rates and poor health outcomes for women, children and girls are the major public health concerns in Pakistan. In 2013, Pakistan's rates of maternal, neonatal and child mortality were ranked as the third highest in the world, with a maternal mortality ratio of 260 deaths out of every 100,000 births and an under-5 mortality rate of 86 deaths per 1000 live births in 2012¹. The challenges in improving this picture have been compounded by intermittent episodes of political instability, climate change and a sustained period of rapid population growth, which has strained service provision in some parts of the country. As a result, Pakistan's progress against the Millennium Development Goals (MDGs) 4 and 5, which address women's and child health, has been slow and it is currently behind the target.

The burden of mortality and poor health outcomes is highly variable across Pakistan, and falls disproportionately on the poorest areas of the country. Further support is urgently required to improve the situation of women and children from these vulnerable communities and to help them lead healthy and productive lives. High rates of maternal mortality, and the barriers that prevent vulnerable or isolated women accessing quality health care, are also symptomatic of low levels of female empowerment. Part of the rationale behind DFID's Delivering Reproductive Health Results (DRHR) through Non-State Providers programme is that enabling women to take greater control over their fertility can be an important step towards redressing this gender imbalance, in addition to improving the reproductive health (RH) of women.

Low contraceptive prevalence rates in parts of Pakistan may be attributed to a range of factors. Limited access to modern contraceptives is certainly a part of the problem. Oxford Policy Management's (OPM's) DRHR programme baseline report in 2013 indicated that over a quarter of married women of reproductive age (MWRAs) in the sample could not identify their nearest source of RH services, and that 'stock outs' of contraceptives were common in the health facilities sampled (mostly small dispensaries, health posts and pharmacies). However, the reason why many couples do not use contraceptive methods is more nuanced than simply low access. The report found that women were much more likely to identify barriers such as health concerns, cultural or religious restrictions, or their partner being opposed to contraceptives use. This points to a deeper issue of low demand for contraceptives, which suggests the need for behaviour change communication (BCC) interventions alongside efforts to improve the supply of RH services.

1.2 Background on family planning service provision context in the study sites

Pakistan has experienced a high rate of population growth since its inception. In 2015 Pakistan was ranked sixth most populous country of the world². The country initiated interventions around population welfare efforts in the 1950s under the rubric of the Family Planning Association of Pakistan, a national level non-governmental organisation (NGO). In 1990 the Ministry of Population Welfare was created to look after population welfare affairs. Village-based family planning (FP) workers were inducted to take FP services to the doorsteps of people in urban and rural areas. In 2010, as a consequence of the 18th amendment in the constitution, the Ministry of Population

¹ UN Children's Fund (UNICEF) 2013 country data, available here: http://www.unicef.org/infobycountry/pakistan_pakistan_statistics.html.

² World's 50 Most Populous Countries: 2015. Downloaded from <http://www.infoplease.com/world/statistics/most-populous-countries.html>.

Welfare ceased to exist and its functions were devolved to the provinces under Population Welfare Departments.

Under the rubric of the Population Welfare Department, in Sindh, there are 628 Family Welfare Centres (FWCs) - the pillars of Population Welfare Program³. They serve as static facilities to about 7,000 people, whereas operating through its satellite clinics and outreach facilities it covers a population of around 12,000. The scope of work of the FWCs includes the provision of FP and maternal and child health (MCH) services and treatment of minor ailments. In addition, there is a cadre of male village-based workers, introduced for deployment at Union Councils. Through this intervention, the programme is able to reach all the Union Councils of Sindh. Two social mobilisers are also present (male/female) in each FWC, to undertake social mobilisation around FP and RH. The second level of facilities is the Reproductive Health Services (RHS) 'A' centres, which provide a full range of RH services, including comprehensive FP services. These include conventional and clinical methods as well as male and female contraceptive surgery facilities in static units and in extension service camps with safe and sterile conditions. There are 67 RHS A centres, located in district and tehsil level hospitals. There are also 14 RHS-B centres, which are based in hospitals /clinics of NGOs, and private clinics. These centres are either bring run by the Population Welfare Department centre or in collaboration with Greenstar Social Marketing (GSM), Marie Stopes Society Pakistan (MSS) and Pakistan Voluntary Health & Nutrition Association (PAVHNA).

In Punjab province, the Population Welfare Department has also established a very large network of population welfare centres, which include 119 Family Health Clinics (previously Reproductive Health Services Centres), 1,500 FWCs, 117 Mobile Service Units, 1,665 male mobilisers, and 54 RHS-B centres. The total sanctioned strength of employees of the Department is 10,635.

In keeping with its commitment to control population growth, Pakistan introduced another programme into its health system the Lady Health Workers Programme, in 1994. The programme was launched under the leadership of then Prime Minister, Benazir Bhutto, and 110,000 lady health workers (LHWs) were recruited from all over Pakistan to serve primarily rural and peri-urban areas. The programme was later renamed the National Family Planning and Primary Health Care Program, the primary focus of which is promoting the idea of family spacing and small families. LHWs create awareness through door-to-door meetings and provide short-term, modern FP supplies to women who express an intention to adopt FP. For long term and permanent methods, LHWs refer clients to the appropriate level of Population Welfare Department centres.

The LHW programme has been very successfully implemented in Punjab province. Punjab has also approved the Integrated Reproductive Maternal, Newborn and Child Health and Nutrition Programme in 2013 (Government of Punjab, 2013), which also focuses on FP and RH. The LHW programme in Sindh has continued to suffer from rapid changes of leadership, poor ownership, too much focus on polio and resource constraints.

International NGOs (INGOs), local NGOs and community-based organisations (CBOs) are also very active in Pakistan in promoting population health and FP. These INGOs, NGOs and CBOs include, for example, the World Population Foundation Pakistan (Rutgers WPF), Rahnuma – the Family Planning Association of Pakistan, the National Institute of Population Studies, the National Trust for Population Welfare (NATPOW), MSS, the Population Association of Pakistan (PAP), the Family Welfare Co-operative Society, the Pakistan Family Welfare Council, UN Population Fund (UNFPA) Pakistan, the Youth Advocacy Network (YAN) and GSM and DKT International. NGOs work in different capacities such as social marketing (SM), referrals, ongoing training of LHWs, advocacy, networking, the formation of women's associations, service delivery, quality improvement, social franchising (SF), research and evaluation, and demand creation⁴.

³ Downloaded from <http://www.pwdsindh.gov.pk/address/msu.htm> on 25 February 2016.

⁴ This information has been downloaded from <http://loqpati.blogspot.co.id/2014/06/program-policy-analysis-lady-health.html>.

1.3 The DRHR programme

In July 2012 DFID signed two contracts with Marie Stopes International (MSI) and Population Services International (PSI) under the DRHR programme to deliver RH results over four years (2012 to 2016), with an option to extend for 12 months. The programme aims to improve maternal and new-born health by ensuring the supply of high quality RH services and commodities through non-state providers in poor and under-served urban and rural areas in selected districts of southern Punjab, northern Sindh, Khyber Pakhtunkhwa (KP) and the Federally Administered Tribal Areas (FATA) in the initial phase, with a possibility of covering other parts of the country in future. MSI aims to scale up access to and use of modern FP services through SF whilst PSI uses SM approaches.

1.3.1 SF model

MSI were contracted by DFID to supply socially franchised clinical services and commodities, provide advice on RH (mainly FP) and undertake demand-side interventions through the non-state providers (for profit and not for profit organisations) in under-served urban and rural areas. The DRHR programme addresses both the supply side and demand side of RH services (with a focus on FP) through a SF approach. Supply-side services include ensuring the availability of a wide range of modern FP methods, and related information, and overcoming barriers to access and effective utilisation. Demand-side services include raising awareness and understanding of FP options, and community sensitisation to address social and cultural barriers that prevent service utilisation.

The geographical scope of the SF programme as proposed at inception included 10 districts in Punjab (projected total of 1,246 providers over four years), three districts in Sindh (410 providers) and 10 districts in KP (949 providers). MSI's approach to SF uses the MSS 'Suraj' SF model, which addresses supply-side aspects of quality, equity, affordability and availability of services through a network of Suraj franchises (Suraj A and Suraj B clinics) supported by reproductive health private providers (RHPPs) (called Pehli Kiran), which are supplied with contraceptives to provide short-term and intrauterine device (IUD) services. Suraj clinics are essentially partnerships with local private health services providers located in peri-urban and rural areas at an average distance of 40km–80km from district headquarters. On the other hand, Pehli Kiran are service providers particularly targeting far-flung rural communities. RHPPs represent more than 90% of the total number of providers in the franchised network. A third component was added in Year 3: ROSHNI mobile vans started providing FP services in far-flung hard to reach areas across 15 districts.

On the demand side, the project trains SF providers on client-centred services, counselling and side-effect management. In addition, the project trains the Pehli Kiran in community mobilisation and Information, Education and Communication (IEC) utilising existing materials. The principle is that subsidised services and commodities catalyse demand for FP in the short term and precipitate behaviour change that can maintain demand in the long term. Additionally, the BCC and marketing activities distribute vouchers for FP services through a network of field worker marketing agents (FWMs) to help increase health seeking behaviour of the poor and under-served for FP services. FWMs conduct door-to-door visits to market the Suraj brand and services, mobilise the community, generate referrals and distribute vouchers to potential clients based on a poverty assessment. Those in receipt of vouchers are entitled to obtain an IUD insertion for free at franchised providers. Those who do not qualify for the voucher pay the full amount (200 Pakistani rupee (PKR)).

1.3.2 SM model

PSI was contracted by DFID to supply SM commodities, products and advice for RH (mainly FP) along with demand-side interventions in under-served urban and rural areas. Similar to SF, the SM approach addresses both the demand side and the supply side of FP/RH commodity and service provision. PSI is working in this programme through their affiliate GSM.

GSM uses a total market approach, which makes markets work for the poor by ensuring all segments of society are reached with high quality products and services according to their ability to pay and includes commercial for profit products, social marketed subsidised products and products that are free for the poorest. GSM, as part of its marketing strategy, used total market analysis (Year 1) to ensure that products were available at multiple price points. A voucher scheme was initially planned but later removed with DFID's agreement.

In order to deliver the programme activities GSM partnered with the Rural Support Programmes Network in southern Punjab (19 DRHR programme districts), Health and Nutrition Development Society in Northern Sindh (12 DRHR districts) and Community Appraisal and Motivation Program (CAMP)⁵ in FATA (13 DRHR districts) and KP (24 DRHR districts), with a focus on under-served urban and rural areas. GSM and its partners mapped providers and commercial outlets in these areas in Year 1 and have ensured that FP products and services are available in these outlets. Of note is the fact that the mid-term review noted the undue focus on urban providers and outlets (60% vs 40% rural), which led to intensified efforts in Year 3 to redress this imbalance.

To empower women and girls to make healthy reproductive choices, GSM and its implementation partners reach young women, men and key influencers like health care providers, husbands and mothers in law with evidence informed messages promoted through interpersonal communication. In Year 3 GSM also launched its toll-free helpline, which provides FP and sexual and RH information, counselling and referrals.

1.4 DRHR programme impact evaluation

This impact evaluation of the DRHR RH programmes implemented by MSI and PSI sheds light on the extent to which women in targeted regions of Pakistan are accessing and using high quality RH services. It also provides evidence on which strategies to improve RH outcomes are effective in this context, given the particular combination of constraints that women face. It is anticipated that the findings of the report will be useful in informing future policy and practice in this area, with a view to delivering improved health outcomes for vulnerable women and infants in Pakistan. It is widely acknowledged that there is still a need to develop rigorous studies to evaluate the effects of SF on access to and quality of health services in low and middle-income countries⁶.

The report is complemented by a study of the value-for-money of the two programmes (OPM, 2016).

⁵ CAMP was no longer a partner in Year 3 due to the government having rescinded its permission to operate in the FATA region (February 2015).

⁶ Beyeler N, York De La Cruz A, Montagu D (2013) The impact of clinical social franchising on health services in low- and middle-income countries: A systematic review. *PLoS ONE* 8(4): e60669.

2 Impact evaluation methodology

2.1 What the evaluation is measuring

This impact evaluation assessed the DFID-funded RH programmes in Pakistan implemented by MSI and PSI against their main objectives of expanding delivery and utilisation of high quality RH services and commodities in under-served areas. These objectives of these programmes can be summarised by improvement in four key impact areas:

- access to RH services;
- utilisation of RH services;
- equity of access and utilisation based on poverty status; and
- quality of RH services.

The evaluation used a combination of propensity score matching (PSM) and differences in differences methods to estimate the changes in outcomes that can be attributed to the programmes. These are quasi-experimental techniques that aim to remove the bias that can result from a simple comparison between households that are exposed to the programme and those who are not, in cases when the placement of the programme is not random. The hypotheses that were tested are listed in .

Table 1.

In relation to access, our expectation was that the interventions would increase access to RH services by expanding the network of providers, which can be assessed through indicators relating to time, cost and awareness of providers by households.

Interventions to increase supply but also demand should increase utilisation through increased affordability of RH products and services, increased knowledge and self-efficacy of clients, and increased acceptability of services. This was expected to lead to increased coverage, a higher contraceptive prevalence rate and reduced unmet need for contraception.

Given the pro-poor design of the programmes (both in terms of area targeting and use of vouchers to increase access for under-served groups), our expectation was that households from poorer quintiles would benefit disproportionately from the programmes, which would be assessed by disaggregating indicators for access and utilisation by socio-economic status.

Finally, the programme includes a strong emphasis on quality improvement, which is a critical pathway for increasing acceptability, coverage and ultimately impact. A number of indicators from the provider survey are used to examine the extent to which the interventions improved different facets of quality, such as hygiene, availability of staff and drugs, and knowledge of providers on FP.

Table 1: Key impact indicators

IMPACT AREA	HYPOTHESIS	RISKS/ ASSUMPTIONS	INDICATORS
Access	PSI/MSI are increasing the supply of RH services to either franchised providers or a network of retailers. This should help to overcome supply-side constraints to access to RH services	We assume that PSI/MSI focus their interventions on currently under-served areas, i.e. they focus on remote/rural settings rather than 'easy wins' – this is a risk given that pay for performance indicators do not specify where in programme districts service levels must rise	<ul style="list-style-type: none"> • Average time for household to reach nearest RH provider (<i>HH survey</i>) • Average out-of-pocket expenditure on transport to nearest RH provider (<i>HH survey</i>) • Knowledge of RH providers in community or nearby (<i>HH survey</i>)
Utilisation	PSI/MSI interventions both seek to solve supply-side constraints. In addition, PSI has a small component for BCC – active BCC should go some way towards spreading information on RH best practices as well as challenging existing practices and cultural norms	The BCC component is only small and there are no other efforts at demand generation. If uptake is constrained by demand rather than supply there may be no increase in utilisation of services	<p><i>Impact</i></p> <ul style="list-style-type: none"> • Contraceptive prevalence rate (<i>household (HH) survey</i>) • Unmet need for contraception (<i>HH survey</i>) <p><i>Knowledge</i></p> <ul style="list-style-type: none"> • Client perception of contraceptive attributes (<i>HH survey</i>) • Client self-efficacy in regard to contraceptive use (<i>HH survey / exit interview</i>) <p><i>Coverage:</i></p> <ul style="list-style-type: none"> • % of targeted population using subsidised FP methods (<i>HH survey</i>) • % of MWRA exposed to RH related BCC activities (<i>HH survey</i>) • % of MWRAs (or someone on their behalf) who have visited a RH provider in the last three months (<i>HH survey</i>) <p><i>Household costs:</i></p> <ul style="list-style-type: none"> • % of MWRA (or someone on their behalf) spending any money to obtain the contraceptive method last time (<i>HH survey</i>) • Mean spending on obtaining the contraceptive method last time (<i>HH survey</i>) <p><i>Acceptability and barriers to use</i></p> <ul style="list-style-type: none"> • % of MWRAs received service from female provider in last visit (<i>HH survey</i>) • % of women reporting not using due to social, marital, religious restrictions, health concerns or due to inability to afford or access contraception
Equity	The SF intervention includes a voucher entitling beneficiaries targeted through a poverty scorecard to receive free FP services. In addition, both SF and SM interventions are targeted at increasing the supply of FP services in currently under-served areas likely to have higher populations of poorer households	Neither PSI/MSI are using a scorecard that is nationally recognised (e.g. the National Income Support Programme (NISP) poverty scorecard) so we cannot be sure of their effectiveness in targeting the poorest	<ul style="list-style-type: none"> • Access to SF and SM services by socio-economic status (<i>household survey</i>) • Utilisation of SF and SM services by socio-economic status (<i>household survey</i>) • Knowledge of methods by socio-economic group (<i>household survey</i>)
Quality	Both PSI/MSI specify eligibility criteria for the	Quality of service is not specified in the pay for	<ul style="list-style-type: none"> • % of providers with no reported stock outs (<i>provider survey</i>)

	<p>providers/retailers that work with them. Furthermore, they are supposed to provide some level of training to those working with them before they can enter the programme, which could lead to an increase in quality</p>	<p>performance contract of either MSI/PSI and therefore there may be lower incentives to maintain high quality services</p>	<ul style="list-style-type: none"> • % of RH providers fully staffed (<i>provider survey</i>) • % of RH providers having standards of service displayed at FP delivery point (<i>provider survey</i>) • % of RH staff with adequate knowledge (<i>provider survey</i>) • General hygiene of provider (<i>provider survey</i>) • Prices of RH products and services (<i>reported by outlets and facilities</i>) • Range of products (<i>at outlets</i>) • Changes to household perceived quality (satisfaction, waiting times, information on side effects) (<i>HH survey</i>)
--	---	---	--

2.2 How the impact was measured

There are two stages to our quantitative data analysis of the effects of the DRHR initiative. Whilst the core of the analysis consists of a quasi-experimental estimation of programme impact, this is preceded by a descriptive analysis of treatment and control groups' inter-temporal trends.

The first analytical step is therefore informed by a descriptive exploration of patterns and changes over time in the key outcome indicators in each treatment group. This is meant to help us to understand what, if anything, has changed for women and households in our sample since the time of the baseline survey. This type of analysis can suggest whether or not changes over time are consistent with the DRHR programme having been effective, but is not suitable for making causal statements about the impact of the programme since it does not explicitly account for how outcomes may have evolved in the absence of the interventions. It is instead intended to provide a backdrop to the main impact analysis, and to generate preliminary hypotheses about the possible effects of the programme and intermediate outcomes that it may have contributed to. The descriptive comparison of treatment and control groups is also of interest, in order to compare the evolution of their inter-temporal trends and to assess whether or not they followed similar paths.

The second stage of analysis represents the core of the impact evaluation and is based on a quasi-experimental technique, PSM, which allows us to robustly determine the impact of the programme on key outcome areas.

In order to isolate the effect caused by DRHR programme interventions it is not sufficient to simply compare outcomes between women who were exposed to the programme (in treatment areas) and those who were not (in control areas). This is because assignment to the programme was not determined at random, and therefore there may be systematic differences between women in areas where the programme is working and areas where it is not. Such 'selection bias' can invalidate naïve comparisons in outcomes, if there are a set of underlying characteristics that are associated with both exposure to the intervention and the level of final outcomes.

PSM overcomes this problem of the counterfactual by constructing a comparison group for the treatment groups that resembles as closely as possible the comparison group that might have been selected if the programme had been assigned at random. In other words, it seeks to eliminate systematic differences in characteristics of the sample in each treatment group that are

the source of selection bias. This is done by using information from the baseline data to construct a propensity score for each household, which estimates the likelihood that this household is exposed to DRHR programme activities conditional on its observable characteristics. These scores are constructed from a combination of variables found to predict both exposure to the programme and final outcomes of interest (the key outcome that we focused on was current use of contraceptives).

The choice of variables we included in our propensity score model for this evaluation was based on a combination of theory and data-driven approaches. On the theoretical side we selected variables for the model that were known to be associated with contraceptive use amongst married women in the context of the evaluation. This included, for instance, women's self-reported desire for more children, the age at which they got married, and the poverty status of the household they belong to. Additional variables were then added to the model on the basis of a 'data-driven' algorithmic approach, in which all variables in the dataset were tested for their suitability for the model after running a series of tests. The propensity scores were generated using baseline data in order to rule out the possibility that any of the matching variables selected could have been affected by exposure to the programmes.

The impact analysis was then performed by comparing households in the treatment groups with households in the control group that have similar propensity scores. A range of balance diagnostic tests were carried out to confirm that differences in characteristics between MWRA before the programme began were indeed reduced in the 'matched' sample compared with the overall (unmatched) sample. A reduction in bias in the post-matching setting is an indication that the PSM technique successfully approximates a randomised controlled trial in terms of the absence of systematic differences between treatment and control groups.

A number of different matching algorithms were also performed as robustness and sensitivity tests to identify the optimal matching model. Findings from our preferred model are presented in the main body of the report. We show in the annexes the results from an additional specification of the matching model, as a robustness check on the final results.

When done well, PSM successfully removes the potential confounding influence of observable characteristics that are systematically different across treatment groups. However, the risk remains when using PSM techniques that there may be systematic differences in unobservable characteristics that persist even after matching. This includes variables like motivation, emotional intelligence and female empowerment that are not easily measured through quantitative surveys. In order to address this possibility, our PSM analysis is combined with a differences in differences approach. This differences in differences approach involves comparing the difference in outcomes in treatment and control groups before the DRHR RH programmes began, with the difference that is observed afterwards. By making a comparison of outcomes both before and after the programmes were implemented, the influence of any remaining underlying differences between the samples that would otherwise bias the results is removed.

2.3 Survey instruments

We used the same data collection instruments that were used in the baseline (OPM, 2013). This enabled us to compare household outcomes against the same indicators over time. The survey instruments were as follows:

- a household survey covering KP, southern Punjab and northern Sindh⁷. This was a panel survey, and included a woman's questionnaire applied to a randomly selected MWRA in each household, covering their access to and use of RH services;
- a facility questionnaire was implemented in a sample of Suraj A and Suraj B franchises as well as RHPPs/Pehli Kiran providers, to assess the quality of service provision in the MSI SF intervention. The facility questionnaire was also implemented in a sample of control facilities (i.e. non-franchisee facilities); and
- an outlet questionnaire for a sample of retailers and pharmacies, to assess the quality of service provision in the PSI SM intervention. The outlet questionnaire was also implemented in a sample of control retailers and pharmacies (i.e. that are not receiving supplies from PSI).

2.3.1 Household panel survey

The questionnaire included:

- a household roster;
- basic household characteristics;
- a questionnaire for MWRAs;
- Benazir Income Support Programme poverty scorecard questions; and
- modules to address our impact areas (as per Table 1)

2.3.2 Facility and retailer questionnaires

The provider questionnaire was designed to assess the impact of the SF intervention implemented by MSI. The questionnaire included questions relating to the following areas:

- standards at the health facilities (basic amenities and hygiene);
- commodities provided;
- prices of RH services;
- quality of care; and
- staffing and stock levels.

An outlet questionnaire was developed to address the SM interventions implemented by PSI. The questionnaire included questions on:

- commodities provided and stock levels; and
- prices of RH commodities.

For brevity, tools are not included in this report but are available on request from the project manager.

2.4 Sampling strategy

The sampling strategy determined at baseline was based on randomly selecting evaluation clusters from one of three treatment groups.

Treatment 1: clusters where both MSI and PSI were operating.

⁷ We excluded FATA from the sampling frame because of the security situation in that region.

Treatment 2: clusters where only PSI was operating.

Control: clusters where neither RH programmes were operating.

Through these four treatment arms the evaluation aimed to measure both the joint impact of the MSI and PSI programmes and the effect of the PSI programme in isolation at the tehsil (sub-district) level. This makes it possible to assess the value added of having MSI operating in areas where PSI works, allowing the evaluation to draw some conclusions about the relative effectiveness of the two service providers, as requested in the terms of reference. Although it would have been desirable to include a treatment arm for clusters where only the MSI project was operating, it was found during the baseline period that the PSI project operates in all the areas where MSI works.

For the choice of controls, selection started from the sample frame containing tehsils in which no DRHR programme activities were being implemented. Additional areas were removed from the sampling frame if they were associated with security concerns, or expected for other reasons to be outliers in terms of their demographic, socio-economic or geographical characteristics relative to the rest of the sample. Control tehsils were then selected randomly from this sample frame from the full list contained in the 1998 census.

Since DRHR programme activities are implemented in a large proportion of tehsils in KP, Sindh and Punjab, after purposefully removing areas affected by the issues mentioned above, the sampling frame of control tehsils was disproportionately small relative to the sampling frame of treatment areas. This means that the control sample was necessarily selected from a relatively smaller number of districts than the treatment sample, and was not widely dispersed.

We did not remove from the sampling frame areas in which there were other programmes known to be occurring. The reason for this is that doing so would have caused bias in the results. By restricting the control sampling frame further we would have been left with a small pool of potential controls likely to be systematically different from the rest of the areas targeted by both the DRHR programme and other health related programmes. Indeed, as long as the activities of other programmes working in RH are evenly distributed among treatment and control areas, and not concentrated only in control areas, then the effect of this other programming is a part of the context for the intervention that we want to capture in our results.

Choosing control areas on the basis of matching with treatment areas was not done. This is partly due to the fact that it is difficult to access rich, high quality data at the tehsil level on which to perform matching across relevant dimensions. Secondly, the matching that was planned at the individual level (together with differences in differences) is sufficient to address the differences in underlying characteristics between treatment and control groups that would otherwise bias impact estimates. This was confirmed at the analysis stage at endline, when the woman-level PSM model achieved balance between treatment and control groups.

2.5 Sample sizes

The intended sample size was the same as what was achieved at baseline. The original sample was calculated to be the smallest required to allow the evaluation to detect a given change in key indicators associated with the RH programmes with a high degree of statistical confidence (OPM, 2013).

As with any panel survey, there was a risk that field teams would be unable to find some survey respondents, or that some households might have moved away. In addition, since the sampling

strategy involved identifying the same women for the follow-up survey as were interviewed at baseline, there was an additional risk that sampled women could not be re-surveyed even if the household itself was successfully located. In cases where it was not possible to re-interview the same woman from a sampled household, another MWRA was chosen at random from the household in order to try and preserve as much of the sample size as possible. These two sources of attrition are discussed further in Section 2.9.

The final sample size achieved at baseline and at endline is presented in Table 2 below.

Table 2: Overall sample sizes by survey round and treatment group

Baseline	Clusters	Households	Health facilities	Outlets
PSI-only	124	2436	n/a	224
PSI+BCC	50	988	n/a	88
MSI+PSI	126	2502	125	222
Control	100	1962	94	201
Total	400	7888	219	735
Endline	Clusters	Households	Health facilities	Outlets
PSI-only	124	2006	n/a	191
PSI+BCC	49	824	n/a	78
MSI+PSI	123	2008	110	184
Control	100	1498	78	167
Total	396	6336	188	620

Note: Sample size of households refers to the sample of households where an eligible woman (married, and between the ages of 12 and 49) was available for interview and gave consent. Sample size of outlets refers to outlets where the interview was reported as completed.

The sample size used for the core of our analysis departs from the above in two important respects. The first is that we have omitted study of treatment group two: PSI+BCC. At the time of writing it had not been possible to obtain clarity regarding the content of the BCC activities planned under this intervention package, and how it differs from the communication activities that are already part of the main interventions. A lack of knowledge about the implementation plans for this intervention would compromise the ability to interpret the results. We therefore concentrate on the core treatment options: PSI-only and PSI+MSI.

Secondly, the matching analysis is conducted only using the sample of women who were successfully re-interviewed at baseline, which results in a smaller sample size than had the 'replacement' women not been excluded. This is because the analysis methodology rested on the ability to compare changes over time in outcomes for each sampled woman. Including all women interviewed at endline in the analysis would have necessitated the assumption that the 'replacement' women are sufficiently similar to the original women interviewed to justify treating them as comparable. Differences in underlying characteristics between the samples of women would make it impossible to separate changes in average outcomes from baseline and endline that were due to these differing characteristics of the sample from changes caused by the intervention. A sensitivity analysis was conducted to understand the similarity between the original sample of women interviewed at baseline and the 'replacement' women. On the basis of the results it was decided that the sample of women who replaced original baseline women in cases where they could not be re-interviewed were not sufficiently similar across a range of demographic characteristics to justify treating them as equivalent during analysis. The results of this sensitivity analysis are reported in Annex C.

Therefore the final effective sample for analysis (at both baseline and endline) is given in Table 3. This amounts to an attrition rate of approximately 30% of women from the original sample selected at baseline. Of the 6,900 women originally interviewed for the PSI-only, MSI+PSI and control groups, only 4,798 were included in the final analysis.

Table 3: Sample size used for analysis of the household data

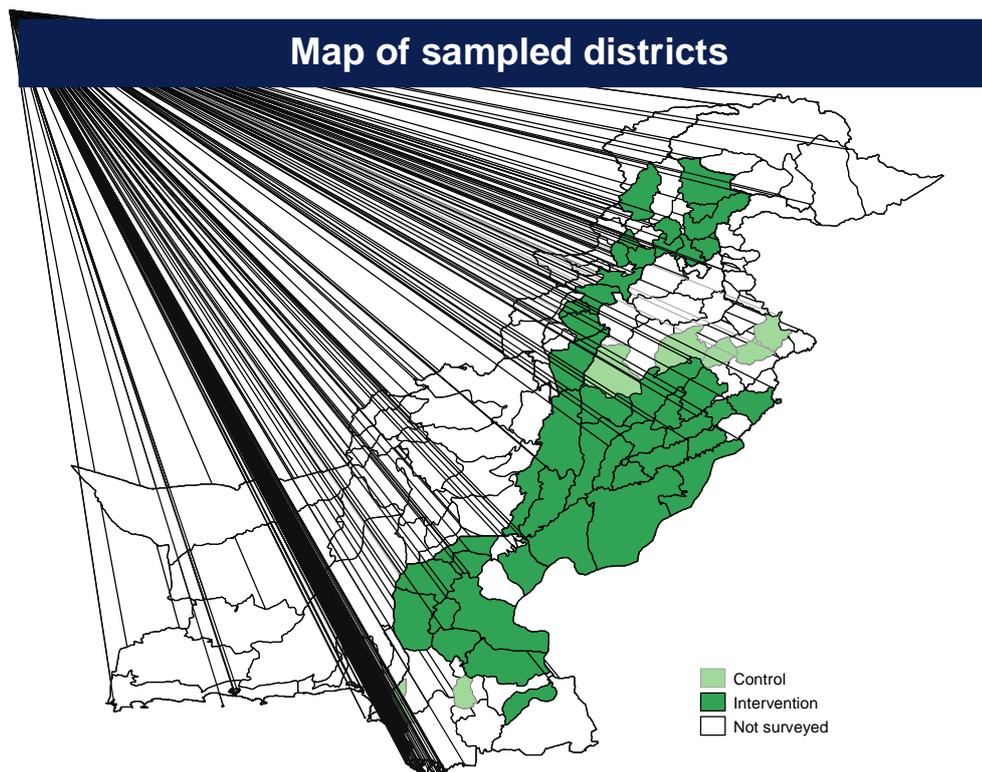
Baseline and endline	Clusters	Households
PSI-only	124	1750
MSI+PSI	123	1755
Control	100	1293
Total	347	4798

2.5.1 The location of the sample

The household panel survey was carried out in northern Sindh, southern Punjab and KP provinces. Figure 1 below illustrates how the different treatment types were distributed across the sample area. Since MSI and PSI operate at the level of the tehsil or taluka (sub-district), there can be an overlap of different treatment types within a given district. In the figure below we consider a district to be covered by the DRHR programme if there are tehsils falling into any of the three treatment groups (PSI-only, PSI+BCC or PSI+MSI) and no 'control' tehsils. Control districts in the figure below are those containing 'control' tehsils only⁸.

This figure illustrates that control districts are not evenly distributed throughout the study area, but are instead clustered together. To the extent that neighbouring districts may be somewhat similar, this suggests a possible risk that households in different treatment groups may also exhibit similar characteristics in the absence of the RH programmes.

⁸ There is one district in Sindh where there are both tehsils marked as 'control' and tehsils with DRHR programme activities taking place. This district is marked as 'Intervention' in Figure 1.

Figure 1: Sampled districts according to treatment status


2.5.2 The facility and outlet samples

The survey of health facilities is used to assess the quality of service provision in the MSI SF intervention. Surak A and Suraj B franchises, as well as RHPPs/ Pehli Kiran providers were sampled from the communities in which the household survey was conducted in MSI+PSI treatment group tehsils and control tehsils. Table 4 below summarises the composition of this sample at the time of the follow-up survey, in terms of the levels of health facilities contained in each group. Significantly more Type I health centres were sampled in the treatment group (corresponding to dispensaries, sub-centres and health posts with limited curative and preventative care facilities), and significantly fewer Type II facilities (which contain some in-patient beds and can offer curative and preventative care).

Table 4: Types of sampled health facilities

Indicator	MSI + PSI	Control	Difference	P value of difference	MSI + PSI sample size	Control sample size
	Proportion of health facilities by level					
Health centre Type I ²	76.4	53.8	22.6***	0.001	110	78
Health centre Type II ³	12.7	28.2	-15.5***	0.008	110	78
Hospital	10.9	17.9	-7	0.17	110	78

*= significant at 10% level, **= significant at 5% level, ***=Significant at 1% level (1) dispensary, sub-centre, health post e.g. no beds, limited curative and preventative care; staffed by nurse or auxiliary. (2) e.g. some in-patient beds, curative and preventative care; staffed by team of professionals and auxiliary staff.

The outlet survey was carried out in a sample of retailers in all treatment groups and the control group to measure the quality of service provision through the PSI SM intervention.

Table 5: Types of sampled outlets

Indicator	Treatment 1: PSI-only	N	Treatment 2: PSI+BCC	N	Treatment 3: MSI+PSI	N	Control	N
Proportion of retailers by type								
Small shop	1.2***	170	8	75	8.2	183	11.5	165
General store	13.5***	170	18.7*	75	18.6**	183	30.3	165
Kiryana store	37.6	170	33.3	75	23.5**	183	35.8	165
Pharmacy	45.3***	170	37.3**	75	47***	183	22.4	165
Others	2.4***	170	2.7**	75	2.7**	183	0	165

Notes: Asterisks indicate significant differences from the control group. *= significant at 10% level, **= significant at 5% level, ***= significant at 1% level.

2.6 Survey implementation

OPM's Pakistan office carried out the survey. Eight field teams were engaged for data collection. The field team members were thoroughly trained prior to the data collection. The training included classroom training, field practice and piloting.

The fieldwork was completed in six weeks during November–December 2015.

2.7 Data management and analysis

Data was collected using computer assisted personal interviewing (CAPI) and periodically sent to the OPM Islamabad office using a secure Dropbox folder. Data were then thoroughly examined for missing values, answers out of range, and logical inconsistencies and the cleaned data were then converted from SPSS into Stata.

Intermediate datasets were then created in which the data were prepared for analysis. This involved creating outcome indicators and matching covariates. All variables were created according to the principles adopted for the baseline analysis to ensure full comparability of results. Observations of women who did not consent to interview, and households where an eligible MWRA could not be selected for interview, were removed from the data.

Weights were then created at the household and women level. The use of weights adjusts for the fact that women living in different areas may have had slightly different probabilities of being selected into the sample, due to the differing sizes of evaluation clusters. This is required to ensure that the final results are fully representative of the evaluation area.

Poverty scores were estimated in the baseline data, based on the NISP poverty scorecard. This is an index score which approximates the poverty and welfare status of households based on a weighted combination of observable characteristics, such as asset ownership and the ratio of dependents in the household (children and the elderly) relative to the overall household size. Households are characterised as poor or not poor depending on whether they fall above or below a certain cut-off, which we have defined in line with the Pakistan Living Standards Measurement Survey (PLSM) (2008/09).

2.8 Ethical consideration

OPM's Ethical Review Committee (ERC) reviewed and approved this research project. Local institutional approval was obtained from the Pakistan Bureau of Statistics. All the respondents provided informed consent prior to data collection.

2.9 Limitations of the analysis

2.9.1 Attrition

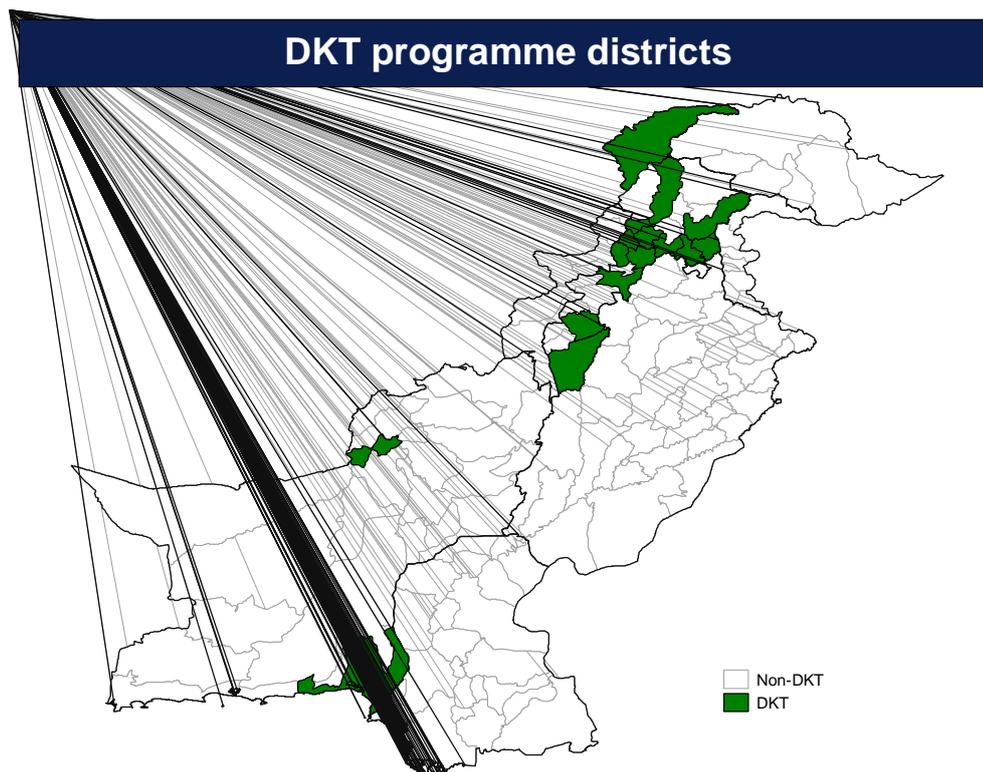
As described in Section 2.5 above, the data for the endline survey were affected by two types of attrition. Attrition at the household level occurred where the household interviewed at baseline could not be located at endline, or where an interview with an eligible MWRA could not be completed. In addition, attrition at the woman level occurred where the household was found but the original woman surveyed at baseline could not be interviewed in that household.

As a result of these two sources of attrition, and the decision of the analysis team not to incorporate data from women interviewed at endline as 'replacements' for originally sampled women, the final sample used for the matching analysis contains only 70% of the original women interviewed at baseline. This loss of sample, while not unexpected for a panel survey conducted at the individual level, nonetheless means that the power of the study to detect statistically significant changes in outcomes is reduced. This may mean that the findings are not able to detect small changes in outcomes caused by the intervention.

2.9.2 Contamination

Contamination refers to the situation in which households in control areas are also exposed to programme activities, meaning that the control group is no longer a true 'comparison' group for the groups where interventions were implemented. In this case the impact evaluation is likely to underestimate the true effects of the programmes, since control group households also experience changes as a result of interventions they receive.

A potential source of contamination to our sample may come from DKT International's activities, which also aim to promote RH in Pakistan through SM and SF approaches. DKT International works in KP, where MSI and PSI also operate. Based on the list of DKT programme areas provided to the evaluation team, and graphed in Figure 2, there are no overlaps between DKT and the control areas for the MSI PSI evaluation.

Figure 2: Districts where DKT International operates

2.10 How to read tables in this report

Two main kinds of tables are presented in this report. The first kind are descriptive tables, which show changes between the baseline and endline surveys in key indicators, for each treatment group. These tables follow a standard format. For each indicator, we present the mean value of the indicator for each treatment group at baseline and endline. We then present the difference between baseline and endline, and indicate whether the difference over time is statistically significant using asterisks (*). If no asterisks are given it means that the estimates are statistically similar. The level of significance is denoted as follows: three asterisks (***) indicate the difference is significant at the 99% level of confidence; two asterisks (**) indicate a 95% level of confidence; one asterisk (*) indicates a 90% level of confidence. The presentation of statistical significance is accompanied by p-values. Sample sizes are also reported.

The second kind of table presents impact estimates. These tables present the size of the estimate (known as the average treatment effect on the treated), which is the magnitude of the change in outcomes associated with the intervention. Asterisks indicate whether the treatment effect is statistically significant, this time measured at only the 95% or 99% level since the 90% level is considered to indicate only very weak significance for the impact evaluation.

3 Demographic characteristics

The demographic characteristics of the sample at baseline are described in the MSI PSI baseline report of 2013, which reveals a youthful population, with high rates of fertility, relatively low levels of education and an incidence of poverty that is higher than the national average, according to the NISP scorecard.

For the purposes of the follow-up survey, we concentrate here on examining whether there are salient differences between the populations in the different treatment groups under investigation. The comparability between different treatment groups is pertinent for the impact evaluation methodology since it provides an indication as to how unbalanced the samples are before matching is undertaken. It also sheds light on our descriptive results, which may be driven in large part by pre-existing differences in the nature of the samples in each group.

3.1 Age distribution of sampled MWRA

The age distribution of the household sample was investigated in the baseline report, where it was shown that 43% of the overall household population was aged under 15 years and only 3% was aged over 65. In Figures 3 and 4 below we report on the age range of sampled MWRA in particular (who had to be aged between 15 and 49 to be eligible for the main sample).

These figures show that the samples are quite balanced according to age, with the PSI-only group containing a slightly younger sample of MWRA than the control group.

Figure 3: Ages of sampled MWRA in PSI-only and control groups

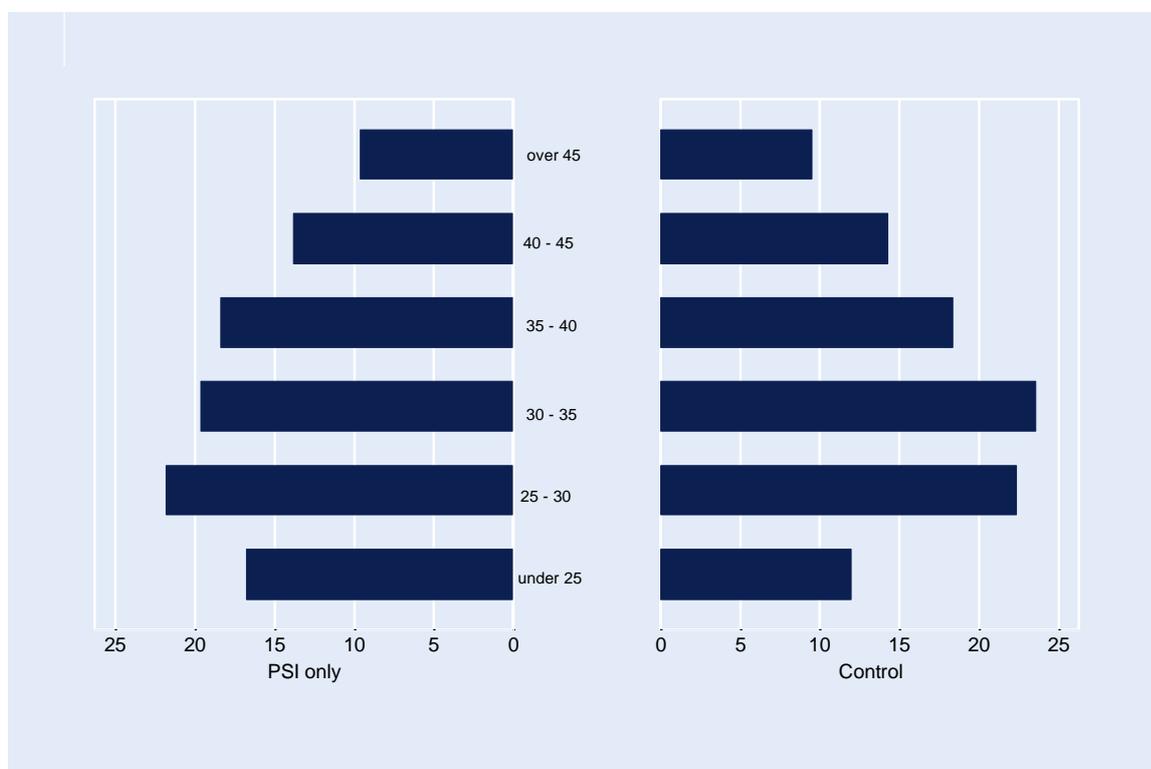
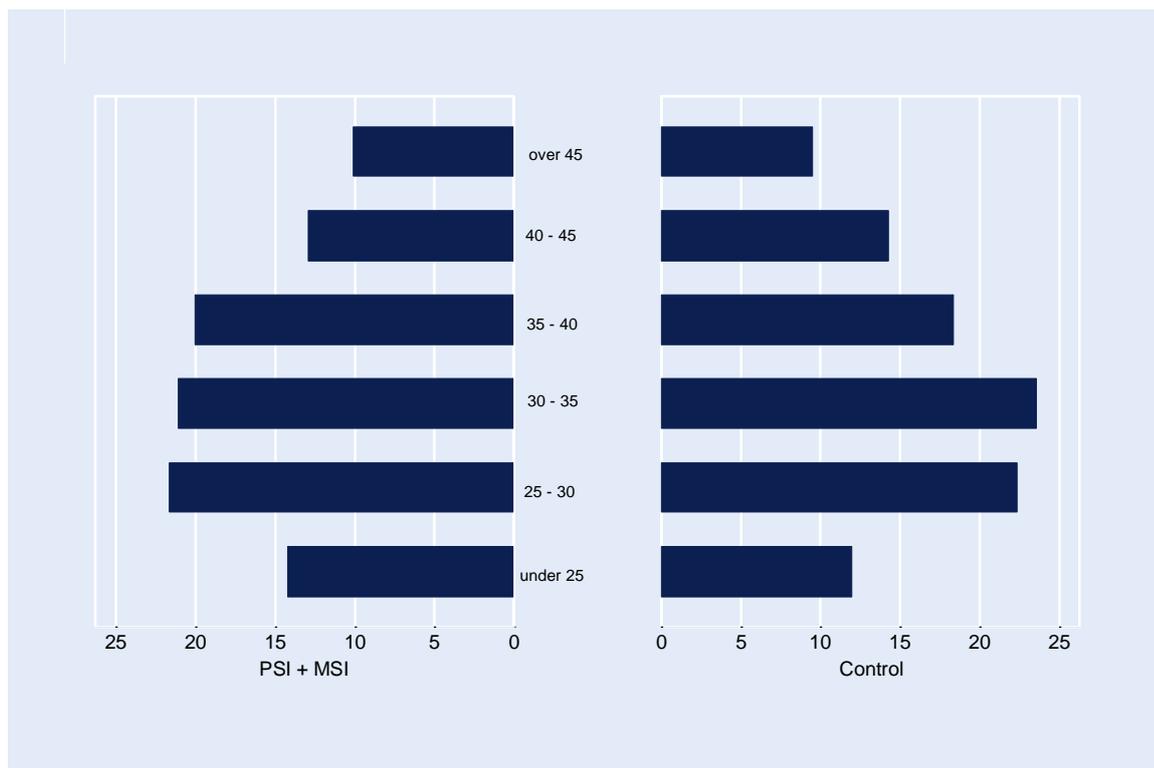
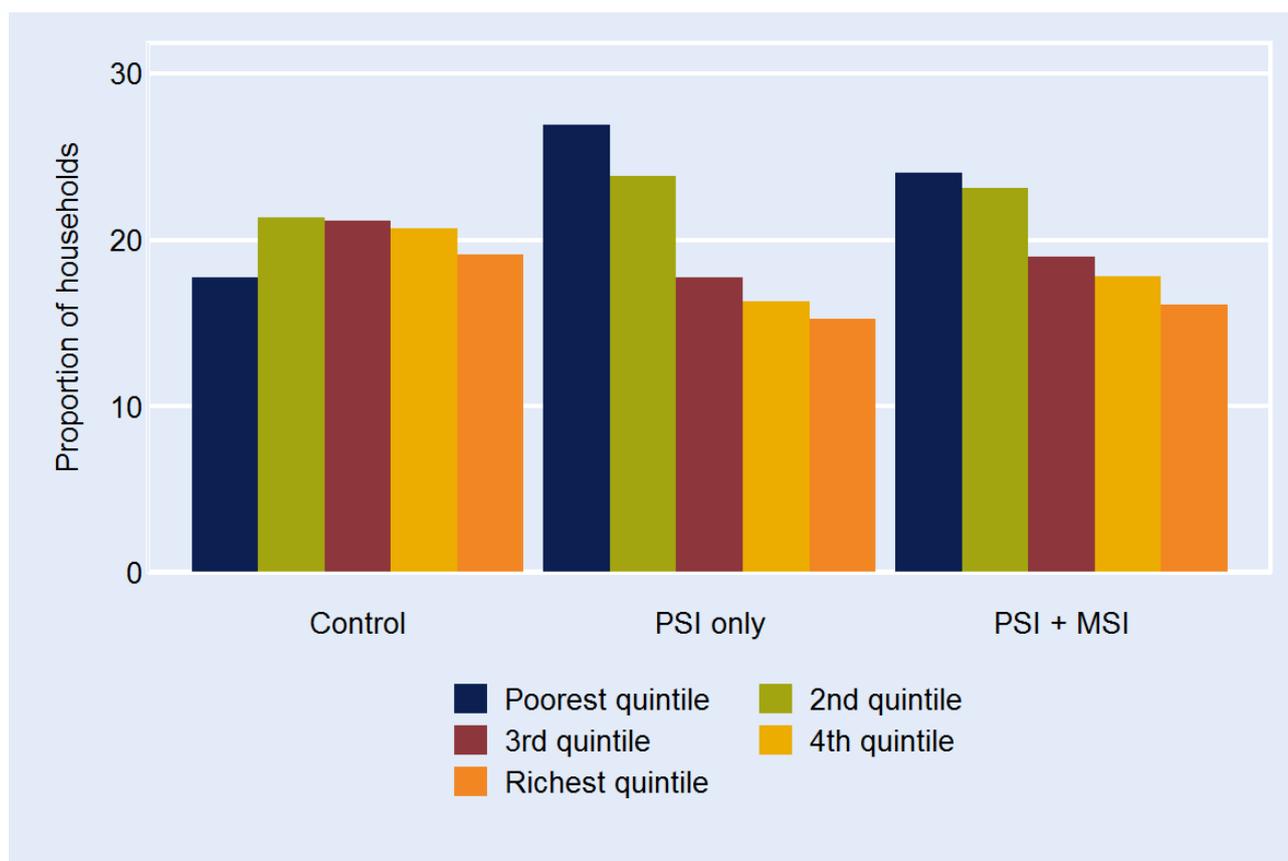


Figure 4: Ages of sampled women in PSI+MSI, and control groups



3.2 Poverty index quintiles

The poverty status of sampled households has been estimated by applying the NISP poverty scorecard to baseline data. These poverty scores can be used to show where sampled households fit into the national poverty quintiles of the NISP poverty scorecard, as used in the 2008–2009 PLSM survey. Figure 5 shows the distribution of these quintiles in each treatment group for the households that were successfully followed up at endline. The figure indicates a modest degree of targeting toward households falling into the poorest poverty quintile, especially so for the PSI-only treatment group.

Figure 5: Poverty quintiles

Notes: The NISP poverty score is a Proxy Means Test (PMT) score calculated using 23 variables. A PMT takes a weighted index of a set of observable indicators to approximate the level of poverty of the household. The poverty quintiles presented here are based on applying the NISP poverty scorecard to the data from our follow-up survey. Weights are applied at the household level.

3.3 Characteristics of households and MWRA respondents at endline

Tables 6 and 7 below tabulate a set of demographic characteristics of households and MWRAs at the time of the follow-up survey.

The PSI-only group is found to have significantly larger household sizes (with a mean of eight persons), compared to the two other groups, as well as more MWRAs in the household. The PSI+MSI group had a significantly higher proportion of female-headed households (over 9%). Both treatment groups had a significantly higher incidence of households classified as poor according to the NISP poverty scorecard at baseline than the control group. This again suggests some degree of targeting of the programmes towards poor households.

Table 6: Household characteristics at endline

	Treatment 1: PSI-only	N	Treatment 3: PSI+MSI	N	Control	N
Mean household size	8.03**	1992	7.11	1973	7.32	1283
Mean number of MWRAs	1.34***	1992	1.18	1973	1.17	1283
Poor as defined by NISP poverty scorecard	31.6***	1978	19.9**	1973	10	1272
Proportion of household heads who are female	4.3	1992	9.2**	1973	5.3	1283
Urban location	12.3	1992	17.9	1973	17.8	1283

Source: DFID RH endline survey, November–December 2015. Asterisks indicate that the estimate is significantly different to the control comparator. *= significant at 10% level, **= significant at 5% level, ***= significant at 1% level.

In relation to MWRAs, a significantly lower proportion were literate in the PSI-only group (at 24%, this was less than half that of the control, which had 50% literacy among MWRAs) (Table 7), which is matched by significantly fewer years of school in the PSI-only group (fewer than two years on average). Correspondingly, the PSI-only group reported significantly earlier age at first marriage and age at first pregnancy, compared to the two other groups.

Table 7: Characteristics of MWRA at endline

	Treatment 1: PSI-only	N	Treatment 3: PSI+MSI	N	Control	N
Mean age in years	31.21	1976	32.96	1963	32.34	1275
Proportion who are literate	24***	1976	37**	1963	50	1275
Mean years of schooling	1.82***	1976	2.77***	1963	4.3	1275
Occupation						
Agriculture	11.4	1976	11.5	1963	8.5	1275
Business	0.3	1976	0.1	1963	0.1	1275
Service	1	1976	1.3	1963	1.5	1275
Skilled labour	2.6	1976	2.9	1963	2.7	1275
Unskilled labour	3.1***	1976	3.1	1963	0	1275
Unemployed	0.1	1976	0	1963	0	1275
Housewife	80.9	1976	80.5	1963	84.5	1275
Others	0.1	1976	0.1	1963	0.5	1275
Mean age at first marriage	17.74***	1975	18.56**	1963	19.5	1273
Mean age at first pregnancy	18.99***	1876	19.72***	1861	20.6	1197
Mean number of live births	4.26*	1844	4.38***	1822	3.9	1166

Source: Asterisks indicate that the estimate is significantly different to the control comparator. *= significant at 10% level, **= significant at 5% level, ***= significant at 1% level

4 Access to RH services

Table 8 shows the changes to travel time and costs to reach the nearest RH provider (which includes all outlets for the SM programme as well as health facilities for the SF one) across the three groups, as proxies for changing access. There has been little change, descriptively, in the time and cost required to reach the nearest RH provider. There are some sizeable fluctuations in the mean spend on transport to reach the nearest RH provider, but these findings may be misleading – standard errors are reasonably large for this indicator in particular, and sample sizes are more restricted, leading to large confidence intervals.

In terms of changes to knowledge of nearest RH provider (Table 9), across all groups we see a decrease in the mean number of RH providers known to respondents in the community, driven by a decrease in awareness of government providers. This is accompanied by a small and marginally significant increase in the percentage of MWRAs who are aware that their nearest RH provider is a non-state provider.

Table 8: Average travel time and cost to reach nearest RH provider

	PSI-only						MSI+PSI					
	Baseline mean	Endline mean	Difference	P-value	Baseline N	Endline N	Baseline mean	Endline mean	Difference	P- value	Baseline N	Endline N
Mean time to reach nearest RH provider (in minutes)	27.5	30.46	2.96	0.47	1229	1369	19.59	18.19	-1.40	0.32	1300	1396
Of those visiting RH provider, % that spend any money on transport	52.5	55.4	0.03	0.46	1233	1369	28.3	31.4	0.03	0.44	1294	1396
Mean spend on transport to each nearest RH provider (PKR)	96.0	159.6	63.61	0.43	636	700	127.7	70.6	-57.1	0.12	430	495
Among MWRAs obtaining contraceptives methods or advice in the past three months, time in minutes taken to reach provider	30.294	21.902	-8.392*	0.076	123	136	17.268	14.978	-2.29	0.474	134	142

Notes: Asterisks indicate that the difference between the baseline and endline value is statistically significant. *= significant at 10% level, **= significant at 5% level, ***= significant at 1% level. Providers include health facilities and other outlets (for the SM programme).

	Control					
	Baseline mean	Endline mean	Difference	P-value	Baseline N	Endline N
Mean time to reach nearest RH provider (in minutes)	18.2	15.9	-2.28	0.25	857	937
Of those visiting RH provider, % that spend any money on transport	26.7	21.6	-0.05	0.22	840	937
Mean spend on transport to each nearest RH provider (PKR)	140.4	75.8	-64.67**	0.05	287	262
Among MWRAs obtaining contraceptive methods or advice in the past three months, time in minutes taken to reach provider	18.375	17.837	-0.538	0.878	75	91

Notes: Asterisks indicate that the difference between the baseline and endline value is statistically significant. *= significant at 10% level, **= significant at 5% level, ***= significant at 1% level. Providers include health facilities and other outlets (for the SM programme).

Table 9: Access to RH services in the community and nearby

	PSI-only						MSI+PSI					
	Baseline mean	Endline mean	Difference	P-value	Baseline N	Endline N	Baseline mean	Endline mean	Difference	P-value	Baseline N	Endline N
Mean number of RH providers known by the respondents in the community, of which:												
Government	0.57	0.38	-.196***	0.001	1721	1648	0.93	0.74	-.186***	0	1616	1616
Private	0.24	0.23	-0.015	0.699	1721	1648	0.37	0.36	-0.007	0.898	1616	1616
Total	0.81	0.6	-.212***	0.004	1721	1648	1.3	1.1	-.193***	0.01	1616	1616
Percentage of MWRAs who reported that nearest RH provider was...												
Government	43.6	47.2	3.6	0.525	1719	1723	67.7	60	-7.7***	0.005	1711	1711
Private	17	22.5	5.4*	0.066	1719	1723	18.5	25.5	7*	0.068	1711	1711
% of respondents who do not know what is the nearest RH provider	39.4	30.3	-9.1	0.147	1719	1723	13.8	14.5	0.7	0.793	1711	1711

Notes: Asterisks indicate that the difference between the baseline and endline value is statistically significant. *= significant at 10% level, **= significant at 5% level, ***= significant at 1% level. Private providers include a private hospital, a private clinic, a private practitioner, pharmacy, shop, FWC or mobile services unit. Government providers include a government hospital, Reproductive Health Centre (RHC), Basic Health Unit (BHU), LHW, lady health visitor and a community midwife.

	Control					
	Baseline mean	Endline mean	Difference	P-value	Baseline N	Endline N
Mean number of RH providers known by the respondents in the community, of which:						
Government	0.86	0.84	-0.021	0.817	1102	1093
Private	0.37	0.42	0.057	0.506	1102	1093
Total	1.23	1.26	0.035	0.803	1102	1093
Percentage of MWRAs who reported that nearest RH provider was...⁽¹⁾						
Government	64	70.6	6.5	0.139	1102	1100
Private	18.5	17	-1.5	0.714	1102	1100
% of respondents who do not know what is the nearest RH provider	17.4	12.5	-5	0.228	1102	1100

Notes: Asterisks indicate that the difference between the baseline and endline value is statistically significant. *= significant at 10% level, **= significant at 5% level, ***= significant at 1% level.

5 Utilisation of RH services

5.1 Current use of contraceptive methods

The proportion of MWRAAs who had ever used contraception and any modern method showed significant increases in all groups (Table 10). As this is an indicator that increases over time, this is anticipated. In relation to current use and unmet need, there was no significant change, with rates remaining high across all three groups (38% for the PSI-only group at endline, compared to 28% for PSI+MSI and control) (Table 11). This compares with 20% for the country as a whole in the 2012–13 Demographic and Health Survey (DHS).

Table 10: Ever used contraceptives

Indicator	PSI-only						MSI+PSI					
	Baseline mean	Endline mean	Difference	P-Value	Baseline N	Endline N	Baseline mean	Endline mean	Difference	P-Value	Baseline N	Endline N
% MWRAs who have ever used any contraceptive method...												
Any method	38	50.1	12.2***	0	1692	1723	60.5	68.7	8.2***	0	1686	1711
Any modern method	32.4	38.7	6.3**	0.02	1692	1723	47.6	58.2	10.6***	0	1683	1711
Female sterilisation	7.2	8.8	1.6	0.139	1715	1723	11.4	15.2	3.8***	0	1700	1711
Male sterilisation	0.1	0.1	0	0.994	1711	1723	0.6	0.4	-0.2	0.675	1699	1711
Pill	10.3	11.9	1.6	0.505	1706	1723	13.8	18.3	4.5***	0	1690	1711
IUD	6.7	8.1	1.5	0.112	1707	1723	10.4	13.3	3***	0.001	1694	1711
Injectable	12.1	15.3	3.2	0.159	1712	1723	14.8	19.7	5***	0.003	1696	1711
Implants	0.4	0.4	0	0.899	1712	1723	1	0.8	-0.2	0.621	1694	1711
Condoms	9.9	18.2	8.2***	0	1706	1723	21.1	26.8	5.7***	0	1696	1711
Emergency pill	0.6	1.9	1.3**	0.012	1717	1723	0.6	2.2	1.6***	0	1696	1711
Any traditional method	15.2	26.5	11.3***	0	1705	1723	32.4	35.3	2.9	0.368	1694	1711
Rhythm	7	12.5	5.4***	0.004	1717	1723	16.6	15.4	-1.2	0.647	1696	1711
Withdrawal	11.8	19.6	7.8***	0	1710	1723	23.8	28.5	4.7**	0.038	1693	1711

Notes: Asterisks indicate that the difference between the baseline and endline value is statistically significant. * = significant at 10% level, ** = significant at 5% level, *** = significant at 1% level.

	Control					
Indicator	Baseline mean	Endline mean	Difference	P-Value	Baseline N	Endline N
% MWRAs who have ever used any contraceptive method...						
Any method	60.4	72.4	12***	0	1092	1101
Any modern method	47.5	58.7	11.2***	0	1091	1101
Female sterilisation	10.4	14.7	4.3***	0	1100	1101
Male sterilisation	0.3	0.1	-0.2	0.544	1098	1101
Pill	10.2	11.2	1	0.653	1099	1101
IUD	11.4	16.6	5.2	0.234	1099	1101
Injectable	11.3	14.8	3.5*	0.052	1097	1101
Implants	0.7	1	0.3	0.503	1102	1101
Condoms	25.9	39.3	13.4***	0	1098	1101
Emergency pill	1.1	2.2	1.2	0.232	1093	1101
Any traditional method	36.8	44.2	7.3**	0.036	1093	1101
Rhythm	13.1	18.9	5.8*	0.072	1098	1101
Withdrawal	30.4	35	4.6	0.156	1097	1101

Notes: Asterisks indicate that the difference between the baseline and endline value is statistically significant. *= significant at 10% level, **= significant at 5% level, ***= significant at 1% level.

Table 11: Currently using contraceptives

Indicator	PSI-only						MSI+PSI					
	Baseline mean	Endline mean	Difference	P-Value	Baseline N	Endline N	Baseline mean	Endline mean	Difference	P-Value	Baseline N	Endline N
% MWRA currently using contraceptives ⁽¹⁾	26	29.1	3	0.104	1473	1568	43.8	42	-1.8	0.442	1466	1552
% MWRA currently using contraceptives, excluding those trying to get pregnant, who report they cannot get pregnant or are menopausal	35.7	39.6	3.9**	0.049	1087	1143	57	57	0	0.992	1165	1164
Unmet need for contraception ⁽²⁾	39.8	38.2	-1.6	0.668	1715	1724	27.5	28	0.5	0.807	1705	1711
Among current users of contraception, % of MWRAs who are currently using:												
Any modern method	75.3	73.8	-1.5	0.669	500	589	76.1	74.2	-1.9	0.581	573	659
Female sterilisation	29.8	27.9	-1.9	0.566	501	597	30.9	32	1.1	0.642	573	668
Male sterilisation	0.2	0.1	-0.1	0.252	501	597	1.7	0.5	-1.2	0.287	573	668
Pill	5	3.5	-1.5	0.263	501	597	4.3	5.8	1.5	0.198	573	668
IUD	9.1	8.1	-1.1	0.449	501	597	9.5	7.8	-1.8	0.155	573	668
Injectable	12.6	13.5	0.9	0.743	501	597	8.7	8.4	-0.3	0.822	573	668
Implants	0.2	0.4	0.2	0.359	501	597	0.5	1.6	1.1	0.211	573	668
Condoms	18.2	19.8	1.6	0.655	501	597	20.5	17.5	-3	0.246	573	668
Any traditional method	25	26.2	1.3	0.714	500	589	24	25.8	1.9	0.592	572	659
Rhythm	3.3	6.8	3.5**	0.046	501	597	3.3	5.2	1.9	0.231	573	668
Withdrawal	21.6	19.3	-2.3	0.519	501	597	20.6	20.4	-0.3	0.94	573	668

Notes: (1) Defined as the proportion of MWRAs (or partners) who are currently practising contraception. (2) As defined by <http://www.measuredhs.com/topics/upload/Figure-2-Revised-unmet-need-definition-flowchart-Bradley-et-al-AS25.pdf>. Asterisks indicate that the difference between the baseline and endline value is statistically significant. * = significant at 10% level, ** = significant at 5% level, *** = significant at 1% level.

Indicator	Control					
	Baseline mean	Endline mean	Difference	P-Value	Baseline N	Endline N
% MWRAs currently using contraceptives ⁽¹⁾	45.1	48.9	3.8	0.385	957	1011
% MWRAs currently using contraceptives, excluding those trying to get pregnant, who report that they cannot get pregnant or are menopausal	59	60.6	1.7	0.687	722	765
Unmet need for contraception ⁽²⁾	26.4	28.1	1.7	0.559	1100	1102
Among current users of contraception, % of MWRAs who are currently using:						
Any modern method	62.2	70.5	8.3	0.234	420	457
Female sterilisation	26.5	27.1	0.6	0.851	420	463
Male sterilisation	0.9	0	-0.9	0.281	420	463
Pill	4.4	1.6	-2.7**	0.045	420	463
IUD	7.3	8	0.8	0.755	420	463
Injectable	3.3	2	-1.3	0.284	420	463
Implants	1.4	0.2	-1.3	0.178	420	463
Condoms	18.4	31	12.5**	0.049	420	463
Any traditional method	37.8	29.5	-8.3	0.234	420	457
Rhythm	6.6	3.7	-2.9	0.442	420	463
Withdrawal	31.2	25.5	-5.6	0.314	420	463

Notes: (1) Defined as the proportion of MWRAs (or partners) who are currently practising contraception. (2) As defined by <http://www.measuredhs.com/topics/upload/Figure-2-Revised-unmet-need-definition-flowchart-Bradley-et-al-AS25.pdf>. Asterisks indicate that the difference between the baseline and endline value is statistically significant. *= significant at 10% level, **= significant at 5% level, ***= significant at 1% level.

5.2 Knowledge

Knowledge of any FP method was almost universal and has not changed. Awareness of modern methods appears to be increasing, and possibly displacing knowledge of traditional methods (Table 12). All groups showed improved overall understanding of the efficacy and preventative effects of contraception, although in the PSI+MSI group there was found to be a small but significant reduction in those knowing that contraception prevented unwanted pregnancies (Table 13).

Table 12: Trends in knowledge of MWRAs regarding contraceptive methods

Indicator	PSI-only						MSI+PSI					
	Baseline mean	Endline mean	Difference	P-Value	Baseline N	Endline N	Baseline mean	Endline mean	Difference	P-Value	Baseline N	Endline N
% MWRAs who have heard about:												
Any method	89.1	91.8	2.7	0.403	1722	1723	99.2	99.4	0.3	0.32	1707	1711
Any modern method	88.4	89	0.6	0.807	1722	1723	99	99.4	0.4	0.199	1707	1711
Female sterilisation	70.4	66.2	-4.2	0.174	1722	1723	92	89.7	-2.4*	0.084	1705	1711
Male sterilisation	24.2	21.7	-2.5	0.146	1716	1723	30.3	36.8	6.5**	0.024	1703	1711
Pill	83.4	82.6	-0.7	0.69	1718	1723	93.7	93.8	0.1	0.951	1704	1711
IUD	66.5	62.9	-3.7	0.131	1720	1723	86.7	87.7	1	0.588	1704	1711
Injectable	81.5	79.5	-2	0.449	1721	1723	94.7	96	1.3	0.236	1704	1711
Implants	21.3	33.5	12.2***	0	1720	1723	26.7	50.8	24.1***	0	1699	1711
Condoms	57.6	61.9	4.2	0.26	1720	1723	78.6	81.9	3.3	0.133	1704	1711
Emergency pill	11.1	17	5.9**	0.01	1720	1723	15.9	25.4	9.5***	0	1703	1711
Any traditional method	56.9	60.6	3.7	0.564	1719	1723	78.4	71.2	-7.2***	0.009	1703	1711
Rhythm	43.2	30	-13.2***	0.002	1720	1723	61.7	37.7	-24***	0	1701	1711
Withdrawal	44.5	53.5	9.1	0.179	1719	1723	66.3	63	-3.3	0.259	1703	1711
Mean number of methods that MWRAs can name	5.033	5.088	0.055	0.722	1722	1723	6.456	6.629	0.173	0.132	1707	1711
Mean number of modern methods that MWRAs can name	4.157	4.253	0.095	0.307	1722	1723	5.181	5.622	.441***	0	1707	1711
Mean number of traditional methods that MWRAs can name	0.877	0.835	-0.042	0.631	1719	1723	1.284	1.007	-.276***	0	1703	1711

Notes: Asterisks indicate that the difference between the baseline and endline value is statistically significant. *= significant at 10% level, **= significant at 5% level, ***= significant at 1% level.

Indicator	Control					
	Baseline mean	Endline mean	Difference	P-Value	Baseline N	Endline N
% MWRAs who have heard about:						
Any method	99.9	99.7	-0.2	0.399	1102	1101
Any modern method	99.8	99.7	-0.1	0.689	1102	1101
Female sterilisation	95.2	93.2	-2.1	0.112	1102	1101
Male sterilisation	48.2	41.9	-6.3	0.104	1102	1101
Pill	94.1	93	-1.2	0.359	1102	1101
IUD	85.8	85.1	-0.7	0.817	1101	1101
Injectable	94.4	94.7	0.3	0.812	1102	1101
Implants	22.2	40.6	18.4***	0	1102	1101
Condoms	84.4	87.2	2.8	0.241	1101	1101
Emergency pill	16.6	31.1	14.5***	0.004	1102	1101
Any traditional method	73.5	77.9	4.4	0.109	1101	1101
Rhythm	48.6	48.5	-0.1	0.985	1100	1101
Withdrawal	62.3	68.6	6.3*	0.074	1101	1101
Mean number of methods that MWRAs can name	6.52	6.84	.321**	0.013	1102	1101
Mean number of modern methods that MWRAs can name	5.41	5.67	.258***	0.002	1102	1101
Mean number of traditional methods that MWRAs can name	1.109	1.17	0.061	0.359	1101	1101

Notes: Asterisks indicate that the difference between the baseline and endline value is statistically significant. * = significant at 10% level, ** = significant at 5% level, *** = significant at 1% level.

Table 13: Trends in knowledge of MWRAs regarding efficacy and preventative effects of contraceptive methods

Indicator	PSI-only					MSI+PSI						
	Baseline mean	Endline mean	Difference	P-Value	Baseline N	Endline N	Baseline mean	Endline mean	Difference	P-Value	Baseline N	Endline N
% MWRAs who hold correct beliefs on the efficacy of contraception in preventing...												
Unwanted pregnancy	60	54.2	-5.8	0.344	1721	1723	78.3	72.8	-5.5**	0.018	1708	1711
Sexually transmitted disease	13.9	15	1.1	0.436	1720	1723	26.9	21.6	-5.3	0.202	1702	1711
Skin disease	14.7	33.6	18.8***	0	1720	1723	21.2	52	30.8***	0	1703	1711
Diarrhoea	13.1	33.3	20.2***	0	1716	1723	18.1	51.2	33***	0	1698	1711
Heart disease	13.5	33.8	20.3***	0	1721	1723	15.5	53.6	38***	0	1699	1711
% MWRA who do not know whether contraception is effective in preventing...												
Unwanted pregnancy	36.2	34.4	-1.8	0.746	1721	1723	16.4	12.4	-4.1**	0.011	1708	1711
Sexually transmitted disease	71.8	48.2	-23.6***	0	1720	1723	53	26.2	-26.8***	0	1702	1711
Skin disease	76.4	55	-21.4***	0	1720	1723	61.9	30.7	-31.2***	0	1703	1711
Diarrhoea	80.6	57.8	-22.8***	0	1716	1723	70	34.6	-35.4***	0	1698	1711
Heart disease	80.6	58	-22.6***	0	1721	1723	76	32.1	-43.8***	0	1699	1711

Notes: Asterisks indicate that the difference between the baseline and endline value is statistically significant. * = significant at 10% level, ** = significant at 5% level, *** = significant at 1% level.

	Control					
Indicator	Baseline mean	Endline mean	Difference	P-Value	Baseline N	Endline N
% MWRAs who hold correct beliefs on the efficacy of contraception in preventing...						
Unwanted pregnancy	76	72.5	-3.5	0.433	1101	1100
Sexually transmitted disease	22.4	22.5	0.1	0.981	1099	1100
Skin disease	23.5	49.1	25.6***	0	1098	1100
Diarrhoea	22.6	44.3	21.8***	0	1095	1100
Heart disease	24.5	44.5	20***	0.003	1100	1100
% MWRAs who do not know whether contraception is effective in preventing...						
Unwanted pregnancy	17.8	12.4	-5.5	0.126	1101	1100
Sexually transmitted disease	51.7	32.1	-19.6***	0.001	1099	1100
Skin disease	64	35.2	-28.8***	0	1098	1100
Diarrhoea	67.4	41.4	-26***	0	1095	1100
Heart disease	66.5	42.2	-24.3***	0	1100	1100

Notes: Asterisks indicate that the difference between the baseline and endline value is statistically significant. *= significant at 10% level, **= significant at 5% level, ***= significant at 1% level.

5.3 Coverage

The proportion of MWRAAs who reported visiting an RH provider in the last three months dropped significantly for the PSI+MSI group (from 12% to 9%) (Table 14), with no significant change for the two other groups. However, uptake of subsidised FP did increase for all three groups from low starting values, and significantly by 12% points for the PSI-only group (albeit from a very low starting value of 3%).

In relation to exposure to BCC activities, there were significant increases reported across all groups, but especially for the PSI+MSI group and the control.

Table 14: Coverage of FP and BCC services

Indicator	PSI-only						MSI+PSI					
	Baseline mean	Endline mean	Difference	P-Value	Baseline N	Endline N	Baseline mean	Endline mean	Difference	P-Value	Baseline N	Endline N
% MWRAs who have visited an RH provider in last three months	8	6.7	-1.4	0.163	1661	1723	11.9	8.7	-3.2**	0.02	1677	1711
Of those visiting an RH provider, % who visited a ...												
Government provider	48.4	39.7	-8.6	0.228	157	150	51.9	65	13.1*	0.093	161	157
Private provider	51.6	60.3	8.6	0.228	157	150	48.1	35	-13.1*	0.093	161	157
% targeted beneficiaries using subsidised FP methods	2.8	14.9	12**	0.035	78	89	0.9	11	10.2	0.184	60	53
% targeted beneficiaries who do not know whether the last contraceptive method received was subsidised	70	72.1	2.1	0.842	78	89	86.8	69.7	-17.1	0.124	60	53
% MWRAs who have seen any communication campaign (meetings, road show) on contraception in the past one month	1.7	2.7	1	0.162	1695	1706	1.4	6.3	4.9***	0	1667	1693
% MWRAs who have seen any communication materials (billboard, poster) on contraception in the past one month	2.9	4.6	1.7	0.166	1692	1710	3.3	8.1	4.7***	0	1667	1685
% MWRAs who have seen or read anything on contraception in the media, such as TV, radio or newspaper in the past one month	13.1	16.5	3.4*	0.082	1675	1705	20.5	30.9	10.5***	0	1660	1692

Notes: Asterisks indicate that the difference between the baseline and endline value is statistically significant. *= significant at 10% level, **= significant at 5% level, ***= significant at 1% level.

Indicator	Control					
	Baseline mean	Endline mean	Difference	P-Value	Baseline N	Endline N
% MWRAs who have visited an RH provider in the last three months	9.8	13.7	3.9	0.285	1070	1100
Of those visiting an RH provider, % who visited a ...						
Government provider	52.5	60.5	8.1	0.473	98	116
Private provider	47.5	39.5	-8.1	0.473	98	116
% targeted beneficiaries using subsidised FP methods	10.4	21.3	10.8	0.318	44	52
% targeted beneficiaries who do not know whether the last contraceptive method received was subsidised	50.3	65.3	15	0.467	44	52
% MWRAs who have seen any communication campaign (meetings, road show) on contraception in the past one month	0.7	6.6	5.9***	0.001	1094	1092
% MWRAs who have seen any communication materials (billboard, poster) on contraception in the past one month	9.4	13.2	3.8**	0.048	1089	1085
% MWRAs who have seen or read anything on contraception in the media, such as TV, radio or newspaper in the past one month.	22.6	30.1	7.5**	0.03	1083	1069

Notes: Asterisks indicate that the difference between the baseline and endline value is statistically significant. * = significant at 10% level, ** = significant at 5% level, *** = significant at 1% level.

5.4 Household costs

In relation to household expenditure on contraception, there was a significant reduction in the proportion paying for the last method they had used in the control group, but a significant increase in the PSI-only group (

Table 15). However, sample sizes for this set of indicators are relatively small, so these figures should be treated with some caution.

Table 15: Household FP costs

	PSI-only						MSI+PSI					
	Baseline mean	Endline mean	Difference	P-value	Baseline N	Endline N	Baseline mean	Endline mean	Difference	P-value	Baseline N	Endline N
% MWRAs spending any money to obtain contraceptive method last time	53.2	69.5	16.3**	0.041	136	141	43.8	34.5	-9.3	0.278	133	142
Mean spend on obtaining contraceptive method last time (PKR)	254.03	121.282	-132.75**	0.031	74	61	146.892	152.896	6.004	0.918	54	40

Notes: Asterisks indicate that the difference between the baseline and endline value is statistically significant. *= significant at 10% level, **= significant at 5% level, ***= significant at 1% level.

	Control					
	Baseline mean	Endline mean	Difference	P-value	Baseline	Endline N
% MWRAs spending any money to obtain contraceptive method last time	50.8	24.6	-26.2*	0.063	84	109
Mean spend on obtaining contraceptive method last time (PKR)	68.9	578.96	510.065	0.242	37	25

Notes: Asterisks indicate that the difference between the baseline and endline value is statistically significant. *= significant at 10% level, **= significant at 5% level, ***= significant at 1% level.

5.5 Acceptability and reduction of barriers

Many women prefer to see a female provider and the gap between these preferences and their actual experience is one indicator of the extent to which services are acceptable to MWRAs.

For the PSI-only group, the gap narrowed over the project period, though largely through a reduction in preferences for a female provider (Table 16). For the control group, the gap widened, largely through a decrease in those who saw a female provider at their last visit. Again, sample sizes for this class of indicators are too small to make conclusive statements about what has changed.

Among MWRAs not using contraception, the reasons for non-use are found to be relatively stable over time. There are significant increases in women reporting that they cannot fall pregnant or are menopausal, and a decrease in women reporting that they want more children. These effects may be due to the time interval between baseline and endline.

Table 16: Acceptability

	PSI-only						MSI+PSI					
	Baseline mean	Endline mean	Difference	P-value	Baseline N	Endline N	Baseline mean	Endline mean	Difference	P-value	Baseline N	Endline N
% of MWRAs who saw a female provider at the last visit	41.9	42.5	0.6	0.943	150	142	34.2	25.5	-8.7	0.24	157	157
% MWRAs who would have preferred to have a female provider	73.5	52.9	-20.5**	0.024	77	144	62.4	69.3	6.9	0.473	169	151
Among MWRAs not currently using contraceptives, the proportion who reported the main reason for not using as being...												
Wants to get pregnant	31.4	26.3	-5.1	0.34	960	973	34.8	27.8	-7***	0.006	854	886
Partner not willing	4.3	9	4.7	0.267	960	973	4.3	3.9	-0.4	0.812	854	886
Husband away	6.6	6.6	0	0.973	960	973	11.8	11.4	-0.4	0.85	854	886
Social restriction	3.1	2.4	-0.7	0.807	960	973	0.1	0.9	0.8*	0.08	854	886
Religious restriction	3	5	2	0.116	960	973	2.3	2.3	-0.1	0.949	854	886
Cannot afford	2	1	0.7*	0.06	960	973	0.9	0.5	-0.4	0.346	854	886
Contraceptive method not available	1.5	1.4	-0.1	0.946	960	973	0.6	0.6	-0.1	0.891	854	886
Health concerns/ problems	8.7	18.3	9.5*	0.072	960	973	13.6	15.8	2.1	0.398	854	886
Cannot fall pregnant	2.4	3.9	1.5	0.166	960	973	1.8	8.1	6.3***	0	854	886
Menopausal	2.4	7.3	4.9***	0	960	973	4.4	9.5	5.1***	0	854	886
Currently breast feeding	22.8	15.1	-7.7	0.109	960	973	15.6	15.9	0.3	0.912	854	886

	Control					
	Baseline mean	Endline mean	Difference	P-value	Baseline N	Endline N
% of MWRAs who saw a female provider at the last visit	53.2	26.9	-26.3*	0.058	91	111
% MWRAs who would have preferred to have a female provider	76.3	71.9	-4.4	0.657	93	113
Among MWRAs not currently using contraceptives, the proportion who reported the main reason for not using as being...						
Wants to get pregnant	35.1	19.1	-16***	0	540	553
Partner not willing	4.1	7.2	3.1	0.479	540	553
Husband away	11.3	10.6	-0.6	0.891	540	553
Social restriction	0.1	0.4	0.3	0.27	540	553
Religious restriction	0.5	1.9	1.4*	0.07	540	553
Cannot afford	0.5	1.6	1.1	0.287	540	553
Contraceptive method not available	0	0	0		540	553
Health concerns/ problems	12.5	16	3.5	0.322	540	553
Cannot fall pregnant	1.7	10.2	8.4***	0.003	540	553
Menopausal	5.9	9	3	0.124	540	553
Currently breast feeding	16	13.8	-2.1	0.594	540	553

6 Equity

In the tables below we examine whether there are differential trends in key outcomes for households who are classified as poor, compared with those who are not. 'Poor' households are defined as those whose NISP poverty score falls below a certain cut-off, as estimated using the 2008–2009 PLSM survey. To the extent that poverty status may have changed as a result of exposure to the interventions, in the tables below the results are disaggregated according to the estimated poverty status at baseline to ensure full comparability across the groups.

Disaggregating changes to utilisation by poor/non-poor households, a pro-poor pattern of change was found for the PSI-only group in relation to ever using contraception (perhaps partly linked to lower starting levels for the poorer households – 26%, as against 45% for non-poor) (

Table 17). However, the control group also showed a pro-poor increase in ever using contraception (i.e. a bigger increase in reportedly ever using contraception for poor than for non-poor households), even though the starting level here was higher (56%, as against 61% for the non-poor) (

Table 18).

It appears as though increases in ever using modern contraceptive methods over time accrue to non-poor households faster than they do to poor households. Amongst non-poor households, the increase in ever using modern methods was large and always significant, while the corresponding point estimate for poor households was smaller and only significant in the PSI-only treatment group. Conversely, poor households appear to be more inclined to use traditional methods. In the PSI-only group, the increase in ever using traditional methods over time was 14.4% for poor households, compared with 9.5% for non-poor households.

Table 17: Descriptive evidence of trends in equity of utilisation among treatment groups

Indicator	PSI-only					MSI+PSI				
	Baseline mean	Endline mean	Difference	P-Value	N	Baseline mean	Endline mean	Difference	P-Value	N
% MWRAs ever used contraceptive method	38	50.1	12.2***	0	1723	60.5	68.7	8.2***	0	1711
<i>Poor households</i>	26.3	39.5	13.2***	0.008	434	55.7	62.2	6.5**	0.047	379
<i>Non-poor households</i>	44.9	56.3	11.4***	0	1277	61.8	70.3	8.5***	0	1324
% MWRAs ever used modern method	32.4	38.7	6.3**	0.02	1723	47.6	58.2	10.6***	0	1711
<i>Poor households</i>	24.5	27.4	3	0.553	434	43.7	50.6	6.9**	0.04	379
<i>Non-poor households</i>	37.1	45.2	8.1***	0	1277	48.6	60	11.4***	0	1324
% MWRAs ever used traditional method	15.2	26.5	11.3***	0	1723	32.4	35.3	2.9	0.368	1711
<i>Poor households</i>	5.9	20.3	14.4***	0	434	30	27.6	-2.4	0.648	379
<i>Non-poor households</i>	20.7	30.1	9.5***	0	1277	33.2	37.4	4.2	0.189	1324
% MWRAs currently using contraceptives	26	29.1	3	0.104	1568	43.8	42	-1.8	0.442	1552
<i>Poor households</i>	16.3	19.1	2.8	0.525	391	40.4	39.5	-0.9	0.827	344
<i>Non-poor households</i>	31.8	34.6	2.8	0.195	1169	44.9	42.5	-2.4	0.331	1200
Unmet need for contraception	39.8	38.2	-1.6	0.668	1724	27.5	28	0.5	0.807	1711
<i>Poor households</i>	49.3	48.9	-0.5	0.924	434	35	32.5	-2.5	0.617	379
<i>Non-poor households</i>	34.1	32	-2.1	0.593	1278	25.3	26.9	1.6	0.456	1324

* = p value < 0.1, ** = p value < 0.05, *** = p value < 0.001

Table 18: Descriptive evidence of trends in equity of utilisation among the control group

Indicator	Control				
	Baseline mean	Endline mean	Difference	P-Value	N
% MWRAs ever used contraceptive method	60.4	72.4	12***	0	1101
<i>Poor households</i>	57.3	76.5	19.2**	0.026	183
<i>Non-poor households</i>	60.9	71.8	10.9***	0	908
% MWRAs ever used modern method	47.5	58.7	11.2***	0	1101
<i>Poor households</i>	47.3	58	10.7	0.191	183
<i>Non-poor households</i>	47.5	58.8	11.3***	0	908
% MWRAs ever used traditional method	36.8	44.2	7.3**	0.036	1101
<i>Poor households</i>	23.3	44.8	21.6**	0.017	183
<i>Non-poor households</i>	39.3	44.2	4.9	0.215	908
% MWRAs currently using contraceptives	45.1	48.9	3.8	0.385	1011
<i>Poor households</i>	49.7	44	-5.7	0.682	167
<i>Non-poor households</i>	44.5	49.7	5.2	0.237	834
Unmet need for contraception	26.4	28.1	01.7	0.559	1102
<i>Poor households</i>	33.4	39.4	5.9	0.627	183
<i>Non-poor households</i>	24.9	26.1	1.2	0.649	909

We also investigated whether changing underlying awareness was equitable (Tables 19 and 20). Gains to awareness of modern methods accrue quite evenly to poor and non-poor MWRA in the PSI+MSI group. However, in the control group the increases are seen in the non-poor group only, which may indicate some success in raising awareness among the poor on the part of PSI and MSI in combination.

Table 19: Descriptive evidence of trends in awareness of contraceptive methods, by poverty status, among the PSI-only and MSI+PSI groups

Indicator	PSI-only					MSI+PSI				
	Baseline mean	Endline mean	Difference	P-Value	N	Baseline mean	Endline mean	Difference	P-Value	N
% MWRAs who have heard about:										
Any method	89.1	91.8	2.7	0.403	1723	99.2	99.4	0.3	0.32	1711
<i>Poor</i>	75.9	82	6.1	0.562	434	98.3	99.5	1.1	0.197	379
<i>Not Poor</i>	96.6	97.5	0.8	0.646	1277	99.4	99.4	0	0.904	1324
Any modern method	88.4	89	0.6	0.807	1723	99	99.4	0.4	0.199	1711
<i>Poor</i>	73.9	78.3	4.4	0.52	434	98	99.5	1.5	0.191	379
<i>Not Poor</i>	96.6	95.1	-1.5	0.361	1277	99.2	99.4	0.2	0.599	1324
Any traditional method	56.9	60.6	3.7	0.564	1723	78.4	71.2	-7.2***	0.009	1711
<i>Poor</i>	40.3	51.5	11.2	0.495	434	71.3	65.8	-5.4	0.236	379
<i>Not Poor</i>	66.3	65.8	-0.5	0.922	1277	80.4	72.9	-7.5***	0.007	1324
Mean number of methods that MWRAs can name	5.033	5.088	0.055	0.722	1723	6.456	6.629	0.173	0.132	1711
<i>Poor</i>	3.426	3.602	0.176	0.639	434	5.945	6.236	.291*	0.098	379
<i>Not Poor</i>	5.945	5.945	0	1	1277	6.594	6.742	0.148	0.229	1324
Mean number of modern methods that MWRAs can name	4.157	4.253	0.095	0.307	1723	5.181	5.622	.441***	0	1711
<i>Poor</i>	2.807	2.973	0.167	0.359	434	4.861	5.352	.491***	0.002	379
<i>Not Poor</i>	4.922	4.99	0.068	0.523	1277	5.268	5.699	.431***	0	1324
Mean number of traditional methods that MWRAs can name	0.877	0.835	-0.042	0.631	1723	1.284	1.007	-.276***	0	1711
<i>Poor</i>	0.62	0.628	0.009	0.968	434	1.084	0.884	-.2***	0.005	379
<i>Not Poor</i>	1.025	0.955	-0.07	0.309	1277	1.338	1.043	-.295***	0	1324

Notes: Asterisks indicate that the difference between the baseline and endline value is statistically significant. * = significant at 10% level, ** = significant at 5% level.

Table 20: Descriptive evidence of trends in awareness of contraceptive methods, by poverty status, among the control group

	Control					
Indicator	Baseline mean	Endline mean	Difference	P-Value	Baseline N	N
% MWRAs who have heard about:						
Any method	99.9	99.7	-0.2	0.399	1102	1101
Poor	100	99	-1	0.268	183	183
Not poor	99.8	99.8	0	0.904	909	908
Any modern method	99.8	99.7	-1	0.689	1102	1101
Poor	100	99	-1	0.268	183	183
Not poor	99.7	99.8	0.1	0.667	909	908
Any traditional method	73.5	77.9	4.4	0.109	1101	1101
Poor	68.7	77.2	8.5	0.214	182	183
Not poor	74.3	78.2	3.9	0.189	909	908
Mean number of methods that MWRAs can name	6.52	6.84	.321**	0.013	1102	1101
Poor	6.471	6.236	-0.236	0.499	183	183
Not poor	6.524	6.949	.425***	0	909	908
Mean number of modern methods that MWRAs can name	5.41	5.67	.258***	0.002	1102	1101
Poor	5.404	5.18	-0.224	0.284	183	183
Not poor	5.409	5.756	.347***	0	909	908
Mean number of traditional methods that MWRAs can name	1.109	1.17	0.061	0.359	1101	1101
Poor	1.074	1.056	-0.018	0.925	182	183
Not poor	1.114	1.193	0.079	0.24	909	908

Notes: Asterisks indicate that the difference between the baseline and endline value is statistically significant. *= significant at 10% level, **= significant at 5% level.

In relation to access, we found a substantial and significant reduction in mean spend in transport to reach the nearest RH provider for the poor in the MSI+PSI group, and for the non-poor in the control. There was also a significant reduction in time taken to reach this provider in the PSI-only group for the non-poor (Table 21).

Table 21: Average travel time and cost to reach nearest RH provider, by poverty status

	PSI-only					MSI+PSI				
	Baseline mean	Endline mean	Difference	P-value	N	Baseline mean	Endline mean	Difference	P-value	N
Mean time to reach nearest RH provider (in minutes)	27.5	30.46	2.96	0.47	1369	19.59	18.19	-1.40	0.32	1396
<i>Poor</i>	39.06	49.53	10.47	0.153	318	28.3	23.941	-4.359	0.112	291
<i>Not poor</i>	24.107	22.57	-1.533	0.334	1045	17.367	16.811	-0.556	0.685	1099
Of those visiting RH provider, % that spend any money on transport	52.5	55.4	0.03	0.46	1369	28.3	31.4	0.03	0.44	1396
<i>Poor</i>	66.7	69.6	2.9	0.586	318	36.9	35.6	-1.3	0.808	291
<i>Not poor</i>	48.2	49.3	1.1	0.751	1045	26.2	30.4	4.2	0.308	1099
Mean spend on transport to reach nearest RH provider (PKR)	96.0	159.6	63.61	0.43	700	127.7	70.6	-57.1	0.12	495
<i>Poor</i>	88.903	244.6	155.694	0.26	188	172.323	60.068	112.3**	0.038	111
<i>Not poor</i>	99.129	111.4	12.249	0.772	507	111.259	73.558	-37.701	0.213	383
Among MWRAs obtaining contraceptive methods or advice in the past three months, time in minutes taken to reach provider	30.294	21.902	-8.392*	0.076	136	17.268	14.978	-2.29	0.474	142
<i>Poor</i>	44.033	42.13	-1.906	0.876	24	22.18	29.25	7.069	0.429	32
<i>Not poor</i>	27.511	18.52	-8.996*	0.081	111	16.33	12.2	-4.133	0.209	108

Notes: Asterisks indicate that the difference between the baseline and endline value is statistically significant. *= significant at 10% level, **= significant at 5% level, ***=significant at 1% level. Providers include health facilities and other outlets (for the SM programme).

	Control				
	Baseline mean	Endline mean	Difference	P-value	Endline N
Mean time to reach nearest RH provider (in minutes)	18.2	15.9	-2.28	0.25	937
<i>Poor</i>	26.101	25.583	-0.518	0.871	156
<i>Not poor</i>	16.85	14.359	-2.491	0.206	772
Of those visiting RH provider, % that spend any money on transport	26.7	21.6	-0.05	0.22	937
<i>Poor</i>	35.2	27.6	-7.6	0.407	156
<i>Not poor</i>	25.3	20.6	-4.7	0.32	772
Mean spend on transport to each nearest RH provider (PKR)	140.4	75.8	-64.67**	0.05	262
<i>Poor</i>	188.114	124.951	-63.164	0.469	52
<i>Not poor</i>	128.763	64.955	-63.81**	0.027	206
Among MWRA obtaining contraceptive methods or advice in the past three months, time in minutes taken to reach provider	18.375	17.837	-0.538	0.878	91
<i>Poor</i>	27.854	25.786	-2.068	0.736	15
<i>Not poor</i>	17.296	16.627	-0.669	0.852	74

Notes: Asterisks indicate that the difference between the baseline and endline value is statistically significant. *= significant at 10% level, **= significant at 5% level, ***= significant at 1% level. Providers include health facilities and other outlets (for the SM programme).

7 Quality of RH provision

The results for quality of care come largely from the facility and outlet surveys, although some client responses from the household survey are also analysed at the end of this section.

7.1 Health facilities

7.1.1 Infrastructure

Comparing the functionality of basic infrastructure in facilities in the combined intervention group with those in the control, most facilities scored highly for power, running water and latrines. There was a significant improvement in functioning fridges in the control group. For the PSI+MSI areas, only 60% were found with functioning fridges. Both groups saws deteriorating functioning of telephones (

Table 22).

Table 22: Health facility infrastructure

	Control						MSI+PSI					
	Baseline mean	Baseline N	Endline mean	Endline N	Difference	P-value	Baseline mean	Baseline N	Endline mean	Endline N	Difference	P-value
Proportion of health facilities which had the following in working condition on the day of assessment:												
Power source	98.9	94	1	78	1.1	0.319	99.2	125	100	110	0.8	0.318
Running water	97.9	94	94.9	78	-3	0.306	95.2	125	93.6	110	-1.6	0.606
Toilet or latrine	93.6	94	97.4	78	3.8	0.221	98.4	125	97.3	110	-1.1	0.559
Fridge	57.4	94	71.8	78	14.3**	0.049	66.4	125	60	110	-6.4	0.313
Freezer	20.2	94	20.5	78	0.3	0.961	3.2	125	8.2	110	5	0.105
Telephone	96.8	94	76.9	78	-19.9***	0	98.4	125	68.2	110	-30.2***	0

7.1.2 Standards, hygiene and stock outs

There was a significant reduction in the proportion of facilities displaying standards of practice between baseline and endline for the control group, while this increased significantly in the intervention (

Table 23). However, intervention groups showed a significant reduction in the proportion cleaned daily with a disinfectant. There were significant reductions in stock outs for all contraceptive types in control facilities and for condoms and injectables (but not oral contraceptives) in intervention area facilities.

Table 23: Facility standards of care and stock out

	Control						MSI+PSI					
	Baseline mean	Baseline N	Endline mean	Endline N	Difference	P-value	Baseline mean	Baseline N	Endline mean	Endline N	Difference	P-value
Proportion of health facilities with standards of practice posters/ messages displayed	59.6	94	33.3	78	-26.2***	0.001	40	125	56.4	110	16.4**	0.015
Proportion of health facilities cleaned with disinfectant daily	81.9	94	76.9	78	-5	0.425	76.4	125	60.9	110	-15.5**	0.013
Proportion of health facilities with a stock out in the last 12 months...												
Oral contraceptives	46.8	94	9	78	-37.8***	0	13.6	125	11.8	110	-1.8	0.683
Condoms	50	94	3.8	78	-46.2***	0	24	125	9.1	110	-14.9***	0.002
Injectables	46.8	94	9	78	-37.8***	0	16	125	7.3	110	-8.7**	0.035

Notes: A stock out refers to a 'complete' stock out – i.e. if facility reports stock out of brand A condoms and brand B is available, this is not seen as a stock out. In line with the baseline analysis, facilities reporting 'don't know' or 'not applicable' have been counted as a stock out.

There were significant improvements in all types of staffing across the intervention facilities (Table 24)

Table 24: Health facility staffing

	Control						MSI+PSI					
	Baseline mean	Baseline N	Endline mean	Endline N	Difference	P-value	Baseline mean	Baseline N	Endline mean	Endline N	Difference	P-value
Proportion of health facilities that are lacking in the following staff...												
Physician / medical officer	2.1	94	1.3	78	-0.8	0.668	11.6	112	0.9	110	-10.7***	0.001
Clinical officer	1.1	94	0	78	-1.1	0.319	7.1	112	0	110	-7.1***	0.004
Nurse	7.4	94	5.1	78	-2.3	0.532	13.4	112	3.6	110	-9.8***	0.009
Health assistant	1.1	94	0	78	-1.1	0.319	10.7	112	4.5	110	-6.2*	0.084
Social worker	6.4	94	2.6	78	-3.8	0.221	12.5	112	4.5	110	-8**	0.034
Proportion of health facilities in which all FP staff have received specific FP training in last one year	30.9	94	28.2	78	-2.6	0.707	35.7	112	48.2	110	12.5*	0.057

7.1.3 Fees

In relation to fees charged by facilities, sample sizes restrict our analysis but there does appear to be a significant reduction in general FP consultation fees in intervention area facilities (Table 25).

Table 25: Health facility fees

	Control						MSI+PSI					
	Baseline mean	Baseline N	Endline mean	Endline N	Difference	P-value	Baseline mean	Baseline N	Endline mean	Endline N	Difference	P-value
Mean fee charge per consultation... (PKR)												
General FP consultation	83.226	93	84.231	78	1.005	0.957	34.6	125	18.182	110	-16.42**	0.028
IUD insertion	310.494	81	359.24	62	48.748	0.301	291.09	119	323.112	98	32.02	0.294
Implant	383.33	6	1483.33	6	1100.0*	0.081	283.33	3	1637.5	4	1354.17	0.1
Sterilisation	1187.26	23	2437.5	8	1250.23 ₉	0.171	2166.67	15	3928.57	7	1761.91	0.123

7.2 Outlets

7.2.1 Range of FP products

Outlets in PSI-only areas showed a small but significant reduction in FP products supplied over the period, including for injectables. In PSI+MSI areas, there was no significant overall change but there was an increase in oral contraceptives supplied (from 29 to 39%) and a small decrease in condoms. In control areas, there were significant decreases overall and for condoms (Table 26).

Table 26: FP products sold by outlets

	PSI-only						MSI+PSI					
	Baseline mean	Baseline N	Endline mean	Endline N	Difference	P-value	Baseline mean	Baseline N	Endline mean	Endline N	Difference	P-value
Proportion of retailers that sell...												
Any FP product	0.991	225	0.934	183	-.057***	0.004	0.95	221	0.929	170	-0.021	0.397
Oral contraceptives	0.484	225	0.481	183	-0.004	0.943	0.294	221	0.394	170	.1**	0.04
Emergency contraceptives	0.382	225	0.372	183	-0.011	0.826	0.344	221	0.324	170	-0.02	0.673
Condoms	0.804	225	0.721	183	-.083*	0.051	0.887	222	0.818	170	-.07*	0.057
Injectables	0.511	225	0.399	183	-.112**	0.023	0.326	221	0.341	170	0.015	0.75
Pregnancy test kit	0.48	225	0.399	183	-0.081	0.101	0.321	221	0.359	170	0.038	0.439

	Control					
	Baseline mean	Baseline N	Endline mean	Endline N	Difference	P-value
Proportion of retailers that sell...						
Any FP product	1	201	85.5	165	-14.5***	0
Oral contraceptives	30.8	201	29.1	165	-1.8	0.716
Emergency contraceptives	26.4	201	26.7	165	0.3	0.949
Condoms	99	201	81.8	165	-17.2***	0
Injectables	28.4	201	23.6	165	-4.7	0.305
Pregnancy test kit	30.3	201	24.8	165	-5.5	0.241

7.2.2 Costs of FP products

Prices at outlets varied in their levels and trends, but the most striking change was an increase across all groups for injectables (Table 27).

Table 27: Prices of FP products

	PSI-only						MSI+PSI					
	Baseline mean	Baseline N	Endline mean	Endline N	Difference	P-value	Baseline mean	Baseline N	Endline mean	Endline N	Difference	P-value
Mean price for FP products												
Oral contraceptives	23.93	107	23.5	86	-0.43	0.799	28.1	63	20.7	65	-7.3	0.136
Emergency contraceptive	29.26	84	17.3	66	-11.9*	0.085	19.1	76	18.2	53	-0.91	0.663
Condoms	6.9	180	6.6	129	-0.38	0.501	6.6	195	6.52	138	-0.08	0.906
Depovera injectables	64.9	89	80.87	66	16.0***	0	61.7	59	86.97	50	25.2***	0
Norijest injectables	43.21	53	73.2	40	30.0***	0	40.2	41	63.45	28	23.18***	0
Pregnancy test kit	15.3	104	13.6	73	-1.68	0.158	14.4	70	15.5	60	1.1	0.499
	Control											
	Baseline mean		Baseline N		Endline mean		Endline N		Difference		P-value	
Mean price for FP products												
Oral contraceptives	31.4		62		21.01		48		-10.34***		0.004	
Emergency contraceptive	15.6		53		15.88		44		0.262		0.807	
Condoms	6.55		199		8.50		133		1.95***		0.007	
Depovera injectables	61.7		44		84.1		30		22.41***		0	
Norijest injectables	40.7		35		69.8		17		29.2***		0	
Pregnancy test kit	9.7		60		11.8		39		2.0**		0.043	

7.3 Quality of service at the source of last contraceptives

No significant change was registered in waiting times for MWRAs in the intervention arms, while a significant increase did occur in the control areas (though the level remained lower than in PSI-only sites, for example). There was no significant change in any arm for the proportion told about possible side effects, which remains relatively low (27% PSI-only, 38% MSI+PSI, 34% control) (Table 28).

There was a substantial increase in those told about alternative methods in MSI+PSI areas (from 27% to 46%), which corresponded to a similar increase in the proportion of MWRAs rating the service as good in these areas. There was also a significant but less large increase in 'good' ratings in PSI-only areas. Control areas saw significant increases in 'good' ratings and highly significant increases in 'excellent'.

Table 28: Quality of service at the source of last contraceptives obtained

	PSI-only						MSI+PSI					
	Baseline mean	Baseline N	Endline mean	Endline N	Difference	P-value	Baseline mean	Baseline N	Endline mean	Endline N	Difference	P-value
Average waiting time before consulting the provider	19.696	112	21.29	128	1.597	0.849	5.69	123	7.447	138	1.757	0.443
Proportion of MWRAs told about side effects	21.2	163	26.6	151	0.05.4	0.45	31.2	173	38.1	160	6.9	0.434
Among MWRAs told about side effects, proportion told what to do in case of side effects	72.3	29	73.4	53	0.01.1	0.945	62	34	72.4	52	10.4	0.616
Proportion of MWRAs told about alternative methods	27.1	159	30.2	151	0.03.1	0.717	26.7	173	46.3	160	19.6**	0.03
Proportion of MWRAs who rated the service received as¹:												
Excellent	5.2	191	8.4	151	3.2	0.395	14.4	212	13.6	160	-0.8	0.857
Good	45.9	191	64.3	151	18.4*	0.052	43.1	212	76.2	160	33.1***	0
Bad	2.6	191	0.1	151	-2.4	0.226	0.3	212	2.9	160	2.6	0.259
Poor	0.9	191	0.6	151	-0.3	0.753	1.8	212	0	160	-1.8	0.179

Notes: (1) Does not sum to 1 as some women reported no opinion.

	Control					
	Baseline mean	Baseline N	Endline mean	Endline N	Difference	P-value
Average waiting time before consulting the provider (minutes)	4.982	65	15.339	83	10.4**	0.024
Proportion of MWRAs told about side effects	36.4	104	34.4	116	-2	0.797
Among MWRAs told about side effects, proportion told what to do in case of side effects	66.1	25	90.4	45	24.4	0.14
Proportion MWRAs told about alternative methods	34.7	103	45.8	116	11.1	0.174
Proportion MWRAs who rated the service received as ¹ :						
Excellent	0	134	21.9	116	21.9***	0.004
Good	41.6	134	57.6	116	16*	0.064
Bad	2	134	0.3	116	-1.7	0.193
Poor	3	134	3.9	116	3.7	0.342

Notes: (1) Does not sum to 1 as some women reported no opinion.

8 Impact estimates

As a result of the largely convergent trends noted in the descriptive analysis provided above, the impact estimates, which used PSM to remove hidden selection bias, found no significant impact for the two intervention groups, compared to the control, for increased access to RH (reduced time and payments to reach providers) (Table 29) and contraceptive uptake (Table 30).

Table 29: Impact estimates for contraceptive access

	PSI-only vs control				PSI+MSI vs control			
	Average treatment effect	Standard error	T statistic	N	Average treatment effect	Standard error	T statistic	N
Mean time to reach nearest RH provider (in minutes)	1.12	1.427	0.782	1610	-2.22	1.31	-1.692	1670
Of those visiting RH provider, % that spend any money on transport	-0.02	-0.481	-0.481	1600	-0.01	0.031	-0.424	1649

Table 30: Impact estimates for contraceptive utilisation

	PSI-only vs control				PSI+MSI vs control			
	Average treatment effect	Standard error	T statistic	N	Average treatment effect	Standard error	T statistic	N
% MWRAs ever used contraceptive method	-0.02	0.024	-1	2532	-0.03	0.022	-1.133	2524
% MWRAs ever used modern method	-0.01	0.022	-0.663	2534	0	0.021	-0.119	2522
% MWRAs ever used traditional method	-0.04	0.029	-1.285	2545	-0.05	0.028	-1.72	2528
% MWRAs currently using contraceptives ⁹	0.02	0.026	0.794	2354	-0.01	0.025	-0.306	2341
% MWRAs currently using modern methods of contraception	0.02	0.022	0.882	2574	-0.01	0.021	-0.52	2554
% MWRAs currently using traditional methods of contraception	0.02	0.018	0.887	2574	0.03	0.017	1.619	2554

* = | t-statistic | > 1.96, ** = | t-statistic | > 2.58

There was some evidence of increased exposure to communication materials in both intervention groups (

⁹ The results for this indicator may be interpreted as being able to reveal whether or not the interventions have led to an increase in new users of contraception in the treatment groups, where a new user is defined as a woman who was not practising contraception at the time of the baseline survey but is practising it by the time of the endline survey. The reason why the findings on contraceptive use may be interpreted in this way is due to the fact that our analysis methodology includes a differences in differences element. This means that we are comparing a difference between baseline and endline in each indicator for women in one of the treatment groups with the corresponding difference or women in the control group. If the impact on current use of contraception is found to be positive, the interpretation would be that there are a greater number of new users in the treatment group than in the control group at the time of the follow-up survey, which is attributable to the interventions.

Table 31). Impact estimates also found a significant increase in the mean number of modern methods named by MWRAs in the PSI+MSI group, compared to the control (Table 32)

Table 31: Impact estimates for awareness of communication campaigns on FP

	PSI-only vs control			PSI+MSI vs control		
	Average treatment effect	Standard error	T statistic	Average treatment effect	Standard error	T statistic
% MWRAs who have seen any communication materials on contraception in the last month	0.09**	0.028	3.166	0.13**	0.027	4.883
% MWRAs who have seen any communication campaigns on contraception in the past month	-0.02	0.013	-1.844	-0.01	0.012	-0.95
% MWRAs who have seen anything on contraception in the media in the past month	0	0.018	0.25	0.02	0.017	0.989

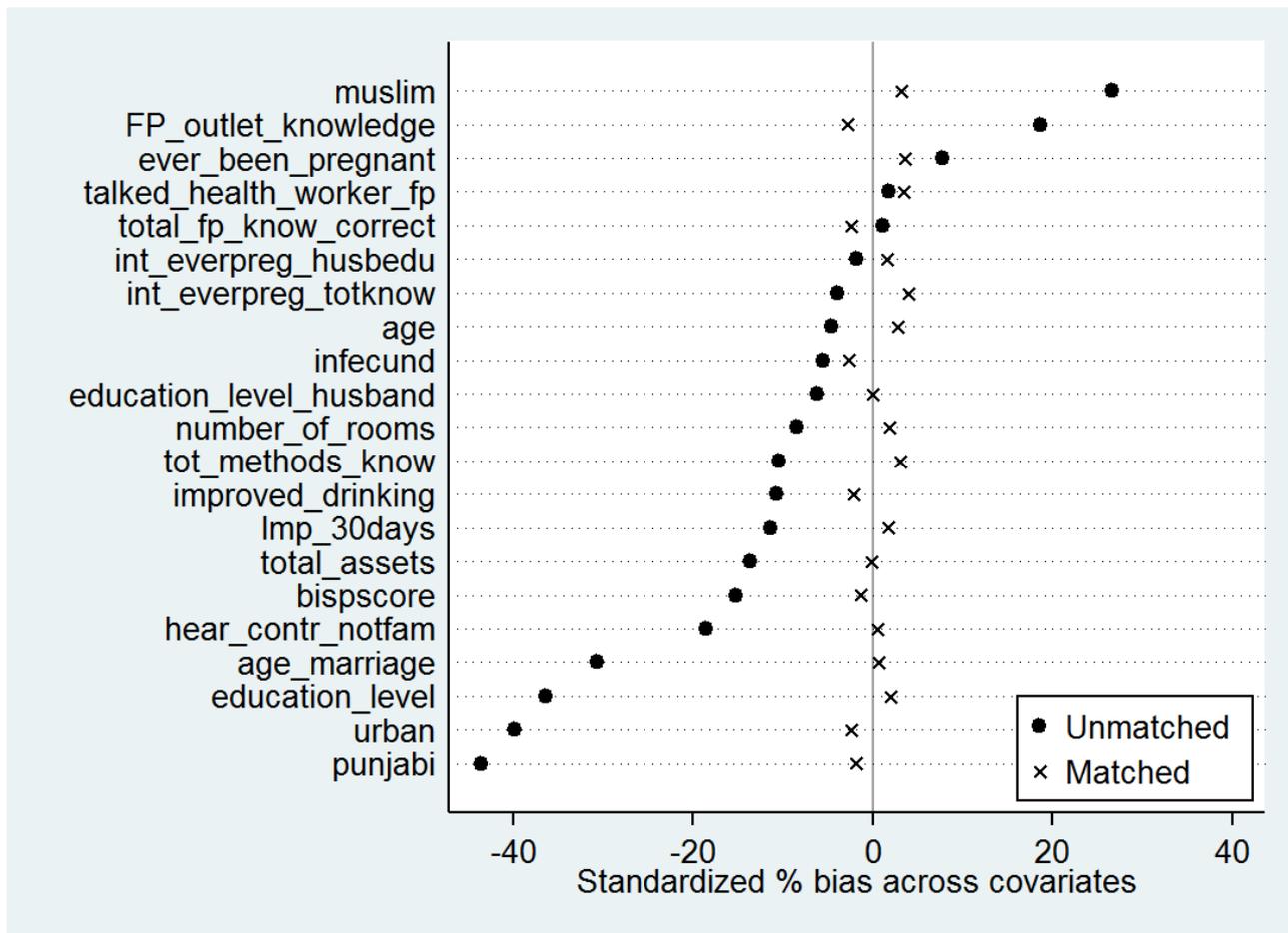
* = | t-statistic | > 1.96, ** = | t-statistic | > 2.58

Table 32: Impact estimates for contraception awareness

	PSI-only vs control			PSI+MSI vs control		
	Average treatment effect	Standard error	T statistic	Average treatment effect	Standard error	T statistic
Mean number of methods that MWRAs can name	-0.17	0.122	-1.381	0.01	0.114	0.051
Mean number of modern methods that MWRAs can name	-0.01	0.095	-0.106	0.23*	0.088	2.562
% MWRAs who cannot say which method is most effective	-0.01	0.028	-0.316	0.01	0.026	0.504

The PSM model specification used to generate impact estimates is described in Annex A, and full results associated with the main outcome (current use of contraceptives) are given in Annex B. The matching algorithm selected performed very well in terms of achieving statistical balance between treatment and control groups after matching, according to their overall characteristics. An illustration of the improvement in standardised bias associated with each variable after matching is given in Figure 6 below. Prior to matching, the figure shows that some variables were associated with a high degree of bias; however, after matching is performed the balance across these variables is shown to be markedly reduced.

Figure 6: Covariate balance for the main model before and after matching



Note: This figure shows standardised bias associated with all the covariates included in the PSM model. It corresponds to the analysis performed to analyse current contraceptive use, in the PSI+MSI group vs the control group. The matching algorithm used is kernel matching (Gaussian), with trimming set at 5 and a bandwidth parameter of 0.03.

9 Discussion and conclusion

The overall finding of a lack of a significant effect of the interventions on access to and utilisation of contraception, using PSM to eliminate hidden biases between intervention and control samples, is disappointing. There is evidence that the intervention has improved awareness on the part of MWRAs, compared to control MWRAs, and in some of the descriptive analysis we see a relatively higher performance in some of the domains that should have a positive effect on utilisation – for example, in the combined MSI+PSI areas we found increased exposure to BCC compared to controls, pro-poor gains in awareness of FP and reduced transport costs to reach RH providers, more alternatives being offered to women, and (for both PSI-only and MSI+PSI) increased staffing of facilities. However, on most indicators, performance (change since baseline) was either comparable to or worse than in the control areas. Current use of FP across all groups and unmet need were relatively constant, as was access.

This evaluation was designed to establish impact, not to examine the process of programme implementation, and the absence of other accompanying research tools such as qualitative investigations makes it hard to be conclusive about the reasons behind our results. One set of explanatory factors may lie in the high levels of general mobilisation, which explain upward trends in FP indicators across all areas, including in the controls. Data from the three DHS surveys conducted in Pakistan over the past two decades show an increase of 17 percentage points in the use of modern contraceptive methods, from 9% in 1990–91 to 26% in 2012–13. However, while this ‘rising tide’ may hold true for some indicators that we examine, the absence of clear gains in the core access and utilisation indicators suggest this is not the main explanation.

As described in the context section, many public programmes and NGOs are active across the country. In a ‘real world’ setting we are not able to control for all other investments in control areas, so confounders are possible, but preliminary investigations suggest that these programmes should be broadly equivalent across the areas, and therefore not likely to fully explain the lack of significant ‘value added’ from the interventions.

It is also possible that ‘contamination’ from the intervention has spilled over into control areas, bringing gains that mask project benefits. This is not very plausible, however, given the nature of the interventions, which (with the exception of the media messages) require a presence on the ground.

Another possibility to consider is that programme activities were carried out in restricted areas, and their effectiveness is diluted in sampling at the tehsil level. The programme was designed to be effective at a sub-district level, which is why the evaluation sampling was constructed accordingly, but it is possible that the activities were not tehsil-wide – an explanation that can be discussed with the implementers.

A more fundamental question is whether the programme was designed to address the barriers that were most significant to women, and whether it addressed them effectively. When women were asked in the baseline survey why they were not using contraception less than 2% reported access-related reasons (affordability/availability), with health concerns and social/religious concerns being more important. It is possible that the programme did not succeed in reducing these concerns. We found that reasons for non-use remained relatively stable over the period.

Another hypothesis could be that the programme was well designed but poorly implemented – for example, through lack of oversight of franchised outlets and clinics. However, the evidence on changing quality of care gathered here does not particularly support this hypothesis.

A recommendation for future impact evaluations might be to include some qualitative components, such as interviews with implementers and clients, in order to understand better the sometimes unexpected results generated by surveys.

Annex A Econometric methods

A.1 Selecting the PSM model

The objective of the PSM model is to deliver statistical balance between treatment and control groups in terms of their observable characteristics after they are matched. The choice regarding what model to use to achieve this involves both selecting the covariates on which observations will be matched, and the algorithm to do so.

The matching covariates were constructed from the baseline data, in order to rule out the possibility that any could have been affected by exposure to the intervention. Variable selection was based on an approach that integrated both a theoretical understanding of the context for the interventions and RH issues in the region, and a 'data-driven' approach. On the theoretical side, we began by considering which factors are anticipated to be related to both exposure to DRHR programme activities and contraceptive use in women (the main outcome variable), based on the experience of the evaluation team and knowledge of related literature. We then implemented a data-driven selection model to test all available variables for suitability for the matching model, to ensure that no potential confounding variables were excluded. This involved running a series of tests on the baseline data, first discarding sets of highly collinear variables, then implementing selection algorithms (forward and backward stepwise regression) on the remaining variables to isolate ones that emerged as being significantly correlated with the outcome variable. Higher order terms and interaction terms were also tested in the same way for suitability for the matching model. The result of this set of selection models was the inclusion of a core set of variables in the matching model that were consistently selected across the majority of the different methods employed.

The next step was to select our preferred matching model specification. We tested both kernel and nearest neighbour approaches, and within each class also tested a wide range of small refinements to identify the best performing model (for example, adjusting the bandwidth and trimming parameters in the kernel matching model).

Our preferred model is kernel matching, in which each treated observation is matched with a weighted average of outcomes of control observations on common support. This algorithm has the advantage of preserving information from control group observations in the sample, thus lowering the overall variance. An alternative model that we did not select is nearest neighbour matching, in which each treatment group observation is matched to a defined number of control observations that are closest in terms of their propensity scores. In this process, individual observations that are not suitable matches are discarded. We found that this model appeared to perform worse than the kernel model in terms of the final balance of the matched sample. Results from the nearest neighbour model are included as a robustness check to the main results in Annex B.

A.2 Additional diagnostic test results

The full set of results from our main matching model is reported in Tables 33 and 34 below. The main body of the tables capture information about how effectively the matching procedure has improved the balance of the sample (in other words, the comparability between treatment and control groups given their initial underlying differences). At the bottom of the table, the impact estimates are reported, showing the average treatment effect on the treated (ATT) associated with the variable being tested, the standard error of the estimate and the T-statistic.

Covariates used to construct propensity scores are given in the first column on the left. The next column reports the coefficients from the first stage of analysis, in which each variable is tested for association with exposure to the treatment group and the propensity score for each observation is calculated. Standard errors are reported in parentheses, together with asterisks to indicate whether the variables are significantly related to treatment assignment or not. The next set of columns show the mean value of each covariate in the sample, in the treatment and control group. This is reported first on the unmatched sample, and then again on the sample after observations were matched according to their calculated propensity scores. In each case, asterisks are presented to indicate whether the difference in means between the treatment and control groups is statistically significant. These asterisks give an indication of how comparable the two samples are. The difference in means in the unmatched sample is shown to be highly significant for many of the variables, whereas fewer of the variables emerge as being significantly different from one another in the matched sample. This is a signal that the matching procedure effectively improved the balance of the sample.

The next set of columns report a measure of bias associated with each of the matching covariates. The column titled ‘% change’ shows how bias has been altered as a result of matching: a positive number indicates that the balance of the sample has been improved.

The final two columns report p-values associated with the difference in mean values of each covariate for treatment and control groups in the matched and unmatched samples. These are the basis on which the asterisks are presented: a P-value of less than 0.1 means that the difference is significant at the 10% level (*), P value < 0.05 means significance at the 5% level (**) and $p < 0.01$ means significance at the 1% level (***).

At the bottom of the table, before the final impact estimates are presented, there are two additional diagnostic test results: Rubin’s B and Rubin’s R. These are two summary measures of covariate imbalance before and after matching. According to standard practice, the sample is well matched if Rubin’s B is less than 25 and Rubin’s R is between 0.5 and 2.

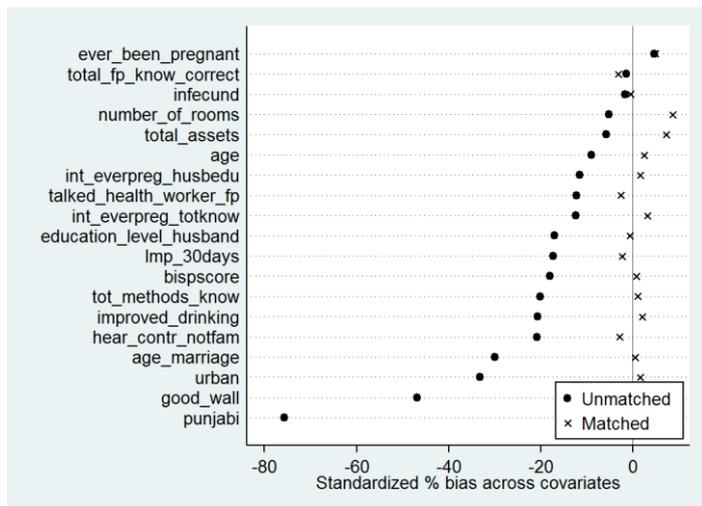
Table 33: Main matching model specification and results: PSI-only vs control

This table reports the results from the main matching model testing for the impact of the PSI-only group on current use of contraceptives amongst MWRAs, compared with the control group. The model specification is kernel matching (Gaussian), with bandwidth set at 0.03 and trimming at 5.

Covariates	First stage probit coefficients	Means unmatched		Means matched		Obs.	Standardised bias			P-value of test		
		Treatment	Control	Treatment	Control		Unmatched	Matched	% change	Unmtch	Mtch	
Number of rooms in the household	-0.013 (.024)	2.14	2.21	2.13	**	2.01	2288	-5.3	8.5	-60.7	0.203	0.026
Good wall materials	-0.332 (.076)***	0.66	0.86	0.69		0.69	2288	-46.9	1.6	96.6	0	0.717
Household has improved drinking source of water	-0.381 (.141)***	0.93	0.98	0.96		0.95	2288	-20.7	2	90.6	0	0.621
Total number of household assets	0.073 (.017)***	3.42	3.57	3.4	*	3.22	2288	-5.9	7.1	-20.9	0.158	0.077
MWRA's mother tongue is Punjabi	-0.881 (.061)***	0.3	0.65	0.31		0.31	2288	-75.7	0.5	99.3	0	0.894
Age at marriage	-0.022 (.008)***	18.48	19.58	18.55		18.54	2288	-30	0.4	98.5	0	0.91
MWRA has ever been pregnant	0.084 (.308)	0.93	0.92	0.93		0.92	2288	4.5	4.7	-4.4	0.272	0.241
Last menstrual period within the last 30 days	-0.216 (.074)***	0.76	0.83	0.77		0.78	2288	-17.3	-2.4	86.2	0	0.563
MWRA first heard about contraception from health personnel or a communication campaign.	-0.134 (.062)**	0.25	0.35	0.26		0.28	2288	-20.8	-3	85.5	0	0.437
MWRA talked to health worker about contraception in the past one month	-0.025 (.082)	0.12	0.16	0.12		0.13	2288	-12.3	-2.7	77.7	0.003	0.475
Correct beliefs about attributes of contraceptives	0.064 (.047)	1	1.01	1		1.02	2288	-1.4	-3.2	-125.8	0.733	0.419
Total number of FP methods MWRA has heard about	-0.011 (.044)	6.23	6.65	6.28		6.26	2288	-20.2	1	95.1	0	0.806
Urban location	-0.487 (.068)***	0.2	0.34	0.21		0.2	2288	-33.3	1.6	95.3	0	0.671
Infertile	-0.157	0.09	0.09	0.09		0.09	2288	-1.8	-0.6	66.9	0.667	0.882

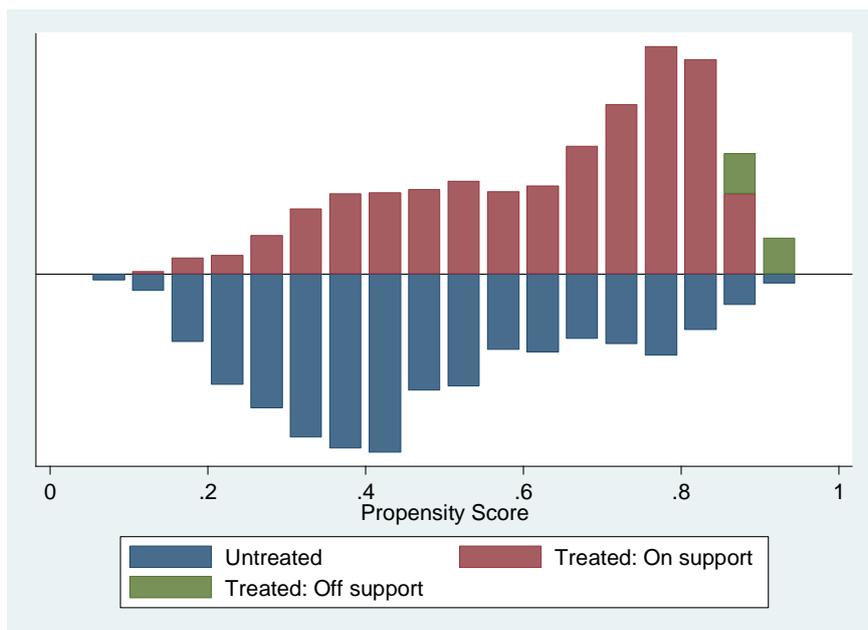
	(.109)												
Ever been pregnant* total contraceptive methods known	0	5.91	***	6.23	5.95		5.87	2288	-12.4	3	75.6	0.003	0.447
	(.046)												
Ever been pregnant* education level of husband	0.047	1.99	***	2.12	2		1.98	2288	-11.6	1.6	86.2	0.005	0.687
	(.098)												
Husband's education level	-0.047	2.13	***	2.32	2.15		2.16	2288	-17.1	-0.8	95.6	0	0.85
	(.095)												
NISP poverty score	0.002	29.41	***	32.4	29.53		29.41	2288	-18.1	0.7	95.9	0	0.856
	(.003)												
Age	0.002	32.06	**	32.74	32.07		31.89	2288	-9.1	2.3	74.5	0.03	0.568
	(.004)												
	Before (U)	After (M)											
Rubin's B	94.49	14.4											
Rubin's R	0.99	0.88											
ATT	0.02												
Standard error	(.026)												
T statistic	.794												
Off common support	66												

Figure 7: Covariate balance for before and after matching



Note: This figure shows standardised bias associated with all the covariates included in the PSM model. It corresponds to the analysis performed for analysis of current contraceptive use, in the PSI-only group vs the control group. The matching algorithm used is kernel matching (Gaussian), with trimming set at 5 and a bandwidth parameter of 0.03.

Figure 8: The distribution of propensity scores in the matched and unmatched sample: PSI-only vs control



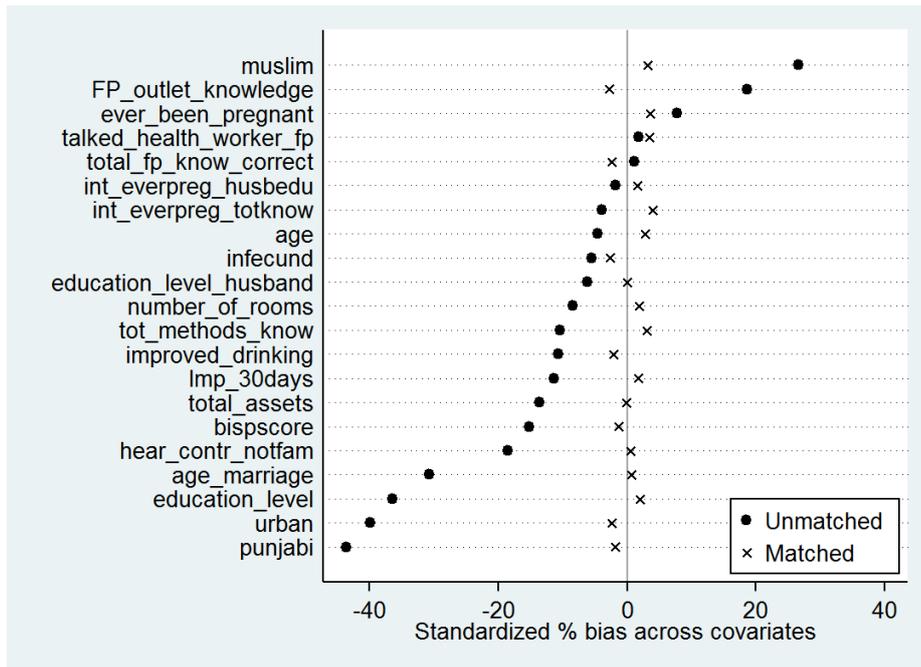
Note: This figure corresponds to the analysis performed to analyse current contraceptive use, in the PSI-only group vs the control group. The matching algorithm used is kernel matching (Gaussian), with trimming set at 5 and a bandwidth parameter of 0.03.

Table 34: Main matching model specification and results: PSI+MSI vs control

This table reports the results from the main matching model testing for the impact of the PSI+MSI group on current use of contraceptives amongst MWRA, compared with the control group. The model specification is kernel matching (Gaussian), with bandwidth set at 0.03 and trimming at 5.

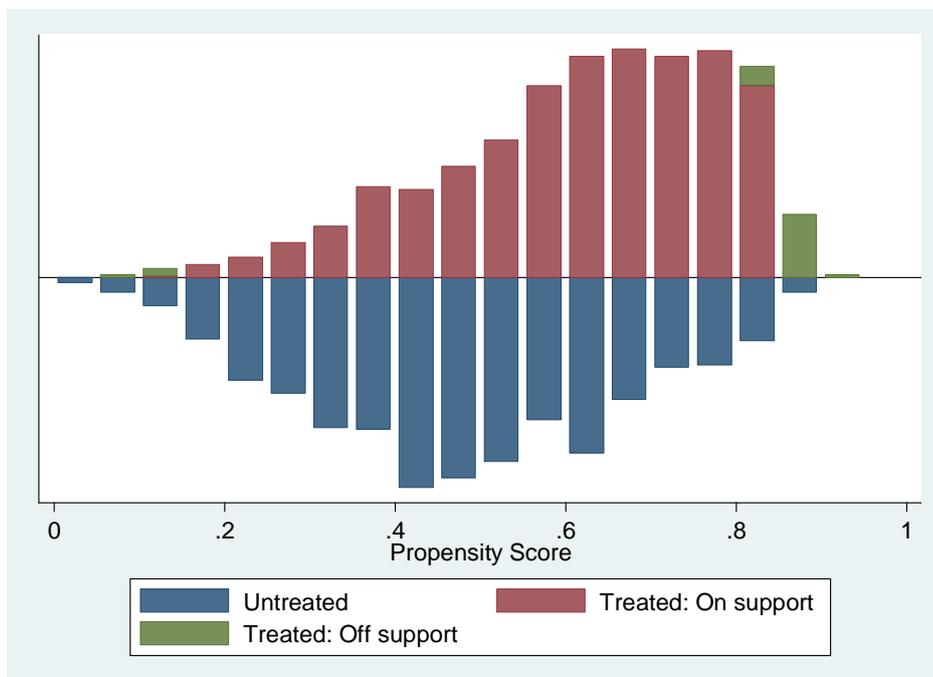
Covariates	First stage probit coefficients	Means unmatched		Means matched		Obs.	Standardised bias			P-value of test		
		Treatment	Control	Treatment	Control		Unmatched	Matched	% change	Unmtch	Mtch	
Number of rooms in the household	-0.024 (.024)	2.09	**	2.2	2.1	2.07	2276	-8.4	1.9	77.6	0.043	0.629
Household has improved drinking source of water	-0.237 (.159)	0.96	**	0.98	0.96	0.97	2276	-10.8	-2.1	80	0.011	0.61
Total household assets	0.022 (.017)	3.22	***	3.56	3.23	3.23	2276	-13.6	-0.2	98.4	0.001	0.956
Education level of MWRA	-0.182 (.037)***	1.55	***	1.9	1.57	1.55	2276	-36.4	2	94.6	0	0.59
Islam	1.111 (.156)***	0.99	***	0.93	0.99	0.98	2276	26.6	3.2	88.1	0	0.18
Mother tongue is Punjabi	-0.469 (.058)***	0.44	***	0.65	0.46	0.47	2276	-43.6	-1.8	95.8	0	0.654
Age at marriage	-0.029 (.009)***	18.44	***	19.55	18.56	18.53	2276	-30.7	0.6	98.1	0	0.88
MWRA has ever been pregnant	0.494 (.333)	0.94	*	0.92	0.94	0.93	2276	7.8	3.6	53.7	0.06	0.361
Last menstrual period within the last 30 days	-0.146 (.075)*	0.78	***	0.82	0.8	0.79	2276	-11.3	1.7	84.9	0.007	0.676
MWRA first heard about contraception from health personnel or a communication campaign	-0.183 (.061)***	0.26	***	0.35	0.27	0.27	2276	-18.5	0.6	97	0	0.887
MWRA talked to health worker about contraception in the past one month	0.049 (.077)	0.17		0.16	0.17	0.15	2276	1.8	3.4	-90.3	0.664	0.385
Correct beliefs about attributes of contraceptives	0.057 (.045)	1.01		1.01	1.01	1.03	2276	1.2	-2.4	-105.4	0.781	0.546
Total number of FP methods MWRA has heard about	-0.009 (.044)	6.44	**	6.65	6.48	6.42	2276	-10.5	3	71	0.012	0.442
MWRA is aware of at least one	0.372	0.81	***	0.73	0.8	0.81	2276	18.6	-2.9	84.5	0	0.445

Figure 9: Covariate balance for the main model before and after matching: PSI+MSI vs control



Note: This figure corresponds to the analysis performed to analyse current contraceptive use, in the PSI+MSI group vs the control group. The matching algorithm used is kernel matching (Gaussian), with trimming set at 5 and a bandwidth parameter of 0.03.

Figure 10: The distribution of propensity scores in the matched and unmatched sample: PSI+MSI vs control



Note: This figure corresponds to the analysis performed to analyse current contraceptive use, in the PSI+MSI group vs the control group. The matching algorithm used is kernel matching (Gaussian), with trimming set at 5 and a bandwidth parameter of 0.03.

Annex B Robustness check

The matching model used for the main analysis is kernel matching, with a bandwidth parameter of 0.03 and trimming set at 5. This is the model that was found to perform best over a range of balance and diagnostic tests during the preliminary stages of analysis preparation. As a robustness check, we present here the results of the PSM analysis using an alternative specification. These tables are based on a nearest neighbour model, in which each treatment group observation is matched to a weighted average of the five control observations that are closest in terms of their assigned propensity scores.

Our findings are found to be robust to this alternative model.

Table 35: Nearest neighbour matching model specification and results: PSI-only vs control

Covariates	First stage probit coefficients	Means unmatched		Means matched		Obs.	Standardised bias			P-value of test	
		Treatment	Control	Treatment	Control		Unmatched	Matched	% change	Unmtch	Mtch
Number of rooms in the household	-0.013 (.024)	2.14	2.21	2.14	2.07	2354	-5.3	5	4.7	0.203	0.188
Good wall materials	-0.332 (.076)***	0.66	0.86	0.66	0.64	2354	-46.9	4.9	89.5	0	0.268
Household has improved drinking source of water	-0.381 (.141)***	0.93	0.98	0.93	0.92	2354	-20.7	5.7	72.4	0	0.241
Total number of household assets	0.073 (.017)***	3.42	3.57	3.42	3.3	2354	-5.9	4.9	16.8	0.158	0.219
MWRA mother tongue is Punjabi	-0.881 (.061)***	0.3	0.65	0.3	0.29	2354	-75.7	0.9	98.8	0	0.806
Age at marriage	-0.022 (.008)***	18.48	19.58	18.48	18.53	2354	-30	-1.3	95.6	0	0.726
MWRA has ever been pregnant	0.084 (.308)	0.93	0.92	0.93	0.92	2354	4.5	5.8	-28.5	0.272	0.136
Last menstrual period within the last 30 days	-0.216 (.074)***	0.76	0.83	0.76	0.77	2354	-17.3	-3.5	79.6	0	0.386
MWRA first heard about contraception from health personnel or a communication campaign	-0.134 (.062)**	0.25	0.35	0.25	0.26	2354	-20.8	-2.4	88.3	0	0.512
MWRA talked to health worker about contraception in the past one month	-0.025 (.082)	0.12	0.16	0.12	0.13	2354	-12.3	-3.2	74	0.003	0.387
Correct beliefs about attributes of contraceptives	0.064 (.047)	1	1.01	1	1.02	2354	-1.4	-3.9	-172.9	0.733	0.313
Total number of FP methods MWRA has heard about	-0.011 (.044)	6.23	6.65	6.23	6.2	2354	-20.2	1.5	92.7	0	0.708
Urban location	-0.487 (.068)***	0.2	0.34	0.2	0.19	2354	-33.3	0.2	99.3	0	0.945
Infertile	-0.157 (.109)	0.09	0.09	0.09	0.1	2354	-1.8	-2.6	-47.4	0.667	0.5
Ever been pregnant* total contraceptive methods known	0 (.046)	5.91	6.23	5.91	5.78	2354	-12.4	5	59.8	0.003	0.196
Ever been pregnant* education level	0.047	1.99	2.12	1.99	1.99	2354	-11.6	0.1	99.4	0.005	0.985

of husband												
	(.098)											
Husband's education level	-0.047	2.13	***	2.32	2.13	2.17	2354	-17.1	-3.8	77.7	0	0.331
	(.095)											
NISP poverty score	0.002	29.41	***	32.4	29.41	29.58	2354	-18.1	-1.1	94	0	0.786
	(.003)											
Age	0.002	32.06	**	32.74	32.06	31.98	2354	-9.1	1	88.4	0.03	0.792
	(.004)											
	Before (U)	After (M)										
Rubin's B	94.49	15.68										
Rubin's R	0.99	0.82										
ATT	0.02											
Standard error	(.028)											
T statistic	.874											
Off common support	0											

Table 36: Nearest neighbour matching model specification and results: PSI+MSI vs control

Covariates	First stage probit coefficients	Means unmatched		Means matched		Obs.	Standardised bias			P-value of test			
		Treatment	Control	Treatment	Control		Unmatched	Matched	% change	Unmtch	Mtch		
Number of rooms in the household	-0.024 (.024)	2.09	**	2.2	2.09	2.08	2339	-8.4	1	88.2	0.043	0.795	
Household has improved drinking source of water	-0.237 (.159)	0.96	**	0.98	0.96	0.96	2339	-10.8	-1.5	86.2	0.011	0.734	
Total household assets	0.022 (.017)	3.22	***	3.56	3.22	3.2	2339	-13.6	0.8	94.5	0.001	0.842	
Education level of MWRA	-0.182 (.037)***	1.55	***	1.9	1.55	1.53	2339	-36.4	2.2	94	0	0.538	
Islam	1.111 (.156)***	0.99	***	0.93	0.99	0.98	2339	26.6	1.2	95.6	0	0.637	
Mother tongue is Punjabi	-0.469 (.058)***	0.44	***	0.65	0.44	0.44	2339	-43.6	0.6	98.7	0	0.888	
Age at marriage	-0.029 (.009)***	18.44	***	19.55	18.44	18.36	2339	-30.7	2.2	93	0	0.561	
MWRA has ever been pregnant	0.494 (.333)	0.94	*	0.92	0.94	*	0.92	2339	7.8	7.3	6.4	0.06	0.061
Last menstrual period within the last	-0.146	0.78	***	0.82	0.78	0.79	2339	-11.3	-2.9	74.8	0.007	0.477	

30 days												
	(.075)*											
MWRA first heard about contraception from health personnel or a communication campaign	-0.183 (.061)***	0.26	***	0.35	0.26	0.26	2339	-18.5	0.4	97.7	0	0.908
MWRA talked to health worker about contraception in the past one month	0.049 (.077)	0.17		0.16	0.17	0.16	2339	1.8	2.5	-35.7	0.664	0.527
Correct beliefs about attributes of contraceptives	0.057 (.045)	1.01		1.01	1.01	1.03	2339	1.2	-2.6	-123.7	0.781	0.5
Total number of FP methods MWRA has heard about	-0.009 (.044)	6.44	**	6.65	6.44	6.36	2339	-10.5	4.2	60.4	0.012	0.276
MWRA is aware of at least one provider of FP in the community	0.372 (.07)***	0.81	***	0.73	0.81	0.81	2339	18.6	-0.4	98	0	0.921
Urban location	-0.489 (.068)***	0.17	***	0.34	0.17	0.18	2339	-39.9	-1.4	96.4	0	0.68
Infertile	-0.208 (.109)*	0.08		0.09	0.08	0.08	2339	-5.5	-1.4	73.7	0.186	0.703
Ever been pregnant* total contraceptive methods known	-0.029 (.046)	6.13		6.23	6.13	5.98	2339	-3.9	6.1	-54.5	0.345	0.112
Ever been pregnant* education level of husband	-0.058 (.101)	2.1		2.12	2.1	2.07	2339	-1.8	2.6	-41.1	0.663	0.505
Husband's education level	0.122 (.098)	2.25		2.31	2.25	2.26	2339	-6.1	-1.1	81.6	0.141	0.77
NISP poverty score	0.003 (.003)	29.91	***	32.33	29.86	30.21	2339	-15.1	-2.1	85.8	0	0.575
Age	0.003 (.004)	32.38		32.73	32.38	32.21	2339	-4.6	2.3	49.8	0.276	0.569
	Before (U)	After (M)										
Rubin's B	80.78	11.7										
Rubin's R	0.81	0.86										
ATT	0											
Standard error	(.027)											
T statistic	.032											
Off common support	2											

Annex C Sensitivity analysis

C.1 Purpose of the sensitivity analysis

The survey for the PSI and MSI evaluation was designed as a panel survey, in which the same women would be interviewed at baseline and endline. In practice, it was not always possible to re-interview exactly the same woman at endline, either because the entire household could not be found or interviewed, or because the woman in the household was not available. Attrition was relatively higher in urban areas than in rural areas, and the cotton picking season in some districts of Punjab also led to many women being unavailable for interview.

In households where the original women could not be interviewed, the survey team randomly selected a different MWRA to interview instead. Therefore the final sample included some women who had not been interviewed previously. This is shown in Table 37 below.

Table 37: Sample sizes and attrition

	Baseline	Endline – all women	Endline – women who were re-surveyed from baseline
Treatment 1: PSI-only	2436	2006	1750
Treatment 2: PSI+BCC	988	824	716
Treatment 3: MSI+PSI	2502	2008	1755
Control	1962	1498	1293
Total	7888	6336	5514

Sensitivity analysis was first conducted to understand how different women who could not be interviewed again at endline were from the women who replaced them, in households that were re-interviewed. Additional sensitivity analysis was then carried out to determine whether attrition at the household level is likely to have compromised the representativeness of the sample

C.2 Sensitivity analysis considering attrition at the woman level

A sensitivity analysis was conducted to understand whether the women who were interviewed at endline only (the 'replacement' women) were similar in their underlying characteristics to women who were originally interviewed at baseline. The purpose of this analysis was to determine whether it would be reasonable to consider the replacement women as equivalent to the original sample of baseline women for the purposes of analysis. If the women were found to be similar, analysis of the whole sample could be conducted, on the assumption that differences in average outcomes between baseline and endline were due to the RH programmes alone, rather than changes in the underlying characteristics of the sample.

This analysis consisted of comparing outcomes at endline between women who were successfully re-interviewed and women who were interviewed for the first time. As this involved endline data, a comparison was only drawn across variables that could not plausibly be influenced by the RH programmes in order to rule out the possibility that observed differences in the samples were due to the effects of PSI/ MSI rather than reflecting their underlying characteristics.

The results indicate significant differences between the two samples in four out of 12 household-level variables tested, and four out of six woman-level variables. These differences were

considered too great to allow the full sample to be used in the analysis without risking bias in the results.

Table 38: Sensitivity analysis: Differences at endline in the households where a replacement woman was interviewed compared with the households where the same woman was re-interviewed

Variable	Households in which the same woman was re-interviewed at endline	Households in which a replacement woman was interviewed at endline	P value of the difference
Baseline NISP poverty score	28.838*	30.468	0.078
Total assets	3.453***	4.002	0
% households owning their house	88.9	91.4	0.187
% households with good floor materials (ceramic tiles, marble, cement, carpet or bricks)	47.5	47.9	0.889
% households with good roof materials (iron sheets, asbestos, T-iron, wood, brick, cement or RCC)	74.1	73	0.848
% households with good wall materials (stone bricks, backed bricks or cement)	69.9	73.6	0.216
% households with an improved source of drinking water ⁽¹⁾	93	93.3	0.834
% households with an improved toilet facility ⁽¹⁾	55.2	58.3	0.495
% households that are urban	14.4	15.5	0.632
% households with land	46.2	49.8	0.575
% households with electricity	81.7**	86.7	0.013
Number of rooms in the household	2.105***	2.457	0

Note: (1) As defined by www.who.int/water_sanitation_health/monitoring/oms_brochure_core_questionsfinal24608.pdf.

* = p value < 0.1, ** = p value < 0.05, *** = p value < 0.001. Differences calculated by applying weights constructed over the sample of women interviewed at endline. Comparisons conducted using endline data.

Table 39: Sensitivity analysis: Differences at endline in women that were interviewed for the first time at endline compared with women who were re-interviewed.

Variable	Women that were re-interviewed at endline	Women who were interviewed for the first time at endline	P value of the difference
% women who are Muslim	97.6	97.7	0.974
% women whose mother tongue is Punjabi	35.7	36.8	0.731
Age	32.601***	29.096	0
% women who can read in any language	31.5**	40.5	0.015
% women who can write in any language	29.3**	36.4	0.029
Age at marriage	18.189***	19.064	0

* = p value < 0.1, ** = p value < 0.05, *** = p value < 0.001. Differences calculated by applying weights constructed over the sample of women interviewed at endline. Comparisons conducted using endline data.

C.3 Sensitivity analysis considering attrition at the woman level

A second sensitivity analysis was carried out to understand the extent to which attrition at the household level (i.e. the households that could not be located or interviewed at all at endline) compromised the representativeness of the sample. The original sample was chosen to be representative of women in the evaluation areas. Representativeness would be lost if the households who were not successfully re-interviewed were systematically different from those who were.

The analysis involved comparing outcomes at baseline between the households who were re-interviewed at endline and those that were not. Since this analysis involved baseline (i.e. pre-intervention) data, it was not necessary to limit the comparison only to variables that would not respond to the effects of the intervention. The analysis found that there were significant differences between the two samples in one out of 12 household-level variables tested, and four out of 12 woman-level variables.

Table 40: Sensitivity analysis: Differences at baseline between households that were successfully re-interviewed at endline and households that were not re-interviewed.

Variable	Households that were re-interviewed at endline	Households that were not re-interviewed	P value of the difference
NISP poverty score	29.3	30.009	0.523
Total assets	3.288	3.409	0.29
% households owning their house	94.7***	89.9	0.001
% households with good floor materials (ceramic tiles, marble, cement, carpet or bricks)	44.4	46.8	0.464
% households with good roof materials (iron sheets, asbestos, T-iron, wood, brick, cement or RCC)	67.9	68	0.966
% households with good wall materials (stone bricks, backed bricks or cement)	70.3	69.5	0.741
% households with an improved source of drinking water ⁽¹⁾	88.1	88.9	0.54
% households with an improved toilet facility ⁽¹⁾	51.6	53.5	0.513
% households that are urban	16.2	20.8	0.113
% households with land	34.2	30.8	0.197
% households with electricity	93.2	91.4	0.294
Number of rooms in the household	2.292	2.3	0.911

Note: (1) As defined by www.who.int/water_sanitation_health/monitoring/oms_brochure_core_questionsfinal24608.pdf.

* = p value < 0.1, ** = p value < 0.05, *** = p value < 0.001. Differences calculated by applying weights constructed over the sample of women interviewed at baseline. Comparisons conducted using baseline data

Table 41: Sensitivity analysis: Differences at baseline between women from households that were re-interviewed at endline and women from households that were not interviewed at endline

Variable	Women from households that were re-interviewed at endline	Women from households that were interviewed for the first time at endline	P value of the difference
% women who are Muslim	97.3*	94.6	0.088
% women whose mother tongue is Punjabi	35.9	40.2	0.272
% women who have ever been pregnant	90.3	89.2	0.599
Age	30.576***	33.602	0.001
% women who can read in any language	32.4	34.8	0.431
% women who can write in any language	30.2	32.6	0.441
Age at marriage	18.805	20.526	0.244
Average number of contraceptive methods known	5.878	5.6	0.197
% MWRAs ever used a contraceptive method	49.9*	43.8	0.08
% MWRAs currently using contraceptives	35**	29.3	0.011
% MWRAs currently using contraceptives, excluding those trying to get pregnant who report that they cannot get pregnant or are menopausal	47.2	40	0.104
Unmet need for contraception ⁽²⁾	31.9	30.4	0.448
% women who are aware of a source of FP in their community	68.3	65.6	0.454

* = p value < 0.1, ** = p value < 0.05, *** = p value < 0.001. Differences calculated by applying weights constructed over the sample of women interviewed at baseline. Comparisons conducted using baseline data.