ePact
Strengthening evaluation effectiveness and impact

Monitoring, Verification and Evaluation Service Provider for the WASH Results Programme

Randomised Control Trial Research Study Summary Report: *Understanding the role of follow-up (outcome phase) activities in achieving sustainable WASH outcomes*

Oxford Policy Management

EDePo @ IFS
Evaluation of Development Policies at the Institute for Fiscal Studies

itad

e-Pact is a consortium led by Oxford Policy Management and co-managed with Itad in association with the Institute for Fiscal Studies.

This evaluation was commissioned by the UK Department for International Development and was funded with UK aid from the UK government.

UKaid
from the British people

This material has been funded by UK aid from the UK government; however the views expressed do not necessarily reflect the UK government's official policies.

Disclaimer

This report has been prepared by the e-Pact consortium for the named client, for services specified in the terms of reference and contract of engagement. Any use of the information in this report should include a citation that acknowledges the e-Pact consortium as the author of the report, using the citation "*ePact (2018) – WASH Results Programme Evaluation – RCT Research Study Summary Report*".

This impact assessment is being carried out by e-Pact. Britta Augsburg (britta_a@ifs.org.uk) from the Institute of Fiscal Studies is the workstream leader for the randomised control trial. Zach White (zach.white@opml.oc.uk) from Oxford Policy Management is the project manager of the evaluation. Julia Larkin is the evaluation Team Leader.

The contact point for the client is Stephen Lindley-Jones (stephen-jones@dfid.gov.uk). The client reference number for the project is PO6507.

e-Pact
Registered in England: 3122495

6 St Aldates Courtyard
38 St Aldates
Oxford OX1 1BN
United Kingdom

Tel +44 (0) 1865 207300
Fax +44 (0) 1865 207301
Email admin@opml.co.uk
Website www.opml.co.uk

Acknowledgements

This report is the culmination of activities undertaken by the e-Pact Evaluation Team between October 2014 and July 2018. We would like to thank Leonard Tedd and Laura Westcott at the UK Department for International Development (DFID) for their guidance and support during the design phase. The team are also thankful to Meredith Bradbury and Tamour Rashid from DFID Pakistan for their assistance in gaining the necessary permissions for conducting the data collection.

The team would also like to thank all those at Plan International and WaterAid who supported the research process, notably Marc Faux at the design stage, John Dean throughout implementation, and Teia Rogers at endline. The team would also like to thank those at Plan International and WaterAid in Pakistan who supported the research, notably: Abdul Saboor, Mohammed Dawud, and Imran Shami from Plan International, and Imran Khan from WaterAid. This work would not have been possible without their commitment to the research and facilitation of implementation activities.

We would also like to thank our dedicated team of enumerators, and in particular Rizwana Khan (Oxford Policy Management (OPM)), who supervised data collection in the field at baseline, and Aneela Sardar (OPM) at endline. Shafique Arif (OPM) acted as survey director in both survey rounds.

Britta Augsburg (Institute for Fiscal Studies (IFS)) led the design and analysis of this randomised control trial (RCT), with support from Zach White (OPM) and Andrej Kveder (OPM). Sarah Javeed (OPM) and Zara Durrani (OPM) led the qualitative research component. Madhav Vaidyanathan (OPM) supported the quantitative analysis and Marwah Malik (OPM) supported the qualitative analysis.

Executive summary

The WASH Results Programme

The UK Department for International Development's (DFID) Water, Sanitation and Hygiene (WASH) Results Programme is a £111 million seven-year programme (2014 to 2021). In 2017 the programme was extended - the first round of the programme was implemented between May 2014 and March 2018 and aimed to bring equitable and sustainable water and sanitation services and hygiene practices to 4.5 million people in 11 countries[1]. The first round was implemented by three NGO-led supplier consortia[2] contracted by DFID under a Payment by Results (PbR) financing modality.[3] Across all three suppliers, between April 2014 and March 2016 the first phase successfully provided water services to at least 1.1 million people, and sanitation services to at least 3.9 million people, and reached at least 13.4 million people with hygiene promotion messages in 11 countries.

The WRP evaluation

In June 2014, DFID signalled the need for more learning on how PbR programmes can be effectively implemented, and under what circumstances. The evaluation of the WRP falls under this remit.[4] The use of PbR in the WASH sector specifically is still relatively new, and the use of the modality represented one of the first times it was used at scale in the WASH sector. The evaluation of the WRP is therefore primarily being undertaken for learning purposes, not as an accountability exercise[5]. The objectives of the evaluation are to assess:

(i) whether the programme successfully achieved its stated objectives;

(ii) the influence of programme design, including the PbR modality, on this achievement; and

(iii) lessons for applying PbR in WASH programming in future.

Focus of the Randomised Control Trial Research Study within the WRP evaluation

The Randomised Control Trial (RCT) Research Study discussed in this report is a component of the overall evaluation and narrowly focuses on only one of the three Supplier consortia, and in only one geographical area; as such, findings reported here might deviate from experiences of other suppliers. Specifically, the RCT Research Study relates to the South Asia WASH Results Programme (SAWRP) - one of the three WRP Supplier consortia. SAWRP is led by Plan International, with WaterAid, Unilever, and Water and Sanitation for the Urban Poor as implementing partners.[6] SAWRP operates in Bangladesh and Pakistan. The RCT Research Study focuses on SAWRP in Pakistan, and in particular on three out of eight districts covered by SAWRP in Pakistan, located in the provinces of Punjab and Sindh. It is worth highlighting that the districts within Sindh and Punjab have distinct characteristics, most notably with respect to their vegetation and climatic conditions, leading to differences in living conditions. Living standards[7] are significantly lower in Badin (31.1) than in Bahawalpur (77.5) and Rahimyar Khan (75.2) (UNDP, 2017).

[1] South Sudan and Liberia were dropped from the original 13-country programme
[2] The three supplier sub-programmes were (i) SWIFT, led by Oxfam GB, (ii) SAWRP, led by PLAN International, and (iii) SSH4A, implemented by the Netherlands Development Organisation (SNV).
[3] In 2017, the WASH Results Programme was extended for the delivery of additional outputs and outcomes in nine of the original countries. All further mention of the WASH Results Programme however only refers to the first round implemented from 2014 to 2018.
[4] Based on the verified results as reported by the suppliers to DFID.
[5] For the original terms of reference for the evaluation, see Annex A of the Midline Evaluation Report
[6] The Water Engineering and Development Centre (WEDC) and Ipsos MORI acted as advisory partners.
[7] As measured on a scale from 0 to 100 by access to clean water, clean fuel, electricity, adequate sanitation, roof quality and basic household assets

Additionally, the RCT Research Study analyses only the second half of the programme's implementation. The WRP was implemented in two phases. During the first phase of the WRP (the 'output phase') SAWRP was paid by DFID for the delivery of output-level results on sanitation and hygiene in programme communities. During the second phase of the WRP (the 'outcome phase') SAWRP was paid by DFID for the delivery of outcome-level results. The 'outcome phase activities' included door-to-door follow-up visits, monitoring, and community-level promotion (e.g. street-plays and community meetings). The RCT Research Study assesses the effectiveness of the outcome phase implementation and does not directly assess the impact of output phase implementation. Given that the RCT Research Study focuses on assessing the effectiveness of the outcome phase activities related to rural sanitation and hygiene, it only speaks to whether the programme successfully achieved a sub-set of its stated objectives, i.e. evaluation objective (i) stated above, as opposed to any other, wider evaluation questions.

The RCT Research Study research questions

The RCT Research Study was designed to answer the following three questions, of which the first is the primary and the other two are secondary research questions:

(1) *What is the impact of outcome phase activities on the sanitation behaviour of output phase beneficiaries?*
(2) *What is the impact of outcome phase activities on the hygiene behaviour of output phase beneficiaries?*
(3) *How do the above impacts differ depending on the output phase modality chosen by implementing partners?*

Interest in the sustainability of sanitation and hygiene behaviour has been growing over recent years amongst policymakers, practitioners, and academics alike. At the time of design of the RCT Research Study (Q4 2015), several rigorous impact evaluations that aimed to understand the impact of sanitation programmes on health outcomes were ongoing or had just been completed. Little, however, was known about the relative effectiveness of different WASH interventions and their sustainability. To the best of our knowledge at the design stage there were no other rigorous impact evaluations being conducted that evaluated approaches with a particular focus on the specific role of follow-up activities in achieving sustainable WASH outcomes. As such it was expected that this study would fill a key gap within the literature[8], and make a contribution to creating knowledge relevant to future WASH programme design. Since the study was first designed, recent evidence syntheses of rural sanitation programmes have stressed the need to more clearly focus on the effectiveness of specific activities. A recent review concluded: '*This review confirms positive impacts of sanitation on aspects of health. Evidence gaps remain and point to the need for research that rigorously describes sanitation implementation and type of sanitation interventions.'* (Freeman *et al*., 2017).

The research questions were established in consultation with DFID and the implementing partners. It was decided early on that an assessment of the overall PbR could not be a focus of the RCT Research Study because the establishment of a control group was deemed infeasible. Several potential research topics were considered in relation to evidence gaps identified by DFID[9]. The research questions were selected based on: i) the extent to which they addressed evidence gaps identified by the DFID; ii) their relevance to answering questions surrounding sustainability; iii) how robustly they could be addressed by methods that provided an unbiased measure of 'impact'[10]; and iv) their fit with supplier implementation approaches and timelines.

[8] Particularly that employing experimental designs.
[9] See DFID (2013) 'WASH evidence paper'.
[10] Which was interpreted by the evaluation team in the context of the Organisation for Economic Co-operation and Development (OECD) Development Assistance Committee (DAC) criteria, which situate 'impact' as the causal effect of the programme on indicators of interest. The programme's effect can relate to medium-term 'outcomes' or to longer-term 'impacts'. In both cases we would refer to these as 'impact' evaluations.

Study design: A mixed-methods approach

A mixed-methods approach, consisting of a quantitative and qualitative component, was chosen to answer the three research questions outlined above. Under the quantitative component the research questions were addressed using experimental methods (specifically an RCT). RCTs are an increasingly popular approach among development economists as they provide a robust measure of 'impact' through building a credible counterfactual, and hence are considered a 'gold-standard' to answer questions such as research questions 1 and 2 in particular. In order to be able to gain a deeper understanding and to explain patterns, trends, and mechanisms (the 'what' and the 'why'), which cannot be readily addressed with an RCT, the quantitative component was coupled with a complementary, sequential qualitative research study.

The quantitative evaluation design is a cluster RCT. Communities in which output phase activities had been implemented formed the population from which the RCT sample of 123 communities was randomly drawn. Of these, 61 were randomly assigned to a treatment arm in which the 'outcome phase activities' were implemented (hereafter referred to as 'treatment' communities). The remaining 62 communities were assigned to a control arm, where no activities took place. The baseline survey for the RCT Research Study was implemented in Q1 2016 – at the point at which the programme switched between its Output- and outcome phases. Activities were assessed through an endline survey, which took place two years after the start of outcome phase activities, in Q1 2018, and coincided with the end of the outcome phase.

In total, 1,191 households were interviewed as part of the quantitative data collection, 95% of which were interviewed in both survey rounds. Interviewed households are a random sample of exclusively programme beneficiaries from the output phase. The quantitative statistics are therefore representative of programme beneficiaries in SAWRP villages in the two study districts, but not of the general population living in these programme villages. Using monitoring data provided by the implementing agencies, this report demonstrates compliance with the randomisation process and shows that attrition is low and unrelated to treatments. This allows any observed changes and difference between treatment and control communities and households to be attributed to the outcome phase activities.

The qualitative research component was designed as a behaviour change study at the household and community level. It took place in Q2 2018, shortly after the quantitative endline survey, so that its findings could feed into the design of the qualitative study. The main purpose of the complementary qualitative research was to better explain behaviour change, patterns, motivations, and underlying impact mechanisms by providing an in-depth understanding of the knowledge, attitudes, practices, and determinants of sanitation and hygiene behaviour change in focal communities. It also sought to examine contextual factors that affected response to the programme and unanticipated consequences resulting from programme delivery.

In total, the qualitative research component conducted 24 key informant interviews (KIIs), 40 in-depth interviews (IDIs), and 43 focus group discussions (FGDs). Sampling for the qualitative research was purposive – communities were selected within both implementation provinces and at both the high and low end of percentage of beneficiaries who openly defecate, and percentage of beneficiaries who had a toilet at the time of the endline survey. A total of 12 villages (eight treatment villages and four control villages) were selected for the qualitative research from amongst the RCT sampling frame. In each community six KIIs and three FGDs took place with community-level programme staff. In the household interviews, no distinction was made between beneficiary and non-beneficiary households. This provides wider, community-level information on the perceptions, uptake, implementation, and sustainability of the intervention, which the quantitative component could not speak to. As will become clear throughout the report, this information turned out to be crucial in making conclusions and recommendations about the programme and its effectiveness.

Significance of features of the research design in relation to interpreting the findings

The research design was premised on there being 'slippage' – that is, a certain degree of reversion to open defecation was expected and the research focused on the degree to which outcome phase activities were able to reduce this slippage. It is important to note that if there was no slippage back to undesired sanitation and hygiene behaviours – even without follow-up activities (i.e. in control communities) – then it would not be possible to detect any impacts as the desired outcome would already be the status quo. In other words, positive impacts could only be detected if slippage occurred. Since slippage is a well-known and recognised phenomenon, also acknowledged and expected by SAWRP implementers, this legitimises the outcome phase activities and this impact assessment. When interpreting the results it is very important to keep in mind the assumed level of slippage. Specifically, that the slippage rates were dramatically different between provinces, and the team detects different levels of impact between the two provinces.

Figure 1: Graphical representation of research design

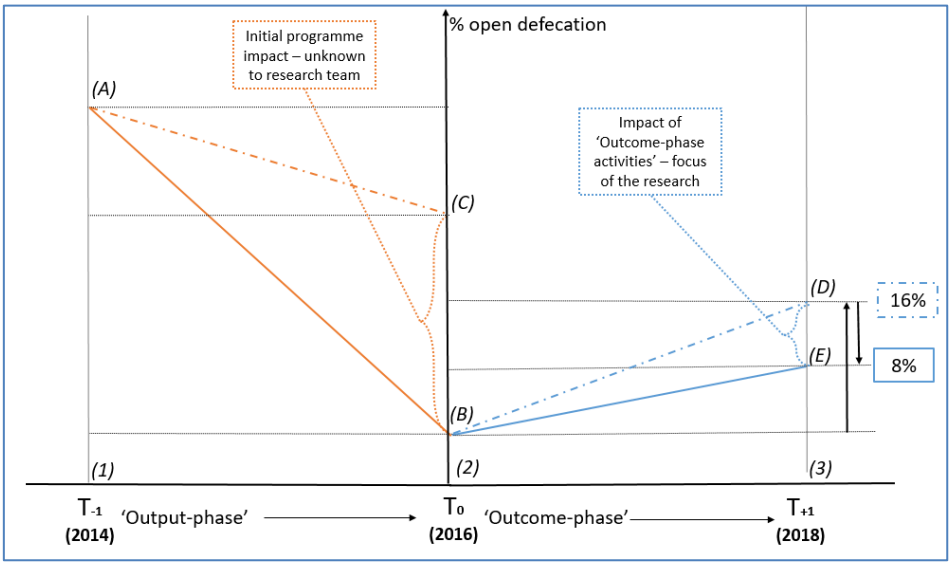


Figure 1 presents a graphical representation of the research, highlighting some of the features outlined above and the key assumptions made. The y-axis represents the percentage of beneficiaries defecating in the open. The y-axis represents time: *(1)* is the start of the output phase; *(2)* is the end of the output phase[11] and the start of the outcome phase; and *(3)* the end of the outcome phase[12]. Point *(2)*, which was midway through the programme, was the start of the RCT Research Study and the start of data collection. Our data do not allow us to make a judgement about the impact of the output phase activities.

The fundamental focus of this research design is to assess the *additional* benefit of the outcome phase activities with respect to key outcome-level indicators. The impact assessment discussed in this report speaks therefore solely to the right-hand side of Figure 1. Point *(B)* was when the outcome phase started and the baseline survey took place. In the design a number of assumptions have to be made, as well as a decision about the minimum impact one wants to be able to detect. In the context of this study, the decision was to adopt a design that would make it possible to detect a minimum reduction in slippage of 8 percentage points (the difference between points *(D)* and *(E)* - assuming that, without the outcome phase activities, a slippage of 16% would occur[13] (the difference between points *(B)* and *(D)*).

[11] The first phase of the programme – implemented between 2013/14 and December 2015.
[12] The first phase of the programme – implemented between January 2016 and March 2018.
[13] The assumed slippage rate drew on existing literature; particularly Tyndale-Biscoe et. al. (2013).

Key findings

(1) What is the impact of outcome phase activities on the sanitation behaviour of output phase beneficiaries? (primary research question)

- Already when starting the outcome phase, about 17% of beneficiary households had at least one member aged five and above defecating in the open as the study's baseline survey reveals. This percentage doubled to 34% in the two years up when the endline survey was conducted. Across both provinces, on average, the outcome phase activities did not reduce this slippage at a statistically significant level.

- The average hides differences at province level: In Sindh the programme activities successfully reduced slippage by 14 percentage points, a finding statistically significant at a 7% significance level. This impact seems to have been achieved by reducing the degree to which toilets became dysfunctional. Latrine collapse due to rains was reported as a key reason for latrines becoming non-functional in communities in Sindh, but less so in treatment communities.

- Where latrines remained functional there was very little slippage in both treatment and control areas. This indicates that latrines becoming non-functional was one of the most significant reasons for reverting to open defecation and the key dimension along which outcome phase activities could have acted.

- The qualitative data indicate that a big incentive for constructing latrines relates to female family members' privacy and honour issues, which are perceived to be violated by open defecation practices. A large number of community members reported that due to high population density and depletion of agricultural fields, there are now fewer private spaces in the villages.

(2) What is the impact of outcome phase activities on the hygiene behaviour of output phase beneficiaries? (secondary research question)

- Outcome phase activities have no significant impact on the reported hygiene behaviours. At the end of the output phase almost 90% of beneficiaries reported washing hands with soap after defecation – by the end of the outcome phase, this percentage had dropped slightly, by a few percentage points, in both treatment and control groups. Much fewer beneficiaries (around 45%) reported improved handwashing practices before having dinner, both at outcome phase baseline and endline, with no detectable difference between those treated and the control group.

- The level of hygiene knowledge was generally found to be high at endline. Almost 90% of female respondents reported that it is important to wash hands before dinner, and 75–80% reported that it is important to do so after defecating. As with behaviour, outcome phase activities did not affect knowledge significantly.

- The outcome phase activities also did not affect the availability of handwashing facilities (with or without soap). Nonetheless, a general positive trend is noted here – with the number of handwashing facilities increasing in both treatment and control communities.

- The results do not differ by output phase modality and by most beneficiary characteristics. We only find some indication of greater effects among more highly educated beneficiaries.

- The qualitative interviews suggest that in many places awareness creation and other related behavioural change campaign activities conducted during the output phase were impactful and lasting. This could imply that there was not much room for outcome phase activities to impact hygiene behaviour over and above the output phase activities.

(3) How do the above impacts differ depending on the <u>output phase modality</u> chosen by the implementing partners?

- We find very limited and inconclusive evidence that variation in output phase implementation modalities led to different outcome phase results, in either the sanitation or hygiene results. Though, as highlighted above, there were significant differences by province[14]. However, these differences are seen to be related primarily to environmental conditions, rather than differences in implementation approach.

- While the quantitative results suggest that outcome phase activities were more effective in villages where demonstration toilets had been provided, the qualitative results indicate that this was driven by the nature of the communities which received the toilets, rather than by the effectiveness of the demonstration toilets themselves.

Key lessons and recommendations

It should be stated that the use of the terminology 'output phase' and 'outcome phase' is programme-specific. In terms of generalising the results, these can also be thought of as activities related to the initial promotion and the subsequent construction of latrines ('output phase activities'), and activities related to promotion and support following the construction of latrines ('outcome phase activities').

The report highlights that about 17% of beneficiary households had at least one member defecating in the open at the start of the outcome phase, and this percentage was much larger when the endline survey was conducted. We also find that this is primarily (though not always) driven by these households not having access to a functioning toilet[15]. This suggests there was some scope for improvement in the output phase promotion and mobilisation efforts and that not all output phase beneficiaries can be assumed to be using a latrine.

Most fundamentally, the findings confirm a need to focus on initial construction quality (in terms of latrines becoming non-functional over time) as basic latrines reportedly collapsed due to heavy rains. This is not a surprising finding, and is in line with the recent literature on community-led total sanitation (CLTS)[16] from other countries. While there was significant slippage in Sindh, the results suggest that the outcome phase activities were effective in reducing the rate of slippage in this province. The task of ensuring initial construction quality through programming is complex, but opportunities exist. CLTS focuses on encouraging households to construct latrines themselves. The survey results confirm that the targeted populations were largely poor[17] – a critical factor in considering the capacity of rural households to construct durable latrines.

- *Recommendation 1.1: Ensure in programming that initial latrine construction quality is sufficient to withstand likely environmental shocks; and relatedly*
- *Recommendation 1.2: Post-construction follow-up should be retained as a key programme element as this study provides evidence that this is effective in the context of widespread latrine collapse. Furthermore, follow-up support should be focused on areas where latrine collapse is likely due to environmental conditions – as this is where it is likely to be most effective.*

[14] Different partners were operating in different provinces.
[15] The baseline report (Section 3.2.5) discussed possible reasons why beneficiary households might not have had access to a toilet (although by definition they should have). Rather than repeating these here, we will discuss constraints – and related programme design recommendations – that households reported with respect to the construction (and maintenance, hence relevant with respect to sustainability) of toilets.
[16] See Crocker *et al.* (2017) for evidence related to CLTS implementation in Ethiopia and Ghana.
[17] The average beneficiary household in our sample group comprises seven household members, earning around US$1,000 per annum.

Two important constraints to toilet construction that were reported by households were lack of finance and lack of access to toilet construction materials. Time and again, households reported that the construction of a quality toilet demands an investment beyond their financial capability – and during the qualitative research the expectation of subsidy was found to be high. Construction of very basic and low-cost latrines was one of the solutions used by local-level implementers to encourage community members to overcome the self-reported financial constraints. It should be stated that the construction of this type of basic latrine is likely to have contributed to the initial programme impact[18], and later in re-constructing collapsed latrines. As such, instead of proposing significant departures in programming, the recommendation would be to focus on complementing the approaches used to overcome specific constraints[19]. Regarding the former, given the controversies around subsidies for sanitation, the use of microfinance for sanitation investment has been tried as an alternative (and also as a complement to subsidies, see Augsburg et al., 2018) and is being used more and more in different contexts.

- *Recommendation 2.1: Ways to reduce the cost of toilet construction should be explored. Such exercises have been undertaken in other contexts, such as by WaterAid in Nigeria.*
- *Recommendation 2.2: The potential use of microfinance for sanitation investment should be further explored. It is increasingly being used as a complement to programming in other contexts. Nonetheless, the effectiveness of providing microcredit is largely dependent on the financial situation of households and it is possible that many households may be ineligible or require a different type of support.*

Respondents to the qualitative research in the three districts further highlighted that they did not have easy access to hardware shops selling sanitation equipment and materials.

- *Recommendation 3: These findings suggest scope for improving access to sanitation markets and materials. While the programme included the set-up and facilitation of sanitary marts, at least in communities visited for the qualitative research there was low levels of awareness of these. While we are not in a position to generalise these findings, they suggest that a closer look to, and integration of, supply improvements is an important programme ingredient in terms of improving both outputs and outcomes. In cases where it is not realistic for a WASH programme to actively engage with the broader market for materials an explicit focus should be put on making use of the available materials to the best effect, including ensuring there is sufficient capacity built-in, even at the lowest implementation level, for durable construction.*

The qualitative research confirms the important role men play in investment decisions, as well as the relatively greater self-reported benefits of improved sanitation for women in terms of their privacy, safety, convenience, and honour. In discussing the findings programme staff highlighted that in their programme approach there was an explicit and implicit focus on targeting women. However, the evidence suggests that men are less likely to use a toilet – even when they have access. Indicating that this is one area in which the programme could have improved and is an important aspect to consider in future programmes.

- *Recommendation 4: In designing promotion activities ensure that they reach all members of a community. This may include conducting them at times when men are not working in the fields, and developing promotion messages targeted at different sub-groups within a population. Reaching men with promotion activities, while acknowledged to be more difficult, has important implications for women where investment decisions rest with men in the household. In the context of rural Pakistan women's modesty and dignity are important motivations for constructing latrines.*

[18] Though it is out of the direct scope of work of the research team to comment on this

[19] In line with the recent literature on a need to complement CLTS with other approaches, see especially Venkataramanan *et al.* (2018).

Finally, the strategy of providing demonstration toilets to deserving beneficiaries that show a willingness to change behaviour seems recommendable – though care needs to be taken to avoid creating the expectation of subsidy. In some of the communities in Sindh, the local programme implementers reported that construction of demonstration latrines became an issue because other households also started demanding free latrines. On refusal to facilitate such demands, the project had to counter disinformation campaigns and hostility from a few segments of the community.

Lessons and recommendations related to 'outcome phase activities'

Some of the lessons and recommendations above are relevant not only to the achievement of output phase targets, but also to their sustainability. There are two further areas of investigation, particularly pertinent to the outcome phase, which the research suggests would improve the sustainability of outcomes.

The first is that the intensity of follow-up visits in the outcome phase was perceived to be low and their targeting too specific (not just in terms of gender but also in terms of pockets and sub-groups within communities). It seems that within the context of this intervention the structure and workload of the programme's social organisers (SOs), in particular, had a lot to do with the lack of intensity of outcome phase activities. SOs and CRPs also highlighted significant logistical constraints. These concerns could have influenced the degree and quality of social mobilisation activities conducted on the ground, and are likely to have led to suboptimal implementation in some communities. A substantial number of SOs also highlighted a need for further training related to outcome phase implementation.

- *Recommendation 5: In programme design and in planning post-construction follow-up activities sufficient attention should be paid to staff capacity to execute these plans. This relates to additional training or capacity-building (if necessary) to reflect a shift in programme focus, as well as ensuring field staff targets are realistic and that they are adequately resourced to conduct their work.*

The second important point emerging from the qualitative research is that the benefits of the programme were not always community-wide. While a number of communities were reportedly certified open defecation free (ODF) by government; the qualitative findings suggest that this does not always imply community wide-coverage, or that ODF status was sustained. The qualitative research also highlights somewhat flexible definitions of the term ODF as different standards were applied both across and within districts. Aiming for community-wide impacts has significance with regards to the mechanisms of behaviour change (collective social pressure), and equity considerations (by definition it is more inclusive).[20] There is increasing evidence that such an aim is important in order to ensure health impacts (there is increasingly some experimental evidence to support this view[21]).

- *Recommendation 6: Future programmes should include an explicit focus on achieving community-wide impacts. In CLTS ODF status is often used, though this can be an unreliable indicator due to varying definitions, and it usually requires government certification. Where it is not feasible to use ODF as an indicator, programmes should be designed to include a measure related to the extent of programme coverage within communities.*

[20] See especially Cronin *et al*. (2017) for a discussion of the importance of focusing on community-wide impacts in reference to the Sustainable Development Goals (SDGs) – authored by senior global WASH leaders within the United Nations Children's Fund (UNICEF).

[21] See especially Okullo *et al*. (2017), who conclude that ODF status has a significant impact on water quality; and Harris *et al.* (2018). who conclude that in their study child height-for-age had a significant and positive linear relationship with community latrine coverage, while child weight-for-age and household water quality had non-linear relationships. Child growth and water quality were not associated with individual household latrine ownership.

Table of contents

Acknowledgements ii
Executive summary iii
List of tables, figures, and boxes xiii
List of abbreviations xv
1 Introduction 1
1.1 The WRP 1
1.2 Objectives of the overall WRP evaluation 2
1.3 Focus of the RCT Research Study 2
1.4 Contribution of the RCT Research Study to the global literature 3
1.5 Research questions of the RCT Research Study 4
1.6 Related evaluation outputs 6
1.7 Target audience for the evaluation 6
1.8 Engagement with evaluation stakeholders 6
1.9 Structure of this report 8
2 Intervention context 10
2.1 WASH in Pakistan 10
2.2 Institutional context 10
2.3 Geographical focus and context 11
3 The intervention 13
3.1 Intervention logic 13
3.2 Significant features of programme design 13
3.3 Implementation structure 15
3.4 Implementation activities by phases 16
3.5 Social mobilisation process 19
3.6 Programme monitoring data 20
4 Overview of the RCT Research Study design 23
4.1 Why a mixed-methods approach was adopted 23
4.2 Mixed-methods sequencing in the analysis 24
4.3 Evaluation and implementation timelines 25
4.4 Quantitative evaluation component design 28
4.5 Qualitative evaluation component design 36
4.6 Significant variations from the 2015 ‘RCT Research Study Design Document’ 40
4.7 Limitations of the research 41
5 Village profiles and socioeconomic characteristics of beneficiary households 45
5.1 Village profile 45
5.2 Beneficiary households 46
6 Sanitation practices and behaviours in communities 47
6.1 Defecation behaviour 48
6.2 Discussion of the key determinants of latrine usage 60
7 Hygiene knowledge and behaviour (handwashing at critical times) in communities 73
7.1 Hygiene – self-reported behaviour 74
8 Key findings, lessons, and recommendations 80

8.1 Key findings 80
8.2 Lessons and recommendations 82
References 86
Annex A Evaluation stakeholders engagement 88
Annex B Revisions of power analysis assumptions 91
Annex C Definition of key indicators 93
Annex D Robustness Defecation behaviour - robustness check on reporting 96
Annex E Descriptive statistics tables 101
Annex F Heterogeneous impacts by output phase modality 104
Annex G Details of SAWRP outcome phase activities 111
Annex H Terms of reference 118

List of tables, figures, and boxes

Tables

Table 1: The WRP Supplier programmes 1
Table 2: Research questions and relevant sections in the report 9
Table 3: Development Indicators Pakistan, study Provinces and districts 12
Table 4: output phase results in Pakistan 16
Table 5: Outcome phase targets 16
Table 6: Activity reporting by partners – implementation in treatment areas 21
Table 7: Experimental design – number of clusters 31
Table 8: Number of respondents per district and treatment arm 32
Table 9: Attrition 33
Table 10: Sample balance 35
Table 11: Qualitative instruments administered for each participant group 37
Table 12: Impact estimates – Defecation behaviour (household level) 48
Table 13: Heterogeneous impact estimates, defecation behaviour (household level): by province 52
Table 14: Heterogeneous impact estimates, toilet ownership: by province 53
Table 15: Heterogeneous Impact estimates, defecation behaviour (random respondent): by respondent characteristics 56
Table 16: Heterogeneous Impact estimates, defecation behaviour (household level): by output phase modalities 58
Table 17: Impact estimates – hygiene behaviour (main respondent) 74
Table 18: Impact estimates – hygiene knowledge (main respondent) 77
Table 19: Impact estimates, hygiene indicators (household) 78
Table 20: Key indicators related to defecation behaviour 93
Table 21: Construction of Hygiene indicators – self-reported behaviour 94
Table 22: Construction of Hygiene indicators – observable indicators 94
Table 23: Overview of survey modules 95
Table 24: Descriptive statistics – Defecation behaviour (random respondent) 96
Table 25: Impact estimates - Defecation behaviour (random respondent) 97
Table 26: Heterogeneous Impact estimates, defecation behaviour (random respondent): By Province 97
Table 27: Descriptive statistics – hygiene indicators (random respondent) 98
Table 28: Descriptive statistics – hygiene indicators (random respondent) 99
Table 29: Impact estimates, hygiene indicators (random respondent) 100
Table 30: Descriptive statistics – Defecation behaviour (household level) 101
Table 31: Descriptive statistics – hygiene behaviour (main respondent) 101
Table 32: Descriptive statistics – hygiene behaviour (random respondent) 102
Table 33 : Impact estimates – Hygiene behaviour (random respondent) 102
Table 34 Descriptive statistics – hygiene knowledge (main respondent) 103
Table 35 Descriptive statistics – hygiene indicators (household) 103
Table 36: Heterogeneous Impact estimates, defecation behaviour (household level): by NRSP community presence 104
Table 37: Heterogeneous Impact estimates, defecation behaviour (household level): By receipt of demonstration latrine 104
Table 38: Heterogeneous Impact estimates, defecation behaviour (household level): By presence of community sanitation promotion 105
Table 39: Impact estimates – Toilet ownership 105
Table 40: Heterogenous Impact estimates, hygiene behaviour (random respondent): By Province 106
Table 41: Heterogeneous Impact estimates, hygiene behaviour: by output phase modalities 106

Table 42: Heterogenous Impact estimates, hygiene behaviour (random respondent): By NRSP community presence 107
Table 43: : Heterogenous Impact estimates, hygiene behaviour (random respondent): By receipt of demonstration latrine 107
Table 44: Heterogeneous Impact estimates, hygiene behaviour (random respondent): By presence of community sanitation promotion 108
Table 45: Heterogenous Impact estimates, handwashing after defecating (random respondent): By respondent characteristics 108
Table 46 Heterogenous Impact estimates, handwashing before eating (random respondent): By respondent characteristics 109
Table 47: Comparison of the research themes again the study's feasibility and aims 127
Table 48: Assumptions stated to be underlying DFID's TOC 130

Figures

Figure 1: Graphical representation of research design vi
Figure 2: Geographical focus of the RCT Research Study 11
Figure 3: Overview of the WRP timelines 25
Figure 4: Graphical representation of research design 27
Figure 5: Descriptive statistics – defecation behaviour (household-level) 49
Figure 6: Descriptive statistics – defecation behaviour , by province 52
Figure 7: Descriptive statistics – ownership of functioning toilet that is used, by province 54
Figure 8: Marginal effects by age – defecation behaviour of random respondent 56
Figure 9: Descriptive statistics – hygiene behaviour (main respondent) 75
Figure 10: Descriptive statistics – hygiene knowledge (main respondent) 77
Figure 11: Descriptive statistics – hygiene indicators (household) 79
Figure 12: Descriptive statistics – hygiene indicators (random respondent) 98
Figure 13: Descriptive statistics – hygiene indicators (random respondent) 99
Figure 14: The WASH Results Programme Theory of Change 131

Boxes

Box 1: Other relevant evaluation outputs 6
Box 2: Partner implementation activities 18
Box 3: Summary findings on sanitation practices and behaviours 47
Box 4: Qualitative data and generalisability vis-à-vis quantitative results 47
Box 5: Defecation behaviour – robustness check on reporting 49
Box 6: Presenting the heterogeneity analysis 51
Box 7: Defecation behaviour – robustness check on reporting 53
Box 8: The use of 'demonstration latrines' in programming 60
Box 9: Summary findings on sanitation practices and behaviours 73
Box 10: Hygiene behaviour – robustness check on reporting 76

List of abbreviations

BBCMs	Broad-based community meetings
BCC	Behaviour change communication
BISP	Benazir Income Support Programme
CAPI	Computer-assisted personal interviewing
CCU	Country Coordinating Unit
CLTS	Community-led total sanitation
CRP	Community Resource Person
DFID	UK Department for International Development
DRR	Disaster risk reduction
E&I	Equity and inclusion
FGD	Focus group discussion
HH	Household
ICC	Inter-cluster correlation
IDI	In-depth interview
IFS	Institute of Fiscal Studies
KII	Key informant interview
LSO	Local support organisation
M&E	Monitoring and evaluation
MDGs	Millennium Development Goals
MoU	Memorandum of understanding
MVE	Monitoring, verification, and evaluation
NRSP	National Rural Support Programme
NGO	Non-governmental organisation
O&M	Operation and maintenance
ODF	Open defecation free
OECD DAC	Organisation for Economic Co-operation - Development Assistance Committee
OPM	Oxford Policy Management
PATS	Pakistan Approaches to Total Sanitation
PbR	Payment by results
PEW	Programme Evaluation Workstream
PHEDs	Provincial Public Health Engineering Departments
PKR	Pakistani rupees
RAG	Red-amber-green
RCT	Randomised control trial
SAWRP	South Asia WASH Results Programme, a consortium led by Plan International
SDGs	Sustainable Development Goals
SOs	Social organisers
ToC	Theory of change
ToR	Terms of reference
UC	Union Council
UNICEF	United Nations Children's Fund
VFM	Value for money
WASH	Water, sanitation, and hygiene
WEDC	Water and Engineering Department
WRP	WASH Results Programme
WSP	Water and Sanitation Programme

1 Introduction

DFID contracted three suppliers to deliver the WRP and thereby to provide sustainable WASH services to at least 9.3 million people between April 2014 and March 2018 in 11 countries.[22] The suppliers were the Consortium for Sustainable WASH in Fragile Contexts (SWIFT), SAWRP, and Sustainable Sanitation and Hygiene for All (SSH4A).

DFID also contracted e-Pact to deliver the monitoring, verification, and evaluation (MVE) services for the 12-country WRP. The evaluation component is led by OPM, and the monitoring and verification component is led by ITAD. This report presents the results of an RCT and accompanying qualitative research component, which are part of the wider evaluation component, and it focuses on the effectiveness of activities aimed at increasing the sustainability of sanitation and hygiene outcomes achieved in rural areas of two provinces of Pakistan.

1.1 The WRP

DFID's WRP, a £111 million seven-year programme (2014 to 2021). The programme was divided into two rounds. The first round of the programme, which was budgeted at £70 million and took place between May 2014 and March 2018, and aimed to bring equitable and sustainable water and sanitation services and hygiene practices to 4.5 million people in 11 countries. Implemented was by three NGO-led supplier consortia contracted to DFID under a PbR financing modality.[23] Table 2 provides an overview of the three supplier sub-programmes.

Table 1: The WRP Supplier programmes

Programme and budget	Lead supplier and consortium partners	Countries of operation	Programme components
SAWRP **£24,995,906**	**Lead**: Plan International **Consortium partners**: WaterAid, Water Sanitation for the Urban Poor (WSUP) and Unilever. Water Engineering and Development Centre (WEDC) and Ipsos MORI provided supp on monitoring and knowledge management **Implementing partners**: various national NGOs	Pakistan, Bangladesh	Sanitation; hygiene, including a substantial school hygiene promotion component; and a modest water supply component (all rural)
SSH4A **£24,032,477**	**Lead**: SNV (not a consortium) **Implementing partners**: various national NGOs and government partners	Ghana, Ethiopia, Tanzania, Uganda, Kenya, Mozambique, Zambia, Nepal (Also South Sudan originally, but dropped in 2016 due to fragile-state context)	Rural sanitation and hygiene only
SWIFT **£19,668,078**	**Lead**: Oxfam **Consortium Partners**: Tearfund, WSUP; the Overseas Development Institute (ODI) supported monitoring, evaluation, accountability, and learning **Implementing partners**: Practical Action, Sanergy, Concern Worldwide, and various national NGOs and public water utilities	DRC, Kenya. (Also Liberia originally, but this was terminated in 2014 due to the Ebola crisis)	Water supply, sanitation and hygiene (urban and rural)

[22] Two countries were dropped from the original 13-country programme.
[23] In 2017, the WASH Results Programme was extended for the delivery of additional outputs and outcomes in nine of the original countries. All further mention of the WASH Results Programme however only refers to the first round implemented from 2014 to 2018.

1.2 Objectives of the overall WRP evaluation

In June 2014, DFID made a commitment to implement greater use of PbR for the funding of international development programmes, and DFID signalled the need for more learning on how PbR programmes can be effectively implemented, and under what circumstances.

The use of PbR in the WASH sector specifically is still relatively new, and the use of the modality represented in the WRP was one of the first times it was used at scale in the WASH sector. Where it has been applied, this has mostly been in the form of results-based aid to a government agency. The use of PbR for the contracting of international NGOs to deliver community-based WASH projects at scale was, at the time the WRP was designed, rare, and still remains relatively uncommon. Lessons from the evaluation are likely to be of great interest not only to DFID and the contracted suppliers, but also to WASH sector stakeholders and international NGOs generally.

The evaluation of the WRP is therefore primarily being **undertaken for learning purposes**, not as an accountability exercise[24]. The objectives of the evaluation are to assess:

i. whether the programme successfully achieved its stated objectives;

ii. the influence of programme design, including the PbR modality, on this achievement; and

iii. lessons for applying PbR in WASH programming in future.

1.3 Focus of the RCT Research Study

The RCT Research Study is a sub-component of the overall evaluation, focusing on one specific aspect of implementation of one of the three supplier consortia. The RCT Research Study relates to one of the three suppliers – the SAWRP consortium, which is led by Plan International, with WaterAid, Unilever, and Water and Sanitation for the Urban Poor as implementing partners; and the Water Engineering and Development Centre (WEDC) and Ipsos MORI as advisory partners. SAWRP operates in Bangladesh and Pakistan. The RCT Research Study focuses on SAWRP in Pakistan.

SAWRP operates in Bangladesh and Pakistan, but the RCT Research Study only covers SAWRP in Pakistan – and within Pakistan focuses on the implementation of one local implementing partner[25]. Within Pakistan, SAWRP operates in a total of 12 districts. Plan International is active in nine districts, while WaterAid works in three districts. The RCT Research Study focuses on three districts: two in Plan International implementation areas and one in WaterAid implementation areas.

The RCT Research Study focuses only on the second half of the programme implementation in Pakistan. As is discussed in further detail in Sections 3 and 4.3 the programme was implemented in two phases. During the first phase (the 'output phase') SAWRP was paid by DFID for the delivery of output-level results. During the second phase of the programme (the 'outcome phase') SAWRP was paid by DFID for the delivery of outcome-level results. The RCT Research Study focuses on the outcome phase implementation only. The RCT Research Study focuses on assessing the effectiveness of the outcome phase activities related to rural sanitation and hygiene behaviour, and as such speaks only to objective (i) stated above**.** It does not seek to address objectives (i) and (iii), i.e. aspects related to the effectiveness of using PbR as a contracting modality. This is considered by the evaluation under a different workstream – the Programme Evaluation Workstream (PEW).

[24] For the original terms of reference for the evaluation, see Annex A of the Midline Evaluation Report.
[25] The justification for this is addressed in the 'RCT Design Document' with regards to methods. Specifically that the RCT Research Study's design is seen to be more useful when focusing on a more homogenous implementation approach.

Justifications for the focus of the RCT Research Study are discussed across Section 1 of the 'RCT Design Document', which also discusses how the RCT Research Study fit into the overall evaluation of the WRP, and the research questions in relation to the OECD DAC evaluation criteria.

Annex H presents information related to the focus of the RCT Research Study with respect to the terms of reference (ToR). Annex H reproduces the original ToR; Annex H.1 presents the clarifications to the ToR as agreed with DFID; and Annex H.2 documents the rationale underpinning the selection of the research questions, as later agreed with DFID and formalised in the 'RCT Design Document'. It is important to note that the original ToR included the requirement to conduct an 'impact evaluation'[26]. The timing of the contracting was such that it was not possible to conduct an impact evaluation related to the output phase implementation – this was one of the factors that led to the focus on the outcome phase implementation.

It is important to state from the outset that the RCT Research Study has a narrow focus with respect to the WRP as a whole. It is best viewed as a sub-component of the evaluation focused on specific research questions as opposed to an evaluation of the programme as a whole. As such, there are many aspects of supplier implementation not reflected in the assessment. Notably, the RCT Research Study focuses on implementation in the outcome phase and therefore does not assess the impact of output phase implementation[27]

1.4 Contribution of the RCT Research Study to the global literature

Interest in the sustainability of sanitation and hygiene behaviour has been growing over recent years amongst policymakers, practitioners, and academics alike. This increased interest is at least partly driven by a set of disappointing impact evaluation studies in the field of sanitation that failed to show health impacts[28]. These sparked an intensified focus on improving achievements in sanitation uptake and continued behaviour change with respect to usage and handwashing – perceived to be necessary first-stage conditions for health improvements to be achieved.

At the time of the design of the RCT Research Study, several rigorous impact evaluations that aimed to understand how to effectively increase sanitation had just been completed or were ongoing[29]. A number of studies have analysed factors that are correlated with the sustainability of WASH outcomes, primarily focusing on factors that sustain the ODF status of communities. These studies provide some guidance as to what factors support and hinder achievement of sustained ODF status. One such example is a study conducted by the Water and Sanitation Programme (WSP) of the World Bank in Bangladesh in 2011, which assessed the correlation between programmatic and social factors and sustained sanitation[30] outcomes. The study's results stress, in particular, the

[26] Which was interpreted by the Evaluation team in the context of the OECD DAC criteria which situate 'impact' as the causal effect of the programme on indicators of interest. The programme's effect can relate to medium-term 'outcomes' or to longer-term 'impacts'. In both cases we would refer to these as "impact" evaluations.
[27] Though as will be discussed later the data collected reflect some aspects of implementation in both phases.
[28] See for example the evaluation studies accompanying the Scaling Up Rural Sanitation Programme launched in 2007 by WSP, in collaboration with local and national governments of India, Indonesia, and Tanzania.
[29] Examples include: (1) Guiteras, Levinsohn and Mobarak (2015), which tested through an RCT three different approaches aimed at increasing the use of hygienic latrines; (2) three impact evaluations conducted by Plan International in partnership with the University of North Carolina Water Institute and its local offices in Kenya, Ethiopia and Ghana to test (through RCTs) three modified approaches to their currently implemented, standard CLTS model; (3) a study conducted by WaterAid in collaboration with the Institute for Fiscal Studies (IFS), London, on the effectiveness of CLTS in the Nigerian context as well as the scope of augmenting the CLTS modality with a sanitation marketing approach; (4) a study conducted in India by WSP, again in conjunction with the IFS, testing the role that microcredit for sanitation can play and the importance of creating awareness in parallel.
[30] Another example is the 'WASH Benefit Study' undertaken by the Gates Foundation in Bangladesh (and Kenya), but this may be less relevant – as it does not explicitly focus on sustainability and follow-up or monitoring.

importance of 'follow-up visits'[31], the poverty status of the household, and access to private sector providers.

However, little is known about the relative effectiveness of specific activities, as postulated in another study conducted by WSP (2012) in Indonesia, which concludes that '*while a number of factors can be associated with ODF outcome achievement and sustainability, no single factor […] guaranteed ODF achievement. It is not possible to rank them in terms of importance, although some are associated and reinforce each other.*' The factors referred to here are regular monitoring and follow-up activities, high social capital, quality of triggering, triggering response, access to information about affordable sanitation products, and access to latrine supplies.

To the best of our knowledge at the design stage of the RCT Research Study there were no other rigorous impact evaluations being conducted that evaluated approaches with a specific focus on the specific role of follow-up activities in achieving sustainable WASH outcomes. As such, this study was expected to fill a key gap within the literature[32]. Furthermore, since the study's design, recent meta-analyses of rural sanitation programmes have stressed the need to more clearly focus on the effectiveness of specific activities. A recent systematic review by Freeman *et al*. (2017) concluded: '*This review confirms positive impacts of sanitation on aspects of health. Evidence gaps remain and point to the need for research that rigorously describes sanitation implementation and type of sanitation interventions.*'

1.5 Research questions of the RCT Research Study

The study concentrates on the effectiveness of the outcome phase activities of SAWRP in rural sanitation, with a secondary focus on hygiene. We describe these activities in more detail in Section 3. While the RCT Research Study's focus is on the outcome phase activities, we also, primarily through qualitative work, draw a link to the role of the output phase.

The primary research question of the RCT is:

(1) What is the impact of outcome phase activities on the sanitation behaviour of output phase beneficiaries?

Our secondary research questions are:

(2) What is the impact of outcome phase activities on the hygiene behaviour of output phase beneficiaries?

(3) How do the above impacts differ depending on the output phase modality chosen by implementing partners?

As discussed above, answering these research questions is expected to fill a key gap in the literature surrounding widely used promotion approaches. The research questions were established in consultation with DFID and SAWRP. Annex H.2 provides details of some of the decision-making underpinning the focus. For the thematic focus of the research study, we considered several options. The choice of the proposed research is the result of a careful process of interaction between stakeholders, striking the balance of meeting four key aims:

i) the research questions should be relevant to the sector and contribute to the current status of the literature – specifically, they should speak to one or more of the key knowledge gaps identified in the DFID (2013);

[31] In the context of rural sanitation implementation this is broadly understood to mean promotion activities taking place with households that had constructed latrines in response to initial promotion, i.e. focusing on latrine use.
[32] Particularly that employing experimental designs.

ii) the research questions should link to the WRP implementation;

iii) the design should be rigorous so as to be able to make a significant contribution; and

iv) the design should fit the implementing partners' timeline and should imply only minimal adjustments to their implementation approach.

Early on, it was decided that an assessment of the overall PbR modality could not be a focus of the RCT Research Study: at least in part because the establishment of a control group was deemed infeasible and also did not fit with the timings of discussions. Two important challenges were faced at the design stage. First, the majority of outcome phase activities were undefined. This was the case since consortium partners were very engaged in the output phase and had limited time to think about the outcome phase. In discussing the design, we hence had to strike a balance between identifying research areas that met the above-mentioned criteria, and working with limited information. The second constraint was that output phase activities had already started, which posed challenges in identifying areas in which certain activities had not yet taken place in order to introduce the randomisation of activities in relation to treatment and control. Despite these significant challenges, detailed discussions amongst all stakeholders led to the identification of five possible research questions. The areas covered by these questions are as follows:

1. the role of water points in achieving sustained sanitation outcomes;
2. the role of demonstration latrines in inducing uptake and moving up the sanitation ladder;
3. sanitation marketing and its role in supporting uptake and moving up the sanitation ladder;
4. the role of CRPs in achieving sustainable sanitation; and
5. the role of follow-up activities in achieving sustainable sanitation.

Ultimately, Research Question 5 was selected via a process that considered the research questions in relation to: i) the extent to which they spoke to key gaps in the literature identified by the DFID WASH evidence paper; ii) their relevance to answering questions surrounding sustainability; iii) how robustly they could be addressed by methods that provided an unbiased measure of 'impact'[33]; and iv) their fit with the supplier implementation approach and timelines.

1.5.1 Links between the RCT Research Study and the overall WRP evaluation

It is important to stress at the outset this RCT Research Study forms just one of a number of components of the overall evaluation of the WRP. As such, it does not seek to answer the PEW evaluation questions directly, but rather the design was focused, as agreed with DFID, on addressing the research questions. The linkages between the research questions and the PEW evaluation questions are addressed broadly in Section 1.3 of the 'RCT Design Document'. Further details of the links with the other evaluation activities are detailed in Sections 1.1 to 1.4 and Annex B of the 'RCT Design Document'. Results from the RCT Research Study will primarily feed into the learning around effectiveness, sustainability, and future recommendations.

[33] As defined above

1.6 Related evaluation outputs

Throughout this report references are made to other key outputs related to the research. For clarity, Box 1 provides an overview of these documents and when they were produced.

Box 1: Other relevant evaluation outputs

Outputs related directly to the RCT Research Study

The 'RCT Design Document' – final submission February 2016 – This document outlines the RCT Research Study's design. Importantly, it contains details of the process behind the selection and finalisation of the research questions. It was submitted to DFID in Q4 2015, then a revised submission was submitted in Q1 2016 following a SEQAS review and a review by the DFID WASH Policy team.

The 'RCT Baseline Report' – final submission June 2016 – This details the findings from the baseline data collection. With respect to the RCT Research Study's design the most significant factor was to check that sample balance was achieved across all key indicators.

The 'RCT Implementation report – February 2016' – This report details the RCT Research Study baseline data collection.

The 'RCT Implementation report – April 2018' –This report details the RCT Research Study endline data collection. Reflections are included in relation to data quality based on reports from the field teams.

1.7 Target audience for the evaluation

As outlined in the 'RCT Design Document' DFID and the key suppliers of the WRP are the main stakeholders of the RCT Research Study and the evaluation overall; they are also the main target audience for the evaluation findings. In addition, the learning and findings on the efficacy of PbR modalities for sanitation and hygiene promotion are likely to be relevant to a wider sector audience, both in Pakistan and more internationally. The contribution of the RCT Research Study to the sector literature has been outlined in Section 1.4.

1.8 Engagement with evaluation stakeholders

1.8.1 Engagement to date

The focus of the RCT Research Study was established through a number of consultation rounds with SAWRP and DFID to identify research questions, the study area, and the logistics of study implementation. As part of this process, two scoping trips to Pakistan were undertaken, one in January 2015 and the second in November 2015. During these trips detailed discussions were held with the Country Coordinating Unit (CCU), Plan International, and WaterAid, and in November 2015 discussions were also held with the local implementing partner National Rural Support Programme (NRSP). One of the key activities of the second trip was a mapping of SAWRP outcome phase activities, which was crucial in finalising the RCT design.

Parallel to the stakeholder engagements and discussions, the sampling frame for the study was chosen and the relevant data for sample selection were obtained from the implementing partners. A draft of the RCT Design Document was submitted for review in December 2015 and then, following a review by DFID and SEQAS, was revised and finalised in February 2016.

Following finalisation of the design an MoU was signed between the research team, SAWRP, and the implementing partner (NRSP) – formalising the expectations and responsibilities on both sides. This MoU is included in A.1 and provided a framework for engagement throughout the research.

During RCT Research Study baseline data collection in 2016 there was coordination with SAWRP field staff to identify beneficiary households. At endline, in line with the MoU, data collection was largely conducted independently (although in coordination with the local implementing partners). This was feasible given that sufficient data to locate panel households had been obtained previously. Following each round of data collection a draft of the 'survey implementation' report was sent to SAWRP to review prior to submission to DFID.

Community-level programme implementation staff were also approached and included in the qualitative research study where feasible, after coordination with the CCU and NRSP district offices. This is explained further in Section 4.5.

The RCT Research Study baseline report was shared with partners for comment prior to submission to DFID, and an event was held at the Plan International offices with headquarters staff where the research team presented the key results for further discussion.

A draft of this report was shared with the partners prior to submission to DFID. In addition, events were held in Pakistan and the UK with implementing staff, where the research team presented the key findings, and all parties present discussed these. These events were a very important step in further contextualising the findings and finalising this report.

1.8.2 Planned future engagement

The dissemination of the findings will take place within the evaluation's[34] broader dissemination approach. This is detailed in Section 8 of the 'Evaluation Design Document'. Within e-Pact there is a learning and dissemination team that coordinates dissemination activities. The level of public dissemination and attribution to organisations (in line with the MoU – Annex A.1) will be discussed following finalisation of the report. In addition, the RCT Research Study team are keen to pursue academic publications related to the findings.

[34] Including managing the e-Pacts blog site related to the WRP: https://washresultsmve.wordpress.com/

1.9 Structure of this report

The remainder of the report is structured as follows:

- Section 2 provides details of the implementation context;
- Section 3 takes a closer look at the intervention studied, including its various phases and implementation structures and processes;
- Section 4 provides an overview of the study design;
- Section 5 presents details on village profiles and socioeconomic characteristics of beneficiary households based on data collected as part of the survey rounds and community qualitative research;
- Sections 6 presents the findings related to defecation behaviours;
- Section 7 presents the findings related to hygiene knowledge and practice; and
- Section 8 summarises key findings and presents lessons and recommendations;

- Annex A reproduces the MoU signed by the research team and the implementing partners;
- Annex B details revisions to the power analysis
- Annex C provides details of the definition of key indicators
- Annex D provides details of the analysis of indictors included to check the robustness of the main indicators used in the analysis;
- Annex E provides the data tables for descriptive statistic presented graphically in the report;
- Annex F provides the analysis table related to heterogeneous impacts discussed in the report;
- Annex G provides further details of the outcome phase activities as reported to the research team by the implementing partners; and
- Annex H reproduces the original ToR, clarifications to the ToR; and details the decision making process that led to the identification of the research questions.

1.9.1 Mapping of research questions to report sections

Table 2 (overleaf) provides a mapping of the research questions to the relevant sections in the report under which they are discussed.

Table 2: Research questions and relevant sections in the report

Research question	Sub-questions/themes	Relevant sections in report	Source of information
1. What is the impact of outcome phase activities on the sanitation behaviour of output phase beneficiaries?	**What is the impact on the study's primary outcome, the '*Proportion of output phase beneficiary households in which at least one household member age five or above does not use the toilet the household has access to*'?**	**Section 6.1**	Quantitative impact estimations and qualitative findings
	We further analyse impacts on other sanitation-related indicators, in particular open defecation behaviour of beneficiaries and household toilet ownership and functionality information.	*Section 6.1.2*	
	We explore the key determinants of latrine usage in the study areas.	*Section 6.2*	Primarily qualitative findings
2. What is the impact of outcome phase activities on the hygiene behaviour of output phase beneficiaries?	**What is the impact on the study's secondary outcome, namely the '*Proportion of households that continue to report improved hygiene behaviour before eating and after defecation*'?**	**Section 7.1.1**	Quantitative impact estimations and qualitative findings
	We further investigate impacts on indicators related to hygiene knowledge	*Section 7.1.2*	
	We also analyse impacts of the hygiene-related indicators in relation to individual characteristics (e.g. age, gender).	*Throughout Section 7.1*	
	We further investigate impacts on additional variables capturing the presence of handwashing facilities and observed cleanliness.	*Section 7.1 and Annex D.3*	
3. How do the above impacts differ depending on the output phase modality chosen by implementing partners?	**Impacts on primary and secondary outcomes are analysed by province and implementing agency, as well as by output phase implementation modality and intensity.**	**Section 6.1.2 to 6.1.4 and throughout Section 7.1**	Quantitative heterogeneity analysis and qualitative findings
	What were field-level staff and beneficiaries' experience of the intervention?	*Section 3 and throughout Sections 6 and 7*	Primarily qualitative findings

2 Intervention context

This section outlines the sectoral and institution context in which SAWRP was implemented in the country, before the actual intervention is discussed in detail in Section 3.

2.1 WASH in Pakistan

When programme operations started in 2014, the UNICEF-WHO Joint Monitoring Programme update showed that Pakistan was significantly off-track against the sanitation Millennium Development Goal (MDG) target of 64%, but on-track for the water supply target of 93%. By 2015, the sanitation MDG target had just been met, and the water supply target had almost been met, though total access to improved water supply had not increased over the previous five years. In rural areas, access to improved sanitation had reached 51% and, while this was still low, it represented a 10% increase over the 2010–2015 period.

These data do not reflect major disparities between locations, the lowest access rates being in the poorest and most remote districts. It is also common in many parts of the country for extended families to live in one compound and share a single toilet. Against this backdrop, meeting the Sustainable Development Goals (SDGs) in Pakistan will be a huge challenge.

2.2 Institutional context

Pakistan has a decentralised federal government system and WASH is a provincial subject. There is currently no national water supply or sanitation policy or strategy; neither is there a nationwide sector monitoring system. Instead, each province is responsible for making its own arrangements within a decentralised government framework. That said, there is a *de facto* rural sanitation strategy known as Pakistan Approaches to Total Sanitation (PATS). Based on CLTS but with some additional components, PATS emerged from a programme known as Rural Sanitation in Flood-Affected Districts (RuSFAD), which began in 2010 following major floods that year. Both Plan International and WaterAid played a role in that programme, which was led by UNICEF.

Responsibility for WASH at sub-national (provincial) level is not clearly defined. Provincial Public Health Engineering Departments (PHEDs) have for decades played a lead role in the development of new rural water supply schemes (though they have been less active in supporting operations and maintenance) but until recently they were not active players in sanitation. Meanwhile, there is broad consensus that, following decentralisation, local governments should take on responsibility for WASH, or at least for supporting the long-term use and functionality of facilities post-installation, but there is a long way to go. The anticipated role includes a role in post- ODF follow-up to encourage long-term latrine use, including the maintenance, repair, and upgrading of existing toilets. Local health departments, via Lady Health Workers, already have a role in promoting sanitation and hygiene, but the extent to which this actually happens varies enormously, being more common in the context of externally-supported projects. UNICEF, while working closely with the PHEDs, is helping to develop the role of local governments at district and union (sub-district) levels, particularly in WASH planning and coordination and providing a supportive environment for sustainability.

Government implementation capacity in WASH has been weak for many years – particularly in sanitation and hygiene promotion – and external agencies have relied heavily on NGO partners. A notable exception here is Punjab, the only province where the PHED has complemented externally-funded projects by developing its own PATS programme and investing some US$4 million of

government funds in the initiative. In Sindh a large WASH programme is about to begin, funded through a multi-donor trust fund coordinated by the World Bank: the Multi-Sector Action on Nutrition programme will run in Sindh, with US$26 million in World Bank funding until 2020. Though it is multi-sector in name the focus is overwhelmingly on sanitation[35].

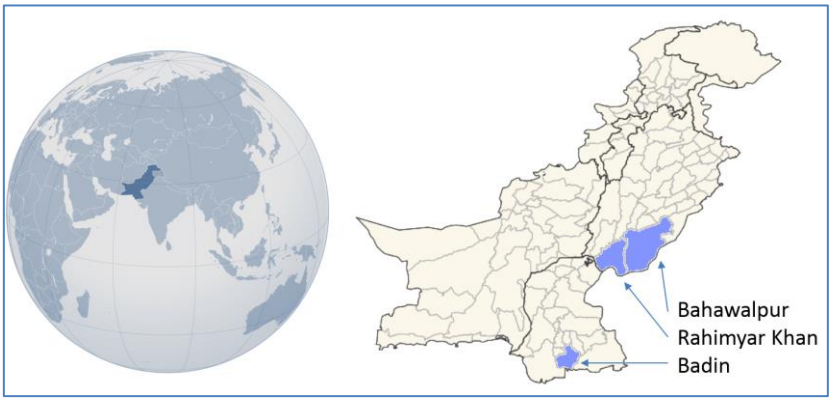
It is important to note that programme activities were largely implemented independently of the government structures – though with coordination at the provincial and district levels, and deeper engagement at the Union Council (UC) level. However, the implementation approach used (predominantly CLTS) was well aligned with the Pakistan government's own approach (PATS).

2.3 Geographical focus and context

The RCT Research Study focuses on three out of the eight districts covered by NRSP under the WRP in Pakistan (details of the sampling approach are given in Section 4.4.3); these are Badin in the province of Sindh and Bahawalpur and Rahimyar Khan in the province of Punjab.

In recent years, Pakistan has been subject to repeated humanitarian emergencies, not least the devastating floods of 2010. Many of the locations where development agencies work have been affected by emergencies, and occasionally still are. As a result, external support has fluctuated between disaster relief and support with a longer-term perspective; this partly explains why hardware subsidies remain a feature of many sanitation and hygiene projects despite the adoption of CLTS. Some SAWRP project locations are flood-prone and while none have been affected by flooding on the scale seen in 2010, there were some damaging flash floods over the course of the programme, particularly in parts of Punjab Province. In contrast to this, Sindh Province (where the programme also operates) is more arid and subject to greater water stress than the other locations where SAWRP operates.

Figure 2: Geographical focus of the RCT Research Study



It is important to keep in mind throughout reading this report that the districts within Sindh and Punjab have distinct characteristics[36]. Most notably, Bahawalpur is located within Pakistan's green belt, Rahimyar Khan is in a desert area, and Badin is located in a coastal zone. Badin is more temperate throughout the year, while Bahawalpur and Rahimyar Khan experience higher summer temperatures between May and August, preceding the monsoon. All three districts experience less than an average of 10 mm of rain in the months January to June and October to December, with an intense period of rainfall in July, August, and September. This is markedly more intense in Badin than is the case in both the districts in Punjab.

[35] See details here: http://projects.worldbank.org/P158769?lang=en *[accessed July 2018]*.
[36] It should be noted this was not part of the sampling strategy – the selection of the districts is discussed in Section 4.4.3

Badin is poorer and less developed on a number of dimensions than those in Punjab. For example the 2015 Human Development Index (HDI) was with 0.412 significantly lower in Badin (Sindh) than in the two Punjabi provinces Bahawalpur and Rahimyar Khan (0.645 and 0.625 respectively) and in line, living standards, as measured on a scale from 0 to 100 by access to clean water, clean fuel, electricity, adequate sanitation, roof quality and basic household assets, was also significantly lower in Badin (31.1) than in Bahawalpur (77.5) and Rahimyar Khan (75.2) (UNDP, 2017). These figures are presented in Table 3, which also places them into the context of study provinces and Pakistan more generally.

Table 3: Development Indicators Pakistan, study Provinces and districts

	Pakistan	Sindh		Punjab		
		Province	Badin	Province	Bahawalpur	Rahimyar Khan
HDI	0.681	0.640	0.412	0.732	0.645	0.625
Immunisation rate	82.1	73.0	73.1	89.0	83.5	83.4
living standard	74.5	67.6	31.1	83.0	77.5	75.2

3 The intervention

This section outlines the interventions under SAWRP, focusing on: salient features of the programme design, the programme's implementation structure, a description of programme activities by intervention phase and the reasoning behind this, detail on the social mobilisation processes, and the programme's overall achievements as per the monitoring data.[37]

3.1 Intervention logic

SAWRP in Pakistan did not have an explicit theory of change (ToC). Similarly, while the DFID developed a ToC for the overall WRP (all three suppliers) this was very generic in nature. 0 reproduces the DFID ToC, with annotations made at the design stage of the PEW. This ToC does not provide a useful basis against which to appraise the specific causal mechanisms related to the intervention studies. Nonetheless it does outline the basic definitions of output- and outcome-level indicators – which are used as a framework in this study. However, the fundamental logic[38] underpinning the intervention is that continued promotion after households construct a latrine (the 'outcome phase activities') related to sanitation and hygiene contributes to sustaining and furthering behaviour change (the DFID outcome-level indicators in their ToC). This logic is based in the broader literature.

3.2 Significant features of programme design

Given that the RCT Research Study is taking place in a wider evaluation context, the PEW, we discuss in this sub-section some of the relevant findings from that workstream as they relate to the RCT Research Study's findings. The PEW findings draw on interviews with global and country-level management staff of the lead suppliers and interviews with all of the implementing partners in Pakistan - two rounds of PEW data collection mirrored the RCT timeline.

Output-/outcome phase split: Neither WaterAid nor Plan International had routinely included a specific outcome phase in their earlier WASH projects in Pakistan and Bangladesh, and there was no clear roadmap for delivering the outcomes when SAWRP began, and indeed for a long time into the output phase. Nevertheless, programme managers considered it very positive that SAWRP had an additional two years to focus on consolidating behaviour change and fostering sustainability. It is worth noting that there was a planned reduction in staffing between the two phases. This is supported from the information obtained from the SOs during the FGDs that while the outcome phase of the programme demanded more behaviour change communication (BCC) and mobilisation activities, the level of effort by the SOs decreased between phases. Most programme logframes or results frameworks in WASH anticipate the delivery of both outputs and outcomes, but having a dedicated

[37] We note that analysis of the WASH Results Programme's alignment with the Paris Declaration principles was neither an objective of this nor of the programme's more general evaluation. Thus, such evidence was not sought out and the evaluators did not come across evidence that the suppliers were following these principles explicitly. Overall programme evaluators nonetheless considered that the WASH Results Programme delivered in line with the Declaration's five principles in the following ways: (i) Ownership: The WASH Results Programme worked within the policy frameworks of developing countries and sought to enhance these by catalysing the WASH in countries; (ii) Alignment: As stated above, the WASH Results Programme worked within local systems and national WASH objectives to the extent feasible; (iii) Harmonisation: The suppliers of each of the sub-programmes are experienced WASH providers for DFID, and in some cases also had, or were at the time of the WASH Results Programme, implementing WASH programmes for DFID in the same or similar areas. There is potentially room for DFID to further harmonise with other donors to further reduce country-specific barriers to sustainable WASH outcomes; (iv) Results: The WASH Results Programme aimed to bring equitable and sustainable water and sanitation services and hygiene practices to 4.5 million people; (v) Mutual accountability: DFID, together with its delivery agents, was accountable for the development results.

[38] In the view of the RCT Research Study team.

outcome phase is rare, and its inclusion was felt by many respondents to be a significant and positive feature. Being accountable for delivering against the outcome indicators reportedly helped to ensure that suppliers remained active in community support and engagement to facilitate the transition from outputs to outcomes up to the programme end.

Output phase timelines: During the midline case study in Pakistan it was the suppliers' perception that they were bound by output phase targets (which were aligned to the MDGs target deadline) for December 2015, and that DFID was pushing them to achieve MDG targets rather than focusing on the sustainability elements of the programme. The general consensus is that the December 2015 deadline focused people's minds on the results, but created a certain amount of pressure on all those involved in delivering the output phase results. Consortium partners and implementing partners estimated that at least six-months were needed to implement the output phase implementation activities, which closely follow the widely adopted CLTS approach, from 'triggering'[39] to the construction of facilities. Through careful scheduling of inputs, implementing partners were able to achieve their 'outputs' in a consistent manner. However, implementing partners indicated that in exceptional circumstances, such as working in a new community, or working in areas where 'direct incentives' have been the 'norm' in past WASH interventions, achieving full coverage took longer, between six and 18 months. In cases where ensuring community engagement was difficult, the implementing partners sought political support through the UC Chairmen or through religious leaders (in mosques). At midline Local Government Department in Sindh remarked: '*[The] SAWRP outcome phase is going to be more crucial and challenging as it will require more engagement with all stakeholders to lead the process towards making the interventions sustainable, especially in relation to the behavioural change component.'*

To an extent, the verified results confirm that the programme targets were indeed achievable. In fact, the level of over-achievement indicates that the programme could have been even more ambitious. Lastly, it is also important to note that at both midline and endline the SAWRP staff in Pakistan highlighted the pressure created by the tight timelines.

At the midline PEW case study in Pakistan all organisations (suppliers and implementing partners) agreed that with hindsight they would have increased the size of their teams, especially the monitoring team, and the overall number of CRPs. It was evident from the field visits and interviews with district-level staff – Plan International in Bahawalpur, Lodhran Pilot Project in Lodhran, Muslim Aid in Rajanpur, and NRSP in Bahawalpur – that the burden of meeting tight deadlines for submitting monitoring reports, providing evidence, and dealing with other quality assurance issues fell on the shoulders of the monitoring officers. Monitoring staff highlighted that working long hours (10- to 12-hour days), was one of the coping mechanisms used.

During the endline case study in Pakistan the pressure that was placed on staff during the output phase was again raised. For example, Lodhran Pilot Project highlighted that it changed its office opening hours policy to facilitate working into the evenings. Additionally, some implementing partner staff reflected that the long hours and increased pressure was more of an issue for female staff – whom in Pakistan often have to balance their work responsibilities against family responsibilities.

Implementation approaches chosen at design stage: The programme objectives were relevant in that they were closely aligned with DFID's design and priorities. In Pakistan, much of the work undertaken by SAWRP represented the extension, expansion, or replication of recent or ongoing projects managed by consortium partners and built upon operational approaches and partnerships tested and improved under these initiatives. The consortium was therefore in a good position to deliver the expected results at scale. Senior managers also highlighted to the evaluators that '*the*

[39] The initial step in the social mobilisation process under CLTS

PbR modality in itself was not an important factor in the decision to engage or not with the WRP'. For them the payment modality was a secondary issue – it was 'a means to an end'.

Operationalisation of the PbR modality: Plan International was the holder of the contract with DFID, and between the three lead partners (Plan International, WaterAid, and Unilever) there was a PbR arrangement where partners were paid pro-rata against achievement, in line with a fixed price per beneficiary. It is important to note that in the case of SAWRP in Pakistan, the PbR modality was not cascaded to local implementing partners, who operated under grant agreements and were required to carry out financial and activity reporting against a workplan, in addition to complying with the results monitoring and verification requirements.

3.2.1 Linkages to other programmes

The implementation of the WRP in Pakistan was largely independent of other initiatives, though the programme was implemented in coordination with provincial government and in line with the PATS approach (see Section 2.2 for details). Additionally, UC-level staff were engaged directly by the programme. Under the PEW, provincial-level government staff were interviewed at the evaluation's midline assessment at the end of the output phase (Q1 2016) and evaluation's endline assessment at the end of the outcome phase (Q1 2018); they noted the good coordination practices of both partners.

3.3 Implementation structure

Plan International in the UK held the contract with DFID and was the consortium lead. Between Plan International and WaterAid in the UK there was a PbR arrangement whereby the partners were paid pro-rata at a fixed price for the results achieved. The flow of funds from DFID to the national offices of Plan International and WaterAid was via the UK offices of each organisation. At the national level in Pakistan Plan International and WaterAid each had a full-time programme manager allocated to the programme. In addition, there was a CCU, which was responsible for aggregating the results data and some aspects of programme management.

In-country, Plan International and WaterAid collaborated with four implementing partners. For both Plan International and WaterAid these were also NGOs, and in all cases the implementing partners were implementing under grant agreements, as opposed to being contracted using a PbR modality. Both WaterAid and Plan International assigned full-time staff to support the implementation at the district level. During the outcome phase the implementing partners were given detailed workplans and produced activity and financial reports for Plan International and WaterAid, as well as collecting the 'results' data, which were then externally verified and reported to DFID for payment under the PbR contract Plan International holds with DFID.

Implementing partners worked through SOs in implementing the intervention. An SO had responsibility for implementation in a number of UCs and the villages/communities within those UCs. Each SO supervised and worked with a network of CRPs, who implemented promotion activities at the community level together with the SOs. The CRPs were from the villages and were volunteers, and the SOs also managed the relationships with the CRPs in their areas.

3.4 Implementation activities by phases

3.4.1 Output phase activities and results

During the output phase CLTS activities were conducted at community level to stimulate the uptake of new facilities by individuals and households. In some programme areas the supply side was stimulated through sanitation marketing, a further key component of the PATS approach, along with school WASH. Table 4 provides an overview of the output phase results for SAWRP in Pakistan.

Table 4: output phase results in Pakistan

Indicators		Targets	Actual achievement
1.1	Number of poor people accessing safe and reliable drinking water sources	90,000	110,341
2.1	Number of poor people accessing basic or improved household latrines	1,000,000	1,205,582
3.1	Number of poor men, women, and children reached by handwashing promotion **in villages and schools**	1,786,000	2,020,793

Source: Monitoring, Verification and Evaluation Framework for Project Outcomes (MVOC) – SAWRP, April 2016.

3.4.2 Outcome phase strategy

It is worth noting that SAWRP's outcome phase strategy was not developed and finalised until Q2 2016, i.e. two quarters into the outcome phase. The late development of the strategy by programme management staff in the UK is attributed to the pressure that staff were under during the output phase. Once the strategy and approach were finalised the implementing partners were given detailed workplans. In addition to *implementing* the agreed-upon activities detailed in the workplans, their grant agreements required them to *report* on these activities, as well as on financial expenditure, and they were required to monitoring their results, which SAWRP would then submit to DFID for payment following third-party verification. In the case of Plan International, the research team did not see an explicit written strategy – rather the interviews and workplans signalled the focus of the outcome phase strategy, and these are detailed below. For WaterAid a specific written strategy was developed, which also contained an overview of the partner workplans. In the outcome phase, Plan International and WaterAid received payments based on three targets. These are detailed in Table 5.

Table 5: Outcome phase targets

Area	Indicator	Measurement
Water	90% poor people across the project districts continue to use reliable, safe drinking water sources	Red/amber/green (RAG) ratings based on reported water availability throughout the year and observed
Sanitation	75% poor people across project districts continue to use basic or improved latrines	RAG ratings based on self-reported frequency of toilet use and observed toilet condition
Hygiene	The percentage point increase of poor men, women, and children who continue to practice handwashing with soap at critical times	Increase in knowledge of all five critical occasions
		Self-reported practice
		Knowledge of critical times (for Plan International and WaterAid). Observation of behaviour (for Unilever)

Source: Monitoring, Verification and Evaluation Framework for Project Outcomes (MVOC) – SAWRP, April 2016.

3.4.3 Outcome phase activities

During the RCT Research Study design phase the research team had limited information on the outcome phase activities, given that the outcome phase strategies of the SAWRP partners were still being finalised. However, Plan International and WaterAid had a reasonable idea of what these would include. At design stage, SAWRP provided assurances that the level of implementation (household, village, and other institutional settings, such as with local support organisations (LSOs) at the level of the UC) was set for each of those four categories and would not change. The activities were divided into four broad distinct groups, the first three of which it was suggested could/would be randomised for the RCT Research Study:

1. *Door-to-door follow-up visits at the household level*. SOs were to be provided with a database of output phase beneficiaries for the area they worked in. SOs were then to put mechanisms in place for households on this list to be approached on a regular basis, to deliver hygiene messages, follow-up on sanitation facility updates, and maintenance.
2. *Village-level activities*: Similar to the door-to-door activities, but not targeted at the household specifically, these were instead activities planned with schools, community groups, mosques, etc. that aimed again at sustaining sanitation and hygiene-related behavioural improvements achieved during the output phase.
3. *ODF declaration*: The outcome phase activities were to support the process of ODF status declaration. The specific term 'declaration', as opposed to 'certification', was used as the facilitation of ODF certification was to involve government engagement, which was to fall under the fourth set of outcome phase activities.
4. *Governance-related activities*: These were planned activities that aimed at strengthening governance structures to achieve sustainable outcomes. The planned activities included the establishment of networks with local government departments etc. and aimed, among other aspects, at prioritising ODF certification. In WaterAid districts district ODF plans were reportedly developed with the local government.

The research team gained a more detailed understanding of the activities during the course of implementation. Box 2 (overleaf) provides an overview of these activities and monitoring data pertaining to their implementation are presented in the next section, which will also detail which activities were randomised and hence form part of the RCT Research Study. However, as compared to the research team's understanding at design stage there are some important points to note:

(i) The implementation approach remained broadly in line with what was envisaged at the time the RCT Research Study was being designed – though we later gained a greater understanding of the precise activities within the four broad categories.

(ii) It is worth noting that both Plan International WaterAid provided community-level incentives for achieving ODF status. These included: constructing additional water points in the community; adding sand filtration systems to existing water points; and providing septic tanks networked to nearby houses.

(iii) It is also worth noting that the monitoring activities of the SAWRP implementing partners were intensive during the outcome phase. The outcome phase results were assessed via two survey rounds, though during this period the partners continued to do intensive monitoring focused on output-specific, rather than outcome-specific, indicators, partially as part of normal programme implementation but also as a risk mitigation strategy against unexpected survey results (on which payment was contingent).

Box 2: Partner implementation activities

Key Plan International programme activities

- **Follow-up visits at the village level** in its eight districts to ensure sustainable usage of the latrines constructed during the output phase. During these visits household latrines were visited and health and hygiene messages were also reinforced.
- **ODF certification** and ODF held celebrations at UC level.
- **Conducting health and hygiene sessions** to reinforce the hygiene messages amongst beneficiaries. Complemented by community-wide activities and some mass media, such as street theatre, sports galas, radio programmes etc.
- **Water point follow-up visits** were conducted to ensure functionality and usage of water points constructed during the implementation phase.
- **Celebrated World Toilet Day and Global Handwashing Day** events in project districts to inculcate the importance of sanitation and hygiene. The World Toilet Day events focused on the theme of 'Toilets and Jobs'. The Global Handwashing Day theme of 2016 events was 'Make handwashing a habit'.
- **Follow-up sessions were conducted in government schools**.

Key WaterAid programme activities

WaterAid key programme activities included the following:

- **Health and hygiene sessions (community-level meetings)** to extend awareness-raising efforts regarding personal, domestic, and environmental hygiene conditions. WaterAid has introduced a new version of health and hygiene awareness sessions, which has a very systematic approach to include the maximum community.
- **Village committee meetings** to plan, meet the community, and mitigate the sanitation-related issues of communities.
- **CRP meetings** (monthly) to review their progress as per their routine workplans and assigned tasks.
- **Sanitary entrepreneurs meetings** were organised to ensure the provision of low-cost sanitary goods at community level so the demand could be met on time.
- **ODF follow-up visits** were conducted to maintain the ODF status of certified villages and to mitigate slippage on open defecation behavioural issues.
- **Global Hand Washing Day and World Toilet Day** were also celebrated within different pockets of communities to highlight the importance of this day.
- **WaterAid also organised street theatres** in the community, providing a unique infotainment kind of methodology which not only amused the audience but also triggered them to fulfil their basic hygiene-related responsibilities.
- **Radio jingles** were aired through different FM channels, along with **newspaper messages** published in local and national-level newspapers.
- **Quarterly broad-based community meetings (BBCMs)** at community levels (this also included village walks by the teams and community together).
- **Positive reinforcement through 'Ideal Mother Groups'** at household and community levels.

3.5 Social mobilisation process

This section provides insight into the social mobilisation process, based on FGDs conducted with the SOs and other community-level interviews. WaterAid and Plan International used a similar strategy for social mobilisation and implementation activities, which in the output phase included one male and one female SO for each UC. In all three districts where data were collected, both suppliers had NRSP as the implementing partner, which explains the almost similar approach in terms of building community structures for strengthening collective social responsibility, like Village Organisations, Community Organisations and finally UC-level LSOs. In the majority of UCs where WRP was implemented, especially in Punjab districts, NRSP already had some kind of presence, mostly through its microfinance network.

In addition to the SOs, a second tier of frontline workers – CRPs – were also part of the social mobilisation process. CRPs were local influential people with knowledge about local conditions and the potential to influence people's behaviour. The CRPs reported to the SO who was accountable to the District Coordinator in Badin and District Officers in Punjab.

According to the SOs, their main responsibilities covered three components: (1) health, hygiene, and sanitation community sessions; (2) construction and use of latrines; and (3) provision of clean drinking water. To achieve these objectives, a range of activities were carried out by the SOs, with support from the CRPs. However, while multiple activities were mentioned, including theatre performances, awareness sessions, household visits, school wash programmes, distribution of printed material like posters, banners and brochures, community events like sports galas (Bahawalpur), these did not appear to be standardised across various communities.

Each SO was responsible for between 25 and 30 villages and were given targets of 50 households within each village. This structure implied that they had to cover one or even two communities in a day in order to ensure monthly visits to target communities. However, this frequency (which might not be considered sufficient) could typically not be achieved. The following limitations were mentioned, leading to recommendations related to staff workload, payments, and other incentives:

(i) the programme did not provide separate transport for the SOs and they had to share a vehicle for field visits with other SOs, which wasted a lot of time and reduced SOs' productive working hours;

(ii) no field allowance was provided, limiting their ability to turn to alternative transport modes;

(iii) according to the SOs their workload was sometimes more than their physical capacity and at times they reached their homes at late hours. A few female SOs in Sindh said that a few of their colleagues had left the job because of this reason and many of them faced problems at home due to late hours in the field; and

(iv) considering the workload of and level of effort by the SOs, they also felt that their salaries were low.

With regard to hygiene promotion sessions, the majority of respondents in the qualitative study confirmed that they were aware of or had attended awareness sessions on health and hygiene. While respondents in Badin generally referred to NRSP CRPs/SOs as the organisers of these sessions, there was less awareness about the organisational affiliations of people behind campaigns in both districts in Punjab. Nonetheless, most respondents did mention the presence of NRSP in their community at one point or another, and credited the organisation with creating awareness on aspects of health and hygiene. Only in one treatment community in Rahimyar Khan

did respondents across the board claim that no organisation had ever conducted awareness sessions on health and hygiene.

> *No male or female is working for health and cleanliness in our community. We have only one lady health worker who comes for polio drops but she has never told us anything about health and cleanliness.* (Female FGD participant, Rahimyar Khan, T1)

This point is confirmed by the quantitative data. The main respondents were asked whether they were aware of any activities in their villages 12 months prior to the interview that focused on the promotion of improvements in sanitation (hygiene was not specifically mentioned). 55% of the main respondents in treatment areas stated that they had heard of activities in their village that promoted improvements in sanitation, and 42% said someone in their household had attended these. While significantly lower, still a sizeable percentage of control respondents said they had heard of (35%) or attended (25%) sessions within the last year. Reported awareness and attendance was significantly higher in Sindh than in Punjab and this was also recollected to have taken place more recently in the former province.

FGDs with programme SOs also reveal that other entry points with communities were used for mobilisation. For instance, clerics of the local community mosques were involved in an attempt to use religion as a motivating factor for maintaining health and hygiene. Other ways included brining doctors on board the awareness campaign, by including notes about health and hygiene in their prescriptions.

3.6 Programme monitoring data

The RCT Research Study drew on SAWRP monitoring data for two main purposes: existing secondary monitoring data collected by SAWRP during the output phase were used as the sample frame for the RCT Research Study; and monitoring data on activities were also collected by Plan International and WaterAid for the specific purpose of the RCT study to assist the research team in understanding implementation intensity in the villages sampled. These data are presented in this section.

3.6.1 Activity monitoring data collected for the RCT Research Study by partners

Table 6 (overleaf) provides a summary of the monitoring undertaken in treatment communities for the specific purposes of the RCT Research Study. It is important to note that these differ slightly from what is presented in Box 2. Annex G provides further detail of the activities summarised in Table 6.

Table 6: Activity reporting by partners – implementation in treatment areas

Org	Level	Area	Activity	# of communities	% of communities	# of times per community	Average # people reached per visit
Plan International	Community	Sanitation	ODF declared	32	100%		
			ODF certified	32	100%		
	Household	Sanitation	Follow-up of reported latrines	32	100%	39	6
			Basic to improved latrine	32	100%	29	4
			Basic to improved latrine though CIF and	10	31%	17	10
		Hygiene	Health and hygiene sessions	32	100%	40	22
			Handwashing demonstrations	32	100%	16	22
			Ensuring availability of handwashing	32	100%	30	2
	Community	Hygiene	BCC activities[40]	32	100%	31	37
		Water	Safe drinking water	15	47%	24	9
			Follow-up of hand pump functionality	15	47%	21	9
			Water quality testing	15	47%	3	8
		WASH clubs	WASH club	15	47%	19	22
			Health and hygiene sessions in schools	14	44%	18	36
		Melas/ celebrations	Celebration of international days related to	6	19%	6	278
			Celebration international events	5	16%	2	390
			Sanitation 'Melas'[41]	2	6%	3	385
		SanMark	Mobile sanitation marts	0	0%		
WaterAid	Community	Hygiene	Hygiene session	29	100%	9	216
		Mixed	BBCMs	29	100%	3	147
		Sanitation	ODF celebrations[42]	0	0%	n.a.	n.a.
			Street theatres	3	10%	1	74
		Capacity	Training on DRR	18	62%	1	5
			Training on equity and inclusion (E&I)	18	62%	1	5
			Training on operations and maintenance	10	34%	1	1

[40] Hygiene sessions, street theatre, magic shows, puppet shows, friendly cricket matches for youth.
[41] Community wide events
[42] Following a provisional analysis of the monitoring data WaterAid clarified that ODF was celebrated at BBCMs – though based on the data it remains unclear as to the proportion of treatment communities that were certified ODF.

The Plan International activity reports highlight that in treatment communities the following activities took place:

- 100% ODF declaration and certification was reported.
- Household-level activities were relatively consistently implemented in all study communities. These included follow-up on reported latrines, upgrading of latrines, and behavioural change campaigns. Household-level hygiene promotion reportedly happened in all communities, and repeatedly. For example, 40 health and hygiene sessions were reported per community over the two year period. This implies approximately one visit every two to three weeks, somewhat higher than what SOs interviewed as part of the qualitative survey reported.
- Community-level water activities happened in about half of the treatment communities.
- Activities related to WASH clubs were, after ODF declaration, the most frequent community-level activity (47% of communities reported on WASH clubs). The 'international days' celebrations happened in very few communities (about 20%) – but reached more people than other activities. There were no Melas reported in the treatment communities.

The WaterAid activity reports highlight that in treatment communities the following activities occurred:

- ODF declaration was not reported on (no data were included in the activity report for any communities) – though ODF celebrations was an activity.
- Again, household-level activities were the focus, and were implemented in all communities, though with notably less frequency (~of times per community), but with considerably more people reached per visit.
- The street theatre plays were not a widespread part of the promotion – reportedly happening in 10% of the communities. Trainings and capacity development happened in around 60% of communities for DRR and E&I – but in about half the number for O&M.

Plan International did not report that any activities took place in control communities in the outcome phase[43], though WaterAid reported via the RCT monitoring templates that there was some implementation by other organisations in control communities as follows:

- In 11 of the 32 communities they note: '*Community Investment Fund/Community Livelihood Fund (CIF/CLF), Activities other than WASH are in progress with Local Support Organization/Village Organization/Community Organization structures, organizations other than NRSP are also working on*.'
- In the other 21 they note: '*In accordance with Government's directives, Lady Health Workers of Health Department visit villages to: aware the community about importance of Safe Drinking Water, child nutrition and importance of health. These LHWs also deliver Health & Health messages*.'

Lastly, it should be noted that there are some challenges in interpreting the monitoring data supplied by SAWRP. **SAWRP**. In one follow-up visit a SO may have covered multiple topics (and therefore be classified as engaging in different activities in the monitoring data). Whereas Plan International included 'N/A' against all cells where they did not have data, WaterAid just left cells blank – sometimes for entire activity groups. Thus, it is not clear if the activity was not undertaken or if it was just not populated with data. For the purposes of the analysis it was assumed that where cells were blank it was because this activity was not undertaken.

[43] It is noted by the research team that monitoring activities of other organisations in the control areas was challenging for partners – a point reflected in the wording of the MoU (see Annex A.1).

4 Overview of the RCT Research Study design

This section gives an overview of the main features of the RCT Research Study design. Further details are available in the 'RCT Design Document'.

4.1 Why a mixed-methods approach was adopted

The approach chosen to answer our research questions outlined above is that of a mixed-methods approach, combining a quantitative RCT with qualitative research. There are two key advantages of using a mixed-method approach for this study. Firstly, in the design, one method can be used to inform the design of another method, in our case the quantitative analysis informed the qualitative study design. Secondly, a mixed-method approach allows us to triangulate the research findings, allowing us to establish and explain patterns, trends, and mechanisms (the 'what' and the 'why').

Under the quantitative component the research questions were addressed using experimental methods. Specifically, we designed and implemented an RCT, which is generally seen as the most robust of all quantitative impact evaluation methodologies – if implemented well and successfully. By randomly allocating an intervention to a sub-set of eligible individuals, households or communities, the assignment process itself creates comparable intervention and non-intervention groups (typically also referred to as treatment and control groups) that are statistically equivalent to one another. This is extremely powerful since a comparison group that is generated through random assignment constitutes a true counterfactual. With a large enough number of observations, the randomised assignment process will produce treatment and control groups that have a statistically equivalent average for all their characteristics, both observable (e.g. wealth or access to markets) and unobservable (e.g. cultural attitudes to sanitation). Thus, if implemented well and on a sufficiently large sample, randomisation creates a comparison group that satisfies three key conditions identified by Gertler *et al.* (2011) as necessary for a credible impact evaluation:

- treatment and control (intervention and non-intervention) groups must share, on average, the same characteristics;
- treatment and control groups should react to the programme in the same way; and
- treatment and control groups should not be differentially exposed to other interventions during the period of the evaluation.

RCTs are an increasingly popular approach among development economists (Deaton & Cartwright, 2018), including to assess effectiveness of WASH interventions (see for example Briceno et al, 2015; Cameron et al., 2014; Guiteras et al., 2015; Pickering et al., 2015). Its popularity lies in its ability to answer the counterfactual question: how would communities that participated in a programme fare in the absence of the programme (or in the absence of a component of the programme)? Or, in the context of our research questions, would the WASH outcomes achieved by the end of the programme's output phase have been more or less sustainable in the absence of the programme's village-level outcome phase activities? Of course, as with any evaluation design, an RCT also comes with drawbacks and risks, which we will discuss throughout the remainder of this section.

While the RCT method is useful in that, if implemented successfully, it provides unbiased estimates of the impact of the intervention studies, its applicability is limited with respect to the generalisability of the results and shedding a light on the mechanisms of intervention impact.

The qualitative research study was intentionally sequenced after the completion of the RCT endline data collection as a complementary qualitative study that aimed to contextualise and establish a catalogue of possible responses to the findings of the survey, and to establish and explain patterns, trends, and mechanisms. Onwuegbuzie and Johnson (2006) provide a useful typology of sources of legitimacy in mixed-methods approaches. The approach used in the present study – primarily sequential-explanatory[44] mixing – addresses a number of important points related to legitimacy from the Onwuegbuzie and Johnson (2006) typology, notably:

- **weakness minimisation** – '*the extent to which the weakness from one approach is compensated by the strengths from the other approach*';
- **sequential** – '*the extent to which one has minimized the potential problem wherein the meta-inferences could be affected by reversing the sequence of the quantitative and qualitative phases*';
- **multiple validities** – '*the extent to which addressing legitimation of the quantitative and qualitative components of the study result from the use of quantitative, qualitative, and mixed validity types, yielding high quality meta-inferences*';
- **sample integration** – '*the extent to which the relationship between the quantitative and qualitative sampling designs yields quality meta-inferences*'; and
- **political** – '*the extent to which the consumers of mixed methods research value the meta-inferences stemming from both the quantitative and qualitative components of a study'*.

More intuitively, a key example of the benefits of our mixed-methods approach is where a quantitative impact finding that might be difficult to interpret can be substantiated or explained with qualitative evidence, providing information on possible reasons and mechanisms behind this finding. In turn, the quantitative study can help shed light on whether this finding is specific to the household with whom detailed discussions were held or whether it is valid for a significant part of the population. Another example is where the qualitative research is able to identify apparent variations across different population groups (e.g. people living with disabilities or specific sub-locations) that are difficult to be explored through the quantitative survey due to sample size restrictions. Such a triangulation exercise with the two approaches increases the validity and credibility of the RCT Research Study.

Our approach to mixing methods is further discussed in relation to the study's limitations in Section 4.7.

4.2 Mixed-methods sequencing in the analysis

In this study, an initial quantitative analysis of the main impact indicators preceded the qualitative research round. These initial quantitative results informed the design of the qualitative research, both in terms of location and respondent selection as well as instrument design. The design, piloting, and data collection using the qualitative tools then followed. We then conducted a light-touch preliminary thematic analysis of the qualitative transcripts, which helped inform the finalisation of the quantitative analysis plan (with a focus on adding areas to be included in the heterogeneity analysis). Following this initial step both the qualitative and quantitative teams conducted their main analysis independently. We then triangulated and synthesised the results through a series of workshops before the final analysis and write-up was finalised. Throughout this report the qualitative and quantitative findings are presented jointly under the relevant thematic headings.

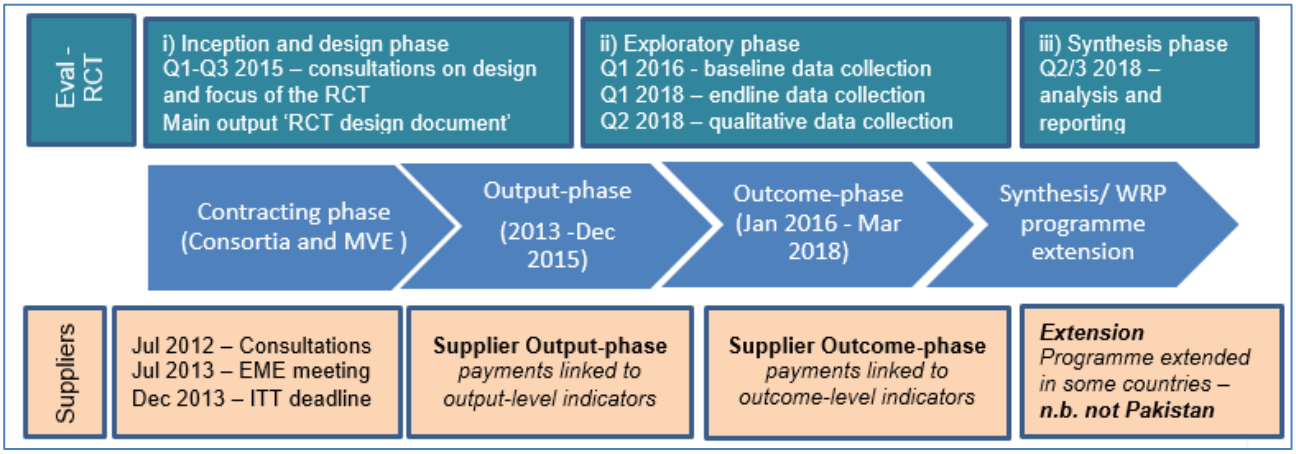
[44] See Ivankova *et al*. (2006) for a discussion of applying this method

4.3 Evaluation and implementation timelines

4.3.1 The RCT Research Study activities and their timing

Figure 3 provides an overview of the timelines for the WRP. The details of the key activities and their timing are provided below. The overall WRP evaluation, including the RCT Research Study, is structured into four phases. The RCTs timeline were aligned with supplier implementation timelines. These are detailed below, with the inception and design phase combined for brevity.

Figure 3: Overview of the WRP timelines



i) Inception and design phase (May – December 2015)

This phase focused on identifying high-level evaluation questions for the overall evaluation via the clarification of supplier activities and of the evaluation objective and scope. This process played a key role in establishing the thematic focus of the RCT.

The focus of the RCT was then established through a number of consultation rounds with SAWRP and DFID, and formalised through the finalisation of the research questions. As part of this process, two scoping trips to Pakistan were undertaken, one in January 2015 and the second in November 2015. During these trips detailed discussions were held with the CCU, Plan International, and WaterAid, and in November also with the local implementing partner NRSP. One of the key activities of the second trip was a mapping of SAWRP outcome phase activities, which was crucial in finalising the RCT design. Parallel to the stakeholder engagements and discussions, the sampling frame for the study was chosen and the relevant data for sample selection were obtained from the implementing partners.

Furthermore, we prepared our team for the study, and ensured that WASH experts would be able to join activities such as questionnaire development and piloting. We conducted initial questionnaire pre-testing in November 2015, informing the household and community questionnaire development. Finally, we obtained ethical approval for this study from OPM's ethics board. Details on these points are provided in the RCT Design Document as well as more generally for the whole evaluation in the Evaluation Design Document.

A draft of the RCT Design Document was submitted for review in December 2015 and then following a review by DFID and SEQAS revised and finalised in February 2016.

ii) Explanatory phase (April 2015 – August 2018)

This phase included the main data collection activities for the RCT, which were concentrated in early 2016 and in late 2017/early 2018. The quantitative baseline survey data collection took place

in February–March 2016, just before the roll-out of SAWRP's outcome phase intervention activities. Details of the survey implementation were included in the 'RCT Research Study Implementation report', which was submitted following the data collection in Q2 2016. During the outcome phase two rounds of activity monitoring were reported by Plan International for the purposes of the RCT. These took place in Q2 2017 and Q1 2018. The quantitative endline survey data collection took place in February–March of 2018, just before the end of SAWRP's outcome phase intervention activities. Details of the survey implementation were included in the 'RCT Research Survey Implementation Report', which was submitted following the data collection in Q2 2018. Following data collection being finalised the data were cleaned and a preliminary quantitative analysis conducted. This informed the design of the qualitative tools, which took place between March and April 2018. Qualitative data collection was completed between April and mid-May 2018 (the start of Ramadan).

iii) Synthesis phase (August–October 2018)

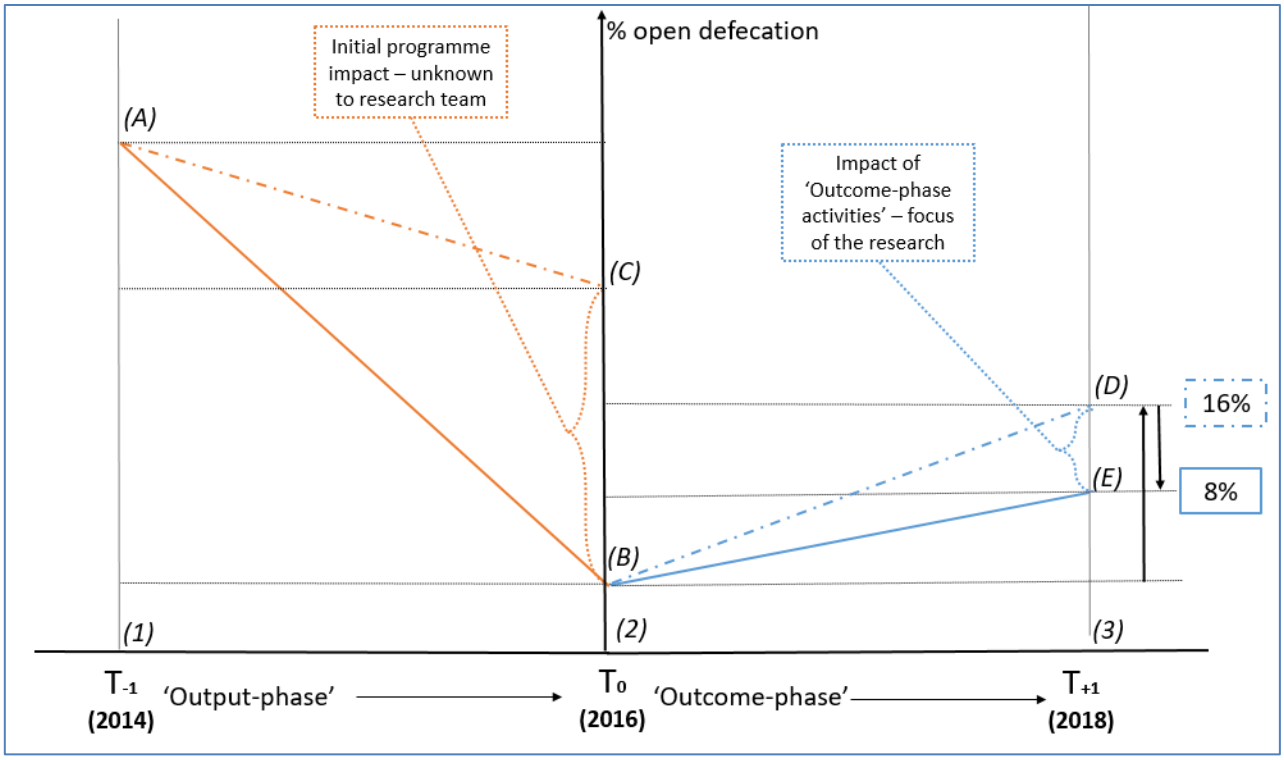
This final stage of the evaluation entails the reporting of the RCT Research Study findings. This report marks the main output of this phase with respect to the RCT workstream. A draft of this report was shared with the partners prior to submission to DFID. In addition, events were held in Pakistan and the UK with implementing staff to discuss the findings.

4.3.2 Significance of features of the research design in relation to interpreting the findings

Figure 4 presents a graphical representation of the research, highlighting some of the features outlined above and key assumptions made. The y-axis represents the percentage of beneficiaries defecating in the open. The y-axis represents time, indicating three key project moments of SAWRP: *(1)* the start of the output phase; *(2)* the end of the output phase[45] and start of the outcome phase; and *(3)* the end of the outcome phase[46]. Importantly, point *(2)*, which is midway through the programme, was the start of the RCT Research Study and the start of data collection. As such, this report cannot say anything about what happened before this point. The figure pictures one possible scenario of what might have happened: when the programme started, a certain percentage of beneficiaries performed open defecation, which is assumed to be A in the figure. Over time, this percentage decreased to *(B)*. Point *(C)* is the assumed percentage had no output phase activities taken place. Point *(C)* in the figure is between *(A)* and *(B)*, hence assuming that some households would have stopped performing open defecation even without output phase activities. Other scenarios are possible. In the extreme case, one could assume that everybody would have continued as usual without the output phase and point *(C)* would be as high as point *(A)*, implying a larger impact of the output phase. **Our data do <u>not</u> allow us to make a judgement about the impact of the output phase activities.**

[45] The first phase of the programme – implemented between 2013/14 and December 2015.
[46] The first phase of the programme – implemented between January 2016 and March 2018.

Figure 4: Graphical representation of research design



The fundamental focus of this research design is to assess the *additional* benefit of the outcome phase activities with respect to key outcome-level indicators. The impact assessment discussed in this report speaks therefore solely to the right-hand side of Figure 1. Point *(B)* was when the outcome phase started and the baseline survey took place. In the design a number of assumptions have to be made, as well as a decision about the minimum impact one wants to be able to detect. In the context of this study, the decision was to adopt a design that would make it possible to detect a minimum reduction in slippage of 8 percentage points (the difference between points *(D)* and *(E)* - assuming that, without the outcome phase activities, a slippage of 16% would occur (the difference between points *(B)* and *(D)*). The assumed level of slippage was based on the available literature in the area – this assumption and its justification are further detailed in Section 4.4.3 and Annex B. The level of the minimum detectable effect decided upon was a case of balancing expected impact against budget; as detecting very small differences implies considerably larger sample sizes. Furthermore, if only very small impacts are anticipated from an intervention the value of conducting research using methods such as RCTs becomes more questionable.

4.4 Quantitative evaluation component design

The quantitative evaluation component is based on a cluster RCT. In what follows, we provide details of the evaluation design, highlighting how it enables us to evaluate the impact of the village-level outcome phase activities.

4.4.1 Experimental design

The research activities were introduced in a manner that both aimed to minimise any potential impact on SAWRP implementation and that allowed us to rigorously assess their impact on outcomes of interest. Specifically, the decision to use a RCT approach, which is generally seen as the most robust of all quantitative impact evaluation methodologies – if implemented well and successfully. By randomly allocating an intervention to a sub-set of eligible individuals, households or communities, the assignment process itself creates comparable intervention and non-intervention groups (typically also referred to as treatment and control groups) that are statistically equivalent to one another. This is extremely powerful since a comparison group that is generated through random assignment constitutes a true counterfactual (Gertler *et al*., 2011). In theory, this should ensure that when we compare the outcomes of treatment and control at the end of the outcome phase, the only difference is due to exposure to these activities and not due to any (un)observed differences between them. It allows one to obtain unbiased estimate of impacts of the treatment on outcomes of interest.

The evaluation design randomised across villages that were included in the output phase, to either receive village-level outcome phase activities (treatment) or not to receive them (control). In total, 123 programme villages were allocated to one of the two experimental arms. The original power analysis provided in the RCT Design Document calculated a needed sample size of 120 villages and 10 households per village (i.e. an overall sample size of 1,200 households). The underlying assumptions and resulting sample sizes were in line with a standardised minimum effect size of 0.2, a commonly used rule of thumb for 'small' effect sizes, first proposed by Jacob Cohen (1977/1988), with empirical support later provided by Mark Lipsey (1990). We update these assumptions in Annex B.

This allocation was done in December of 2015, so that implementing partners could start their planning of the implementation of outcome phase activities. The RCT Research Study endline data were collected at the same time of year as the RCT Research Study baseline survey to: i) align as best as possible with programme implementation timeframes; and ii) minimise the influence of seasonal factors on the measurement of key indicators.

We note that this design implies that the RCT Research Study evaluates the combined impact of three of four outcome phase intervention areas undertaken by the SAWRP consortium in Pakistan (door-to-door follow-up visits at the household level, village-level activities, ODF declaration) within the context of the fourth (governance-related activities). Detailed discussions with the implementing partners revealed that it would be extremely difficult to exclude some villages from the governance-related activities. An alternative would have been to consider randomisation at a higher level (such as UC or *thesil*). However, this was not possible because there were insufficient union councillor *thesils*, since the RCT design requires randomisation over a sufficient number of geographical units (see Section 2.5 of the RCT Design Document for details).

4.4.2 Empirical strategy

This experimental design allows us to answer our research questions by comparing our outcomes of interest in treatment villages (T) relative to that in control villages (C). In technical terms, denoting the outcome as Y, we will compute $\mathrm{E}(\mathrm{Y}|\mathrm{T}=1)-\mathrm{E}(\mathrm{Y}|\mathrm{C}=1)$. In regression terms, we estimate specifications of the following forms:

$$\boldsymbol{Y_{ij}=\alpha_0+\alpha_1 T_j+\theta_j+\varepsilon_{ij}} \quad (1)$$

- Y_{ij} is the outcome variable of interest for household or individual i in village j at the time of the endline survey, when outcome phase activities have been implemented.
- T_j is the treatment indicator, a binary variable equal to 1 if the village was selected to receive village-level outcome phase activities.
- θ_j are strata dummies for the districts.
- ε_{ij} is the error term (which is correlated within communities j, given the cluster RCT design).

In this specification α_1 measures the effectiveness of village-level outcome phase activities in relation to the outcomes of interest relative to the control villages. In addition, we run a specification that accounts for covariates, improving precision. The control variables $\mathrm{X_{ij}}$ are chosen according to those that most explain the variation in our primary outcome variable. To do so we went through an iterative process, regressing the outcome variable onto a list of potential covariates[47]. The variable that reduced the inter-cluster correlation (ICC) most was kept and the process repeated with this variable included. The process continued until the ICC could not be further reduced. In particular, we control for i) the lag of the outcome variable of interest, and ii) whether or not the household has a functional toilet in the dwelling. We further account for the baseline value of the outcomes considered (an ANCOVA analysis), which McKenzie (2012) suggests for outcomes with high autocorrelation, such as toilet ownership. The specification we estimate becomes:

$$\mathrm{Y_{ij}}=\alpha_0+\alpha_1\mathrm{T_j}+\mathrm{X_{ij}}'\beta+\theta_j+\mathrm{v_{ij}} \quad (2)$$

To analyse potential mechanisms behind impacts we furthermore present heterogeneous treatment effects. We note though that the study was *not* powered to have these heterogeneous impacts as primary research questions and not all relevant cross-cutting issues (such as poverty or disability) are addressed in this report. Our approach is as follows: for each sub-group k in the variable for which we want to study heterogeneity in the effect, we define indicators $d_{j\mathrm{k}}$ that take value 1 if household/individual/village i belongs to sub-group k, and 0 otherwise. We are, in particular, interested in differential impacts by output phase implementation modalities (Research Question 3). In this case, di1 equals 1 if WaterAid's output modality was chosen and di2 equals 1 if Plan International's modality was implemented. We then run the following regression:

$$\boldsymbol{Y_{ij}=\alpha_0+\alpha_1 T_j*d_{j1}+\alpha_2 T_j*d_{j2}+X_{ij}'\beta+\theta_j+v_{ij}} \quad (3)$$

In this specification, α_1 and α_2 measure the impacts under the two different output phase modalities. We note that this is the same as considering the impacts by province; this analysis is not able to disentangle whether the geographical or modality specificities are driving any potential findings. The qualitative findings are designed to understand the role of implementation modalities and of individual beneficiary characteristics in more detail.

[47] Those typically chosen in impact analyses such as these relate to household composition and other characteristics, and those that were imbalanced at baseline and the lag of the outcome considered

There are two important points to note. First, a large part of our analysis is conducted at the individual or household level, independent of whether a household was reached by a specific outcome phase activity or not. As such, all estimates are identified through an intention-to-treat (ITT) analysis. Though conservative, this ensures that estimates are not subject to bias arising from selection, i.e. those that choose not to join village-level activities may be different from other beneficiary households in the village. Further, ITT is typically the more interesting parameter, particularly for the purpose of policymaking, since it is reasonable to assume that full compliance will never be achieved and therefore that the ITT is a better measure of the expected benefit of the programme than a measure of impacts on those reached. At the same time, it is worth noting that the implementers made considerable effort to reach every beneficiary with outcome phase activities. Second, we are not conducting any multiple hypotheses testing in this report. Multiple hypotheses testing is relevant when more than one outcome is being considered, and increases in importance with the number of hypotheses tested. If one were to test 20 hypotheses independent of each other at a significance level of 0.05, one would face a chance of 64% of at least one significant result, even if all of the tests are actually not significant. Given that we consider only a few key outcomes, this is not a big concern, and even less so given the lack of impacts we find. However, we raise this point in order to point out an important caveat for the heterogeneity analysis.

4.4.3 Quantitative sampling strategy and sample size

Sampling of clusters

Our evaluation design is a cluster RCT, where we define the village as the cluster unit. By village, we refer to the 'database village', as defined by the consortium partners (from now on simply referred to as 'villages' or 'clusters') and recorded in the output phase beneficiary database, described further below. This is in most cases what is commonly referred to in Pakistan as a 'sub-village' within a 'revenue' village, which was the unit of intervention for many output phase activities (such as CLTS), and is the lowest level geographical unit used in official statistics in Pakistan.

The SAWRP consortium has a number of local implementing partners, but only one (NRSP) works with both WaterAid and Plan International. A choice was made to focus only on NRSP implementation areas ensures a greater degree of homogeneous implementation of outcome phase activities across the whole sample, and also eased management logistics for the study as coordination was thus between the research team and one filed-level implementing agency.

A total of 406 villages were covered by NRSP SAWRP implementation. Our 123 final study villages were selected at random from these. The RCT baseline report provides details on this selection, and replacements and additional villages were added. The main reason for replacement was a low number of beneficiaries in some villages[48].. The final list of study villages includes 118 of the 120 study villages initially selected, plus five of the back-up villages. Two were replaced because they only had one beneficiary household and three were added to increase the sample size as some of the originally selected villages had less than 10 beneficiary households.

The final number of clusters per treatment arm and district is provided in Table 7. The distribution differs by state since we wanted to ensure an (almost) equal number of treatment and control villages per implementing partner (i.e. 61 villages in Badin where WaterAid operates and 62 villages in Bahawalpur and Rahimyar Khan, where Plan International operates). We decided to select the number of treatment and control communities per district in relation to the number of programme villages in these districts, which leaves us with unequal village distribution within districts.

[48] The 'RCT Implementation report – February 2016' provides further details of this process.

Table 7: Experimental design – number of clusters

	Treatment	Control	Total
Badin	31	30	61
Bahawalpur	13	11	24
Rahimyar Khan	18	20	38
Total	**62**	**61**	**123**

The treatment and control allocations were shared with Plan International, WaterAid, and the implementing partner (NRSP) in 26 January 2016, and NRSP suspended ongoing activities in control areas the same week. As mentioned above, these were activities that followed on from the output phase, which were implemented up until the outcome phase strategy was defined. To formalise the need to adhere to the randomisation and not implement activities in control communities, an MoU was signed outlining responsibilities and expectations on both sides – including SAWRP's responsibility to maintain the control communities – at the start of the outcome phase. We are confident that this was honoured by all parties and monitoring data confirm the same.

Sampling of respondent households

Sampling of respondent households was also based on the output phase beneficiary database. This database was used by SAWRP to monitor and report programme performance and hence contains information on every household that constructed a new toilet or improved their existing toilet during the output phase, and which was assumed to be at the same time a hygiene beneficiary. The database was also externally verified using a systems-based approach, as part of the PbR contracting requirements.

Two caveats are worth noting (and will be discussed again further below). The database data do not include information on: (1) households that did not build a toilet during the output phase (and hence were not a 'beneficiary'); and (2) information on usage of the toilet built or improved during the output phase. To address point 1, the research team discussed with DFID the possibility of identifying a random sub-set of non-beneficiary households and expanding the sample frame to include these, but this was decided against for cost reasons (both in terms of identifying as well as interviewing the non-beneficiaries).

At RCT baseline, respondent households were randomly selected from the list of beneficiary households in the RCT study villages. The target was to interview 10 households per study village, implying an overall targeted sample size of 1,230 households in 123 clusters. We therefore aimed to select 12–13 households per study village, so as to be able to replace households that were not available for an interview (which could be for a number of reasons, including migration, unavailability, or refusal). The respondents and replacement respondents were selected randomly, in one draw, through a random number generator, from the set of beneficiary household in a village. Each household was at the same time allocated a random number, which was used to sort households (randomly), with the first 12–13 taken as the main respondents to be targeted and the remaining as back-ups (to be targeted in the order as listed).Table 8 shows the achieved sample size: a total of 1,191 beneficiary households, 594 in control villages, and 597 in treated villages.

Table 8: Number of respondents per district and treatment arm

Province	District	Actual sample size achieved at RCT baseline		
		Control (62 villages)	Treatment (61 villages)	Total
Sindh	Badin	284	298	582
Punjab	Bahawalpur	112	128	240
	Rahimyar Khan	198	171	369
Total		**594**	**597**	**1,191**

Source: Household-level survey

Updated power calculations

In the context of this study, the decision was made for the primary research question to be around the sustainability of sanitation usage. The statistical power analysis therefore focused on the binary variable of whether at least one household member aged five or above does not use the toilet the household has access to. The rationale behind the assumptions made at the time of designing the study are outlined in Annex B. Here, we briefly enumerate and discuss updates to these assumptions.

Assumption 1: The starting RCT baseline value of our key indicator was assumed to be 10%. We found instead that 17% of beneficiary households had at least one member aged five years or above who defecated in the open.

Assumption 2: The ICC was assumed to be 0.1. After accounting for RCT baseline toilet ownership we found the ICC to be 0.14.

Assumption 3: We assumed a slippage rate of 16% in control communities. We found the actual slippage rate to be 17%.

Assumption 4: We assumed that treated households would experience a slippage of 7–8 percentage points. According to descriptive statistics, this rate was 12 percentage points.

Given our sample size of roughly 120 villages and 10 households per village (i.e. an overall sample size of 1,200 household), in order to observe the power of 80%, the slippage in the treatment areas would needed to be below 5.2 percentage points; or given the RCT baseline value of 17.3% the endline value should have been below 22.5%.

4.4.4 Overview of quantitative survey instruments

For the quantitative analysis we worked with two main data sources: community survey data and household survey data. We collected these data specifically for this study and they constitute our primary data. It was necessary to collect these data as the existing data sources were insufficient for answering our research questions. Developed instruments were thoroughly tested and piloted before survey roll-out. Details of these activities are discussed in the RCT Design Document. Details of the data collected specifically for this study are as follows:

***Community survey*:** The village survey was administered to a group of specific community members. It was designed such that different sections of the village survey could be administered to different people within this group: for example, questions on facilities in schools could be addressed to the teacher. In this case, the name of the respondent answering each section was recorded. This survey instrument collected information on the village-level sanitation status; on the main activities of villagers; on the availability and accessibility of infrastructure; on shocks experienced by the village (such as flooding); and on exposure to the output phase activities. At endline, the scope of the village survey was reduced significantly, focusing only on information related to the ODF status of the

village, water availability, and shocks. The rationale for this decision was that village data were primarily intended for contextualisation purposes and for potential heterogeneity analysis.

***Household survey*:** The household questionnaire was administered to the set of selected beneficiary households in the study villages. The questionnaire had either two or three respondents: the main respondent in the household (typically the spouse of the household head or the head him/herself); the household head (if different from the main respondent); and a randomly selected adult household member. The household survey collected information on households' wealth status (information on the dwelling and assets), water sources used, exposure to and knowledge of output phase activities, engagement in village affairs, and some information on financial status (income, debt, and savings). A section on the toilet owned (if applicable) was administered to the household head and covered information on the toilet's type, construction details, information on maintenance, and improvements made and planned. The observation modules covered handwashing facilities as well as toilets owned.

The data collected at RCT baseline and endline used personal (face-to-face) interviewing and computer-assisted personal interviewing (CAPI) on ultra-mobile personal computers and tablets. A CAPI survey design considerably improves the quality and efficiency of any major-scale survey and especially evaluation using complex questionnaires as a standardised measurement instrument. Use of CAPI is of particular importance in panel surveys as electronic questionnaires allow the use of information collected in the previous survey to be used in the follow-up. By feeding back the previously collected information the interviewer can resolve not only inconsistencies within the survey itself but also anomalies with the RCT baseline (for instance, impossible reductions in age of family members, or unlikely changes in household characteristics). Our approach to implementing the surveys comprised three phases; details of these stages are provided in Section 3.2.3 of the RCT Design Document and in the RCT baseline and endline survey implementation reports.

4.4.5 Attrition

Of the 1,191 households interviewed during baseline, i.e. at the start of outcome phase activities, we were able to re-interview 1,132 households, implying an attrition rate of 5% two years down the line. Table 9 lists the reasons why not all households could be re-interviewed two years after the RCT baseline survey. The primary reason (48 households, 81%) is that households relocated and hence could not be tracked.

Table 9: Attrition

Reason	Number of households	
	Baseline	Endline
Successfully interviewed	1191	1132
Relocated	0	48
Merged with another study household	0	4
Duplicate	0	2[49]
Unexplained	0	5
Total	1191	1191

Source: Household-level survey

[49] These households were sampled from two different but adjacent primary sampling units. At the time of the baseline survey, different survey teams covered these two primary sampling units and the household did not report that it was interviewed twice. While preparing for the endline survey, this duplication error was spotted and the household was interviewed only once during the endline survey.

4.4.6 Sample balance

As the allocation of marginal applicants into the treatment and the control group was random, we expected no systematic differences between both groups pre start of outcome phase activities. We confirmed this in the RCT baseline report analysis. Since we were not able to re-interview all households, we repeated the analysis on the post-attrition sample and present the results on key characteristics and indicators in Table 10 (overleaf).

For each variable we present the RCT baseline mean for the control group, as well as the difference in means between the control and treatment group[50], showing in the last column the p-value for a t-test of equality of these means.

The results presented indicate that there are no statistically significant differences between the two groups except a small (2.24) difference in the age of the household head. When conducting a joint significance test for treatment–control balance based on a large set of variables together we find a p-value of 0.27. Thus, there is no systematic overall difference between the two groups and no evidence of imbalance. This result is in line with our finding displayed at the bottom of the table, that attrition is balanced between treatment and control.

[50] We note that characteristics are missing for some households, which is why the sample size is not 1,132 for all variables presented.

Table 10: Sample balance

		Control			Treatment–control	
	Obs.	Obs.	Mean	SD	Diff.	P-value
Panel A: Post-attrition household sample						
Household characteristics:						
Head, age	1129	560	45.59	0.72	-2.24	0.02
Head, male	1129	560	91.25	1.49	0.49	0.80
Head, years of schooling	1129	560	2.99	0.24	0.06	0.84
Head, literate	1128	559	41.50	2.57	0.15	0.97
Number of females	1129	560	3.51	0.10	-0.15	0.24
Number of males	1129	560	3.53	0.11	-0.21	0.12
Number of children under five	1132	561	1.04	0.06	-0.12	0.12
Number of children under three	1132	561	0.60	0.04	-0.07	0.15
Household size	1129	560	7.04	0.18	-0.36	0.12
House is owned	1132	561	92.16	1.76	-1.61	0.55
Number of rooms	1132	561	1.86	0.07	-0.14	0.10
Household yearly income	1115	548	108499.00	7570.00	2554.00	0.79
Benazir Income Support Programme (BISP) poverty score	1129	560	21.51	0.59	0.39	0.62
Latrine facility:						
Improved latrine facility	1092	538	78.07	2.68	-0.09	0.98
Water available in the latrine	1017	504	1.90	0.07	0.05	0.56
Latrine is shared	959	475	23.79	2.63	4.93	0.24
Household with a 5+ year-old member that openly defecates	1132	561	16.58	2.63	0.76	0.84
Hygiene behaviour of main respondent:						
Washes hands before eating	1132	561	88.77	1.61	1.60	0.46
Washes hands before dinner	1132	561	26.02	2.49	-3.96	0.25
Washes hands after defecating	1132	561	73.26	2.89	0.47	0.91
Washes hands after cleaning child that has defecated	635	317	37.54	3.32	-3.89	0.41
Village characteristics						
Is the village's main road a soil road?	119	57	36.84	6.42	11.54	0.21
Is the village's main road in good condition?	119	57	43.86	6.60	-9.99	0.27
Has the village been declared ODF?	117	56	66.07	6.35	-10.33	0.26
Has the village been certified ODF?	71	37	45.95	8.25	4.05	0.74
NRSP been in the community for >2 years	113	55	20.00	5.42	12.76	0.13
Panel B: Attrition						
Not surveyed at endline	1191	594	5.56	1.03	-1.20	0.37

Notes: Unit of observation: household. Panel A: Sample includes only households surveyed at endline. Panel B: Sample includes all households surveyed at baseline, standard error clustered at the village-level.

Source: Household and village-level survey

4.5 Qualitative evaluation component design

This section provides an overview of the design of the qualitative endline research conducted at the community and household level in a selected number of villages in the RCT focal districts. This fed into our mixed-methods approach to the evaluation, as discussed in Section 4.1.

4.5.1 Qualitative workstream design

The qualitative research component was informed by the findings of the quantitative endline survey and aimed to better explain behaviour change, patterns, motivations, and underlying impact mechanisms by understanding the knowledge, attitudes, practices, and determinants of sanitation and hygiene behaviour change in focal communities. It also sought to examine contextual factors that may have affected response to the programme and unanticipated consequences resulting from programme delivery. It was designed as a *behaviour change study* at the household and the community level.

As will be discussed in Section 4.5, the qualitative research was conducted in a total of 12 villages (eight treatment villages and four control villages). To ensure equal representation, four treatment and two control villages were selected from the WaterAid-NRSP communities in Badin district in Sindh, while two treatment and one control village each were selected from Bahawalpur and Rahimyar Khan districts in Punjab where Plan International was working with NRSP to implement programme activities (again, a total of four treatment and two control villages were visited in the province).

The qualitative research was intentionally sequenced to take place after the completion of the RCT endline survey data collection to explain and better understand the major trends in sanitation and hygiene behaviours emerging from quantitative impact findings in the programme's target communities. The qualitative study had two interlocking components, in order to get a holistic understanding of behaviour change at community level:

> (i) **community and household interviews** that, for instance, explored trends and the sustainability of improved sanitation and hygiene behaviours, the difficulties of practising and remembering to practise certain behaviours, variations amongst different individuals and groups, and the influence of social norms and contextual influences, etc; and
>
> (ii) **interviews with community-level implementers** (SOs and CRPs) to understand how local programme modalities may have influenced knowledge, attitudes and practices, and community response to the programme.

These two components were undertaken through one piece of fieldwork completed between April and May 2018.[51]

4.5.2 Overview of qualitative research instruments

The qualitative research utilised three main instruments during fieldwork: KIIs, IDIs, and FGDs. All instruments used both structured and unstructured questions, which allowed for efficient probing of certain aspects of implementation and behaviour change, while also allowing for

[51] Given the programme's operational deadlines, part of the community-level implementers interviews (i.e. discussions with NRSP–Plan International and NRSP–WaterAid SOs) had to be pre-scheduled to take place in January–February 2018 alongside the quantitative survey data collection, while programme staff were still working. NRSP's (extended) contracts with SOs ended on 31 January 2018.

unanticipated and context-specific information to be captured and new assumptions to be explored. These semi-structured interviews were not limited to a set of pre-defined questions and therefore varied between respondents. The interviews were designed to allow a free-flowing conversation, structured around the research questions. Interviewers (eight in total) were trained to encourage the interviewee to seek clarification to increase the quality and reliability of responses. Interviewers were also encouraged to take observations during their community visits and interviews, and to feed these into their conversations, debrief sessions, and the contextualisation of transcripts.

Table 11 summarises the types of instruments that were used to collect information from the different respondents during the qualitative research in each sample village, and the expected totals of such instruments.

Table 11: Qualitative instruments administered for each participant group

Method/ instrument	Respondents	Number per treatment sample village	Number per control sample village	Total sample (12 villages)
KIIs	Community leaders (Lady Health Workers; local counsellors; land owners; community elders) CRPs	2	2	**24 KIIs (16 treatment and 8 control)**
IDIs	Household representatives: (any adult man or woman residing in the house)	4 (2 male and 2 female)	2 (1 male and 1 female)	**40 IDIs (32 treatment and 8 control)**
FGDs	Youth adult (18–39 years)	2 (1 male and 1 female)	2 (1 male and 1 female)	**40 FGDs (32 treatment and 8 control)**
	Middle-aged (40–65 years)	2 (1 male and 1 female)		
	SOs	1 per district		**3 FGDs**
Total per treatment village		**2 KIIs + 4 FGDs + 4 IDIs**		
Total per control village		**2 KIIs + 2 FGDs + 2 IDIs**		
Total across qualitative study sample		**24 KIIs + 43 FGDs + 40 IDIs**		

The various instruments were designed to ensure that triangulation is possible between responses by ensuring that there is an overlap in the questions asked of different respondents (eliciting information on each research question from different perspectives). For example, community members' responses around sanitation and hygiene behaviours and practices can be triangulated against the responses from community leaders and other key informants. In addition, these can be verified and compared to information from local CRPs and SOs working on the implementation side of the programme, in order to get a holistic understanding of the context and underlying dynamics influencing behaviour change.

Field researchers were provided with three days of classroom-based training and practice on qualitative instruments in Islamabad. This was followed by an additional three days of supervised field pilot and debrief in programme communities in Bahawalpur district to test, validate, and finalise the qualitative tools. Piloting in a SAWRP district was prioritised for technical reasons over logistical and budgetary considerations in this regard.[52] The qualitative tools focused heavily on

[52] Piloting in Bahawalpur programme communities, as opposed to other communities accessible from Islamabad, meant that both field teams (including the Sindh team, which then travelled on to Badin) and OPM's qualitative leads were transported to and lodged in Bahawalpur city for several days. Bahawalpur was selected for the pilot because in addition

exploring sanitation and hygiene behaviour change contextualised in programme implementation practices. As such, the full test of instrument flow and administration in context would only be possible with respondents belonging to programme communities, as opposed to interviewing on WASH practices in communities elsewhere in the country.

Selection of communities

Since the qualitative study serves as a complement to the quantitative survey, it draws on the dataset used for carrying out the quantitative survey. Both Punjab and Sindh were included to account for geographical variations in the delivery of the programme, as well as any differences between the engagement strategies of the implementing agencies in the two provinces. Within provinces, communities were chosen from all three study districts (Badin (Sindh), Bahawalpur (Punjab), and Rahimyar Khan (Punjab))

A targeted approach, based on stratified purposive sampling was then used to shortlist a sample of 12 communities, equally divided between the two provinces[53]. Two primary indicators of interest that were used to select communities were: *percentage of beneficiaries who openly defecate*, and *percentage of beneficiaries who have a toilet*.[54] Based on these two indicators, communities at both extremes – high and low percentages – were chosen from each district to ensure a mix of responses, and to allow breadth of the topic and themes to be covered.

To allow for comparison between the two phases of the programme, a total of four control communities were chosen, equally spread across the two provinces[55]. These were communities that were part of the output phase only and hence in theory were not subjected to prolonged engagement strategies. An important consideration in the selection of these communities was proximity to the chosen treatment communities, with close communities selected to cut down on time and logistics requirements in respect of the qualitative study field team.

The above-mentioned method of extreme case sampling of communities with particular characteristics, i.e. high (low) open defecation rates, and high (low) presence of toilets, for the qualitative study allowed the research to explore communities from different contexts in-depth in order to draw out recurring and divergent themes from each case study. Furthermore, it allowed the inclusion of outliers in the study, which would typically be disregarded in quantitative studies.

Sampling of respondents within communities

Sampling for each instrument type (KIIs, IDIs, and FGDs) was done when field teams were present in the selected communities.

- **For KIIs,** the CRP appointed under the programme in each community was identified and interviewed. Where CRPs were not present, a community leader, such as a teacher or village elder, was interviewed instead. In each community, the gender balance of KII respondents was ensured by interviewing an equal number of male and female key informants[56].

to being a programme target district, language barriers were less limiting for members of both the Sindh and Punjab field teams (Urdu is widely understood in communities in Bahawalpur and all field researchers were also fluent in the language) and government permission had already been secured to work in the programme districts.

[53] Of the 12 treatment communities, four were chosen from Badin, two from Bahawalpur, and two from Rahimyar Khan.

[54] We remind readers that this indicator does not provide information on community-level open defecation rates, but is only informative for beneficiary household sanitation behaviours.

[55] Of the four control communities, two were in Badin, one in Bahawalpur, and one in Rahimyar Khan.

[56] Two male and two female key informants were interviewed in each treatment village, and one male and one female respondent was interviewed in each control village.

- **For IDIs**, CRPs and/or community leaders were used to identify households, ensuring the inclusion of a mix of households that practised open defecation and those that did not. In addition to this parameter, IDIs were equally divided between male and female respondents to ensure gender-sensitive responses are accounted for and any gender-based differences can be accounted for.
- **For FGDs**, again, CRPs/community leaders were used to identify a mix of individuals: males (both for the youth (18–39 years) and middle-age (40– 65 years) groups) and females (both for the youth and middle-age females group). Two male researchers facilitated and collected data from male FGD groups, while the female research team members conducted FGDs with female members of the community. For FGD respondents, an attempt was made to include both users and non-users of latrines in each group; however, this was not always possible. Some respondents of IDIs also took part in FGDs.

In order to gain additional insights about the programme, FGDs were also conducted in each of the three districts with SOs appointed under the programme. Although two SO FGDs (one male and one female) had initially been planned per district, given the programme's operational deadlines and limited field team presence at the time when FGDs with SOs were arranged (January–February 2018) it was only feasible to conduct one FGD per district with a mixed group of male and female SOs in each district; as such a total of three FGDs with SOs were held. SOs were invited to a central location (district headquarters or other agreed location) where they participated in FGDs.

An important caveat in respect of respondent selection is that unlike the quantitative RCT survey, which only targeted WRP beneficiary households, the qualitative interviews and discussion were conducted with a wider group of randomly selected respondents/ households in the communities. As such, the information from the qualitative research is not necessarily limited to WRP beneficiaries in the target communities, although it is possible that they were included as part of the randomly selected qualitative respondents.

This was done for both technical reasons related to the design of the qualitative study and due to logistical considerations. For instance, KIIs had to be arranged with community leaders and CRPs regardless of whether they belonged to beneficiary households, to get an accurate understanding of the local context. FGDs were categorised by the gender and age bracket of respondents, while IDIs at the household level were conducted with 'doers' (use latrine) and 'non-doers' (do not use latrines) selected in the community at random. Given that an intensive endline quantitative survey had been conducted just a few months prior to the qualitative fieldwork, this helped avoid the problem of respondent fatigue with the data collection process, as well as potentially limiting bias in terms of what previously interviewed (i.e. for the quantitative survey) respondents may expect the researchers would want to hear about the programme, based on their past interview experience. It also helped provide wider, community-level information on the perceptions, uptake, implementation, and sustainability of the WRP at the village level, which is information the quantitative surveys were not able to provide.

4.6 Significant variations from the 2015 'RCT Research Study Design Document'

This section outlines key design departures in the data collection and analysis from the RCT Research Study Design document agreed in 2015.

Sequencing of qualitative and quantitative components: Initially, the qualitative study component had been sequenced to precede the endline survey and quantitative analysis. This was suggested as at the time of design the research team had only limited information on the intervention approach – not least because the SAWRP consortium was still finalising this. As such, a pre-quantitative endline round of qualitative research was deemed necessary to ensure that the research team had a sufficiently good understanding of the intervention.[57] However, during the outcome phase the research team gained a much more detailed view of the intervention activities through the activity reports produced for the RCT team by Plan International. Additionally, the monitoring data of the programme shed more light on the implementation. As such, it was deemed that a post-endline qualitative round would deliver greater value, especially in terms of helping explain and justify survey data trends and deviations. While this change in sequencing did not change the design of the qualitative approach greatly, some specific changes were made.

Control communities added to qualitative sample: The qualitative sample described in the RCT Design Document comprised only treatment communities (see Section 4 of the RCT Design Document). Three treatment villages were to be selected from each of the three RCT districts, giving a total sample of nine treatment communities for the qualitative study. However, the initial analysis of the endline quantitative survey data suggested fairly limited impact estimates and similar rates of slippage in both treatment and control communities between survey rounds. It was then decided that the (now post-quantitative endline) qualitative research would expand into control communities in order to explore the underlying dynamics and factors leading to these results. As described in Section 4.5, the qualitative sample eventually visited included eight treatment and four control communities across the three target districts.

Qualitative study refocused particularly on behaviour change study at household and community level: The RCT Design Document proposed a two-pronged focus for the qualitative study, looking at: (1) behaviour change at the household and the community level; and (2) conducting a governance and political economy analysis at various levels of programme implementation. As work on the wider PEW workstream was planned and undertaken, it was decided that the governance and political economy analysis was being covered under, and was better suited to, the PEW workstream, rather than in the RCT Research Study. As such, the qualitative study was focused particularly on exploring and understanding behaviours and determinants of change at the household and community level. Nonetheless, as the next point illustrates, community-level implementation influences and stakeholders were included in the research given their programme and supply-side relevance to behaviour change.

Larger sample size and adjustment in sampling determinants and respondent selection: As described previously, the qualitative sample was expanded to 12 communities in order to accommodate and give equal provincial representation to treatment and control communities across the three study target districts. This also led to some adjustments in the number of instruments administered across respondent groups in each community (see Section 4.5). Although purposive selection was still used to shortlist the sample treatment and control communities for the qualitative study, the primary indicators used for selection were adjusted to the *percentage of beneficiaries who openly defecate*, and the *percentage of beneficiaries who have a toilet* (see Section 4.5) after initial

[57] Our understanding of the activities at design phase and the justification for the pre-endline study are detailed in Section 1.6 and Section 4.1 of the design document.

quantitative analysis results came in and due to non-availability of community-level data on the original (somewhat generic) indicators mentioned in the RCT Design Document.

Adjustments in analytical tool (barrier analysis): The RCT Design Document proposed the use of an extensive barrier analysis (sample size of 270 individual interviews) to provide a statistical indication of the barriers and promoters associated with particular behaviours. This would have involved very significant cost, time, and human resource requirements and it was decided that a similar analysis and understanding of determinants of behaviour change could be more efficiently collected by increasing the number and diversifying the target respondents of FGDs, KIIs, and IDIs at the household and community level. In addition, an intensive endline quantitative survey had been conducted just a couple of months prior to the qualitative fieldwork, so this helped avoid the problem of respondent fatigue with the data collection process.

Feasibility of data analysis on NVivoO Software: the RCT Design Document proposed the use of NVivo software for systematic coding and analysis of qualitative data. The methodological efficiency of this software is well established. However, the software has maximum functionality with English language transcripts, while the qualitative data from the field were transcribed (hand-written) in Urdu, given the language command and ease of the field researchers. While translation into English was an option (and was in fact conducted for a sizeable proportion of transcripts), this is a very costly and time-consuming process. The OPM core qualitative team comprised Pakistani nationals who were fluent in both languages and were comfortable working with the Urdu transcripts as well. Therefore it was decided that they would work directly from the transcripts, with an analytical framework developed on MS Word in the interests of time and cost savings.[58]

4.7 Limitations of the research

The key limitations of the research relate to: external validity, 'power', social desirability bias, sampling concerns, recall bias, and language considerations. Details of the limitations, and mitigating actions taken in each instance, are as follows:

External validity (generalisability): This limitation relates mainly to the generalisability of the impact estimations in the quantitative analysis. That is, the estimations are relevant to the particular intervention studied and the context in which it was implemented. The qualitative findings partially mitigate this limitation through providing greater insight into the context and mechanisms of impact. A more detailed discussion of external validity follows, in Section 4.7.

'Power': By design, the quantitative component would only be able to detect impact above a certain threshold. This threshold was established at design stage. Specifically, the study was powered to detect a standardised effect size of 0.2. The key constraint on power is sample size and we relied on assumptions at design stage to establish the reasonable thresholds. Unfortunately, while having made considerably conservative assumptions, our actual power turned out lower than anticipated. We discussed how far they were met, in Section 4.4.3. We do include the lagged value of the dependent variable, which, given its high inter-temporal correlation improves precision (McKenzie, 2012).

[58] As it turned out, the qualitative research had to fit a tight window of opportunity between when the initial quantitative results came in and before the onset of Ramadan and extreme summer temperatures in Pakistan. Given the additional time taken on transcription and translation, it was essential to complete the qualitative coding and analysis processes in time for the first draft report deadline.

Programme implementation in control communities. Based on reporting by SAWRP, the team are confident that there was not implementation of household or community level activities in control communities. However, there remains a possibility that activities of NRSP in the treatment communities may have affected behaviours in control communities (referred to a 'spill over'). This potential mechanism was taken into account in the design phase by checking that treatment and control communities should not be too close to each other. If such spillover nevertheless happened, we would underestimate the program impact. Our data does not allow us to test whether any such spillovers might have taken place. Furthermore there were a set of activities that were implemented at the district and UC levels (notably capacity building and some mass media promotion). This report therefore analyses the effectiveness of community and household level promotion activities that take place in the context of these UC level activities. It is not possible to infer whether the same (or higher) impacts could have been achieved without these UC level promotions.

Qualitative sampling method – indicative results: While qualitative results do not claim to provide statistical validity in the same sense as quantitative findings,[59] it is important to stress that the small sample of communities visited for the qualitative research implies the information provided is indicative. Furthermore, as was described previously in Section 3.4.2.2, due to both technical considerations related to study design and practical concerns the respondents for the various qualitative interviews were selected from a wider group of individuals/households in the programme target communities, and were not necessarily limited to WRP beneficiaries. As such, the qualitative findings report wider community contexts and perceptions around sanitation and hygiene behaviours and experiences with the WRP. The research contrasts experiences within and across communities, without claiming that there is one 'true account or perspective'. Nonetheless, the findings are indicative of how some communities and stakeholders perceive and experience the WRP programme, and as such give insight into sanitation and hygiene behaviour change processes and how these may be improved in the future.

Measurement validity. There are several common limitations that apply to all research instruments (and in our case the tools of both the qualitative and quantitative components) related to measurement validity. That is, how well the data captured by the tools reflect reality. In the case of this study there are several key factors that will have affected measurement validity to some extent in both the qualitative and quantitative research. It is important to state that these measurement limitations are assumed to not be biased in line with treatment and control allocation, and so are not seen to affect the impact estimations.

- **Social desirability bias:** Any research risks response bias, in particular with regard to questions that respondents may interpret as having a 'correct answer'. This is because the research itself may influence the way in which respondents answer questions or speak about the programme, due to power imbalances and the perceived need to say what is expected, rather than what may be the case. The research has mitigated this through triangulating the data using responses from multiple respondents and different instruments. Training also stressed the need for the unbiased delivery of instruments. In addition, during fieldwork the researchers were trained to put the respondents at ease and emphasise that the purpose of the study was not to penalise or directly benefit either them or the programme but to understand perspectives and behaviour change in communities.
- **Recall bias**: As part of RCT Research Study design, both treatment and control communities received output phase interventions between 2014 and 2016, and thereafter treatment communities continued to receive outcome phase interventions between 2016 and 2018. As such, by the time teams for the qualitative study began working in the communities in April–

[59] Its strength is rather to understand each case (in this case a community) in a complete and holistic way. The risk of visiting atypical communities and gaining an incorrect or incomplete understanding of the relevant processes remains, but is mitigated by visiting different types of communities (treatment and control) in different districts, and by paying close attention to ways in which the context of each community may be atypical.

May 2018, treatment communities had not received programme interventions for several months, while it had been a couple of years since control communities had last been visited by the programme. This creates space for recall bias, particularly when reflecting on programme activities and staff behaviour. In addition, the WRP did not have blanket coverage even in the communities in which it worked and second-hand community accounts of behaviours and interventions were sometimes received. These were mitigated and checked through the triangulation approach used in the study, by looking at responses from multiple respondents and different instruments. In addition, researchers were trained to conduct detailed and clear probing on programme phases and activities, though it is unclear how successful this was in working with community respondents.

- **Language:** Although the instruments were originally designed in English, these were translated into Urdu for training purposes and administration in the field was also done in Urdu and/or local languages (Sindhi/Saraiki/Punjabi). During training, the indigene researchers were involved in discussions around terminology, phrasing, and translation – to ensure they had an understanding of the research themes beyond immediate translation. In addition, pre-testing and piloting of instruments in the field helped get the terminology right for some of the most sensitive/challenging questions. This later facilitated precise and standardised delivery during data collection. However, the limitation of *ad hoc* interpretation in probing in order to elicit nuanced responses cannot be completely eliminated. Transcription was also done in Urdu, and the lead researchers at the analysis stage were fluent in both Urdu and English, allowing them to analyse field data and report information with minimal loss or misinterpretation in translation.

Survey data collection. The research team was generally able to work unimpeded to collect the information. The RCT baseline and endline 'survey implementation reports' document the quantitative data collection, with the endline report outlining some issues experienced in data collection and the possible impact on data quality. These were generally regarded as not significant and the key concerns surrounding data quality raised are reflected in the discussion above surrounding measurement validity. Crucially, attrition between survey rounds was lower than anticipated. As data collection was sufficient and in line with what was planned it was not necessary to resort to secondary sources to address gaps in the planned data collection.

Programme staff availability: One component of the qualitative research aimed to capture institutional perspectives and influences on implementation by interviewing programme staff (SOs) at the district level. These discussions were scheduled earlier, in January–February 2018, before the bulk of qualitative research activities took place in April–May 2018, in order to access programme staff before they dispersed at the end of the WRP in March 2018. The research had initially planned to conduct two FGDs with SOs per district (one male and one female FGD). However, when local programme offices were approached, it was learnt that the majority of SO employment contracts with WRP terminated in December 2017–January 2018 and it would be difficult to approach the individuals. As such, given varying programme staff availability, one FGD/group interview was conducted with a (mixed) group of SOs per district.[60]

4.7.1 Generalisability of results

Whilst we believe that the evaluation design as described above provides for internal validity, i.e. that the evaluation is sufficient to provide estimates of impact that can be correctly attributed to programme impact, it is also important to consider the question of external validity or generalisability.

[60] In Badin a mixed FGD was conducted with eight SOs (five males and three females). In Bahawalpur there was also a mixed FGD with six SOs (two males and four females). In Rahimyar Khan a group interview was conducted with the three male SOs who were available.

An evaluation has external validity if the conclusions drawn from the evaluation are appropriate for the total population that will eventually be covered by the WRP. Since we selected our study communities at random from those covered in the study districts, our findings will have external validity for other study communities within these districts. However, they will not be easily extrapolated to non-programme villages within the study districts since SAWRP decided to focus on poor villages in particular. Similarly, the results are not easily extrapolated to other districts given important contextual differences.

Furthermore, the quantitative results will not be generalisable to non-programme beneficiaries – in general but also specifically to non-beneficiaries within our study communities. While this is a limitation of the study, it is one that is not specific to our setting. Extrapolating results from a specific intervention in a specific context is generally a difficult task and so we do not attempt this here.

We are, however, able to rely on the qualitative work to draw conclusions that go beyond the beneficiary population within study communities. This is done with care as the qualitative research component was based on a relatively small sample of 12 communities (eight treatment and four control) across the three districts. However, the fact that the qualitative research was conducted with and without programme beneficiaries makes it possible to put the quantitative findings into the wider village context. For example, we find that even if the large majority of beneficiaries do not conduct open defecation, they often still live in a context where this practice remains widespread.. In some cases it was more appropriate to talk of whether findings were transferable rather than generalisable: researchers needed to investigate context in order to construct an argument that a finding in one setting, or one that was relevant to one group, was likely to apply in another. The risk of visiting atypical communities and gaining an incorrect or incomplete understanding of the relevant behaviour change processes remains, but is mitigated by visiting both treatment and control communities in different districts and contexts, and by paying close attention to ways in which the context of each community may be atypical.

5 Village profiles and socioeconomic characteristics of beneficiary households

The study's RCT baseline report provided a comprehensive picture of the study villages (which are representative of the NRSP villages in the study districts) and the beneficiaries residing within these. We provide here a summary, and refer the interested reader to the 'RCT Baseline Report', which also includes a breakdown by province.

5.1 Village profile

The village-level survey interviews were conducted with, on average, six key participants. During the RCT baseline survey, in 79% of communities, the CRP was present during the interview session, while a council member was present in 30% of interviews, village heads in 26% of interviews, Imams in 14%, teachers in 8%, and masons in 8%. In 56% of the interviews there were also community members other than those listed above present. These interviews were designed to collect information on programme villages as a whole. As such, they provide a picture of the context in which the intervention took place.

Study villages can typically (in 45% of cases) be accessed through paved roads, with concrete or asphalt, with the remaining villages predominantly being accessed by dirt roads (43%). In the majority (56%) of villages it was reported that roads are in a poor condition throughout the year. This percentage increases to 86% when considering only villages that are accessed via dirt roads. Bus services are very uncommon and the most common form of transport to the *tehsil* headquarters is an auto rickshaw (50.9%).

Not surprisingly, then, we find that access to services is far from ideal (with access being defined as within the village or within one hour's travel time). Only 9% of villages were reported to have access to either a primary or community health centre. In Punjab the most commonly accessible medical facility is a private hospital (accessible to 30% of villages), and in Sindh it is a government hospital (13%). In Sindh, none of the villages was reported to have access to a police station, bank, or market. In Punjab, between 11% and 14% of villages have access to these services. Only educational facilities are more easily accessed: 65% of programme villages have access to a primary school. The percentage drops sharply though when one considers educational levels higher than primary education, with only 5% of villages being located within one hour travel of a secondary school.

Turning to water and sanitation, our RCT baseline data reveal that 88% of villages reportedly access improved water sources as their most commonly used source of drinking water (100% in Punjab and 77% in Sindh). In Punjab the most commonly used water source is a hand pump (82%) and in Sindh it is tube wells and boreholes (43%). However, in the qualitative sample it was noted that a noticeable number of households in the research communities in both districts of Punjab and Badin in Sindh complained of brackish or saline potable water in their villages. In one of the treatment communities in Badin, men and women in the FGDs and IDIs reported that the potable water taps installed in their village had been damaged and now they were again dependent on saline water from hand pumps for drinking purposes as well. In another treatment village in Badin, community members said that the tap installed by NRSP now had brackish water in it and as a result they had to walk to the next cluster to collect drinking water.

When conducting the RCT baseline survey, at the end of the output phase activities, two-thirds (66%) of programme villages were self-declared ODF and one-third (29.8%) certified ODF by government staff. In only 7% of villages was it reported that almost no one in their village

uses a latrine. Regarding ODF certified villages, in 100% of such villages in Punjab it was reported that almost everyone uses a latrine. This figure is lower in Sindh (66%), with the other 33% reporting that few to most people use a latrine. Again, in the qualitative sample communities, in two out of four ODF certified communities in Punjab (one each in Rahimyar Khan and Bahawalpur districts), community members said that very poor households still continued to practice open defecation as they could not afford to construct toilets.

A more detailed discussion on open defecation practices is provided in Section 6.2.5, which discusses quantitative and qualitative research findings with respect to our primary outcome of interest, open defecation behaviour. While we also discuss toilet ownership in that section, we want to provide a more detailed picture of the nature and characteristics of latrines in study communities at this point, to provide a comprehensive picture of the context in which the intervention and research is embedded – particularly in view of this having been the focus of the output phase.

5.2 Beneficiary households

Before providing an overview of the average programme beneficiary, we remind the reader that the average programme beneficiary is not representative of the average household living in programme villages. This is because we drew our sample only from beneficiaries and not from villages as a whole, as discussed in Section 4.4.3.

The average beneficiary household in our sample group comprises seven household members, earning around US$1,000 per annum. The household head is typically male and in his mid-forties. Only about 40% of household heads are literate, with more than half having no formal education, the average number of years of schooling being three.

Although, by definition, all programme beneficiaries would be expected to have access to their own latrine given the timing of this RCT baseline at the end of the output phase,[61] our data suggest that this is not the case for 12% of the households we sampled – this is further discussed in the 'RCT Baseline Report'. Where households have their own latrine, almost all of these sanitation facilities are functioning and are used and almost 80% own an improved latrine. Of those households that report not owning a latrine, about one-third reported typically using one located elsewhere, increasing overall latrine usage rates to 91% among the sample household beneficiaries.

We find that women are generally more likely to use a latrine, and that improved latrines – which 58% of our sample households own (49% in Punjab and 69% in Sindh) – are more often used than unimproved ones.

In terms of hygiene behaviour, we find that almost 90% of main respondents in beneficiary households reported washing hands before eating, though only 26% report% reported washing hands before dinner in particular. 73% reported washing hands after defecating.

[61] We remind the reader that a programme beneficiary is defined as an individual who has gained access to a sanitation facility during the Output-phase.

6 Sanitation practices and behaviours in communities

This section presents our findings with respect to our primary research question: '*What is the impact of outcome phase activities on the sanitation behaviour of output phase beneficiaries?*'

Box 3: Summary findings on sanitation practices and behaviours

Research questions addressed in this section

- *What is the impact of outcome phase activities on the sanitation behaviour of output phase beneficiaries?*
- *How do the above impacts differ depending on the output phase modality chosen by implementing partners?*

Summary findings

- Already pre start of outcome phase activities, about 17% of beneficiary households had at least one member aged five and above defecating in the open. This percentage doubled in the two years until the endline survey was conducted, and, on average, the outcome phase activities did not reduce this slippage at a statistically significant level.
- Where latrines remained functional there was very little slippage in both treatment and control areas. This indicates that latrines becoming non-functional is one of the most significant reasons for reverting to open defecation.
- No statistically significant impacts of outcome phase activities on improved sanitation behaviour are detected on average over the study sample.
- In Sindh the programme activities successfully reduced slippage by 14 percentage points. This impact seems to have been achieved by reducing the degree to which toilets became dysfunctional. Latrine collapse due to rain was a key reason latrines became non-functional.
- We find very limited evidence that variation in output phase implementation modalities led to different outcome phase results.
- The qualitative data reflect that one key incentive for constructing latrines relates to female family members' privacy and honour issues, which are perceived to be violated by open defecation practices. A large number of community members reported that due to high population density and depletion of agricultural fields, there are fewer private spaces in the villages, especially for females to use for open defecation, than there used to be.

Box 4: Qualitative data and generalisability vis-à-vis quantitative results

Reader note on the use and nature of qualitative data used in the study

It is deemed important to remind the reader at this point once more that, unlike the quantitative survey instruments which only collected data on WRP beneficiary households, the qualitative interviews and discussion were conducted with a wider group of randomly selected respondents/households in the WRP programme communities. As such, the information from the qualitative research is not necessarily limited to WRP beneficiaries in the target communities, although it is possible they were included as part of the randomly selected qualitative respondents.

6.1 Defecation behaviour

The key indicator to address Research Question 1 and analyse the effectiveness of outcome phase activities is considered to be the 'proportion of output phase beneficiary households in which at least one household member aged five or above does not use the toilet the household has access to'. The results are presented first not conditioned on whether the household has access[62] to a toilet or not, and second restricting the results to those with access to a toilet. As all respondents were sampled from the SAWRP output beneficiary database it was assumed that all would have access to a latrine; similarly it was anticipated that there would be a degree of 'slippage'[63] over the course of the outcome phase. This section first discusses the whole sample, then turns to discussing the results by province (where impact is detected), implantation modality, and individual characteristics.

6.1.1 Results – whole sample

Columns (1) and (2) in the table below show impact estimates for our first indicator ('Household has at least one member aged five or above that openly defecates', as indicated in the first row of the table) and columns (3) and (4) for our second indicator ('Household has at least one member aged five or above that openly defecates despite access to a toilet'). For each indicator we show estimates without accounting for covariates (columns (1) and (3)), and accounting for covariates (columns (2) and (4)).

Table 12: Impact estimates – Defecation behaviour (household level)

	Household has at least one 5+ year-old member that openly defecates		*Household has at least one 5+ year-old member that openly defecates despite access to a toilet*	
	(1)	**(2)**	**(3)**	**(4)**
Impact	-0.05 (0.05)	-0.05 (0.04)	-0.00 (0.02)	-0.00 (0.02)
District Fes	Yes	Yes	Yes	Yes
Control mean endline (EL)	0.34	0.34	0.09	0.09
No. of villages	123	123	123	123
No. of households	1,132	1,132	1,132	1,132

Notes: Standard errors in parentheses; standard errors are clustered at the unit of randomisation (villages)

Note: Controls include the lag (baseline value) of: 1) the outcome variables; 2) whether household has a functional toilet in the dwelling. Stars indicate statistical significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Source: Household-level survey

The estimated coefficient is negative and suggests that outcome phase activities may have been successful in reducing the number of beneficiary households that reverted to open defecation. However, the coefficient is not significant, independent of whether we account for covariates (helping to improve precision) or not. The difference observed for the first indicator is 5 percentage points; it is worth noting that this study was not powered to detect an impact of five percentage points, but of eight percentage points or higher.

[62] 'Access' was used as opposed to 'ownership' as it is a more culturally appropriate indicator. One drawback of these indicators is that they do not reveal anything about the status of toilet ownership. We will explore this in Section 5.3.
[63] That is, people reverting to open defecation after previously using a latrine.

Box 5: Defecation behaviour – robustness check on reporting

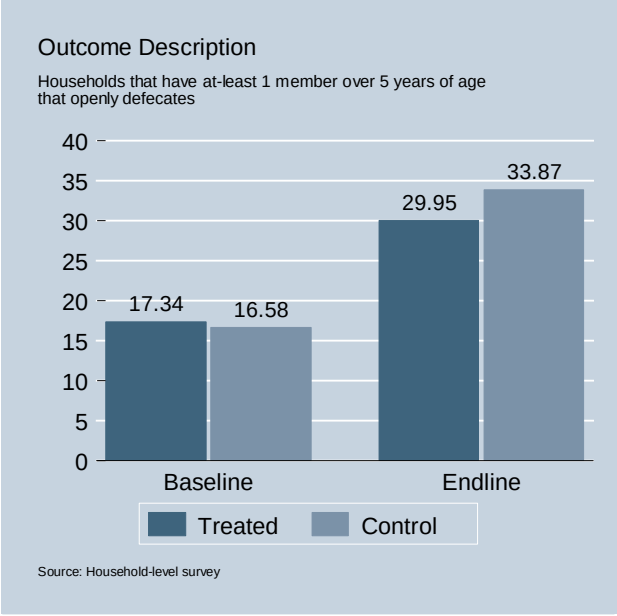
Inclusion of random respondent reported data

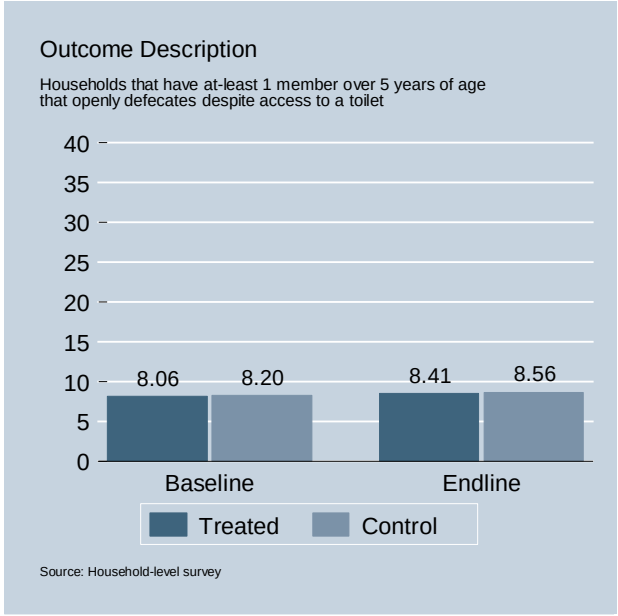
The analysis in this section is based upon the main respondent reporting on the behaviour of other household members. This raises some issues related to measurement validity. The team were well aware of these going into the research and to alleviate concerns we additionally interviewed a random respondent in each household such that the data reported by the main respondent could be triangulated with self-reported behaviour (which is assumed to be more accurate). These data confirm the analysis above. Details of the random respondent analysis are given in Annex D; they are not included in the main body of the report as they confirm the analysis of the main indicator without adding significantly more insight. However, they do lend support to the measurement validity of the main respondent reporting others' defection practices.

The descriptive statistics provide further insight into the results. Figure 5 presents the data – Annex E.1 contains these data in table format, which also presents the standard deviations as well as means by survey round and treatment status. The figure highlights the following:

- At baseline, i.e. pre implementation of outcome phase activities, on average 17% of beneficiary households had at least one member defecating in the open.[64] This difference was well balanced between treatment and control (discussed in the RCT Baseline Report).
- In control areas the percentage of households where members did not exclusively use a toilet doubled to 34%[65] of beneficiary households two years later. In other words, slippage back to open defecation was, without the intervention, 17 percentage points.
- In treatment areas (where outcome phase activities took place) the descriptive statistics indicate that 30% of beneficiary households had at least one member openly defecating when the endline survey took place.

Figure 5: Descriptive statistics – defecation behaviour (household-level)





A further interesting aspect revealed in the figure is that households that have access to a toilet tend to use it. As seen in the right-hand panel, only about 8% of beneficiary households that have access to a toilet have at least one member defecating in the open, a finding that remains constant over time. This suggests two closely linked propositions. First, slippage is likely to be driven by toilets becoming dysfunctional, a point confirmed by both the quantitative and the qualitative

[64] This finding is explored further in the baseline report.
[65] It should be noted this is in line with the assumptions that were used in the power analysis.

research, particularly relevant in Badin, and explored in much greater depth in later sections. Second, where a household has a functional latrine it is generally used by all household members**.** This is supported by positive findings from the qualitative research that indicate significant effort is invested in keeping latrine clean and functional. This is further discussed in Section 6.2.5. Similarly, there are indications that people emptied their pits so they could continue to use them, rather than reverting to open defecation.

However, there remains a percentage of households where not all toilets are used by all household members. The quantitative component estimates this to be the case for 8% of households with a functioning toilet. The qualitative research indicates that this percentage could be significantly higher.

Gender is an important factor with regards to open defecation practices in general, but also specific to those owning a toilet. This is both driven by men working further from home, as well as cultural attitudes. Another related reason highlighted was concerns over the pit filling up too quickly if all members of the household used the toilet. This is also situated within the conversation on gender as it was often given as a reason for men engaging in open defecation while preserving the latrine for use by women and children.

> *'Once [a latrine] facility is available in the house, people don't go out. Men use it, women use it, children use it – all family members use it.'* (Female FGD participant, Rahimyar Khan, T1)

> *'We have a latrine in house. I don't use it because in our area it is a custom that men do not use the same latrine or area that their women use for defecation… There are 10–12 latrines in the village, so those are used but, again, they are only for women and children. Men go to the jungle.'* (Male IDI respondent, Badin, C1)

> *'People generally go out into the jungle or open for defecation. Even those who have latrines in their homes, the male members go out to defecate… I have been using my latrine since I made it. For people who don't have access to latrines, they have no option but to defecate in open.'* (Male key informant, Rahimyar Khan, T2)

This and other factors determining open defecation practices are further explored in Section 6.2.1 (Open defecation practices and gender).

Household respondents interviewed as part of the qualitative research also pointed out the previously discussed slippage. This was most often attributed to factors such as basic, mud, or *katchi* latrines caving in and being destroyed in rains, after which they were not reconstructed. Others were said to collapse due to lack of regular repair or maintenance. This was likely when the households faced monetary constraints or the awareness and attitudinal change was not robust to begin with.

> *'Honestly speaking, half of the houses in the village do not have a latrine. You can visit for yourself. The latrines have come down due to rain. I do not want to mention names here but this is a fact. Most of the people are not building latrines due to bad habits, it is not just [an issue of] affordability.'* (Female FGD respondent, Badin, T4)

> *'There was a latrine in every house when the NRSP visited our village. Gradually overtime, they finished. People did not get them repaired that is why they broke down and collapsed. People have again started using the jungle for defecating. Yes, they know very well about the use of latrines. However, it is only because of their laziness. Of course, poverty is also a factor. But you can also say it [open defecation] is a habit that the men of this village have always possessed since.'* (Female IDI respondent, Badin, T3)

A number of community leaders attributed this slippage to low intervention intensity. Remarks were made in treatment communities that the follow-up activities were not sufficient to sustain significant improvements made in the output phase, and that following ODF certification intervention intensity would fall further. Similarly, there were references to people becoming de-motivated following repeated latrine collapse.

> *'People's attitudes had changed due to this [NRSP]. But now they [NRSP] don't come anymore and neither do people use latrines anymore. All the latrines have caved in. The few that are left, those will fall too, and no one will remake them… People are no longer motivated to make latrines anymore.'* (Male community leader, treatment community, Badin)

> *'There was a change, but I feel all that effort is now going to waste. Lots of people have stopped using latrines. They should have done a system that continuously made people understand again and again, but this wasn't done. Not one, but I'm saying again and again that many people have reverted. Those whose latrines caved in with the rains or because they were not improved, they have reverted to open defecation. Yes, children, women, old everyone is included in this.'* (Male community leader, treatment community, Badin)

6.1.2 Results – by province

Box 6: Presenting the heterogeneity analysis

Reader note on the scope and presentation of the heterogeneity analysis

Scope: We remind the reader that the study was not designed, or powered, to conduct extensive heterogeneity analysis. We therefore present the heterogeneity analysis here solely as suggestive results. One exception is the findings by province. While also here we are limited in power, we are able to detect substantial and significant differences in our data, which the qualitative research further substantiates.

Presentation: Since we consider a relatively large set of variables along which heterogeneity in outcomes is tested, we condensed their presentation at times in this report. Not all impact tables (in particular those related to output phase implementation modality and individual characteristics) follow the same format as presented so far. Instead, we only present regression results where we account for covariates and present several heterogeneity margins in one table. Annex F contains the full set of tables related to heterogeneous impacts by output phase modality. We note that these are, for the large part, insignificant, and so we abstain from conducting multiple hypotheses testing, as already mentioned in Section.0. This applies to heterogeneity analysis of impacts on sanitation behaviour, reported in this section, as well as heterogeneity analysis on impacts on hygiene behaviour.

We find that – when accounting for covariates (column 2) – the intervention had a significant impact on sanitation behaviour in Sindh. Slippage back to open defecation by at least one household member above the age of five is 14 percentage points lower because of the activities conducted in Sindh. As discussed above, this seems to be driven by toilets remaining functional as there is no significant change for household members defecating in the open despite having access to a toilet (the right-hand panel of Table 13), and functionality is seen to be strongly related to latrine collapse due to heavy rains.

Table 13: Heterogeneous impact estimates, defecation behaviour (household level): by province

	Household has at least one 5+ year-old member that openly defecates		Household has at least one 5+ year-old member that openly defecates despite access to a toilet	
Including covariates	No	Yes	No	Yes
Impact in Sindh	-0.13 (0.08)	-0.14** (0.07)	-0.02 (0.04)	-0.02 (0.04)
Impact in Punjab	0.02 (0.04)	0.03 (0.03)	0.02 (0.02)	0.02 (0.02)
District Fes	Yes	Yes	Yes	Yes
Control mean (EL)	0.34	0.34	0.09	0.09
No. of villages	123	123	123	123
No. of households	1,132	1,132	1,132	1,132

Notes: Standard errors in parentheses; standard errors are clustered at the unit of randomisation (villages)

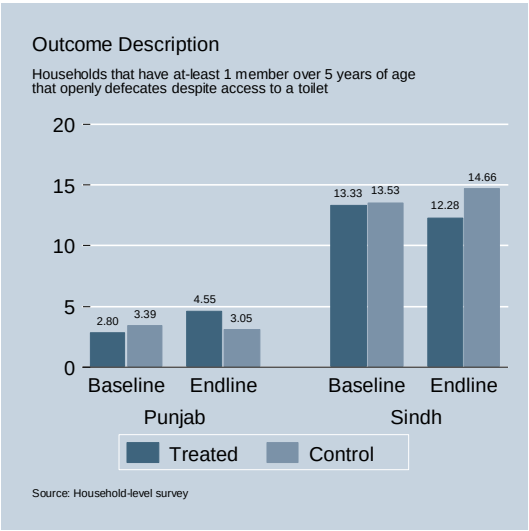
Note: Controls include the lag (baseline value) of: 1) the outcome variables; 2) whether household has a functional toilet in the dwelling. Stars indicate statistical significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

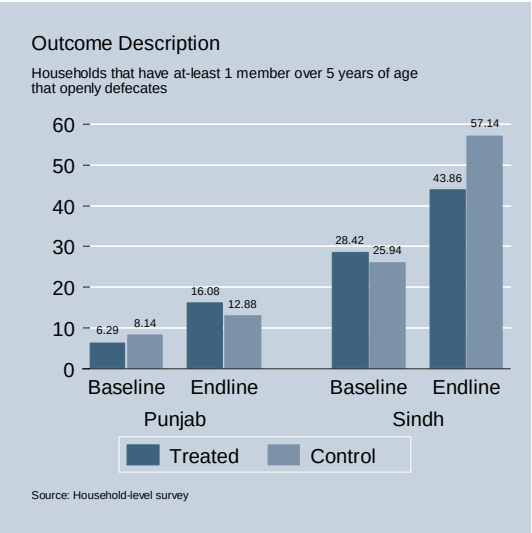
Source: Household-level survey

Descriptive statistics about sanitation behaviour, disaggregated by treatment and province, provide further context to the above findings. These descriptive statistics are presented in Figure 6Figure 6 and reveal a number of interesting points:

- Beneficiary households residing in communities in Punjab are much less likely to have at least one household member still defecating in the open than beneficiary households in Sindh.
- Communities (both treatment and control) in Sindh experience a larger fall-back to open defecation, as indicated by the larger difference between the RCT baseline and endline bars. A potential explanation for the higher slippage came through the qualitative research, noting that, in Sindh, nearly all communities reported some degree of mud latrines falling or collapsing. This was not reported as a problem in communities in Punjab.
- The left-hand panel of Figure 6 shows our regression findings: namely, that treated beneficiary households in Sindh have a lower percentage of fall-back to open defecation than control households.

Figure 6: Descriptive statistics – defecation behaviour , by province





Box 7: Defecation behaviour – robustness check on reporting

Random respondent analysis

We conduct the same analysis for the defecation behaviour reported by the random respondent. We see the same pattern: a reduction in the degree to which households revert back to open defecation in Sindh, and no impact (as in, an estimated coefficient of zero) in Punjab. These results are, however, not statistically significant, even when accounting for covariates, possibly driven by lower precision and power. Annex D.2D.2 presents the results.

As pointed out in the previous section, the impacts in Sindh can be put in close relationship to latrine functionality. In Sindh the intervention led to 13% fewer households reverting back to not having access to a functioning toilet (see Table 14) – though in line with the results above no impact was detected in Punjab.

Table 14: Heterogeneous impact estimates, toilet ownership: by province

	Household has a toilet in the dwelling		Household has a functional toilet in the dwelling		Household has a functional toilet in the dwelling and uses it		Household has access to an improved toilet	
Including covariates	No	Yes	No	Yes	No	Yes	No	Yes
Impact in Sindh	0.11 (0.08)	0.11 (0.07)	0.11 (0.08)	0.12* (0.07)	0.12 (0.08)	0.13* (0.07)	0.10 (0.08)	0.11 (0.07)
Impact in Punjab	0.00 (0.03)	0.00 (0.03)	0.01 (0.04)	0.00 (0.04)	0.01 (0.04)	0.00 (0.04)	0.03 (0.04)	0.03 (0.04)
District Fes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control mean (EL)	0.70	0.70	0.68	0.68	0.68	0.68	0.63	0.63
No. of villages	123	123	123	123	123	123	123	123
No. of households	1,132	1,132	1,132	1,132	1,132	1,132	1,125	1,085

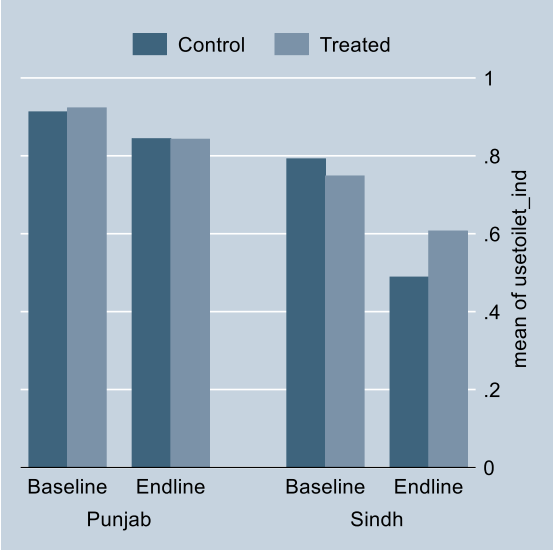
Notes: Standard errors in parentheses; standard errors are clustered at the unit of randomisation (villages)

Note: Controls include the lag (baseline value) of: 1) the outcome variables; 2) whether household has a functional toilet in the dwelling. Stars indicate statistical significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Source: Household-level survey

Descriptive statistics on latrine ownership and usage further contextualise these findings. Figure 7 highlights that while in Punjab over 80% of latrines remained functional, there was lower functionality to begin with, as well as greater slippage, in Sindh. The slippage is, however, visibly lower in treatment than in control communities in Sindh. As this study design is premised on there being slippage arguably there was more limited scope for impact in Punjab given the comparatively lower slippage.

Figure 7: Descriptive statistics – ownership of functioning toilet that is used, by province



Though it is an incredibly intuitive finding that latrine non-functionality emerges as a significant factor, it underscores the importance of focusing on this fundamental aspect in programme design.

As noted in Section 2.3 Sindh experiences more extreme climatic variations throughout the year, with heavier rains and more frequent flooding. This is seen to be a key driver of the higher rates of latrine non-functionality in Sindh as compared to Punjab. Here the quantitative and qualitative results strongly support one another. Respondents in all communities in Badin district reported damage to latrines as a result of rains, because of which the latrines could not be used. In both treatment and control communities visited, respondents reported that the *katchi* mud latrines in at least some of the households had caved or fallen in due to rain and/or flooding. In most cases the latrines were not repaired or reconstructed due to financial constraints, and therefore were no longer functional for use. Members of the household either resorted to open defecation or shared latrines in a neighbour or family member's house, where available and permitted. In treatment areas there was also evidence of people using latrine collapse as an opportunity to upgrade their latrine (discussed in Section 6.2.5).

> *'Every house in our village has a latrine. All the latrines are basic. A lot of latrines got ruined during rain. There are some people who made a latrine again and there are some who use their relative's latrine and say that they will make their own latrine and then there are some like us who resort to open fields and defecate. Yes, there are a lot of people who have started going to the fields to defecate.'* (Male IDI respondent, Badin, T2)

> *'Initially all the households in the village used to have latrines and people would use these and nothing else… now latrine use had decreased – all males defecate in the open now. Even a lot of women openly defecate now. The latrines have fallen. People are no longer motivated to make latrines anymore.'* (Male key informant interview, Badin, T3)

> *'There was a time when everyone in our village was using latrines in their houses. Some people's latrines have now fallen in because of the rains. After that they have now started using the fields. There are also some people who didn't forgo the habit of using latrines, despite not having latrines they use latrines of their close relations.'* (Male key informant interview, Badin, T2)

Further to this, respondents highlighted lack of resources as one of the reasons for not re-building latrines.

> *'We convinced them [community members] financially that they can make the lowest costing latrine. They thought that perhaps a latrine would cost PKR [Pakistani rupees] 10,000 to 12,000 and be made out of bricks according to the standards of the city. We briefed them on how they could make a latrine within PKR 1,000–1,200. We told them to utilise any resource available to them like cow dung, mud and cotton bushes to make the surrounding walls and to install a cheap WC. We told them that they can find expensive WCs in the market and also low costing ones for around PKR 600–700. For people who were very poor, we told them to get WC costing PKR 200–300.'* (SOs FGD, Badin)

> *'The trouble they faced was seasonal rains, which demolished the mud latrines they had constructed out of hard work. These people were not able to purchase materials again due to lack of resources. This was also a reason why they were going towards open defecation.'* (SOs FGD, Badin)

> *'Some people's latrines have fallen. There are five to six houses here whose latrines have fallen and they haven't made them again [Probe: Were they not contacted again or did they not make them?]… I went to their houses and told to make latrines but they said they don't have the money for materials to keep making latrines again and again. They said they are poor and we should ask NRSP to make the latrines for them, but in end the NRSP people stopped coming. But we also told them that NRSP's work was only "motivation".'* (Female CRP interview, Badin, T2)

The SOs overseeing communities in Badin were also aware of this issue of latrines falling in. They confirmed that the majority of the households had mud latrines which had collapsed in rains. SOs reported that many people did not construct these again. Often they could not afford to construct concrete or brick latrines and it was disappointing when their hard work collapsed. SOs claimed that they worked on motivating people in this regard but were not able to provide tangible or monetary assistance. Where households had no access to latrines, either because a latrine had not been constructed or had been damaged and was non-functional, the alternative was usually to defecate in the open. There were a few instances where respondents mentioned sharing latrines, particularly of relatives, in communities in Badin, when someone's latrine had caved in; however, this was not very common. Section 6.2.1 discusses the arrangements, practices, and determinants of open defecation in programme target communities in the three districts. In general, such cases of non-functional latrines were not reported in communities in Punjab, although, as discussed, other problems with pits filling up or pipes blocking were common.

6.1.3 Results – by individual characteristics

Quantitative estimates of the impact heterogeneity along individual characteristics do not suggest any differential impacts – this is despite significant behavioural differences found by the qualitative research for different population sub-groups, and in part with respect to exposure and reaction to the programme. Individual characteristics considered are gender, education, literacy, and age of the random (and hence beneficiary household representative) respondent. This information is summarised in Table 15 for gender, education and literacy, and in Figure 8 for age. None of the estimated coefficients is significant and in Figure 8 all marginal effects by age are estimated to be insignificant. Coefficients are larger for older individuals, but these effects are more imprecisely estimated, as evident from the larger standard errors. The conclusions drawn from the quantitative analysis relate to whether output phase activities affect sub-groups differently, i.e. whether males and females react differentially to outcome phase activities or not. The fact that we do not find any differential reaction among sub-groups might be surprising given some strong qualitative findings on differential behavioural patterns for certain sub-groups.

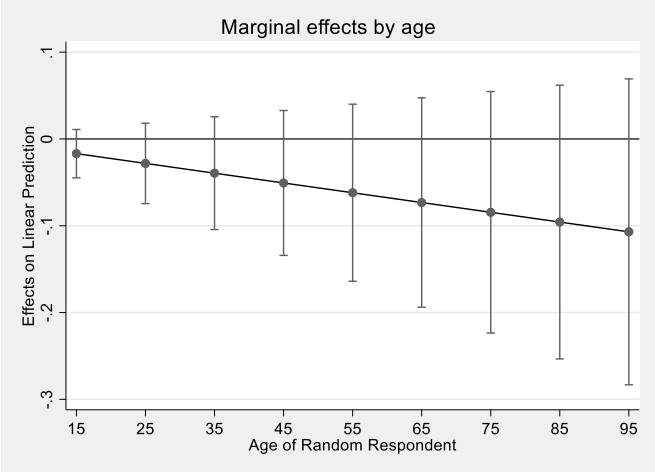
Table 15: Heterogeneous Impact estimates, defecation behaviour (random respondent): by respondent characteristics

	Gender			Education			Literacy		
α_1	*Male*	-0.03	-0.04	*No primary*	-0.05	-0.05	*Illiterate*	-0.05	-0.05
		(0.04)	(0.04)		(0.05)	(0.04)		(0.05)	(0.05)
α_2	*Female*	-0.07	-0.06	*Primary*	-0.02	-0.03	*Literate*	-0.02	-0.02
		(0.05)	(0.05)		(0.06)	(0.05)		(0.05)	(0.05)
District Fes		Yes	Yes		Yes	Yes		Yes	Yes
Control mean (EL)		0.29	0.3		0.29	0.3		0.29	0.29
No. of villages		123	123		123	123		123	123
No. of households		1,079	1,031		1,079	1,031		1,076	1,028

Notes: This table shows the impacts by whether the random respondent is male or female (left-hand panel) for both sanitation behaviour outcomes, whether the random respondent has some primary education or completed primary (and possibly more) education (centre panel), and whether the respondent is literate, i.e. can read and write with understanding in any language, or not (right-hand panel).Stars indicate statistical significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Source: Household-level survey

Figure 8: Marginal effects by age – defecation behaviour of random respondent



In the qualitative research, in all three districts it was reported that men used latrines less frequently than women. Common to all three districts was relating this to men working in the fields. In several communities in Badin and Rahimyar Khan this was also related to cultural beliefs and practices[66] – that is, the non-use by men was usually explained in terms of it being considered 'wrong' for men to use the same space as women for defecation. In addition, privacy concerns and social conservatism around males were also present – these are discussed in Section 6.2.1 in relation to open defecation practices. Where a large number of a household(s) members used the same latrine men also tended to defecate in the open instead.

> *'Men never use the tiled latrines. All men from our village defecate in the open. This is mostly because of habit… The men consider it a sin to use the latrine in the house.'* (Male key informant, Badin, C1)

[66] Though this was not raised in Bahawalpur.

'As we told you it is mostly women and children who use latrines. The men generally go out in the fields. This is because there are more people in the family and only one latrine.' (Male, FGD participants, Rahimyar Khan, T2)

'Some men may use the open and go to the jungle to openly defecate, because men go out of the house to work and earn so they may be doing it in the open.' (Female CRP key informant interview, Badin, T2)

'We do have latrines, but when we go out in the fields for work, then we defecate out there… We do it there [out in the fields] because it would take a lot of time to come back…' (Male FGD participants, Bahawalpur, T1)

While the fact that men are less likely to use toilets leaves a greater margin for change, whether a stronger focus on men would have affected males' defecation behaviour (and would have led to significant programme impacts) remains an open question. In consultations with programme management surrounding the results it was highlighted that the programme did explicitly focus on women in targeting promotion, as well as implicitly as the SOs most often visited communities during office hours. It is also important to highlight that in many cases it would not be practical for a man to return from work to defecate in a latrine at home (especially in the case of farming). Furthermore – arguably – open defecation further from the home presents less of a health risk[67].

There was mixed evidence with regard to age. There is some evidence that, where accessible, elderly members of households prefer the use of latrines for the convenience they afford but in other interviews there were reports of some elderly members of the community being more accustomed to their defecation practices and, as such, preferring to defecate in the open out of habit – which (as with the more educated) led to difficulties in working with this sub-group, as reported by one SO in Bahawalpur.

'Yes there is a difference. Some people use latrines, some go to the fields or to a designated spot for defecation… The elderly now use latrines because they are near and present within the households. If someone doesn't have a latrine, they can use their relative's latrines.' (Male key informant interview, Badin, T2)

'It took us the most time to convince the elderly. Whenever we approached them we always got an answer that it [open defecation] was the practice of their forefathers who never got any kind of illness. It was next to impossible to explain things to them. We approached them in a different way by giving them awareness about importance of their women's privacy and accidents that may occur. This proved as a trigger point.' (SOs FGD, Bahawalpur)

Community respondents reported that households with more educated members normally tended to have latrines as they were more aware of their importance, construction, and usage.[68] In particular, changes in people's behaviours towards defecation and sanitation practices were often attributed to improving education levels and children's schooling. At the same time, in discussion with SOs in Bahawalpur it was flagged that at times it was actually harder to work with a more educated audience since they felt they already knew the sanitation and hygiene messaging.

'About 50% of households in the village have a latrine. Normally those who are educated have latrines at their homes… Education has actually made a lot of improvement, especially in terms of latrine construction and usage. As people are getting education and realising the

[67] The team are cognisant that there is not strong empirical evidence for this statement in the literature. Nonetheless, it is considered a plausible assertion.

[68] It is interesting to note that discussions around education almost entirely emerged in communities and interviews in Punjab, which may link to the generally stronger literacy level in Bahawalpur and Rahimyar Khan vis-à-vis Badin.

importance of latrines they are increasingly switching to latrines and getting them constructed.' (Male KII, Rahimyar Khan, T1)

'We even faced problems with the educated people. Sometimes it is easier to explain things to people who are not educated. Every person has a different mind. When we would talk about the procedure of washing hands, the educated women would comment that they already knew and we need not tell them. But the women who were not educated would listen to us very attentively.' (SOs FGD, Bahawalpur)

6.1.4 Results – by output phase implementation approaches

This section focuses on understanding whether implementation differences might have led to differential programme impacts. We focus on implementation of the output phase rather than the outcome phase implementation as we want to use pre-randomisation indicators to ensure balance along the heterogeneity margin we consider. In particular, we check whether impacts differ along three dimensions:

i. the length of time NRSP has been working in study communities (in particular, whether they started when outcome phase activities started or earlier);

ii. whether a community received a demonstration latrine or not; and

iii. whether someone in the village was responsible for sanitation promotion activities.

In Table 16 we show results on our two household-level sanitation behaviour outcome variables, accounting for covariates. Detailed result tables, in line with the previously presented tables, are provided in Annex F.1.

We generally do not find any strong, consistent indication of differential impacts of outcome phase activities depending on how the output phase was implemented. As mentioned in Box 6, even where some marginally significant impacts are found, we raise caution that these could have occurred at random given the number of hypotheses we test. If one were to test 20 hypotheses independent of each other at a significance level of 0.05, one would face a chance of 64% of at least one significant result occurring, even if all of the tests are actually not significant.

Table 16: Heterogeneous Impact estimates, defecation behaviour (household level): by output phase modalities

	Presence, length of time			Demonstration latrine			Sanitation promoter		
α_1	*<=2yrs*	-0.07	0.00	*No*	0.00	-0.01	*No*	0.06	-0.01
		(0.05)	(0.02)		(0.04)	(0.02)		(0.09)	(0.04)
α_2	*>2yrs*	-0.08	-0.02	*Yes*	-0.11*	-0.01	*Yes*	-0.08*	0.00
		(0.06)	(0.06)		(0.06)	(0.03)		(0.04)	(0.02)
Inc. covariates		Yes	Yes		Yes	Yes		Yes	Yes
District FEs		Yes	Yes		Yes	Yes		Yes	Yes
Control mean (EL)		0.36	0.09		0.36	0.09		0.37	0.09
No. of villages		113	113		119	119		117	117
No. of households		1,050	1,050		1,095	1,095		1,077	1,077

Notes: The table shows heterogeneous impacts by output phase modality. For each heterogeneity considered we show impacts on our two primary defecation behaviour outcome variables. Stars indicate statistical significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Source: Household-level survey

We reference to the areas explored we find as follows:

(i) **The length of time NRSP has been working:** This measure was based on the RCT baseline reporting in the community survey. We do not find any differential impacts along this margin, as can be seen in the first two columns of Table 16.

(ii) **In communities where demonstration latrines were provided, fall-back is suggested to be 11 percentage points lower**. However, this is significant only at the 10% level. Qualitative interviews suggest two possible channels for this result. First, in line with point (ii) above, some responses support that NRSP targeted the neediest with demonstration toilets, possibly achieving impacts by doing so. Second, other respondents reported that the choice of where to place demonstration latrines was rather determined by how responsive and deserving a community had been, implying that higher impacts might not be driven by the toilets but by the fact that the toilets were provided where higher impacts had been achieved.[69] Our data unfortunately cannot identify the relevance or dominance of either of these channels. The quantitative research revealed more generally important heterogeneity in how the demonstration latrine process was rolled out in practice. The number of demonstration latrines reportedly constructed by NRSP in different communities varied from zero to five. In a few cases, the location of the demonstration latrines was decided through consensus and they were provided to the neediest and poorest households in the community. However, there are other reports of demonstration latrines being set up in the CRP[70] or their relatives' houses, or external political pressures linked to their provision. This heterogeneity is likely to imply non-homogeneous impacts, making it harder to identify the same in our quantitative sample.

(iii) **Finally, it is suggested that communities that had a dedicated sanitation promoter had lower slippage, by eight percentage points – again, only significant at the 10% level**. The qualitative findings are not very conclusive about this impact channel. Some support for this channel is provided by the finding that amongst the qualitative communities visited in the district of Badin, for example, nearly all six treatment and control communities where impact results improved reported WASH promotional awareness by a CRP. On the other hand, in one of the treatment communities in Rahimyar Khan where a vast majority of households were reported to be openly defecating, respondents asserted that they did not have a CRP or local promoter of WASH. It is difficult, however, to draw more conclusive inferences from the information available.

[69] The same was reported in terms of water taps provided by the programme during the Output-phase. The quality of ground water was mentioned as a major problem in most communities in the three research districts. To address this issue, after assessing the situation the programme installed one community water tap for every 15 households. These water taps were also used as incentives for more responsive households, which, according to the SOs, were provided with water taps on a priority basis as a reward for their positive behaviour.

[70] In one treatment community in Badin, a female key informant (who was the wife of the CRP) mentioned that NRSP had provided cement and sand for the latrine in their household, while the rest was constructed by community members.

Box 8: The use of 'demonstration latrines' in programming

Qualitative findings regarding demonstration latrines

The inclusion of demonstration latrines was a programme feature specific to WaterAid. The qualitative study found evidence that the distribution and type of demonstrated latrines provided by NRSP in Badin varied quite significantly across communities. The intended impact of demonstration latrines is likely to have varied, alongside this, which makes it more difficult to comment on impact across locations. While respondents in the various treatment and control communities confirmed that NRSP staff had promised to construct demonstration latrines in their villages, there were mixed reports and there was mixed satisfaction with how this process was actually rolled out in practice. In a few cases, the location of the demonstration latrines was decided through consensus and they were provided to the neediest and poorest households in the community. However, there were other reports of demonstration latrines being set up in the CRP's or their relatives' houses, or external political pressures linked to their provision.

> *'For people who were more responsive or most deserving, we also made demo latrines for them under the WASH Result Programme […] which cost around PKR 1,000–1,200. […] When we made a demo latrine for some households, others also demanded that we should make it for them. Some local influentials also brain-washed people to think that the government had given us the required funds, but we were still asking people to make latrines themselves.'* (SOs FGD, Badin)

> *'People made all the latrines at their own cost. NRSP only gave two latrines which they called demo latrines. There are two mohallas [localities] in our village, they made a latrine in each mohalla. The walls of the two latrines were also made by the owners.'* (Male CRP KII, Badin, C2)

While in one or two instances the construction of the demonstration latrines was supervised by NRSP staff themselves, in most cases this was tasked to the CRPs. NRSP often just provided them with a sum of money (variously reported at PKR 1,000–5,000 per latrine), or at times with sanitary hardware material like WCs, cement and pipes, and left it to the CRPs to facilitate latrine construction. Community respondents generally expressed satisfaction with the CRPs' work and facilitation in this process. However, they complained that NRSP facilitation in demonstration latrines was only limited to the actual latrine system itself and households had to make the surrounding walls, floor, and roof at their own cost for the latrines to be usable. Therefore people were expected to, and did, incur significant cost even in cases where they received a demo latrine from the programme. There are also some indications that the inclusion of demonstration latrines created the expectation of subsidy among the community members that did not receive a demonstration latrine.

> *'The NRSP team visited us in the end of 2014. They built four latrines for us. They paid us PKR 1,000 for each latrine. We bought cement and the WC in this PKR 1,000. They also gave us bricks to build holes. This was the only help they provided us in building the latrines. The rest of the boundary wall was constructed by ourselves.'* (Male FGD participant, Badin, T1)

> *'Yes we asked them to complete the demo latrines they had come to install. We requested them to complete the outer walls... But they refused, saying it was not part of their scheme.*' (Male FGD participant, Badin, T4)

6.2 Discussion of the key determinants of latrine usage

This section discusses qualitative findings with respect to the determinants of latrine construction and use. It is divided into key thematic areas. The section begins by discussing the open defecation practices of people within the study communities (i.e. not necessarily output phase beneficiaries), as this sheds light on the practices and motivations behind latrine use. Three sub-sections separately discuss issues related to the infrastructure (motivation to construct; experience of constructing; and maintenance and upgrading practices). The section finishes by discussing the extent to which benefits were community-wide.

6.2.1 Open defecation practices and gender

Though not the focus of the quantitative analysis the initial quantitative analysis conducted ahead of the research highlighted that open defecation remained a widespread practice. As such, the qualitative research, where appropriate, explored open defecation practices. It is again important to note that the qualitative research interviewed non-beneficiaries and beneficiaries, so the findings below do not directly pertain to SAWRP beneficiaries. Nonetheless, they shed important light on the practice of open defecation in the programme area.

There is significant variation between groups in terms of practices – particularly between men and women[71]. Open defecation was raised (by both genders) as presenting a greater challenge for women, who have to go at specific times of the day – either before sunrise or after sunset – with modesty and embarrassment identified as strong reasons for this. Open defecation was generally seen as less of an issue for men who can more freely go throughout the course of the day and further from the home. Lastly, it was noted that seasonal variations influence practices – specifically that during the rains it can become more challenging to find a suitable place. Again, this has differing impacts by gender as women have a more restricted range of places to go.

> *'Women and children use the latrine in their homes. Those who do not have a latrine they go outside to a specified place near their home and defecate there. The rest of the males go out and do it in the jungle.'* (Male FGD participants, Badin, C1)

> *'We, as females have specific timings when we go for defecation such as before dawn and after sunset as we feel embarrassed going out in front of everyone in the day.'* (Female IDI respondent, Rahimyar Khan, T2)

It was also highlighted that to provide a place for women to urinate, or defecate, during the day there is a designated place within the household compound. The space for urination was described as a designated space. The spaces for defecation were described as often having a shallow ditch with a small opening or a cover of bushes or grass on the mud floor, which is used by people to defecate. These somewhat in-between defecation arrangements are reportedly used primarily by women and children though, and not by men. This may be linked to more conservative and stricter privacy concerns around women in the district, compared to communities in Punjab.

> *'We have made a small non-concrete space in the corner of the house for urination. It has walls made of bushes and sticks and we use this space to urinate; we go in the fields for defecation.'* (Female IDI respondent, Badin, T3)

> *'Women and children use the latrine in their homes. Those who do not have a latrine they go outside to a specified place near their home and defecate there. The rest of the males go out and do it in the jungle.'* (Male FGD Participants, Badin, C1)

> *'The men go out to the jungle and the women cordon off a place for themselves with bushes within the household. There is a mud floor and they use that to defecate… First women used to sit anywhere outside to defecate, but now they use a designated space in their houses. This change happened after NRSP came.'* (Female key informant, Badin, T3)

In addition, most respondents that defecate in the open in both treatment and control communities did not appear to be aware of the need or importance of disposing of the faeces correctly. However, there was a widespread emphasis on cleanliness and taking water for washing when defecating. Respondents were aware this made the area dirty, spread germs, and gave off a bad odour, but most believed that the faeces would simply dry up and either get dispersed or eaten by animals. Nonetheless, in a few cases respondents in both treatment and control communities did

[71] Separate FGDs were conducted on gender lines.

report covering the faeces with mud – particularly where people defecated in the vicinity of households or the nearby locality.

> *'You people have come from urban vicinity, I will explain to you how people do it outside. Just sit on the ground wherever you find space in the open fields, then our elders have told us to use stones or mud to scrub our body even if it causes rashes or the skin peels off and then later when possible wash with water.'* (Male FGD participant, Bahawalpur T2)
>
> *'If someone has time, they put soil on it otherwise they just leave it like that and it will turn to waste itself.'* (Male FGD participants, Bahawalpur, T1)
>
> *'Women and children […] put mud on top to cover the faeces because it is very close to the house, just outside the house boundary walls […] The men go off into the bushes in the jungle […] They leave the faeces in the open.'* (Male KII, Badin, C1)

Sharing of toilets was only rarely seen as a viable alternative to open defecation; though there were notable differences between Punjab and Sindh. In Punjab people were generally averse to latrine sharing and had concerns related to overuse ('pits getting filled very quickly') and cleanliness. Though respondents in various communities of Badin raised similar discontents, there was, however, slightly more openness in allowing relatives to share or under certain circumstances.

> *'No it is not like this [that people share latrines]. However, if men are not at home, women can use each other's latrines. In general, such a practice is considered in-decent over here. Therefore, a household's latrine is only used by the members of that household.'* (Male FGD participant, Rahimyar Khan, T2)
>
> *'People do use each other's latrine. No one stops one another. If someone's latrine has collapsed then till the time they don't build their own they can use others latrine, as we are all relatives…Yes we share latrines if we are a guest at someone's place or if our own latrine has fallen down… Or if the latrine is under renovation.'* (Male FGD participant, Badin, T1)

6.2.2 Sanitation practices of babies, toddlers, and children

The focus of the research was on the sanitation behaviour of beneficiaries aged five and above. While this choice was justified at the design stage of the research, it leaves an important gap, namely with respect to the sanitation practices of toddlers and babies. To fill this gap with respect to practices (rather than programme impacts), we turn to information collected as part of the qualitative research.

Defecation practices of small children (zero to three years) were found to be similar across provinces. Children between zero and one year either defecate in their clothes or on a sheet of cloth, which is later washed by their mothers or simply thrown out. As children grow older and are able to walk and squat independently they are directed to a corner in the house for defecation or told to defecate outside the house in the *naala* or the open drains. Almost all respondents mentioned washing the child after defecation with water, though fewer also mentioned the use of soap.

For houses that had latrines which were also functional, most respondents reported that their children between ages two and three use latrines. However, the practice for children aged zero to one year is similar to that of households without latrines. Interestingly, regarding children (older than five years), the qualitative evidence on defecation practices suggests that they tend to follow similar defecation practices as their parents. Mothers are usually responsible for taking their children out to defecate in the open with themselves, and also for helping clean them afterwards, especially when the children are younger. The narrative surrounding the relationship between gender and open defecation discussed in Section 6.2.1 holds here. That is, as children grow older there is a tendency for boys to follow their father's practices and for girls to follow their mother's.

6.2.3 Motivations for constructing and using latrines

The decision to construct a latrine in the household was usually taken by the male heads of the household. However, in most cases, the wife or other females in the household were important in influencing this decision. That said, interviews with SOs and respondents also highlighted that the decision ultimately rested with the man.

> *'I made the latrine for the convenience of the women and children in my house. The decision was my father's. Yes, he consulted the women that NRSP is providing some of the materials [for the latrine], but we will have to bear the rest of the costs ourselves, so should we make it or not. We decided to make it. Once the latrine was made, we had to use it then.'* (Male IDI respondent, Badin, C1)

> *'The final decision to make the latrine was mostly mine. We women faced more difficulty [in defecating in the open] as we had to go too far away… I had to leave my little children at home who would cry a lot. I asked my husband to build me a latrine at home. Then my husband built a basic latrine for us.'* (Female IDI respondent, Badin, T2)

One of the primary influences on households' decisions to construct latrines was to ensure the safety and privacy of female members of the household. Most male respondents across the three districts confirmed that they were uncomfortable – and felt ashamed – about their wives and daughters going out to defecate in the open. This was said to be 'not reasonable or right'. Similarly, women confirmed that they felt much more secure defecating in their own latrines and they no longer risked being 'exposed', 'shamed', or unsafe. As explored above in Section 6.2.1 above this is linked to the greater burden open defecation places on women.

> *'I felt really bad when my family women and children went out for defecation. It is not even safe these days for females and children to go out to isolated spaces to defecate. All sorts of incidents can happen which really scared my brothers and I, so we decided to construct latrines in the house.'* (Male IDI respondent, Bahawalpur, T2)

> *'When we used to go out, all the time there was only one thing on our minds that no one should be able to see us, or no one should come there at that time. Now when we use latrine at home we feel good.'* (Female FGD participant, Rahimyar Khan, C)

One of the main advantages reported for latrines across both treatment and control communities was that they are able to use the latrines in their house whether 'day or night', 'hot or cold', 'rain or sun'. This is particularly so where the latrines are well-constructed with a roof cover and surrounding walls. Where this is not the case, for example where the latrine does not have a roof, although the respondents still claimed to use the latrine – they reported that it is less comfortable when the sun is blazing overhead or rain is pouring in.

> *'The advantage of having a latrine at home is that you can use it round the clock, whenever you want... It was a big problem earlier, if one had to defecate outside and we saw men standing there, we used to come back. After waiting for some time we used to go again and if men were still there we were forced to come back again. Now we don't have to do that.* (Female FGD participant, Rahimyar Khan, T2)

> *'The time doesn't really matter when you have a latrine. Yes, those who go the fields for defecation, they do face problems but having a latrine in your home means you don't have any problems. We do face problems in rain, but winter and summers no it doesn't really affect us.'* (Male IDI respondent, Badin, T1)

In addition, various respondents in all the communities visited in Rahimyar Khan and Bahawalpur complained that it had become increasingly difficult to find a secluded spot for defecation because there has been population growth and most of the local forest cover and fields near the villages has now been cleared away. This caused a lot of inconvenience and as a result several people had turned to the alternative of relatively easy-access and private latrines.

> *'Son, this used to be a jungle earlier. Gradually the jungles vanished. Earlier people also used to visit the bushes for defection. But the bushes also finished. Therefore, people started constructing latrines.'* (Male FGD participant, Rahimyar Khan, T2)

> *'The population has increased so people don't go outside [to defecate] now. It is not possible now to go and defecate outside. There are people all around. Hence latrines are used.'* (Female FGD respondent, Bahawalpur, T2)

The qualitative research indicates quite a high level of general awareness about the importance and benefits of latrines. This awareness was attributed to several sources, such as the work of sanitation promoters in the community, media, and general exposure, increasing levels of education, and also imitation effects in the community. People in all three districts mentioned that they preferred latrines due to perceived health benefits and the protection they afford from diseases. There was a general consensus amongst community respondents that by having latrines and emphasising cleanliness, people's health had improved and the spread of diseases was relatively more contained. Latrines also provide more comfort and facility to elderly and disabled members of households. In addition, latrines allow the owners to maintain a cleaner environment within the house, and at times also have positive spill-over effects in the surrounding community as well. Some respondents reported that it was now easier to walk outside since there were fewer open faeces lying around. However, as discussed in Section 6.2.4, the type of latrines and sewerage exit systems usually constructed in the communities often bring their own set of concerns to discussions on the effects in the wider community environment.

> *'We realised the fact that we were making ourselves sick. NRSP gave us a lot of knowledge. We used to defecate outside and that would enter the streams and nallahs after it rained and that created germs. We now understood this. NRSP representatives filled a glass with dirty water and asked us if anyone of us would drink it. Then they told us we drink this water daily [through defecating openly]. This is how they taught us and we understood, so that is why we decided to build a latrine.'* (Male FGD participants, Badin, C2)

> *'Earlier, we had put two bricks in place of the latrine and we used to sit on those bricks. We used to put soil on the faeces. It used to be very smelly and flies used to sit on it. This would cause illnesses. Then we put this kuppa here. Now when we put water on the faeces, it all washes down the pipes and goes out of the house.'* (Male IDI respondent, Badin, T3)

> *'There is much more cleanliness in the house. We no longer have a bad smell. There are less diseases… My family members are very happy.'* (Male IDI respondent, Badin, C1)

> *'One feels good to see, and to maintain, cleanliness; but these filthy drains and dirty puddles in the streets are the main cause for mosquitoes and diseases'.* (Male FGD participant, Rahimyar Khan, C)

Respondents also clearly highlighted that they had been inspired to make latrines through the work of mobilisers and sanitation and health teams in the area, including those from the NRSP. A few respondents in Rahimyar Khan and Bahawalpur also mentioned additional sources of awareness prior to programme activities, such as the media and greater exposure to the benefits of latrines through travel and time spent outside the village. Once latrines had been introduced in a community, there is some encouraging evidence from communities in Bahawalpur that that also motivated other people to construct and use latrines.

'People from NRSP came here. They build two–three latrines here and told us to make this type of latrine to prevent diseases from spreading… I do not have a latrine in my home, but this all started after NRSP people came… Yes, at first people did not know about latrines.' (Male FGD participants, Badin, C1)

'We have gotten more information from the TV than from XX [the CRP]. She only visited a few houses… People already knew a lot about cleanliness and using the latrine before the NRSP teams came here.' (Female FGD participant, Bahawalpur, T1)

'The decision to make latrine was mine. I had gone to Karachi for a job and used a latrine while I was there. When I came back I decided I would get one so the women and children also do not have to go out.' *(Male key IDI respondent, Rahimyar Khan, T1)*

'Yes, most people got latrines made by looking at each other's actions… There has been an improvement in the people of this area since a while now. They have learnt a lot by observing each other. Those who have TVs in their house have learnt a lot from the TV as well… There has been a change in people's attitudes because of health workers too.' (Female FGD participant, Bahawalpur, C)

In discussions with SOs, they mentioned that some communities had been relatively easy to convince, but they faced resistance in spreading awareness in others. At times, these were simply inception challenges whereby communities had been accustomed to a certain defecation practice and were distrustful of outside messaging. However, in other cases the barriers to mobilisation were said to be more deep-rooted and still persist.

'There were some villages in which people were quite intelligent, we visited once and they were ODF within a week. And then there were other villages which could not be ODF till the end. The village people thought of latrines as something awkward. They thought having a latrine in the house was like keeping garbage in the house… They have been practising open defecation since the times of their ancestors. They said if their ancestors never faced any issues, then why would they?' (SOs FGD, Badin)

'In the beginning it was very difficult because people were very cynical. One day I started to explain to a person and he replied, "You people are very wise but not more than my grandfather. Our forefathers have been defecating in open, do you think they are stupid or are they children to have been doing this?"' (SOs FGD, Rahimyar Khan)

It was interesting to note that in a couple of communities visited for the qualitative research in Rahimyar Khan, latrines were linked to the social standing of the household in the community and the impression it created on others. While this was often not the primary reason for constructing latrines, respondents reported that there was a tendency to look down on people who defecated openly and latrines helped 'put a good impression on incoming guests'. One female community leader also mentioned that in addition to privacy, parents sometimes made latrines in their own houses when their daughters grew up and became of marriageable age as a means of providing social pressure to demand the presence of a latrine for their daughter after marriage. In discussion with SOs in Badin, they also mentioned linking messaging on the construction and use of latrine to everyday practices and customs like marriage and dowry so that '*this [latrines] should also be made a tradition and be sustained and continued*'.

6.2.4 Latrine types and experience of constructing latrines

The RCT baseline results indicated that 88% of households had constructed the latrine themselves and the in-depth qualitative interviews at endline confirm this and provide greater

insight into design. The type of latrine in a household usually varies with the economic conditions and social standing of the household. These communities are usually poor[72] farming and labour communities and the majority of the households with latrines have basic or '*katchi'* latrines – these are usually simple pit latrine systems. Most people's *katchi* or unimproved mud latrines have a WC connected via a pipe to an outside sewerage pit. These sewerage pits are mostly un-cemented and covered by a temporary cover of wooden sticks.

> *'We built latrines in people's houses through our mobilisation. People who could afford it built two permanent latrines and those who could not afford it built temporary latrines. Temporary latrines could be built for PKR 250–350.'*[73] (SOs FGD, Bahawalpur)

> *'The biggest factor is poverty. Even the most basic katchi latrines cost PKR 10,000 to make. Buying the materials, the cost of transporting them to here, then paying the labour and mason. It is very hard for the poor. People who use the fields to defecate these days get very embarrassed but they don't have another choice. They understand everything but they don't have the money to make latrines.'* (Male key informant, Badin T2)

> *'People have made latrines in their homes and have installed a WC and connected it with a pipe that sends all the refuse to a hole they have dug. Those holes are also a part of the latrine. They have built walls around the latrine using bricks and mud or by using wood to make them.'* (Male IDI respondent, Badin T1)

> *'There is no sewerage system, people have dug holes in their houses and the water flows into those pits. There are even some houses which let the water flow to the lower areas.'* (Male, FGD participant, Badin, C1)

> *'X's house had a non-concrete latrine with a WC in it. The WC has not been properly installed in the latrine and is just placed on the top. The women of the house use this latrine to urinate only. They do not use this for defecation because there is no sewerage system and there is no ditch outside. Males of the house defecate under the trees, in the fields or in the bushes outside. Females defecate inside the house in a space surrounded by walls of bushes and sticks.'* (Female IDI respondent, Badin, T3)

> *'First people tried to build a basic latrine with a pit and a concrete slab on top. When they felt the refuse could not be controlled, they moved the pit outside. The flush was made of bricks and connected using a pipe to the drain. We call this flush system the pit latrine.'* (SOs FGD, Bahawalpur)

In a few cases community notables with relatively stronger socioeconomic backgrounds, such as the village dispensers, teachers, or those with government jobs were said to have *pakki* or cemented latrines, and at times even tiled. Some respondents in Bahawalpur and Rahimyar Khan who were financially stronger had more than one improved latrine in their household, with functioning flush systems, somewhat improved septic tanks, handwashing basins, and tiles.

> *'Three latrines in the village are pukka, i.e. they are made with cement (including the demonstration latrines made by NRSP). Some even have tiles fitted in them. The village dispenser has a tiled latrine. The rest, most people don't even have a latrine in their house.'* (Male key informant, Badin, C1)

> *'Sir, we have sewerage lines in our bathroom and we have done tilework (beautification work) as well. We have a complete washroom.'* (Male FGD participant, Bahawalpur, C)

[72] The baseline survey results indicate the average beneficiary household in our sample group comprises seven household members, earning around US$1,000 per annum.
[73] Note that the lowest self-reported cost of latrines communicated by household/community members was PKR 1,000 and went up to PKR 35,000 for permanent, tiled latrines.

NRSP does provide microfinance in some areas. While such integration of multiple programmes can be perceived as 'deepening', and can alleviate the financial constraints faced by households, care needs to be taken as it can also shift the focus and in some cases creates doubts amongst the targeted beneficiaries about the exact motives of the implementers.

> *'No NRSP team has come to our house for any WASH-related information but some other team from NRSP visits our village offering loans. People say that they are even giving loans for constructing latrines. I think they are more interested in giving loans than providing latrines. Because they take an interest over those loans and make money on behalf of poor people.'* (Male FGD, UC 37, Village 37, Bahawalpur)

> *'The only way people can change their defecation practices in my village is if they are given some financial relief as they all are highly indebted and cannot take money out for this purpose.'* (Female community leader, treatment community, Rahimyar Khan)

> *'Some people need a behaviour change only, but some are extremely poor and are helpless. We felt embarrassed asking them to build a latrine when they told us that they could not even afford meals. You only get the true picture when you visit the community.'* (SOs FGD, Bahawalpur)

Interestingly, although the latrines were otherwise described as very basic structures, the presence of a WC or flush was confirmed in almost all communities visited. This is in line with the RCT baseline data- which indicated that over 80% of all households had a pour flush system[74] Respondents in some households in Badin mentioned that their latrines were fitted with 'plastic cuppas' in place of WCs - that were similarly connected via a pipe to an outside pit. The WC or cuppas are usually enclosed within some sort of surrounding walls, often made from mud or bushes and had unbaked floors. Some people also reported hanging cloth curtains where they could not construct walls for privacy in the latrine. Most latrines also have a roof. However, in Rahimyar Khan and Bahawalpur in particular, several households in communities visited for the qualitative research reported their latrines did not have a roof cover, which caused some inconvenience in terms of susceptibility to weather conditions.

> *'All houses have concrete latrines with flushes in them…There is flush on top and holes dug under the ground [for the refuse].'* (Male FGD participants, Bahawalpur, T1)

> *'Our latrine is non-concrete and there is no WC in it. There is a plastic cup ['cuppa'] fixed in it. There are no sewerage lines in the latrine but I have attached a small pipe that flows outside the house towards the backside of the house.'* (Male IDI respondent, Badin, T3)

In the majority of cases in Punjab, people bought their construction materials from the open market, sometimes 10 to 12 kilometres away from their villages. In some of the communities in Badin, people travelled three to four hours to the city of Badin to buy toilet construction materials. To save on transport cost, the vast majority of respondents said they transported the materials by renting a donkey cart, rickshaw, or larger vehicle to move the goods. In a few instances, respondents in Bahawalpur and Rahimyar Khan said they had transported the supplies themselves on their motorbikes; however, this was also often difficult and it ran the risk of damaging the new materials.

> *'We brought our material on the motorbike. I sat behind my husband and held the pot all the way to the village. It was not easy because I was really stressed out that it would fall off and break.'* (Female respondent, IDI, treatment community, BHW T1)

> *'We got the materials for the L from Badin. Materials from there are usually brought in a rikshaw or Suzuki, so we used a rikshaw as well. it was very difficult. …there is no facility for getting supplies in our area. Yes the quality and price was the same as that in the normal market.'* (Male respondent, IDI, control community, BD C2)

[74] Of the remaining ~20%, 12% had no latrine, and the remaining households used either dry systems or other options.

> *'We got the construction material on a donkey cart from the nearby city. It cost us almost 20,000 rupees to get this latrine constructed which we paid from our own pocket.'* (Male respondent, IDI, treatment community, RHY T2)

In one treatment and one control community in Badin, one local CRP implemented an interesting collective action solution: he collected a set sum of PKR 1,000 per household and facilitated the purchasing of materials for all the households:

> *'We collected PKR 1,000 from every house and got all material [to construct their latrines] from Badin. We brought material for five or six latrines and collectively made latrines.'* (Male FGD participant, Badin, T1)

The vast majority of latrines are pit latrines, and therefore require emptying once full. The experience of emptying is discussed in the next section. Some respondents across the three districts mentioned having latrines connected to a communal drainage system that collects the sewerage water and refuse.[75] In the case of a control community in Badin, the main pipe network in the village had been installed by NRSP[76]. However, respondents reported that they had replaced the linking pipes to the main line themselves, since the pipes provided by NRSP had broken. Respondents in two communities (one treatment and one control) in Bahawalpur also reported that fights would break out in the community over improper discharge of refuse into other people's property.

> *'Everyone has similar latrines in their houses. There is a WC fitted and it flows into the main sewerage pipe via a connecting pipe. NRSP installed the main pipe for us which is connected to the latrine in every household.'* (CRP KII, Badin, C2)

The qualitative research found worrying reports in Rahimyar Khan and Bahawalpur of sewerage water being directly drained into wider water bodies or open spaces in the communities, which poses a high health risk. This was reported in at least four communities in the two districts.[77] Respondents complained of poor drainage systems and reported that some people have drainage pipes from their latrines flow directly into the main stream or *naala* running through the village, which means that faeces and other excrement flow openly. Others complained that during the rains, or high flow when the pits fill, sewerage water flows out of the drains into the main walkways and streets in the community, which spreads filth and bad odour, attracts flies and also poses many challenges to mobility. Females in the treatment community in Bahawalpur reported that there had been a recent cholera outbreak in the city due to contaminated water. This problem of poor sewerage systems and excrement draining and contaminating water bodies for general household use was also recognised by SOs working in the communities.

> *'Only 10 households in the community have latrines. Some of these people have made their pipes flow into the naala which runs through the village. That is a bad system because all the faeces flow into the naala and make the area unhygienic. The water from the naala also flows into the fields.'* (Female IDI respondent, Rahimyar Khan, T1)

> *'The taste of the water has changed. It does not taste like it used to earlier. It has changed due to the construction of latrines because the garrhay [pits] are nearby the water pumps and taps at the same place. This is affecting the drinking water and creating a problem.'* (Male FGD participant, Bahawalpur, T1)

[75] This was part of programme design where, as a reward for ODF, the SAWRP partners would in some instances construct a large septic tank and connect several nearby households. This includes both treatment communities in Bahawalpur, one treatment and the control community in Rahimyar Khan, and a control community in Badin.
[76] Reportedly in 2004 – this does not imply contamination.
[77] It includes two communities in Bahawalpur and two communities in Rahimyar Khan.

'There are a few small dirty sewerage lines here from which the refuse water goes into streams… Ever since those streams have been made there is an increase in mosquitoes.' (Male FGD participant, Bahawalpur, T2)

Many SOs voiced similar concerns. In general the perception was that improving sanitation requires a larger financial effort and mobilisation in communities than construction of (basic) latrines alone.

'There are still pending issues. There is no drainage system in most areas. We [the programme] asked people to build latrines but did not tell them about the sewerage system. Open sewers are a big challenge because of germs and bacteria. There is no proper sewerage system and a huge expense is needed to properly eradicate the filth and germs.' (SOs FGD, Rahimyar Khan)

'We know that the distance between the gutter and the water source must be at least 30 feet. In the community every house has its water source placed right next to its latrine. In this way their water gets polluted. Explaining open defecation to people is easy, but to change this set-up is very difficult. People's kitchens were also situated right next to the bathroom. It was very difficult to explain to people that these things were very dangerous for their health.' (SOs FGD, Bahawalpur)

In addition, there were reports that highlighted that some people had constructed their latrines and pits the wrong way due to a lack of knowledge, or even let the refuse exit openly from pipes (as opposed to collecting in a pit or being directed into a drain). As a result the sewerage water tends to mix with the water for other household uses like drinking and cooking, which spreads diseases and discourages peoples from using the latrines.

'There is no [proper] sewerage system. When the drains in the streets fill, then the refuse filth either flows in the streets or it goes back into the latrines. Filthy ponds are formed, mosquitoes are produced. Isn't it better to go out to defecate? At least this would not happen. There are drains and water channels that [now] carry the sewerage water. We are really worried about this.' (Male FGD participant, Rahimyar Khan, C)

'There is no proper system for the refuse water to flush out in some houses. People have dug up pits but the water from these does not go out into the sewerage lines and instead fills up in the pits. There is a severe problem of sewerage lines here.' (Female FGD participant, Bahawalpur, T1)

Finally, in all three districts, water was almost always reported to be available for washing in the latrines or right outside the latrine area. **This is a positive finding in relation to hygiene.** This is interesting in terms of improved sanitation behaviour when compared to open defecation practices discussed earlier, whereby respondents in some communities said they did not carry water with them when they defecated outdoors. Most respondents who owned latrines claimed to fill buckets or containers with water and to store these within the latrine, so that they could be used at all times for washing and for cleaning the latrine. Only a handful of the relatively 'high-end' latrines were said to have running water inside the latrine. Filling and storing water is normally the responsibility of the females of the house.

In general, the majority of the households visited for the qualitative research in Punjab's districts of Bahawalpur and Rahimyar Khan usually had a water source, such as a water tap or tank, hand pump, or boring-hole available. However, access to water within the household compound is less common in Sindh, where water access is more problematic, especially for the poor. Household representatives in Badin said they normally collect water to store in the latrine from the communal taps and streams or from nearby villages. Despite the more cumbersome task, respondents said they make sure water is available as otherwise it is difficult to use the latrine. Water

for other household uses, such as showering, washing, cooking and drinking, is also accessed in a similar manner in Badin.

> *'Having water inside the latrine is of great advantage. You can urinate and if you feel like you can also go for defecation. You don't have to come out to get water. It just saving your body and clothes. If water is available inside the latrine everything would be done quickly.*' (Female FGD participant, Rahimyar Khan, T1)

> *'Water is always available in our latrines because we need water after using the latrine, and even for keeping the latrine clean. If water is not available then it is hard to keep clean.'* (Female FGD participant, Badin, C1)

6.2.5 Maintenance and upgrading of latrines

In light of latrines collapsing or becoming non-functional several respondents – in treatment and control communities in Badin in particular[78] – reported upgrading their latrines.

> *'Yes, our latrine got ruined in rain, so then we used our neighbours latrine. We made our latrine again after some time. Then when it rained again, our walls just fell and the gutter got blocked due to garbage. That is why we got a new latrine made [again]... My family has used a latrine from the beginning. When it rained our latrine used to get ruined, but we used to make it again. We used to use other people's latrine temporarily, but we never went outside.'* (Female IDI respondent, Badin, T1)

I**t is positive to note that several respondents in both treatment and control communities across the three districts confirmed that they did clear out their latrine or build a second pit in order to continue using their latrines.** While some said they cleared the pit themselves, others said this was tasked to sweepers or 'drug addicts' and often cost between PKR 3,000 and 5,000, which was considered to be considerably high. While this is positive with respect to use[79] it does raise some concerns related to the safety of those doing this work and the dumping of waste in the environment.

> *'In terms of cleanliness of latrine we take good care of our latrine. In case the pit (where all of the latrine water goes) gets blocked we call a cleaner to clean it. He basically moves all the water to another pit in the ground and covers it with sand.'* (Female key informant, Rahimyar Khan, T1).

> *'Till date many people are using temporary basic latrines. When one pit fills up they build a second one. This basic latrine does not last long and poses a big problem in the rain.'* (SOs FGD, Bahawalpur)

> *'There is a latrine with a WC fitted inside and a pit outside. If it gets blocked we clean it and still make use of our house latrine. We made the latrine three years ago. We made it ourselves.'* (Female IDI respondent, Badin, C2).

> *'There are latrines, but most of them are not made of concrete. We have just dug up holes and when they get full, we call the bhangies [drug addicts] to clean them. Some take PKR 3,000 for it and some clean it for PKR 5,000. These drug addicts charge too much for cleaning these sewerage pipes… firstly, they are never available and if they are, they charge PKR 10, 000 to clean the village's sewerage pipes.'* (Male FGD participants, Bahawalpur, T1)

[78] Includes two treatment and one control community in Badin and a treatment community in Bahawalpur.
[79] It signals that the household would prefer to make the effort to empty the latrine as opposed to revert to open defecation.

More broadly in the context of general latrine upkeep respondents across the board, in both treatment and control communities visited in the three districts for the qualitative research, were clear on the importance of the upkeep of latrines, and most claimed that cleaning and day-to-day maintenance of their household latrines takes place on a regular basis. There was a general understanding that cleaning of latrines is primarily the responsibility of the women of the house. Various respondents stated that cleaning is primarily important to prevent a build-up of dirt and germs in the latrine, and to avoid bad odours which put people off using latrines. The respondents generally did not report any major challenges in keeping their latrines clean and maintained, However, those with families with younger children face more difficulty in this regard and awareness was mentioned as a problem.

> *'Most of the people here keep their latrines clean because if we do not clean them, it gets very smelly and it is difficult to use it then. We do not feel like using in such a situation then.'* (Male FGD respondent, Bahawalpur, T1)

> *'In our area, there is a lot of dust… The children walk into the latrines with dirty feet, hence we have to keep the latrine clean. Otherwise all the dirt clogs the pipes in the latrines and the latrines gets full.'* (Male FGD respondent, Bahawalpur, T2)

While cleaning and day-to-day maintenance of latrines is quite widespread, there is more mixed evidence on significant repairs or upgrades made to latrines post-construction. Most people continue to use the same basic original latrine structures, claiming it is costly both in terms of finances and time to make such upgrades, and in some cases they lack the knowledge to make upgrades. However, there was evidence in other communities of latrines being progressively upgraded over time.

> *'People are used to latrines now. If someone's latrine has a problem they try to address it themselves because the technology isn't that complicated. Otherwise they get a local plumber to fix the issue.'* (Female CRP KII, Bahawalpur, T1)

> *'We built our latrine five years ago but at first we had a basic latrine and it did not have a WC and it's been five years since we got a WC made. Before this we had a basic latrine and when NRSP arrived they told us about WC and told us to make a latrine. Everyone now has a latrine and a WC in their home.'* (Female IDI respondent, Badin, T1)

6.2.6 ODF prevalence and sustainability in communities

The achievement of ODF at the village level was a focus for the suppliers (in line with best practice in the sector) – though not a deliverable/result or programme objective. The quantitative data cannot speak to this aspect. Given its relevance, we nevertheless dedicate this section to the issue and we report qualitative findings in this regard. Reports from monitoring data and interviews with programme management staff at the country level suggest very high levels of ODF certification[80].

In discussion with groups of SOs during the qualitative study, they also confirmed that all villages in Badin and Rahimyar Khan had been declared ODF, but mentioned a little more variation in Bahawalpur. For instance, in Badin, SOs clarified that all '390 out of 390 [villages] are ODF but only 370 had received a certificate'. Although SOs in Badin mentioned that latrines collapsed at times due to rainfall, they said people later repaired them to a workable condition which

[80] Though it should be noted the certification of ODF is something that is done by the government authorities at the district level under the PATS approach. There is no formal definition of ODF and the certification criteria are known to vary between and within provinces and districts. As such it is seen by the research team to be a 'noisy' indictor and difficult to interpret reliably.

allowed ODF to be maintained. This is supported by the quantitative analysis which did detect impact in Sindh in the context of a large number of latrines becoming dysfunctional in control areas, and to a – statistically significant – lesser degree in treatment areas.

Variation in the definition[81] of ODF was confirmed by the SO FGDs across districts – importantly often not including community-wide latrine coverage as a criterion. In Badin, ODF was described as a broader concept extending beyond latrine construction in households to include aspects of both health and hygiene, such as having fixed points for open defecation and general village cleanliness. Importantly, when 80% of households in a cluster were said to use latrines, this was said to be sufficient for ODF in Badin. In Bahawalpur the emphasis was on achieving 'carpet coverage'[82] in communities in areas that had performed poorly in terms of availability of water and sanitation facilities in the RCT baseline survey. Though the SO FGDs also highlighted that targets and ODF targets could be met without necessarily reaching everyone in a community.

> *'All our target area came under ODF… ODF does not mean that every house is provided with a latrine, it means that people use a latrine. If one person going out into the fields to defecate, the purpose of the ODF was to change him.'* (SOs FGD, Rahimyar Khan)

> *'Our target area is now complete... But I would still say that on average there are 10–20% people who yet need to reach their target [of behaviour change] and still 10–20% percent people who have the chance to revert back to their old habits. It is possible that these people did not listen to us attentively or did not want to listen to us.'* (SOs FGD, Bahawalpur)

Data from the interviews with CRPs present a slightly different picture: the emphasis was very explicitly on every household having a latrine. The CRP interviews support that ODF was a focus of programming, and they confirmed that verification was part of the programme. However, some interviewees highlighted that there was a programme pressure to achieve ODF[83].

> *'For ODF they just went house to house to see the latrines, whether they were clean or not, whether people had actually made them or were they lying. They didn't give us a board or anything for ODF.'* (Female CRP interview, Badin, T2)

Finally, community respondents in Badin reported that no follow-up activities of any sort were conducted in communities that were declared ODF.

> *'Yes there are many villages here that are not ODF. Even in the ones that are ODF, no special activity is taking place.'* (Male CRP interview, Badin, T4)

[81] It should be noted that beyond definition, variation in how the certification was conducted was also clear from the SO interviews.
[82] i.e. 100% latrine ownership.
[83] For instance, a CRP in one treatment community confirmed that after many visits, they had eventually deleted the names of some households that had not made latrines from their village list and then it was declared ODF.

7 Hygiene knowledge and behaviour (handwashing at critical times) in communities

The second overarching research question this report sets out to answer is 'What is the impact of outcome phase activities on the hygiene behaviour of output phase beneficiaries'. This is addressed in this section. We start by analysing impacts on reported hygiene behaviour, followed by hygiene knowledge. We conclude this section by considering the impacts on observable hygiene-related indicators, such as the presence of handwashing facilities and observed cleanliness, as well as the appearance, of the respondent.

Box 9: Summary findings on sanitation practices and behaviours

Research questions addressed in this section

- *What is the impact of outcome phase activities on the hygiene behaviour of output phase beneficiaries?*
- *How do the above impacts differ depending on the output phase modality chosen by implementing partners?*
- *How do the above impacts differ depending on characteristics of the output phase beneficiaries?*

Summary findings

- The outcome phase activities had no significant impact on these reported behaviours. At the end of the output phase almost 90% of beneficiaries reported washing hands with soap after defecation at the end of the output phase – by the end of the outcome phase, this percentage dropped only by a few percentage points in both treatment and control groups. Much fewer (around 45%) beneficiaries reported improved handwashing practices before having dinner.
- The outcome phase activities had no significant impact on the level of hygiene knowledge, which was generally found to be high at endline. Approximately 75% of female respondents reported that it is important to wash hands after defecating and before dinner.
- The outcome phase activities also did not affect the availability of handwashing facilities (with or without soap). However, a general positive trend is noted here – the number of handwashing facilities increased in both treatment and control communities.
- The results do not differ by output phase modality and most beneficiary characteristics. We find some indication that outcome phase activities were more effective at affecting higher educated beneficiaries.
- The qualitative interviews suggest that in many places awareness creation and other related BCC activities conducted during the output phase were impactful and lasting (and potentially led to within- and across-community spillovers, particularly in Badin, Sindh). This could imply that there was not much room for outcome phase activities to impact hygiene behaviour over and above the output phase activities.

7.1 Hygiene – self-reported behaviour

The key indicator we consider analysing in relation to hygiene practices is the '*Proportion of households that report improved hygiene behaviour before eating and after defection*', where improved refers to the use of soap when washing hands.

This is measured using two approaches: we asked the main respondent about her hygiene behaviour and we construct from this information the following indicators: (i) whether the main female respondent reports typically washing her hands with soap after defecating, (ii) and whether she reports doing so before having dinner.

We then create the same indicators for the random respondent, which will present representative reported behaviour in beneficiary households. Unlike in the analysis of defecation behaviour the random respondent data provide more than a robustness check on reporting: with regard to the sanitation behaviour the random respondent data measure the same indicator as the main respondent data (by comparing main respondent-reported data and self-reported data). On the hygiene behaviour indicators the main respondent data for hygiene represent the practices of women and caregivers, while the random respondent data provide a view of behaviour for a cross-section of household members.

7.1.1 Results – self-reported behaviour

Table 17 displays impact estimates for our key hygiene indicators. The results reveal that the intervention had no meaningful or significant impact on reported handwashing practices after defecation and before dinner.

Table 17: Impact estimates – hygiene behaviour (main respondent)

	Main respondent typically uses improved hygiene after defecating		Main respondent typically uses improved hygiene before dinner	
Including covariates	No	Yes	No	Yes
Impact	-0.02 (0.03)	-0.01 (0.03)	0.01 (0.04)	0.01 (0.04)
District FEs	Yes	Yes	Yes	Yes
Control mean (EL)	0.84	0.84	0.45	0.45
No. of villages	123	123	123	123
No. of households	1,132	1,132	1,132	1,129

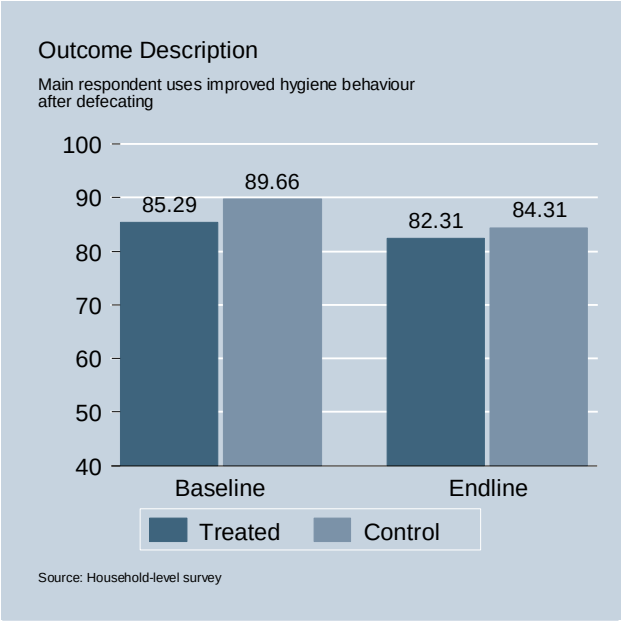
Notes: Standard errors in parentheses; standard errors are clustered at the unit of randomisation (villages)

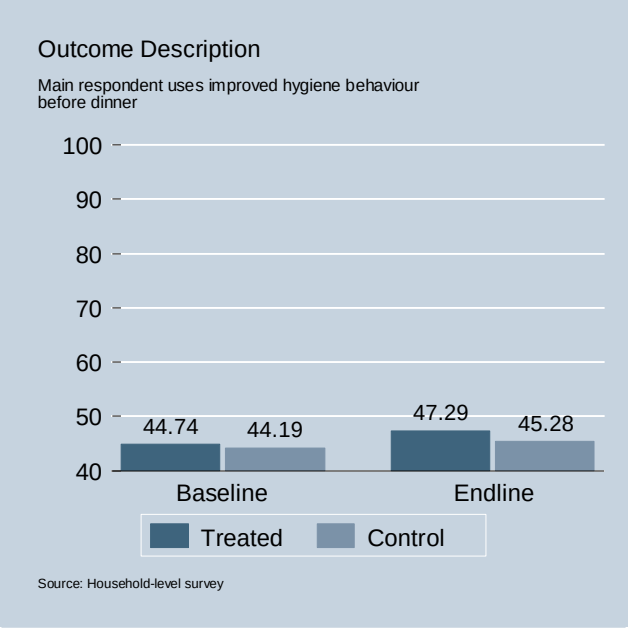
Note: Controls include the lag (baseline value) of: 1) the outcome variables; 2) whether household has a functional toilet in the dwelling. Stars indicate statistical significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Source: Household-level survey

Descriptive statistics reveal important information that places the impact estimates in context. First, improved hygiene practices with respect to handwashing after defecation were reported by over 80% of respondents, leaving little room for improvement. Reported behaviour involving handwashing before meals was significantly lower, at about 45% of respondents. Importantly, both indicators remained constant over time in both the treatment and control communities. The fact that no slippage occurred implies that the intervention could not have avoided slippage.

Figure 9: Descriptive statistics – hygiene behaviour (main respondent)





Turning to the same information but reported by the random respondent, we find very comparable results. Descriptive statistics and impact estimates are presented in Annex 0 and show that also just over 80% of random respondents reported washing their hands with soap after defecating and less than half reported improved hygiene behaviour before eating dinner.

As with sanitation behaviour, a heterogeneity analysis was also conducted for hygiene by province, implementation modality, and individual characteristics for hygiene behaviour. The results suggest a significant impact for respondents in Punjab, being more likely to report washing hands after defecating, but this significance disappears when accounting for covariates (see Annex F.3). The qualitative findings complement this result as no differences in handwashing behaviour were observed across provinces with regards to differences in impacts along the four dimensions related to output phase implementation modality also do not identified differential impacts (see Table 41 in Annex F.3). Similarly, utilising the random respondent data, we do not detect any differences in terms of intervention effectiveness along gender, education, or age, and literacy dimensions: all estimated coefficients are small and insignificant (see Table 45 in Annex F.3).

We find that random respondents with higher education (at least primary education) are 12–14% more likely to report washing their hands with soap before eating. This finding is, however, only significant at the 10% level, but it is corroborated by the findings of the qualitative research that more educated community members are typically more likely to wash their hands. In addition, promotion in schools might have played a role:

> *'There has been a big change due to schools and education now… Our children are also very aware about hygiene and cleanliness through their schools. I remember a few months back, a team from Safeguard soap had visited the school where our children go and given them a session on handwashing and general hygiene.'* (Male FGD respondent, Bahawalpur, T1)

While knowledge of the importance of handwashing was found to be very high, as discussed further in the next section, and while reported handwashing with water was also very frequent, the qualitative findings revealed that the use of soap was variable, depending primarily on availability. Relatively affluent households seem to use soap at all times and have soap lying in the latrine or near the water source. In relatively poor households, washing hands with soap is considered important but availability is variable; whenever soap is available they wash

hands with it. Other alternatives to soap include dish washing soap, cloth washing detergent, or sometimes sand.

Box 10: Hygiene behaviour – robustness check on reporting

Use of proxy hygiene indicators

Self-reported hygiene behaviour is often postulated to suffer from reporting bias. To check the robustness of our reporting we use observable proxy indicators, such as the cleanliness of the respondent's nails. Such observations have been shown to be predictive of child health (Bennett *et al.*, 2018). We analyse the presence of a handwashing facility with soap (presented in Section 7.1.3) and respondents' appearance (such as whether nails are clean and cut). The full results of the respondent appearance indicators are contained in Annex D.3, with a discussion of the analysis and results. In line with the self-reported behaviour we do not detect any impact on these measures. However, it is noted that the observed cleanliness increased in both treatment and control areas for all indicators between survey rounds.

The qualitative results support the quantitative findings about the absence of any differences among respondents/communities based on differences in intervention type and/or the duration of intervention. Hence, regardless of the type of hygiene messaging used and the length of such awareness sessions, the effects were similar. This is partially supported by accounts from control communities where the length of engagement was shorter than treatment communities but respondents reported positive changes to knowledge and attitudes towards hygiene and handwashing.

The findings from the qualitative research shed some light on the relationship between knowledge and practice, and habit formation. While the quantitative results did not detect impact based on individual characteristics, in the qualitative research gender again emerged as a prominent theme. A noticeable number of men said that they wash their hands several times a day out of habit but perhaps use soap only once or twice because they either forget or do not find the soap and do not bother to look for it. This might be because awareness was high even before NRSP started any activities. However, some interviews again emphasised that implementation intensity (as well as progressive improvement), and the targeting of only some households in the communities, also influenced motivations.

> *'Yes, a lady came once and never after that. She talked about hand washing... How can you change habit through one interaction? Had the lady been here every month and talked to us, there would have been many changes.'* (Female FGD participant, Rahimyar Khan, T2)

> *'People already knew a lot about cleanliness and using the latrine before the NRSP teams came here… Those coming and going here did not make much difference because they did not work on all the houses in this area. Madam Azra [the CRP] only used to go to a few houses. She would only visit the houses that did not take care of hygiene and cleanliness [and would inform them about it] […] In our neighbourhood, everyone keeps their houses clean; hence, she did not visit our houses much.'* (Female FGD participant, Bahawalpur, T1)

> *'To be honest with you I wash my hands when I come back from work but I don't use soap every time because I just forget.'* (Male IDI respondent, treatment community, Bahawalpur district)

> *'People didn't care for cleanliness before, now they do. First every house would have a waiyra [place to defecate], now half of them have it. This has been the change. People wouldn't wash hands, now they know the importance and have adopted handwashing as habit. People's thinking has changed and health has improved.'* (Male CRP, treatment community, Badin)

> *'For general hygiene and handwashing practices, it is important to have regular community sessions, with both men and women, where the importance of adopting hygienic practices is emphasised.'* (Female community leader, treatment community, Rahimyar Khan)

Finally, the qualitative research confirms a linkage between hygiene promotion in schools and discussions in the home. Qualitative respondents highlighted that where children were told about improved sanitation practices in school through their teachers and external WASH programmes the children would relay this information back home to their parents. In this manner, households become more aware of the importance and benefits of latrines, which is gradually changing behaviours.

7.1.2 Results – knowledge indicators

In addition to analysing hygiene behaviour, we also looked at intervention impacts on hygiene knowledge. We find that the outcome phase activities did not have an impact on hygiene knowledge – though this is seen to be related to already high knowledge levels at baseline, particularly with respect to handwashing before eating, as shown in Figure 10[84]. It is also noted that hygiene knowledge increased on some indicators (in both treatment and control areas). Table 18 shows the non-impacts on hygiene knowledge – it can be seen that estimated coefficients are small and insignificant.

Table 18: Impact estimates – hygiene knowledge (main respondent)

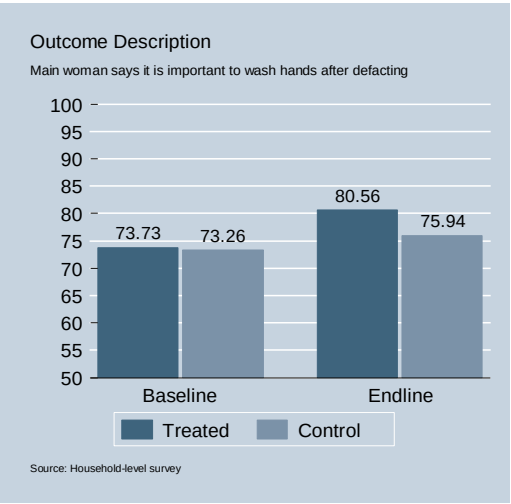
	Main respondent says it is important to wash hands after defecating		Main respondent says it is important to wash hands before eating	
Including covariates	No	Yes	No	Yes
Impact	0.05 (0.03)	0.05 (0.03)	-0.00 (0.03)	-0.00 (0.03)
District FEs	Yes	Yes	Yes	Yes
Control mean (EL)	0.76	0.76	0.79	0.79
No. of villages	123	123	123	123
No. of households	1,132	1,132	1,132	1,132

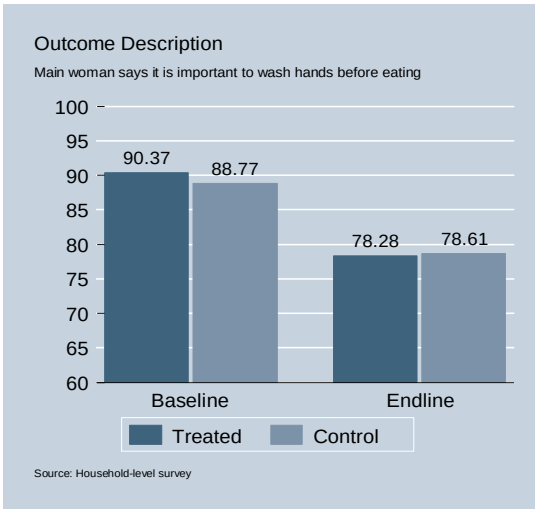
Notes: Standard errors in parentheses; standard errors are clustered at the unit of randomisation (villages)

Note: Controls include the lag (baseline value) of: 1) the outcome variables; 2) whether household has a functional toilet in the dwelling. Stars indicate statistical significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Source: Household-level survey

Figure 10: Descriptive statistics – hygiene knowledge (main respondent)





[84] The differential trends of the two hygiene knowledge indicators may be related to the framing of the question. Respondents were asked to name all the times they thought it was important to wash their hands. This was unprompted, but the enumerators probed until the respondent had listed all of the times they thought important.

The qualitative findings complement the quantitative results. There seemed to be general awareness and understanding among all respondents (females more so than males) about the importance of handwashing for staying healthy and the importance of maintaining the cleanliness and hygiene of the household. Respondents were aware of the fact that germs are transmitted through dirty hands and hence it is important to wash hands, especially before eating and feeding children. The common diseases mentioned were diarrhoea, fever, flu, cough, stomach, and kidney problems, with only a couple of respondents mentioning skin infections transferred through dirty hands. Many female respondents stressed the need to keep children clean by ensuring their own cleanliness, as well as that of children.

Apart from referring to cleanliness as being important for maintaining good health and preventing diseases, many respondents (specifically those from villages in Punjab) referred to religion as being a driving factor for understanding the importance of, and maintaining, cleanliness. Several men and women alike stressed the importance of handwashing and how it was also promoted in Islam to wash hands before and after eating and after using the latrine. In fact, these were the three critical times for washing hands most commonly mentioned by both men and women. Other instances where respondents referred to washing their hands include after house cleaning, after returning from work in the fields, after cleaning babies, after handling cow dung and animals, and before feeding children.

7.1.3 Results – facility observation

Table 19 shows that the outcome phase activities had no impact on the availability of handwashing facilities – with or without soap/detergent available.

Table 19: Impact estimates, hygiene indicators (household)

	Households have handwashing facility		Households have handwashing facility with soap/detergent	
Including covariates	No	Yes	No	Yes
Impact	0.00 (0.04)	0.01 (0.04)	0.02 (0.04)	0.03 (0.04)
District Fes	Yes	Yes	Yes	Yes
Control mean (EL)	0.43	0.46	0.37	0.38
No. of villages	123	122	123	123
No. of households	982	690	1,132	1,022

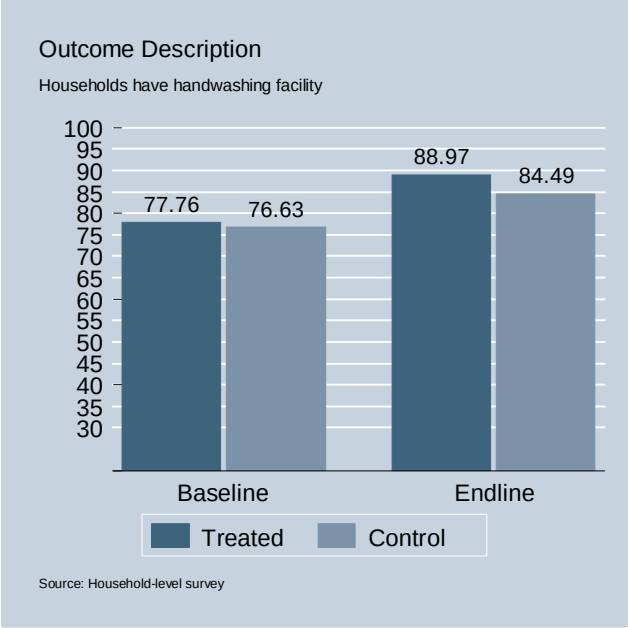
Notes: Standard errors in parentheses; standard errors are clustered at the unit of randomisation (villages)

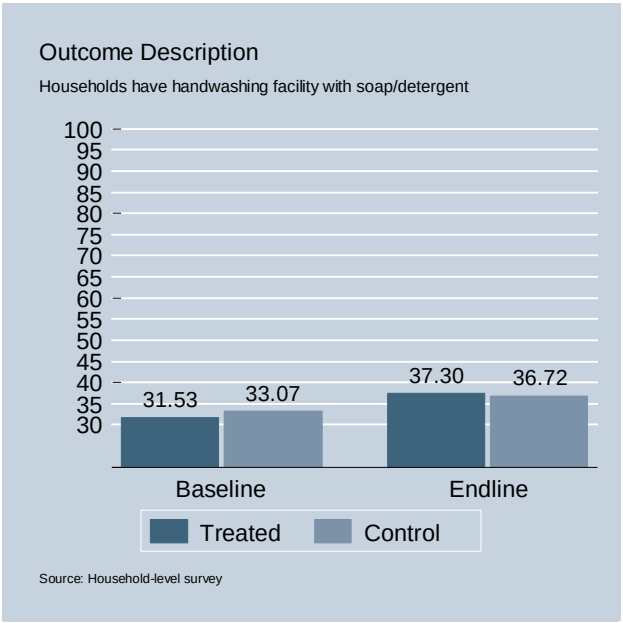
Note: Controls include the lag (baseline value) of: 1) the outcome variables; 2) whether household has a functional toilet in the dwelling. Stars indicate statistical significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Source: Household-level survey

This non-impact finding in terms of availability of facilities might partly be driven by relatively high availability to begin with, as shown in Figure 11. These descriptive statistics reveal that availability of handwashing facilities stood at an average 77% at the time of the RCT baseline survey and increased to, on average, 87%. This relatively high availability of water is related to water being considered integral to handwashing, as well as for maintaining cleanliness. However, when we zoom in and consider how many of the available handwashing facilities have soap present (right-hand panels of Figure 11), the percentage drops significantly. At RCT baseline, 32% of beneficiary households had a handwashing facility and soap when the interviewer came to visit. The percentage increased to 37% at endline.

Figure 11: Descriptive statistics – hygiene indicators (household)





As discussed already discussed, almost all respondents from the qualitative research claimed water was readily available in their homes – this did not necessarily mean that water was sourced from their own homes. A few respondents from Badin commented on drinking water being saline but also did mention that water for other purposes was readily available. The most common source of water across locations is taps/hand pumps. Access to personal taps/hand pumps varies across regions. Almost all respondents from the communities in Punjab (Rahimyar Khan and Bahawalpur – treatment and control alike) have access to personal hand pumps/taps in their own houses. However, villages in Badin do not have their own sources of water and water has to be fetched from communal taps/hand pumps. The ratio of these varies from seven to eight households per tap to 20 households per tap. Other sources of water include wells and/or nearby canals. A few respondents, especially females from Badin, reported fetching water as one of the difficulties faced in practising hygiene. Some households also share water with their neighbours where the neighbours have hand pumps or taps at their own homes.[85]

[85] Among people who do practise open defecation it is common for people to clean themselves with soil or sand if they do not have water and then later wash themselves with water when they reach home. Some respondents carry water in jars when they engage in open defecation. Although in the past it was common for people to clean their hands with soil or sand as well now there seems to be higher awareness and everyone prefers water for cleaning purposes, while sand, ashes, or soil are generally considered an option in the case of water being unavailable.

8 Key findings, lessons, and recommendations

8.1 Key findings

The research design was also premised on there being 'slippage', that is, a reversion to open defecation. Furthermore, given that the sample population was drawn from the output phase beneficiaries it was assumed that all of those sampled for the RCT baseline would have a latrine and use it. It is important to note that if there was no slippage back to undesired sanitation and hygiene behaviours in control communities then it would not be possible to detect any impacts[86]. Intuitively this can be as understood the 'slippage' creating the opportunity for the outcome phase activities to have an impact – without the opportunity to have an impact there can be no impact detected. Though of course, theoretically, if it was known from the start that there would be no slippage in control areas then there would be a less clear rationale for the outcome phase activities in the first place.

The assumed level of slippage transpired to be significant with regards to the results. Specifically, while the research teams assumed one level of slippage for the whole sample, the slippage rates between provinces were dramatically different, and the team detect different levels of impact between the two provinces.

The research was not designed to evaluate the first phase activities, and does not capture the benefit of the output phase activities to beneficiary populations. Our design and the collected data nevertheless reflects aspects of the implementation in both phases. Particularly as latrine construction quality was largely determined in the output phase, but also as the output phase promotion can be seen to have a sustained effect into the outcome phase. Nonetheless; the fundamental focus of this research design was to assess the *additional* benefit of the outcome phase activities with respect to key outcome-level indicators.

(1) What is the impact of outcome phase activities on the sanitation behaviour of output phase beneficiaries? (primary research question)

- Already at the start of the outcome phase implementation, about 17% of beneficiary households had at least one member aged five and above defecating in the open. This percentage doubled to 34% in the two years up when the endline survey was conducted. Across both provinces, on average, the outcome phase activities did not reduce this slippage at a statistically significant level.

- This average however hides differences: in Sindh the programme activities successfully reduced slippage by 14 percentage points. This impact seems to have been achieved by reducing the degree to which toilets became dysfunctional. Latrine collapse due to rains was reported as a key reason for latrines becoming non-functional in communities in Sindh, but less so in treatment communities.

- Where latrines remained functional there was very little slippage in both treatment and control areas. This indicates that latrines becoming non-functional was one of the most significant reasons for reverting to open defecation and the key dimension along which outcome phase activities could have acted.

- No statistically significant impacts of outcome phase activities on improved sanitation behaviour are detected on average over the study sample. This is contrary to expectations. It was

[86] Though of course, theoretically, if there was no should in control areas then there would not be a clear rationale for the Outcome-phase activities in the first place.

hypothesised that some implementation modalities would affect outcome phase effectiveness. However. We note that power issues might be driving this result.

- The qualitative data indicate that a big incentive for constructing latrines relates to female family members' privacy and honour issues, which are perceived to be violated by open defecation practices. A large number of community members reported that due to high population density and depletion of agricultural fields, there are now fewer private spaces in the villages.

(2) What is the impact of outcome phase activities on the hygiene behaviour of output phase beneficiaries? (secondary research question)

- Outcome phase activities have no significant impact on the reported hygiene behaviours. At the end of the output phase almost 90% of beneficiaries reported washing hands with soap after defecation – by the end of the outcome phase, this percentage had dropped slightly, by a few percentage points, in both treatment and control groups. Much fewer beneficiaries (around 45%) reported improved handwashing practices before having dinner, both at RCT baseline and endline, with no detectable difference between those treated and the control group.

- The level of hygiene knowledge was generally found to be high at endline. Almost 90% of female respondents reported that it is important to wash hands before dinner, and 75–80% reported that it is important to do so after defecating. As with behaviour, outcome phase activities did not affect knowledge significantly.

- The outcome phase activities also did not affect the availability of handwashing facilities (with or without soap). Since knowledge and behaviour, rather than facility construction, was the focus of this phase, this does not come as a surprise. Nonetheless, a general positive trend is noted here – with the number of handwashing facilities increasing in both treatment and control communities.

- The results do not differ by output phase modality and by most beneficiary characteristics. We only find some indication of greater effects among more highly educated beneficiaries.

- The qualitative interviews suggest that in many places awareness creation and other related behavioural change campaign activities conducted during the output phase were impactful and – to an extent contrary to expectations – lasting. This could imply that there was not much room for outcome phase activities to impact hygiene behaviour over and above the output phase activities.

(3) How do the above impacts differ depending on the output phase modality chosen by the implementing partners?

- We find very limited and inconclusive evidence that variation in output phase implementation modalities led to different outcome phase results, in either the sanitation or hygiene results. Though, as highlighted above, there were significant differences by province[87]. However, these differences are seen to be related primarily to environmental conditions, rather than differences in implementation approach.

- While the quantitative results suggest that outcome phase activities were more effective in villages where demonstration toilets had been provided, the qualitative results indicate that this was driven by the nature of the communities which received the toilets, rather than by the effectiveness of the demonstration toilets themselves.

[87] Different partners were operating in different provinces.

8.2 Lessons and recommendations

The combination of the quantitative results (which provide strong indications of effectiveness in relation to context) and the qualitative results (which shed light on the mechanisms - the 'what' and the 'why'), allowing us to draw up a number of recommendations related to programming. While the focus of the impact evaluation was the effectiveness of the first output phase, the research revealed recommendations for both components and we start by discussing those related to the output phase and only then turn to the main recommendations with respect to program sustainability.

It should be stated that the division of activities into the categories of 'output phase' and 'outcome phase' is used as short-hand as this relates to the specific implementation of this programe. More broadly this is not a common feature of programming and in terms of generalising the results these are better thought of as activities related to initial promotion and the construction of latrines ('output phase activities), and activities related to promotion and support following the construction of latrines ('outcome phase activities).

8.2.1 Lessons and recommendations related to 'output phase activities'

The report highlights that at the start of the outcome phase about 17% of beneficiary households had at least one member defecating in the open, and this percentage was much larger when the RCT endline survey was conducted. We also find that this is primarily (though not always) driven by these households not having access to a functioning toilet[88]. This suggests there was some scope for improvement in the output phase promotion and mobilisation efforts and that not all output phase beneficiaries can be assumed to be using a latrine.

Most fundamentally, the findings confirm a need to focus on initial construction quality (in terms of latrines becoming non-functional over time) as basic latrines reportedly collapsed due to heavy rains. This is not a surprising finding, and is in line with the recent literature on community-led total sanitation (CLTS)[89] from other countries. While there was significant slippage in Sindh, the results suggest that the outcome phase activities were effective in reducing the rate of slippage in this province. The task of ensuring initial construction quality through programming is complex, but opportunities exist. The promotion approach used – CLTS – focuses on encouraging households to construct latrines themselves. The survey results confirm that the targeted populations were largely poor[90] – a critical factor in considering the capacity of rural households to construct durable latrines.

- *Recommendation 1.1: Ensure in programming that initial latrine construction quality is sufficient to withstand likely environmental shocks; and relatedly*

- *Recommendation 1.2: Post-construction follow-up should be retained as a key programme element as this study provides evidence that this is effective in the context of widespread latrine collapse. Furthermore, follow-up support should be focused on areas where latrine collapse is likely due to environmental conditions – as this is where it is likely to be most effective.*

[88] The baseline report (Section 3.2.5) discussed possible reasons why beneficiary households might not have had access to a toilet (although by definition they should have). Rather than repeating these here, we will discuss constraints – and related programme design recommendations – that households reported with respect to the construction (and maintenance, hence relevant with respect to sustainability) of toilets.
[89] See Crocker *et al.* (2017) for evidence related to CLTS implementation in Ethiopia and Ghana.
[90] The average beneficiary household in our sample group comprises seven household members, earning around US$1,000 per annum.

Study households themselves perceived lack of finance and lack of access to toilet construction materials as important constraints to toilet adoption. Time and again, households reported that the construction of a quality toilet demands an investment beyond their financial capability – and during the qualitative research respondents across the board commented on the need to provide latrines to every household, which they felt the government, or a non-governmental organisation (NGO), should do. Construction of very basic and low-cost latrines was one of the solutions used by local-level implementers to encourage community members to overcome financial constraints. It should be stated that the construction of this type of basic latrine is likely to have contributed to the initial programme impact (though it is out of the direct scope of work of the research team to comment on this), and later in re-constructing collapsed latrines. As such, instead of proposing significant departures in programming, the recommendation would be to focus on complementing the approaches used to overcome specific constraints[91]. These can include financial support as well as innovations that reduce the price of toilets. Regarding the former, given the controversies around subsidies for sanitation, the use of microfinance for sanitation investment has been tried as an alternative and is being used more and more in different contexts.

- *Recommendation 2.1: Ways to reduce the cost of toilet construction should be explored. Such exercises have been undertaken in other contexts, such as by WaterAid in Nigeria.*

- *Recommendation 2.2: The potential use of microfinance for sanitation investment should be further explored. It is increasingly being used as a complement to programming in other contexts. Nonetheless, the effectiveness of providing microcredit is largely dependent on the financial situation of households and it is possible that many households may be ineligible or require a different type of support.*

Respondents to the qualitative research in the three districts further complained, across the board, that they did not have easy access to commercial hardware shops selling sanitation equipment and materials.

- *Recommendation 3: These findings suggest scope for improving access to sanitation markets and materials. While the programme included the set-up and facilitation of sanitary marts, at least in communities visited for the qualitative research there was low levels of awareness of these among households, Community Resource Persons (CRPs), and community leaders.*

  While we are not in a position to generalise these findings, they suggest that a closer look to, and integration of, supply improvements is an important programme ingredient in terms of improving both outputs and outcomes. In cases where it is not realistic for a WASH programme to actively engage with the broader market for materials an explicit focus should be put on making use of the available materials to the best effect, including ensuring there is sufficient capacity built-in, even at the lowest implementation level, for durable construction.

[91] In line with the recent literature on a need to complement CLTS with other approaches, see especially Venkataramanan *et al.* (2018).

The qualitative research confirms the important role men play in investment decisions, as well as the relatively greater self-reported benefits of improved sanitation for women in terms of their privacy, safety, convenience, and honour. In discussing the findings programme staff highlighted that in their programme approach there was an explicit and implicit focus on targeting women. However, the evidence suggests that men are less likely to use a toilet – even when they have access – and are more lenient in adapting their defecation habits, which indicates that this may be an important omission.

- *Recommendation 4: In designing promotion activities ensure that they reach all members of a community. This may include conducting them at times when men are not working in the fields, and developing promotion messages targeted at different sub-groups within a population. Reaching men with promotion activities, while acknowledged to be more difficult, has important implications for women where investment decisions rest with men in the household. In the context of rural Pakistan women's modesty and dignity are important motivations for constructing latrines.*

Finally, the strategy of providing demonstration toilets to deserving beneficiaries that show a willingness to change behaviour seems recommendable. However, our research also suggests that more care needs to be taken in implementation. In some of the communities in Sindh, the local programme implementers reported that construction of demonstration latrines became an issue because other households also started demanding free latrines. On refusal to facilitate such demands, the project had to counter disinformation campaigns and hostility from a few segments of the community.

8.2.2 Lessons and recommendations related to 'outcome phase activities'

Some of the lessons and recommendations above are relevant not only to the achievement of output phase targets, but also their sustainability. Financial constraints are not only relevant for the initial toilet investment but also for maintenance and upgrade; similarly, access to market is necessary for many upgrades and repairs. Since men were found to be more lenient in the consistency of their behaviour change, both with respect to toilet use and handwashing practises, continued engagement that incorporates men seems an important programmatic change to consider. Two further areas of investigation emerged that we believe would improve the sustainability of outcomes.

The intensity of follow-up visits was perceived to be low and their targeting too specific (not just in terms of gender but also in terms of pockets and sub-groups within communities). Particularly research on community led total sanitation intervention seems to suggest that intense follow-up is an important component of successful and significant behaviour change (Abramovsky et al). It seems that within the context of this intervention the structure and workload of social organisers in particular had a lot to do with the lack of intensity in outcome phase activities.

As discussed in Section 3.3, the programme was designed so that social mobilization activities in the communities was carried out by social organisers – who highlighted significant logistical constraints and a high workload. These voiced concerns and perceptions are seen to have influenced the degree and quality of social mobilization activities conducted, and are likely to have led to suboptimal implementation in some communities. A substantive number of SOs in the three districts were also of the opinion that they should have been provided more trainings other than the orientation training they had received at the time of recruitment and one other training of Disaster Risk Reduction (DRR).

- *Recommendation 5: in programme design and planning post construction follow-up activities sufficient attention should be paid to staff capacity to execute these. This relates to additional training (if necessary) to reflect a shift in programme focus, as well as ensuring field staff targets are realistic and that they are adequately resources to conduct their work.*

An important point that came out of the qualitative research is that the benefits of the programme were not community-wide. While a number of communities were declared ODF, it seems that this was not always justified, was not sustainable and many communities remained with a significant fraction of households not having access to latrines. Aiming for community-wide impacts has significance with regards to the mechanisms of behaviour change (collective social pressure), equity considerations[92] (focusing on community-wide impacts by definition is more inclusive), and increasingly there is evidence that it is significant to ensuring health impacts (there is increasingly experimental evidence to support the view that the community-wide sanitation status is an important determinant of health[93]).

- *Recommendation 6: future programmes should include an explicit focus on achieving community-wide impacts. In CLTS Open Defecation Free (ODF) status is often used. Though this can be an unreliable indicator due to varying definitions, and often requires government certification. Where it is no feasible to use ODF as an indicator for those reasons programmes should include a measure related to the extent of programme coverage within communities.*

[92] See especially Cronin et. al. (2017) for a discussion of the importance on focusing on community-wide impacts in reference to the SDGs – authored by senior global WASH leaders within UNICEF

[93]See especially Okullo et. al. (2017) which concludes ODF status has a significant impact on water quality; and Harris et. al. (2018) which concludes that Child height-for-age had a significant and positive linear relationship with community latrine coverage, while child weight-for-age and household water quality had nonlinear relationships. Child growth and water quality were not associated with individual household latrine ownership.

References

Abramovsky, L., B. Augsburg and F. Oteiza (2018). Endline report: Sustainable Total Sanitation - Nigeria.IFS Mimeo.

Augsburg, B. B.Caeyers, B.Malde. 'Complementarities between Micro-Credit and Government Subsidies: Evidence from India's Swachh Bharat Mission'. Forthcoming working paper.

Augsburg, B. and P. Rodríguez-Lesmes (2015). Sanitation and child health in India. IFS Working Paper W15/32. http://www.ifs.org.uk/uploads/publications/wps/WP201532.pdf

Bennett, D. (2012). Does clean water make you dirty? Water supply and sanitation in the Philippines. *Journal of Human Resources*, 47(1), 146–173.

Briceno, B, Aidan Coville, and Sebastian Martinez, "Promoting handwashing and sanitation: evidence from a large-scale randomized trial in rural Tanzania.," Policy Research Working Paper - World Bank, 2015, 7164 (WPS7164), 58–pp.

Cameron, Lisa, Manisha Shah, and Susan Olivia, "Impact Evaluation of a Large-Scale Rural Sanitation Project in Indonesia," Policy Research Working Papers, 2013, 6360 (WPS6360), 1–8.

Crocker et. al. (2017) Sustainability of community-led total sanitation outcomes: Evidence from Ethiopia and Ghana. Int J Hyg Environ Health. 2017 May; 220(3): 551–557. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5475437/

Cronin, A., Gnilo, M., Odagiri, M., and Wijesekera, S. (2017) Equity implications for sanitation from recent health and nutrition evidence. . International Journal for Equity in Health (2017) 16:211. https://equityhealthj.biomedcentral.com/track/pdf/10.1186/s12939-017-0709-5

Deaton, A. and N. Cartwright (2018). "Understanding and misunderstanding randomized controlled trials". Social Science & Medicine, Volume 210, August 2018, Pages 2-21.

DFID (2013a). Water, Sanitation and Hygiene Evidence paper, May 2013.

DFID (2013b). Business Case and Intervention Summary, Water, Sanitation and Hygiene (WASH) Results Programme.

Freeman M.C., Garn J.V., Sclar G.D., Boisson S., Medlicott K., Alexander K.T., Penakalapati G., Anderson D., Mahtani A.G., Grimes J.E.T., Rehfuess E.A., Clasen T.F. (2017). "The impact of sanitation on infectious disease and nutritional status: A systematic review and meta-analysis.", Journal of Hygiene and Environmental Health 2017 Aug;220(6):928-949. doi: 10.1016/j.ijheh.2017.05.007

Gertler, P., Martinez, S., Premand, P. and Vermeersch (2011). Impact evaluation in practice. World Bank.

Guiteras, Levinsohn and Mobarak (2015). Sanitation subsidies. Encouraging sanitation investment in the developing world: a cluster-randomized trial. Science 22 May 2015: Vol. 348 no. 6237 pp. 903-906

Ivankova et. al. (2006) Using Mixed-Methods Sequential Explanatory Design: From Theory to Practice. Field Methods Vol 18, Issue 1, pp. 3 – 20. First Published February 1, 2006. https://doi.org/10.1177/1525822X05282260

Jacob Cohen (1977/1988). Statistical Power Analysis for the Behavioural Sciences. Second Edition, Lawrence Erlbaum Associates.

Lipsey (1990). Design Sensitivity: Statistical Power for Experimental Research; Sage Publications.

McKenzie, D. (2012). *Beyond Baseline and Follow-up: The Case for More T in Experiments*." World Bank Policy Research Working Paper 5639 (2011).Okullo, J., Moturi, W., and Ogendi, G.(2017) Open Defaecation and Its Effects on the Bacteriological Quality of Drinking Water Sources in Isiolo County, Kenya. Environ Health Insights. 2017; 11: 1178630217735539. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5637965/

Onwuegbuzie, A., and Johnson, R. (2006). The Validity Issues in Mixed Research. Research In The Schools Mid-South Educational Research Association Research In The Schools. 13. 48-63.

Paul Tyndale-Biscoe, Matthew Bond and Ross Kidd (2013). “ODF Sustainability Study”.Mark

Pickering, Amy J, Habiba Djebbari, Carolina Lopez, Massa Coulibaly, and Maria Laura Alzua, “Effect of a community-led sanitation intervention on child diarrhoea and child growth in rural Mali: A cluster-randomised controlled trial,” The Lancet Global Health, 2015, 3 (11), e701–e711.

United Nations Development Programme (2017). “Pakistan Human Development Index Report 2017”. UNDP. ISBN: 978-969-8736-19-4.

Venkataramanan, V., Crocker, J., Karon, A., and Bartram, J (2018) Community-Led Total Sanitation: A Mixed-Methods Systematic Review of Evidence and Its Quality. Environ Health Perspect; DOI:10.1289/EHP1965. https://ehp.niehs.nih.gov/ehp1965/

WSP (2012). Achieving and Sustaining Open Defecation Free Communities: Learning from East Java. World Bank

Annex A Evaluation stakeholders engagement

A.1 MoU between research team and implementing partners

MoU between OPM and SAWRP partners– for the Pakistan RCT Research Study

18 Dec 2015

Points of contact

- All parties agree to appoint primary and secondary contact persons at their respective organisations.
- The contact persons at Oxford Policy Management (OPM) in the UK will be primarily **Britta Augsburg** (secondary: Lucrezia Tincani) and **Shafique Arif** in Pakistan (Secondary: Rizwana Aziz).
- The contact persons at PLAN UK will be **Marc Faux** (secondary PLAN UK MEL: John Dean), and **Sajid Ayyub** in the CCU-Pakistan (CCU Pakistan, hosted by PLAN Pakistan), Ifran Aslam (PLAN International Pakistan), Sabir Hussain (WaterAid Pakistan), Shah Hassan (NRSP Bahawalpur), Shahid Iftikhar (NRSP Rahim Yar Khan), and Zaheer Khoso (NRSP Badin).

Responsibilities of OPM

- OPM will lead and manage all data collection activities related to the RCT study, and will ensure that all data collection activities related to the evaluation are conducted to a sufficiently high quality, and in a reasonable timeline.
- OPM will ensure that the involvement of SAWRP Pakistan staff in supporting data collection is kept to a minimum and as specified in SAWRP's responsibilities below.
- OPM will draft data collection instruments *including the translated surveys* in early Jan and share them with PLAN and WaterAid for input (minimum one week turn around). Relevant input will be incorporated by OPM in the final version by the end of January 2016.
- OPM commits to provide regular updates on the RCT study progress and provide opportunities for SAWRP partners to provide input. This will be done through PLAN UK and the CCU Pakistan for further sharing and coordination with SAWRP stakeholders. At the following stages, input will be gained:
    a. Data collection instruments (Baseline and endline)
    b. Procedures for community entry by enumerators
    c. Data analysis & reporting
    d. Policy brief as a result of the study
- OPM commits to informing SAWRP of any changes in the RCT design.
- OPM commits to sharing the findings of the qualitative and quantitative part of the RCT research study with SAWRP before sharing with DFID, providing the opportunity to clarify and explain any findings.
    - We will share the Baseline Report in April/May 2016
    - We will share the Endline report in 2018
    - Findings will be shared in written and oral form (ideally in the form of a face-to-face presentation)
- OPM commits to check with SAWRP whether they would like to be named as an implementing institution before any research findings are published in academic publications at the end of the RCT (around 2018). SAWRP will have the opportunity to opt out if desired. Results will still be shared more widely, but not with any reference to the implementing partner. If joint publication is aimed for, PLAN, WaterAid, and NRSP will be official collaborators and their logos will be placed on the documents.
    - We will share any research articles with you when they have been drafted (in 2018)

Responsibilities of SAWRP Partners

- PLAN and WaterAid will inform OPM on the development of the intervention and will ensure that an approach will be agreed upon by the end of the baseline data collection.
- PLAN, WaterAid and their implementing partners and their implementing partners commit to maintaining the integrity of the design:

1. Maintaining the agreed implementation of the intervention (also known as the outcome phase activities) in the selected study areas until end of March 2018. It is noted that the outcome activities may adapt over time in response to their ability to achieve sustained behaviour change; partners will endeavour to apply a consistent intervention across all treatment villages. PLAN and WaterAid agree to take note of any such changes and inform OPM in such cases.
2. By only implementing the intervention (also known as the outcome phase activities) in selected intervention villages (**treatment villages**) from March 2016 onwards.

- PLAN, WaterAid and their implementing partners and their implementing partners commit to maintaining the integrity of the design, by stopping implementing the intervention (also known as the outcome phase activities) in the **control areas** from 15 Feb 2016 at the latest until endline data collection activities have been completed (expected to be end of March 2018).
- PLAN, WaterAid and their implementing partners commit to providing timely input into data collection instruments, data analysis, draft reports and other outputs.
- PLAN and WaterAid will assist the OPM Pakistan office – if needed – in identifying and entering study villages and selected study households. As indicated above, OPM will ensure that such support is kept to the necessary minimum.
- PLAN and WaterAid commit to inform OPM of any changes to their program implementation and timeline that affect the selected study areas. Occurrences that could trigger such changes include for example floods.
- The Pakistan-CCU commit to inform OPM of any potential 'contaminating' activities undertaken directly by WaterAid / PLAN (media campaigns wouldn't be contaminating activities as long as they involve both treatment and control groups).

- The CCU and SAWRP partners will furthermore inform OPM of any 'contaminating' activities of non-SAWRP members which the CCU become aware of. These could include other WASH programmes (not within WASH Results Programme) or by activities carried out by other community based initiative by the civil society actors, i.e. NGO and/or donors, including any emergency response activities. It is understood that the CCU and SAWRP partners may not have a full overview of the all 'contaminating' activities taking place in Pakistan. It will be OPM's responsibility to assess implications of such activities on the evaluation design.

- Information will also be shared about governance related activities by the government itself (like ODF certification, local government councillor's initiative and MoH activations) which the CCU become aware of as well as the media campaigns. It is acknowledged, that the impact evaluation will take place in the context of such activities as outlined in the design document, implying that such activities are not consider contamination threats.

Signatories

OPM, PLAN Pakistan, WaterAid Pakistan, NRSP

A.2 Key evaluation stakeholders engaged

The table below provides the names of the key evaluation stakehoders engaged by the RCT Research Study team over the course of the research. This is not an exhaustive list as through these points of contact a wider group were engaged to give written feedback on the various reports produced by the RCT Research Study team.

Organisation	Name	Role
DFID	Guy Howard	SRO of WRP and WRP evaluation
	Laura Westcott	Statistics advisor (in post at design and at baseline)
	Stephen Lindley-Jones	WASH advisor (in post at endline only)
Plan International	Marc Faux	Global programme manager – SAWRP (in post design and baseline only)
	Teia Rogers	Global programme manager – SAWRP (in post endline only))
	John Dean	M&E advisor
WaterAid (UK)	Ian Gavin	SAWRP programme manager
	Om Prasad Gautam	Technical Support Manager – Hygiene
	John Knight	Regional Technical Advisor – South Asia
Plan International Pakistan	Imran Shami	Acting Country Director (at the time of endline)
	Dawud, Muhammad	SAWRP Programme manager (Pakistan)
WaterAid (Pakistan)	Siddiq Khan	Country representative
	Sabir Hussain	BCC manager
	Imran Shah	SAWRP programme manager (Pakistan)
CCU (country coordinating unit) Pakistan	Sajid Ayyub	Head of CCU (in post design and baseline only)
	Abdul Saboor	Head of CCU (in post endline only)
NRSP	*Various*	*Field teams liased with NRSP staff during data collection as appropriate.*

Annex B Revisions of power analysis assumptions

B.1 Assumptions made for original power analysis

The starting point for the outcome phase activities assumed that all households constructed a new or improved their existing toilet during the output phase. This was explicit in our design as we sampled from the 'output phase beneficiary database' of Plan International. The aim of the outcome phase was then to sustain usage of these toilets amongst household members. It was expected that without the outcome phase activities some households would be likely to revert back to open defecation behaviour. This could be since their behaviour change was not sustained or because the toilet became unusable and nothing was done about it (pit filled up, pit collapsed, etc). It was hence expected that the percentage of households where at least one household member does not use the toilet would increases in our control group, whereas in the intervention group is would either remain the same, or reduce less (decrease slippage), or, in the ideal case, the usage rates increase.

As there was a certain degree of uncertainty in the starting baseline values of our key indicator, we based our power analysis on standardised effect sizes (SES). Thus the presented power analysis could be considered as relatively robust to different baseline levels of key indicators. We however discussed the information we pulled together and assumptions we made.

To start with, we considered the expected mean of our outcome of interest before the start of the outcome phase activities. Given the assumption that all households in the sampling frame (the programme database) had a toilet, one could consider that everybody also used it, which would provide a baseline of 0% (i.e. there are no output phase beneficiary households in which at least one household member does not use the toilet). We however considered this assumption too optimistic. Unfortunately the programme's monitoring and baseline data did not collect information on sanitation usage rates. We therefore postulated that in 10% of beneficiary households (i.e. those with a toilet constructed or improved during the output phase) there will be at least one household member who does no use it. This was based on some of the best available data available at the time. A study conducted by a team member in India suggested usage rates between 85 and 97%, depending on the location, with rural areas showing lower usage rates (Augsburg and Rodríguez-Lesmes, 2015).

We further needed to get an estimate for the intra-cluster correlation of this variable. Since at the time we did not have data on usage in the programme areas, we turn to other available data. Data from the above mentioned sanitation study on whether a toilet is in use suggests an ICC of 0.07 (7%). Primary data on another sanitation evaluation study in Nigeria suggests a similar ICC of 0.067. Given that this data is from other contexts, we choose to work with a somewhat more conservative ICC of 0.1 in this analysis.

We next discussed our expected change in sanitation usage in the control group, i.e. without the intervention. We turned to secondary literature, particularly to a study conducted Plan International in 2013, titled "ODF Sustainability Study" (Tyndale-Biscoe et. al. (2013)). This study reviewed 96 articles and documents related to CLTS and ODF sustainability[94] and in addition collected primary data of 4,960 households in 116 villages in the four focus countries, namely Ethiopia, Kenya, Uganda and Sierra Leone. The basic take-away message of the report was that 13% of the almost 5,000 interviewed households did not have a functioning toilet anymore when the check was conducted (approximately 2 years after triggering took place – the same timeframe considered in our study), all

[94] According to the authors, out of the 96 documents, only 14 (five in Africa and nine in Asia) described studies that involved a systematic evaluation of some aspect of CLTS, with a structured and clearly described methodology. The mainly focussed however on CLTS effectiveness rather than on its sustainability. Except for studies conducted by WSP in Bangladesh and Indonesia, all studies suffer from limited resources and being small scale.

remaining households still had a functioning toilet. However, when all sustainability criteria were applied, slippage rates increased dramatically to 92%.[95,96] Since our main outcome of interest is that at least one household member does not use the latrine they have access to and not the whole household, we opt to consider the 13% of households whose toilet was not anymore functional, plus those households who showed signs of open defecation around the house, amounting to 8% in Plan International's report. The report would hence suggest a slippage in toilet usage of 21% points if no follow-up activities take place. We believed that the percentage could be slightly lower in this context, where the intervention had a focus on sustainability already in mind. Considering a lower expected slippage is also more conservative for the power analysis. We hence choose to reduce this by 5 percentage points, and therefore assume control households to slip by 16% points on average.

Finally, we needed to make an assumption by how much the intervention could be expected to reduce this slippage rate. Discussion with implementing partners revealed different expectations for both partners. It was said that Plan International concentrated more on quantity within the output phase, which could lead to higher slippage than WaterAid's approach of concentrating on construction of higher quality improved toilets. However, on the other hand, the study district in which WaterAid operates is known to experience flooding, which can trigger slippage. We hence decided that no slippage in the treatment group would be unrealistic and consider that treatment households experience a likelihood of slipping back to OD of 7-8 percentage points (about half of that of the control group).

[95] The report states that: "Eight percent of households with a functioning latrine had visible signs of open defecation around the house indicating that OD was practiced at these households and the overall reversion rate was at least 21%. With respect to hand washing facilities and the presence of soap or ash, only 37% and 25% of households respectively complied. Similarly, adding in the presence of a lid, and the presence of a lid over the hole, reduces the ODF rate to 26% and 19% respectively (virtually all latrines in the study were simple pit latrines – only a handful of households had VIP or pour flush latrines)."

[96] The five criteria applied in PLAN's study were: (1) A functioning latrine with a superstructure, (2) A means of keeping flies from the pit (either water seal or lid), (3) Absence of excreta in the vicinity of the house, (4) Hand washing facilities with water and soap or soap-substitute such as ash, and (5) Evidence that the latrine and hand washing facilities were being used (e.g. a well trodden path).

Annex C Definition of key indicators

This Annex provides details of the indicator construction for the key data used in the report. The variables (indicated in bold) relate to the endline survey questionnaire. The household questionnaire used at endline is available from request – contact details are provided on page ii.

Table 20: Key indicators related to defecation behaviour

Indicator	Numerator	Denominator
Percentage of households with at-least one 5+ year old member that openly defecates	Number of households where at-least one 5+ year old individual openly defecates: **HG01==3, 4 & AGE>=5**	Total number of households
Percentage of households with at-least one 5+ year old member that openly defecates despite access to a toilet	Number of households where at-least one 5+ year old individual openly defecates despite toilet access: **HG01==3, 4 & AGE>=5 & HG02==100**	
Percentage of households where random respondent openly defecates	Number of households where random respondent openly defecates: **HN02==4, 5** *Note: If random respondent and main respondent are the same:* **HN02==HG06**	
Percentage of households that have a toilet in the dwelling	Number of households that have a toilet in the dwelling: **HG02==100**	
Percentage of households that have a functional toilet in the dwelling	Number of households that have a functional toilet in the dwelling: **HG03==100**	
Percentage of households that have a functional toilet in the dwelling and use it	Number of households that have a functional toilet in the dwelling and use it: **HG04==100**	
Percentage of households that have access to an improved toilet	Number of households that have an improved toilet: **HK01==1,2,3,5,6,7,9**	

Table 21: Construction of Hygiene indicators – self-reported behaviour

Indicator	Numerator	Denominator
Percentage of households where main respondent typically used improved hygiene after defecating	Number of households where main respondent typically used improved hygiene after defecating: **HG22==1 & HG23==2**	Total number of households interviewed (main respondent
Percentage of households where main respondent typically used improved hygiene before dinner	Number of households where main respondent typically used improved hygiene before dinner: **(HJ03==1 | HJ03==2) & HJ04==1**	
Percentage of households where random respondent typically used improved hygiene after defecating	Number of households where random respondent typically used improved hygiene after defecating: **HN20==1 & HN21==2**	
Percentage of households where random respondent typically used improved hygiene before dinner	Number of households where random respondent typically used improved hygiene before dinner: **(HN18==1 | HN18==2) & HN19==1**	
Percentage of households where main respondent says it is important to wash hands after defecating	Number of households where main respondent says it is important to wash hands after defecating: **HJ02H==100**	
Percentage of households where main respondent says it is important to wash hands before eating	Number of households where main respondent says it is important to wash hands before eating: **HJ02A==100**	

Table 22: Construction of Hygiene indicators – observable indicators

Indicator	Numerator	Denominator
Percentage of households with handwashing facility	Number of households with handwashing facility: **HP01==100**	Total number of households Total number of households
Percentage of households that have handwashing facility with soap/detergent	Number of households that have handwashing facility with soap/detergent: **HP03==1 | HP03==2**	
Percentage of households where random respondent has clean nails that seem to have been cut in the last 1-2 weeks	Number of households where random respondent has clean nails that seem to have been cut in the last 1-2 weeks: **HN22==100 & HN23==100**	
Percentage of households where random respondents has hair that looks like it has been washed in the last 1-3 days	Number of households where random respondent has hair that looks like it has been washed in the last 1-3 days: **HN26==100**	
Percentage of households where random respondents has hands free of dirt	Number of households where random respondent has hands free of dirt: **HN24==100**	
Percentage of households where random respondents has feet free of dirt	Number of households where random respondent has feet free of dirt: **HN25==100**	

Table 23: Overview of survey modules

Instrument Module		Description
A	**Survey Information**	Household geographical information + basic interview information (interview name, start time, end time etc.)
B	**Identification**	Respondent name and phone number
C	**Introduction**	Description of survey + asking for consent
D	**Household roster**	Information about members of the household
E	**Household characteristics – Main woman in the HH**	Household living conditions + assets
F	**Exposure to programme/interventions – Main Woman in the HH**	Exposure to programme/intervention
G	**Sanitation ownership and usage – Main Woman in the HH**	Information about defecation, latrine usage
H	**Perceptions of other sanitation habits, the community sanitation situation – Main Woman in the HH**	Perceptions/opinions on sanitation habits and community level sanitation state
I	**Water – Main Woman in the HH**	Information on the source of water used by the household + information on sanitation practices w/r to children 0-3 years of age, if present
J	**Hygiene – Main Woman in the HH**	Information on hygiene behaviour of the main respondent + some observations on the hygiene condition of the main respondent
K	**Toiler ownership – Household head**	Information on various details of latrine ownership in the household
L	**N/A in endline instrument**	-
M	**Additional household level information – Household head**	Information on additional household related matters such as access to BISP card, exposure to negative shocks and use of hospitals, etc.
N	**Sanitation use and access – Random household member**	Information on sanitation use, hygiene behaviour + some observations of the hygiene state of the random respondent
O	**Latrine inspection**	Inspection of the latrine, if present in the dwelling – answers to be based on observation
P	**Handwashing facility**	Inspection of household handwashing facility, if present in the dwelling
Q	**Other**	Inspection of presence of animals in and around the dwelling

* The main respondent/woman is the head of the household if female, otherwise the spouse of the male head of household. If neither exists in the household the respondent should be the most knowledgeable adult female household member. Only if the household has no female adult member, the respondent should be the male household head or other knowledgeable household member.

Annex D Robustness Defecation behaviour - robustness check on reporting

D.1 Random respondent – whole sample

The robustness check relates to the fact that the main indicator speaks to behaviour of individual household members as reported by the main survey respondent. Such reporting about others is likely to suffer from some degree of measurement error, the main respondent not having all relevant information.[97] At the same time, it might be more accurate in some cases where the individual itself would have a desire or incentive to provide wrong information about their behaviour.

We can check the degree to which our key indicator might suffer from measurement error by considering self-reported defecation behaviour. The survey was designed to query a random household member about their defecation behaviour. In particular, we consider as an outcome of interest, whether the random respondent reports to openly defecate inside or outside the household compound. We note that this individual level indicator is therefore not directly comparable to the household one since it does not include self-reported behaviour of children age 5.

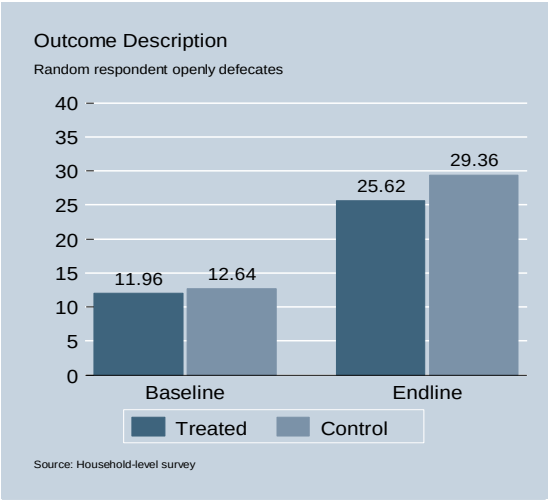
Descriptive statistics on this self-reported variable are presented in Table 24 and the corresponding Figure. It can be seen that, despite the difference in definition, our initial observations drawn from the household level indicator are confirmed using the individual level, self-reported information, a point also observed in the qualitative work: (1) At the start of the outcome phase activities a significant percentage of beneficiary households, about 12%, had at least one adult defecating in the open; (2) this percentage doubled by the time the endline survey was conducted; (3) the increase is slightly lower for households living in treated villages.

Table 24: Descriptive statistics – Defecation behaviour (random respondent)

Unit of Observation: Household

	Random respondent openly defecates		
	Obs.	Mean	SD
Baseline:			
Total	1081	12.3	32.86
Control	546	12.64	33.26
Treated	535	11.96	32.48
Endline:			
Total	1107	27.46	44.65
Control	545	29.36	45.58
Treated	562	25.62	43.69

Source: Household-level survey



Regression results, presented in Table 25 confirm that the difference between treatment and control observed at endline is also at the individual level 5 percentage points but insignificant.

[97] This is also the case for both the quantitative and qualitative research where household, community or programme representatives comment on the behaviour of others. Triangulation across respondents and research instrument was built into the design and analysis of the qualitative study to ensure that themes and behaviours were captured, understood and verified through multiple sources.

Table 25: Impact estimates - Defecation behaviour (random respondent)

	Random respondent openly defecates	
Including covariates	No	Yes
Impact	-0.05 (0.04)	-0.05 (0.04)
District FEs	Yes	Yes
Control Mean (EL)	0.29	0.30
No. of Villages	123	123
No. of HHs	1,107	1,059

Notes: Standard errors in parentheses; standard errors are clustered at the unit of randomisation (villages)

Note: Controls include the lag (baseline value) of: 1) the outcome variables; 2) whether household has a functional toilet in the dwelling. Stars indicate statistical significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Source: Household-level survey

D.2 Random respondent – by province

Table 26: Heterogeneous Impact estimates, defecation behaviour (random respondent): By Province

	Random respondent openly defecates	
Including covariates	No	Yes
Impact in Sindh	-0.10 (0.08)	-0.11 (0.07)
Impact in Punjab	0.00	0.00
	(0.03)	(0.03)
District Fes	Yes	Yes
Control Mean (EL)	0.29	0.30
No. of Villages	123	123
No. of HHs	1,107	1,059

Notes: Standard errors in parentheses; standard errors are clustered at the unit of randomisation (villages)

Note: Controls include the lag (baseline value) of: 1) the outcome variables; 2) whether household has a functional toilet in the dwelling. Stars indicate statistical significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Source: Household-level survey

D.3 Observable hygiene indicators

In this section we check whether variables capturing the appearance of the random respondents has improved due to the intervention. We start by analysing information collected about whether the respondent has clean nails that seem to have been cut in the last 1-2 weeks and whether his/her hair seems to have been washed in the last 1-3 days. Table 27 and Figure 12 show descritve statistics for these variables.

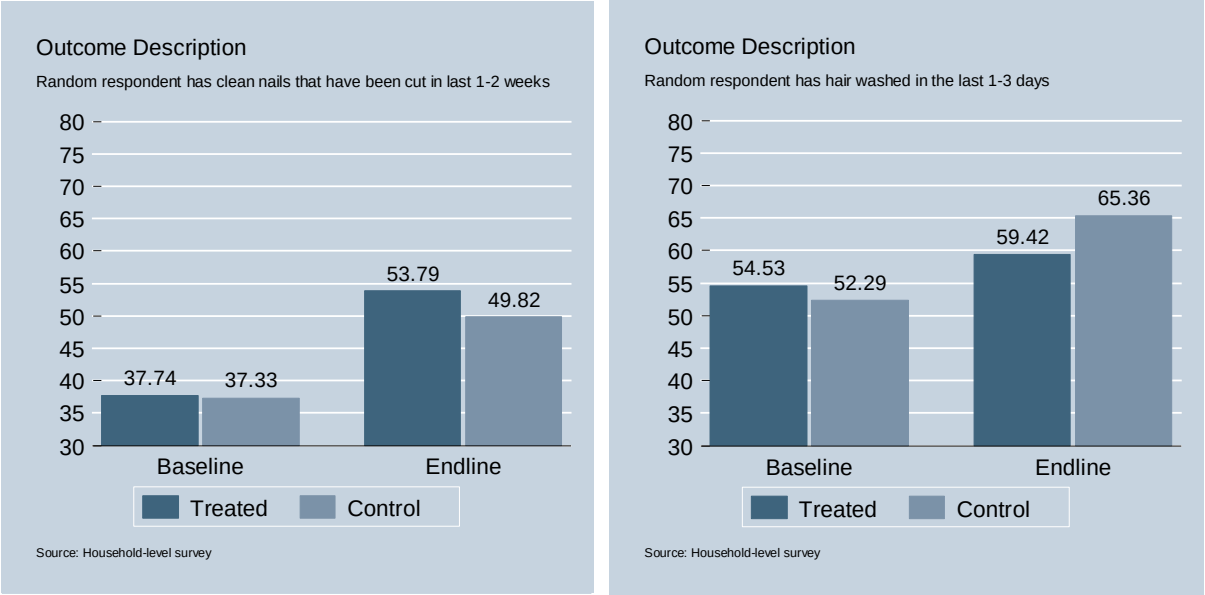
We find that for the first of these two indicators the baseline status was low and balanced – about 37% of the sample had well looked after nails. The percentage of washed hair was higher with on average just over 60%. For both indicators, we find an increase over time, especially so for the appearance of the nails. However, as shown in Table 29 below, the outcome phase intervention had no significant impact on the respondents' behaviour.

Table 27: Descriptive statistics – hygiene indicators (random respondent)

	Random respondent has clean nails that seem to have been cut in the last 1-2 weeks			Random respondent has hair that looks like it has been washed in the last 1-3 days		
	Obs.	Mean	SD	Obs.	Mean	SD
Baseline:						
Total	1039	37.54	48.44	1032	53.39	49.91
Control	525	37.33	48.42	524	52.29	50
Treated	514	37.74	48.52	508	54.53	49.84
Endline:						
Total	1096	51.82	49.99	1026	62.38	48.47
Control	542	49.82	50.05	511	65.36	47.63
Treated	554	53.79	49.9	515	59.42	49.15

Notes: The table provides descriptive statistics for the two primary sanitation behaviour outcomes, showing the number of observations (columns 1 and 4), the mean (columns 2 and 5) and standard deviation (columns 3 and 6) for each variable, both at baseline (upper panel) and at endline (lower panel), for the whole sample and control and treatment areas separately. Source: Household-level survey

Figure 12: Descriptive statistics – hygiene indicators (random respondent)



The other two variables we consider are whether the respondents' hand and feet are free of dirt. Descriptive statistics are shown in Table 28 and Notes: The table provides descriptive statistics for the two primary sanitation behaviour outcomes, showing the number of observations (columns 1 and 4), the mean (columns 2 and 5) and standard deviation (columns 3 and 6) for each variable, both at baseline (upper panel) and at endline (lower panel), for the whole sample and control and treatment areas separately. Source: Household-level survey

Figure 13.Figure 13. These reveal that about 62% of respondents had clean hands at the time of the baseline survey and this percentage increased by about ten percentage points when endline measures were taken. Feet tended to be dirtier, with only 41% of respondents having had clean feet at baseline and about 45% at endline. Again, we do not find that the intervention had any significant impacts on these indicators (Table 40).

Table 28: Descriptive statistics – hygiene indicators (random respondent)

	Random respondent has hands free of dirt			Random respondent has feet free of dirt		
	Obs.	Mean	SD	Obs.	Mean	SD
Baseline:						
Total	1041	62.73	48.38	1038	40.66	49.14
Control	526	63.12	48.29	525	40	49.04
Treated	515	62.33	48.5	513	41.33	49.29
Endline:						
Total	1101	72.75	44.54	1030	45.24	49.8
Control	544	71.14	45.35	507	44.97	49.8
Treated	557	74.33	43.72	523	45.51	49.85

Notes: The table provides descriptive statistics for the two primary sanitation behaviour outcomes, showing the number of observations (columns 1 and 4), the mean (columns 2 and 5) and standard deviation (columns 3 and 6) for each variable, both at baseline (upper panel) and at endline (lower panel), for the whole sample and control and treatment areas separately. Source: Household-level survey

Figure 13: Descriptive statistics – hygiene indicators (random respondent)

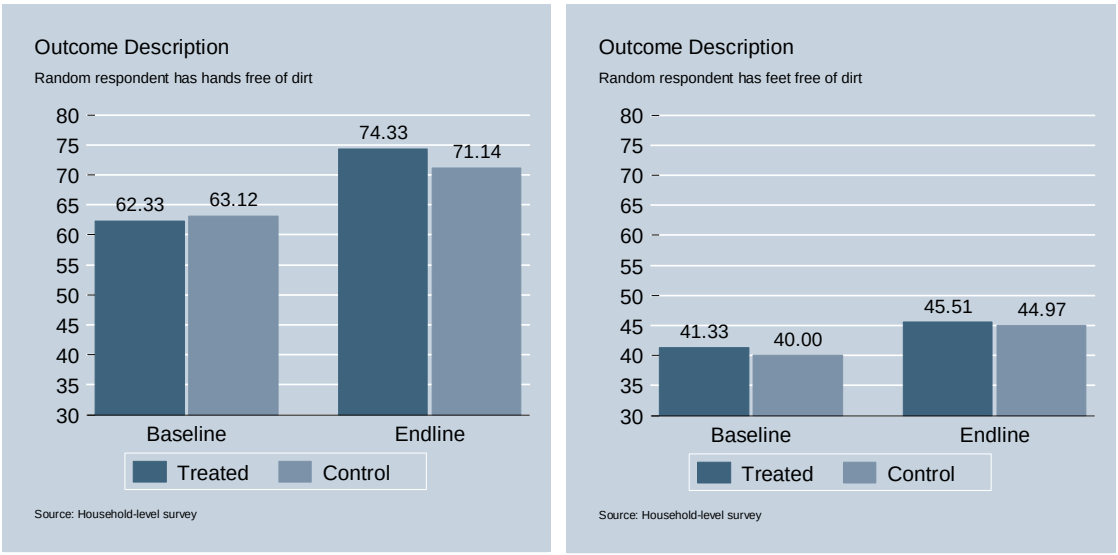


Table 29: Impact estimates, hygiene indicators (random respondent)

	Random respondent has clean nails that seem to have been cut in the last 1-2 weeks		Random respondent has hair that looks washed in the last 1-3 days		Random respondent has hands free of dirt		Random respondent has feet free of dirt	
Including covariates	No	Yes	No	Yes	No	Yes	No	Yes
Impact	0.04 (0.04)	0.05 (0.04)	-0.05 (0.03)	-0.04 (0.03)	0.03 (0.03)	0.04 (0.03)	0.01 (0.04)	0.02 (0.04)
District FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Mean (EL)	0.50	0.49	0.65	0.65	0.71	0.71	0.45	0.44
No. of Villages	123	123	123	123	123	123	123	123
No. of HHs	1,096	1,010	1,026	935	1,101	1,015	1,030	947

Notes: Standard errors in parentheses; standard errors are clustered at the unit of randomisation (villages)

Note: Controls include the lag (baseline value) of: 1) the outcome variables; 2) whether household has a functional toilet in the dwelling. Stars indicate statistical significance: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Source: Household-level survey

Annex E Descriptive statistics tables

E.1 Defecation behaviour – household level

Table 30: Descriptive statistics – Defecation behaviour (household level)

	Household has at-least 1 5+ year old member that openly defecates			Household has at-least 1 5+ year old member that openly defecates despite access to toilet		
	Obs.	Mean	SD	Obs.	Mean	SD
Baseline:						
Total	1132	16.96	37.55	1132	8.13	27.34
Control	561	16.58	37.22	561	8.2	27.46
Treated	571	17.34	37.89	571	8.06	27.24
Endline:						
Total	1132	31.89	46.63	1132	8.48	27.87
Control	561	33.87	47.37	561	8.56	28
Treated	571	29.95	45.84	571	8.41	27.77

Notes: The table provides descriptive statistics for the two primary sanitation behaviour outcomes, showing the number of observations (columns 1 and 4), the mean (columns 2 and 5) and standard deviation (columns 3 and 6) for each variable, both at baseline (upper panel) and at endline (lower panel), for the whole sample and control and treatment areas separately. Source: Household-level survey

E.2 Hygiene behaviour – main respondent

Table 31: Descriptive statistics – hygiene behaviour (main respondent)

	Main respondent typically uses improved hygiene after defecating			Main respondent typically uses improved hygiene before dinner		
	Obs.	Mean	SD	Obs.	Mean	SD
Baseline:						
Total	1132	87.46	33.14	1132	44.61	49.73
Control	561	89.66	30.47	561	44.39	49.73
Treated	571	85.29	35.45	571	44.83	49.78
Endline:						
Total	1132	83.3	37.31	1132	46.29	49.88
Control	561	84.31	36.4	561	45.28	49.82
Treated	571	82.31	38.19	571	47.29	49.97

Notes: The table provides descriptive statistics for the two primary sanitation behaviour outcomes, showing the number of observations (columns 1 and 4), the mean (columns 2 and 5) and standard deviation (columns 3 and 6) for each variable, both at baseline (upper panel) and at endline (lower panel), for the whole sample and control and treatment areas separately. Source: Household-level survey

E.3 Hygiene behaviour – random respondent

Table 32: Descriptive statistics – hygiene behaviour (random respondent)

	Random respondents improved hygiene after defecating			Random respondents improved hygiene before dinner		
	Obs.	Mean	SD	Obs.	Mean	SD
Baseline:						
Total	1044	84.87	35.86	1043	47.36	49.95
Control	526	86.31	34.41	526	46.77	49.94
Treated	518	83.4	37.25	517	47.97	50.01
Endline:						
Total	1108	80.6	39.56	1108	44.95	49.77
Control	546	79.3	40.55	546	43.04	49.56
Treated	562	81.85	38.58	562	46.8	49.94

Notes: The table provides descriptive statistics for the two primary sanitation behaviour outcomes, showing the number of observations (columns 1 and 4), the mean (columns 2 and 5) and standard deviation (columns 3 and 6) for each variable, both at baseline (upper panel) and at endline (lower panel), for the whole sample and control and treatment areas separately. Source: Household-level survey

Table 33 : Impact estimates – Hygiene behaviour (random respondent)

	Random respondents uses improved hygiene behaviour after defecating		Random respondents uses improved hygiene behaviour before dinner	
Including covariates	No	Yes	No	Yes
Impact	0.03 (0.03)	0.03 (0.03)	0.03 (0.04)	0.03 (0.04)
District FEs	Yes	Yes	Yes	Yes
Control Mean (EL)	0.79	0.80	0.43	0.42
No. of Villages	123	123	123	123
No. of HHs	1,108	1,025	1,108	1,024

Notes: Standard errors in parentheses; standard errors are clustered at the unit of randomisation (villages)

Note: Controls include the lag (baseline value) of: 1) the outcome variables; 2) whether household has a functional toilet in the dwelling. Stars indicate statistical significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Source: Household-level survey

E.4 Hygiene knowledge – main respondent

Table 34 Descriptive statistics – hygiene knowledge (main respondent)

	Main respondent says it is important to wash hands after defecating			Main respondent believes it is important to wash hands before eating		
	Obs.	Mean	SD	Obs.	Mean	SD
Baseline:						
Total	1132	73.5	44.15	1132	89.58	30.57
Control	561	73.26	44.3	561	88.77	31.6
Treated	571	73.73	44.05	571	90.37	29.53
Endline:						
Total	1132	78.27	41.26	1132	78.45	41.14
Control	561	75.94	42.79	561	78.61	41.04
Treated	571	80.56	39.61	571	78.28	41.27

Notes: The table provides descriptive statistics for the two primary sanitation behaviour outcomes, showing the number of observations (columns 1 and 4), the mean (columns 2 and 5) and standard deviation (columns 3 and 6) for each variable, both at baseline (upper panel) and at endline (lower panel), for the whole sample and control and treatment areas separately. Source: Household-level survey

E.5 Hygiene – facility observation

Table 35 Descriptive statistics – hygiene indicators (household)

	Households have handwashing facility			Households have handwashing facility with soap/detergent		
	Obs.	Mean	SD	Obs.	Mean	SD
Baseline:						
Total	1022	77.2	41.97	1022	32.29	46.78
Control	505	76.63	42.36	505	33.07	47.09
Treated	517	77.76	41.63	517	31.53	46.51
Endline:						
Total	1132	86.75	33.92	1132	37.01	48.31
Control	561	84.49	36.23	561	36.72	48.25
Treated	571	88.97	31.36	571	37.3	48.4

Notes: The table provides descriptive statistics for the two primary sanitation behaviour outcomes, showing the number of observations (columns 1 and 4), the mean (columns 2 and 5) and standard deviation (columns 3 and 6) for each variable, both at baseline (upper panel) and at endline (lower panel), for the whole sample and control and treatment areas separately. Source: Household-level survey

Annex F Heterogeneous impacts by output phase modality

This annex provides detailed results tables related to estimates by four output phase intervention modalities for sanitation behaviour and hygiene behaviour outcomes respectively. The four output phase modalities considered are: (i) the length of which NRSP has been working in study communities (in particular, whether they started when outcome phase activities started or earlier), (ii) whether NRSP visited the community during output phase activities less than twice a month or twice a month and more, (iii) whether a community received a demonstration latrine or not and (iv) whether someone in the village was responsible for sanitation promotion activities. Section F.1 provides detailed tables for sanitation behaviour and Section F.3 for hygiene behaviour.

F.1 Sanitation behaviour

Table 36: Heterogeneous Impact estimates, defecation behaviour (household level): by NRSP community presence

	Household has at-least 1 5+ year old member that openly defecates		Household has at-least 1 5+ year old member that openly defecates despite access to toilet	
Including covariates	No	Yes	No	Yes
Impact (NRSP in community <=2 years)	-0.06 (0.06)	-0.07 (0.05)	-0.00 (0.02)	-0.00 (0.02)
Impact (NRSP in community >2 years)	-0.11 (0.08)	-0.08 (0.06)	-0.03 (0.06)	-0.02 (0.06)
District Fes	Yes	Yes	Yes	Yes
Control Mean (EL)	0.36	0.36	0.09	0.09
No. of Villages	113	113	113	113
No. of HHs	1,050	1,050	1,050	1,050

Notes: Standard errors in parentheses; standard errors are clustered at the unit of randomisation (villages)

Note: Controls include the lag (baseline value) of: 1) the outcome variables; 2) whether household has a functional toilet in the dwelling. Stars indicate statistical significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Source: Household-level survey

Table 37: Heterogeneous Impact estimates, defecation behaviour (household level): By receipt of demonstration latrine

	Household has at-least 1 5+ year old member that openly defecates		Household has at-least 1 5+ year old member that openly defecates despite access to toilet	
Including covariates	No	Yes	No	Yes
Impact ()	-0.00 (0.05)	-0.00 (0.04)	0.01 (0.02)	0.01 (0.02)
Impact (received demonstration latrine))	-0.11 (0.08)	-0.11* (0.06)	-0.01 (0.03)	-0.01 (0.03)
District Fes	Yes	Yes	Yes	Yes
Control Mean (EL)	0.36	0.36	0.09	0.09
No. of Villages	119	119	119	119
No. of HHs	1,095	1,095	1,095	1,095

Notes: Standard errors in parentheses; standard errors are clustered at the unit of randomisation (villages)

Note: Controls include the lag (baseline value) of: 1) the outcome variables; 2) whether household has a functional toilet in the dwelling. Stars indicate statistical significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Source: Household-level survey

Table 38: Heterogeneous Impact estimates, defecation behaviour (household level): By presence of community sanitation promotion

	Household has at-least 1 5+ year old member that openly defecates		Household has at-least 1 5+ year old member that openly defecates despite access to toilet	
Including covariates	No	Yes	No	Yes
Impact (No-one to promote good sanitation)	0.06 (0.11)	0.06 (0.09)	-0.01 (0.04)	-0.01 (0.04)
Impact (Has person to promote good sanitation)	-0.08 (0.05)	-0.08* (0.04)	0.00 (0.02)	0.00 (0.02)
District FEs	Yes	Yes	Yes	Yes
Control Mean (EL)	0.37	0.37	0.09	0.09
No. of Villages	117	117	117	117
No. of HHs	1,077	1,077	1,077	1,077

Notes: Standard errors in parentheses; standard errors are clustered at the unit of randomisation (villages)

Note: Controls include the lag (baseline value) of: 1) the outcome variables; 2) whether household has a functional toilet in the dwelling. Stars indicate statistical significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Source: Household-level survey

F.2 Latrine ownership

Table 39: Impact estimates – Toilet ownership

	Household has a toilet in the dwelling		Household has a functional toilet in the dwelling		Household has a functional toilet in the dwelling and uses it		Household has access to an improved toilet	
Including covariates	No	Yes	No	Yes	No	Yes	No	Yes
Impact	0.05 (0.04)	0.06 (0.04)	0.06 (0.04)	0.06 (0.04)	0.06 (0.04)	0.06 (0.04)	0.06 (0.04)	0.07* (0.04)
District Fes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control Mean (EL)	0.70	0.70	0.68	0.68	0.68	0.68	0.63	0.63
No. of Villages	123	123	123	123	123	123	123	123
No. of HHs	1,132	1,132	1,132	1,132	1,132	1,132	1,125	1,085

Notes: Standard errors in parentheses; standard errors are clustered at the unit of randomisation (villages). Stars indicate statistical significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Source: Household-level survey

F.3 Hygiene behaviour

Table 40: Heterogenous Impact estimates, hygiene behaviour (random respondent): By Province

	Random respondents uses improved hygiene behaviour after defecating		Random respondents uses improved hygiene behaviour before dinner	
Including covariates	No	Yes	No	Yes
Impact in Sindh	-0.01 (0.05)	-0.01 (0.05)	0.06 (0.06)	0.06 (0.06)
Impact in Punjab	0.05* (0.03)	0.06 (0.03)	0.01 (0.06)	0.01 (0.06)
District FEs	Yes	Yes	Yes	Yes
Control Mean (EL)	0.79	0.80	0.43	0.42
No. of Villages	123	123	123	123
No. of HHs	1,108	1,025	1,108	1,024

Notes: Standard errors in parentheses; standard errors are clustered at the unit of randomisation (villages)

Note: Controls include the lag (baseline value) of: 1) the outcome variables; 2) whether household has a functional toilet in the dwelling. Stars indicate statistical significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Source: Household-level survey

Table 41: Heterogeneous Impact estimates, hygiene behaviour: by output phase modalities

	Presence, length of time			Demonstration latrine			Sanitation promoter		
α_1	<=2yrs	0.05	0.07	No	0.04	-0.01	No	0.05	-0.10
		(0.04)	(0.05)		(0.04)	(0.07)		(0.07)	(0.09)
α_2	>2yrs	-0.04	-0.09	Yes	0.00	0.05	Yes	0.01	0.04
		(0.05)	(0.07)		(0.05)	(0.06)		(0.04)	(0.05)
Inc. Covariates		Yes	Yes		Yes	Yes		Yes	Yes
District Fes		Yes	Yes		Yes	Yes		Yes	Yes
Control Mean (EL)		0.8	0.44		0.8	0.43		0.81	0.44
No. of Villages		113	113		119	119		117	117
No. of HHs		947	946		989	988		973	972

Notes: Standard errors in parentheses; standard errors are clustered at the unit of randomisation (villages)Stars indicate statistical significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Source: Household-level survey

Table 42: Heterogenous Impact estimates, hygiene behaviour (random respondent): By NRSP community presence

	Random respondents uses improved hygiene behaviour after defecating		Random respondents uses improved hygiene behaviour before dinner	
Including covariates	No	Yes	No	Yes
Impact (NRSP in community <=2 years)	0.05 (0.04)	0.05 (0.04)	0.06 (0.06)	0.07 (0.05)
Impact (NRSP in community >2 years)	-0.02 (0.05)	-0.04 (0.05)	-0.07 (0.07)	-0.09 (0.07)
District FEs	Yes	Yes	Yes	Yes
Control Mean (EL)	0.80	0.80	0.44	0.44
No. of Villages	113	113	113	113
No. of HHs	1,027	947	1,027	946

Notes: Standard errors in parentheses; standard errors are clustered at the unit of randomisation (villages)

Note: Controls include the lag (baseline value) of: 1) the outcome variables; 2) whether household has a functional toilet in the dwelling. Stars indicate statistical significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Source: Household-level survey

Table 43: : Heterogenous Impact estimates, hygiene behaviour (random respondent): By receipt of demonstration latrine

	Random respondents uses improved hygiene behaviour after defecating		Random respondents uses improved hygiene behaviour before dinner	
Including covariates	No	Yes	No	Yes
Impact (not received latrine demonstration)	0.03 (0.03)	0.04 (0.04)	-0.00 (0.07)	-0.01 (0.07)
Impact (received latrine demonstration)	0.01 (0.05)	0.00 (0.05)	0.04 (0.06)	0.05 (0.06)
District FEs	Yes	Yes	Yes	Yes
Control Mean (EL)	0.80	0.80	0.44	0.43
No. of Villages	119	119	119	119
No. of HHs	1,071	989	1,071	988

Notes: Standard errors in parentheses; standard errors are clustered at the unit of randomisation (villages)

Note: Controls include the lag (baseline value) of: 1) the outcome variables; 2) whether household has a functional toilet in the dwelling. Stars indicate statistical significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Source: Household-level survey

Table 44: Heterogeneous Impact estimates, hygiene behaviour (random respondent): By presence of community sanitation promotion

	Random respondents uses improved hygiene behaviour after defecating		Random respondents uses improved hygiene behaviour before dinner	
Including covariates	No	Yes	No	Yes
Impact (No-one to promote good sanitation)	0.01 (0.07)	0.05 (0.07)	-0.14* (0.08)	-0.10 (0.09)
Impact (At-least 1 person to promote good sanitation)	0.02 (0.04)	0.01 (0.04)	0.05 (0.05)	0.04 (0.05)
District FEs	Yes	Yes	Yes	Yes
Control Mean (EL)	0.80	0.81	0.45	0.44
No. of Villages	117	117	117	117
No. of HHs	1,053	973	1,053	972

Notes: Standard errors in parentheses; standard errors are clustered at the unit of randomisation (villages)

Note: Controls include the lag (baseline value) of: 1) the outcome variables; 2) whether household has a functional toilet in the dwelling. Stars indicate statistical significance: * p < 0.10, ** p < 0.05, *** p < 0.01.

Source: Household-level survey

Table 45: Heterogenous Impact estimates, handwashing after defecating (random respondent): By respondent characteristics

	Gender			Education			Literacy		
α_1	*Male*	0.01 (0.03)	0.01 (0.04)	*No Primary*	0.03 (0.04)	0.04 (0.04)	*Illiterate*	0.03 (0.04)	0.03 (0.04)
α_2	*Female*	0.04	0.04	*Primary*	-0.02	-0.01	*Literate*	0.01	0.01
		(0.04)	(0.04)		(0.04)	(0.05)		(0.04)	(0.04)
Inc. Covariates		No	Yes		No	Yes		No	Yes
District FEs		Yes	Yes		Yes	Yes		Yes	Yes
Control Mean (EL)		0.80	0.80		0.80	0.80		0.80	0.80
No. of Villages		123	123		123	123		123	123
No. of HHs		1,080	1,025		1,080	1,025		1,077	1,022

Table 46 Heterogenous Impact estimates, handwashing before eating (random respondent): By respondent characteristics

	Gender			Education			Literacy		
α_1	*Male*	0.04 (0.05)	0.05 (0.05)	*No Primary*	-0.00 (0.04)	-0.01 (0.04)	*Illiterate*	-0.01 (0.04)	-0.01 (0.05)
α_2	*Female*	0.03	0.02	*Primary*	0.14*	0.14*	*Literate*	0.12*	0.12*
		(0.05)	(0.05)		(0.07)	(0.07)		(0.06)	(0.06)
Inc. Covariates		No	Yes		No	Yes		No	Yes
District FEs		Yes	Yes		Yes	Yes		Yes	Yes
Control Mean (EL)		0.43	0.42		0.43	0.42		0.43	0.42
No. of Villages		123	123		123	123		123	123
No. of HHs		1,080	1,024		1,080	1,024		1,077	1,021

F.4 Additional details on sanitation behaviour by sub-groups

Box 3 and Section 6.1.2 provide information on defecation practices in our study areas, with information on sub-group specific findings provided in Section 6.1.3. We provide here additional information for some of these sub-groups; information that was not deemed directly relevant to speak to the impact evaluation but that was considered very interesting nonetheless.

F.4.1 The role of (outside,) male labourers

SOs and some FGD respondents, particularly in Bahawalpur, noted that male labourers and outsiders who came to the villages for work tended to defecate in the open in some cases, even if the rest of the community primarily used latrines. This was because they were often poor and did not have access to latrines during their temporary periods of stay in the communities.

> *If someone comes to the village for the harvest of wheat and stays there for one week, the local people will not let them use their toilets for the whole week. [As such] it is impossible to eradicate ODF at these places* (Social Organisers FGD, Bahawalpur)

F.4.2 Defecation practices of the elderly

The qualitative research revealed evidence that, where accessible, elderly members of households prefer the use of latrines for the convenience they afford. Latrines are located within the household and are more user-friendly which make it easier for the elderly to use them given their more limited mobility. If a latrine is not available, elderly members of the household, including elderly male members, were usually permitted to defecate somewhere close to the household compound for this same reason. Unlike the case of younger men, this was not frowned upon and in fact seen as necessary to facilitate the elderly where latrine usage was not an option. In addition, there were also reports of some elderly members of the community being more accustomed to their defecation practices and as such preferring to defecate in the open out of habit.

> *Yes there is a difference. Some people use latrines, some go to the fields or to a designated spot for defecation… The elderly now use latrines because they are near and present within the households. If someone doesn't have a latrine, they can use their relative's latrines* (Male key informant interview, Badin, T2)

F.4.3 The role of household size

Qualitative interviews also suggest that the size of the household or family influences defecation behaviour in some cases. This was particularly evident in Rahimyar Khan. When several people were expected to use a single latrine in a large household or joint compound, often the men, and at times women and children as well, resorted to open defecation practices for ease and to avoid waiting. Interestingly, some communities also mentioned family members deliberately defecating in the open to ease pressure on the latrine due to fears that it would clog or overflow too easily if everyone used it regularly throughout the day. Open defecation despite presence of a household latrine, was also reported at times when guests or relatives were visiting and there was again too much pressure on a single latrine, although sometimes excess individuals were reported to use latrines in a neighbour or close relative's house.

> *When there are large families and there is only one latrine, it is another major issue. In the beginning only adults used to visit latrine, but gradually children have also started using them* (Female FGD participant, Rahimyar Khan, T2)

F.4.4 The role of health and disability

There is some variation in the defecation patterns reported for the invalid, sick or those with a disability or health impairment. In several communities there was said to be no change from their standard defecation behaviour, other than that someone usually supported the invalids in accessing and using the latrine or defecation spot. However, as in the case of the elderly, invalids were usually permitted to defecate within or near the household compound for ease.

At least some respondents in the three districts also narrated that they used a tray or container filled with soil for invalids to defecate into while they remained in their position or in bed. The infirm, like the elderly often found it difficult to walk and/or squat so this mobile arrangement was done to facilitate them. The faeces were then thrown outside along with the used soil. In some cases, respondents mentioned creating special chairs or stools to place over the latrine to facilitate the elderly or disabled in using the facility without squatting.

There were also some reports from communities in Badin and Rahimyar Khan that the invalid and disabled received more grace and leeway from other community members and were allowed to use and share the latrines available in the community while younger, able-bodied individuals were usually refused.

> *When my father was very sick, it was very difficult to take him to the latrine. Therefore, we took a big dish, put some soil in it, and put it under his bed for him. When he was done with his defecation, we would pick it up and throw it out. Then we would wash him with water there because he could not walk to the latrine himself. We would do everything right there for him* (Male FGD participant, Bahawalpur, T1)

Annex G Details of SAWRP outcome phase activities

As described in Section 3.6 SAWRP provided monitoring data to the ePact research team. This annex contains the details of the activities as provided (reproduced verbatim) by the suppliers.

G.1 Plan International Pakistan activities

Activity	Description	Level	Frequency
Follow up of reported latrines	During Follow up Visit CRP, SOs and monitoring teams visit reported latrines (new and rehabilitated reported by CRP, try to meet the HH head otherwise with members of that households) during output phase to ensure continuous use of latrines. Important sanitation messages includes; cleanliness of latrine, importance of use and functionality.	Household	Monthly
Basic to Improved Latrine	As detailed below, those households which constructed basic latrines during output phase, are mobilized (though follow up visits) to construct improve latrines. They are also provided awareness about the nearest SaniMart facilities. There are number of households who have improved their basic latrines from their own resources. They are also provided awareness about the nearest SaniMart facilities. Low cost latrine options are also availble at SaniMart.	Household	Monthly
Basic to Improved Latrine though CIF and Sanitary marts	Majority of the marginalized households opted to construct basic latrines due to financial constraints. In order to facilitate them to improve their basic latrines, CIF and SaniMarts were planned and established. Purpose of ICF is to provide easy access to micro-financing to these households, while main purpose of SaniMarts is to access to low cost sanitation products. Financial support provided to poor people through interest free community investment funds for the improvement of their latrines.	Household	need based
Sanitation Melas	Sanitation and hygiene activities conducted in communities and schools for awareness raising	Community & UC	Once in a year

This impact assessment is being carried out by e-Pact. Britta Augsburg (britta_a@ifs.org.uk) from the Institute of Fiscal Studies is the workstream leader for the randomised control trial. Zach White (zach.white@opml.oc.uk) from Oxford Policy Management is the project manager of the evaluation. Julia Larkin is the evaluation Team Leader.

The contact point for the client is Stephen Lindley-Jones (stephen-jones@dfid.gov.uk). The client reference number for the project is PO6507.

Mobile sanitation marts	IP staff facilitates local entrepreneurs to organize these mobile marts. The major objective of this activity is to aware and provide access to variety of low cost sanitation items/material to those communities located far from main commercial hubs. This facilitates IP in mobilizing the households to improve their basic latrines.	UC	Monthly
Establishment of CIF	Community investment fund are established in 2 UCs of District RYK to provide interest free loan to poorest people to improve their sanitation facilities on easy installments. Local Support Organization (LSO) operates this fund as per agreed terms and conditions. Whoever, intends to avail this facility, contacts local CO/VO who after verification recommends him/her for this facility.	Union Council level	Once in a project
Health and Hygiene Sessions	H&H Sessions conducted by SOs (where required, by CRPs as well) on monthly basis to deliver hygiene messages in community and schools, through display of IEC material & demonstration.	Community	Monthly/weekly
Health and Hygiene sessions in schools and visit of WASH Clubs. Hand washing demonstrations	WASH clubs are formed in all Government schools in communities to spread message of WASH through Children Hand washing demonstrations at community level during health and hygiene sessions	School	Monthly
Ensuring Availability of Hand washing Facility	Social Mobilizers (SOs) have been assigned a cluster of communities. These SOs are mainly responsible for implementing SAWRP Outcome activities in these communities; for instance, conducting follow up visits in which they deliver Health & Hygiene (H&H) messages and monitor whether members of these households are practicing improved hygiene behaviour. Physical verification that if the household latrines are functional and being used by all members of the household all the time, visit water points to assess its functionality and use; make arrangements and conduct numerous BCC activities mainly community and school sessions, street theatres, hold meetings with local community organization and Operation & Maintenance (O&M) committees and hold capacity building sessions for these structures, hold orientation session with local religious leaders and representatives of local bodies. These SOs submit their field visit reports to their supervisors and district M&E maintains an Action Tracker for follow up and to ensure the compliance of agreed mitigation measures against identified issues. It is important to note these follow up visits are not planned separately for each activity. Instead, file plan is developed to ensure that every community/household is accessed at-least once during the quarter. So in one visit usually multiple activities are covered.	Household	Monthly
Sanitation week	Sanitation week were celebrated 2 times in all three UCs. A huge BCC campaign run in the shape of local theatre, WASH activities in schools, matches between different UC along with H/H sessions	Community and School	Once in a year
BCC activities	Includes: Hygiene sessions, Street theaters, magic shows, puppet shows, friendly cricket matches in youth. NRSP working on its three step community structure to make committees at village, Revenue village and UC levels to keep working information in their respective UCs	Community	1 time activities or annually

Celebration of International Days related to WASH	Celebration of global days i.e. World Water Day, World Toilet Day, Global Hand washing Day to raise awareness among community to improve their WASH services at household and community level	Community and school level	annually
Water user committees for the sustainability of water points;	Water user committees are responsible for operation & maintenance of water points and also ensure the sustainability	community level	Regular follow up
Safe Drinking Water	Hand pumps are installed for safe drinking water; water test after 8 months, while field work they meet SDW beneficiaries and community structure for improvement in habits regarding SDW	community	Monthly
Water quality testing;	Quality testing of water points installed under SAWRP to continue using safe water. SO raise awareness about water borne diseases.	Community	every 8 months
Follow up of Hand pump functionality and use of safe drinking water	Social Mobilizers (SOs) have been assigned a cluster of communities. These SOs are mainly responsible for implementing SAWRP Outcome activities in these communities; for instance, conducting follow up visits in which they deliver Health & Hygiene (H&H) messages and monitor whether members of these households are practicing improved hygiene behaviour. Physical verification that if the household latrines are functional and being used by all members of the household all the time, visit water points to assess its functionality and use; make arrangements and conduct numerous BCC activities mainly community and school sessions, street theatres, hold meetings with local community organization and Operation & Maintenance (O&M) committees and hold capacity building sessions for these structures, hold orientation session with local religious leaders and representatives of local bodies. These SOs submit their field visit reports to their supervisors and district M&E maintains an Action Tracker for follow up and to ensure the compliance of agreed mitigation measures against identified issues. It is important to note these follow up visits are not planned separately for each activity. Instead, file plan is developed to ensure that every community/household is accessed at-least once during the quarter. So in one visit usually multiple activities are covered.	Household	monthly
Development of Water Safety Plan	Environmental sustainability, clean environment. Raising awareness about water borne diseases. Functionality, use and ownership of water points	community level	Once, but can be revised on need base
Monthly Meeting with CO/VO/LSO	Monthly meeting are arranged by CO/VO in the village with the objectives of discussing all the project activities (vis-a-vis Water, Sanitation & Hygiene) in their respective village; progress against the planned activities/targets, discuss issues and identify and agree on mitigation measures. Apart from VO/CO members, general members of the community, CRP and SO also attend this meeting.	Community	monthly or need based

G.2 WaterAid Pakistan activities

Activity	Description	Level	Frequency
Hygiene Session	NRSP closely works with communities on hygiene to sensitize them on following key topics 1. Social mobilization team hold hygiene sessions on clean and safe drinking water. Now communities are well aware on safe drinking water. 2. Mobilization teams hold sessions on use of toilets in respective communities which resulted that communities left their practices of open defecation/ feaces in the open area. During discussion NRSP teams ensured that toilets are functional and clean at household level and all toilets are in use and fly free. 3. Social Mobilization teams hold sessions on hand washing practices with soap on critical timings and personal hygiene (In this activity staff performed demonstration during sessions and hold communities engaged for replication at village level) 4. Communities are sensitized during hygiene sessions on cleaning of sewerage lines and information on safe sanitation system in respective communities and ensured the replication of hygiene messages at household level. It is ensured that at least one participants form every HH take part in the Hygiene session. Hygiene sessions are conducted on Monthly basis.On the account of record following evidence kept on board: • Snap shots Depending on Community Consesus • Attendance sheets • Resolution • Proceeding register at CO level.	Community level	Monthly Basis
Broad Based Community Meeting (BBCM)	What BBCM involves/entails: • Village level broad-based community gathering • Engage and involve 3-5 voluntary 'Sughar Maaa' in each target village, based on set criteria, so that they positively promote and reinforcing interest behaviours along with other community wash groups. • Engage, appreciate and excite the communities to build a community-wide momentum for the desired change through wider community participation and representation especially women and youth. • Collective self-analysis by the community to assess & appreciate their progress towards ODF targets so far and what still needs to be done. • Elicit family and community commitments and pledges around areas for improvement • Publically recognize and reward well-performing (compliance) households/individuals and groups from community. • Create social pressure and promote interest behaviours as social norms Entry point intervention • Household and community level • Introduction of staff and project • Objectives of meeting • Problem identification regarding water, sanitation and hygiene • Action planning against identified problems Community self-analysis of the targets, progress achieved and what further needs to be done. • Positive reinforcement of interest behaviours through innovative fun, infotainment and group activities using emotional drivers (promoter and barriers) • Use of a range of printed, audio-video BCC materials Record as evidence • Pictures/photos of the event • Community action plans• Record of Community Wash Groups	Community level	Quarterly Basis

Activity	Description	Level	Frequency
ODF Celebrations	After the series of Hygiene sessions communities constructed their latrines and cleaned their households and surroundings. Community Organizations submitted resolution for the acquiring of ODF status. After the confirmation by project team the application processed towards District Open defecation Free Inspection Committee (OIC). OIC visited the respective village and ensured that whole community is using toilets for defecation. The surroundings of the village cleared from Human Excreta and animal dung. ODF certificate awarded to the communities and Sign board was installed in the village. After declaration of ODF village, ODF celebration organized on achievement of ODF status as the message can be broader replicated and communities of surrounding come to know about the importance of ODF. Attendance of the participants, report and snapshots kept as record as a means of verification.	Community level	One time activity
Training on DRR	Community training organized on DRR, which was participated by male and female members of the COs which benefited in sustaining project schemes and coped with the disaster and reduced affects and made communities able for preparedness for disasters. Snapshots, attendance kept in record as an evidence.	Community Organization level	One time activity
Training on E&I	Community training organized on Equity and Inclusion keeping in view on Water Rights. This training enabled communities on decision making to include the opinion of all community members based on age, gender, religion or any disability while taking any decision. Snap shots and attendance sheet kept as record for evidence.	Community Organization level	One time activity
Training on O&M	Training on Operation and Maintenance (O&M) organized for CO members which was participated by male and female members of the communities. This training benefitted communities for sustaining project schemes. Snapshots, Attendance sheet, Report are kept record as an evidence.	Community Organization level	One time activity
Street Theatres	Street Theatre are organized at community level for sustaining the Open Defecation Free environment and maintain the status of the schemes. Community is sensitized in enlightened manner.Stories are created ressembling the community scenario and benefit of use toilet for defecation and disadvantages of open defecation and consiconses. The event was documented with Theater report, snapshots as an evidence..	Community level	One time activity
Distrrict WASH Forum	District WASH forums are notified by District Governement of both the district Badin and Thatto, District WASH forum meetings are heald on Quarterly Basis in those Meeting Government Line Departments Like viz: Education, Social Welfare, Local Bodies deptt, represntatives from community organizations and organizations working on WASH in the districts sit together to share their learnings and achievemnets. This meeting is chaired by District Local governement Department Head and Social welfare department.	Local Government	Quarterly Basis

Activity	Description	Level	Frequency
Meeting with Local Support Organization(LSO) on WASH	NRSP has three tier mobilization structure which helps in sustainablity and adaptablity of projects. It was also ensured that all the stake holders paly their part on mainstreaming WASH. Meeting with LSO Bodies were arranged to enusre thir part is played effectively. LSo's are taking responsiblity by creating awareness in community and monitoring the project schemes for quality result.	Community Organization level(LSO)	Quarterly Basis

Annex H Terms of reference

Introduction

1. DFID is seeking a service provider to provide an independent and rigorous monitoring, verification and evaluation (MVE) function for the WASH Results Programme (a large delivery programme providing access to water, sanitation and hygiene).
2. These Terms of Reference (TORs) set out DFID's requirements for monitoring, verification and evaluation including the objective, scope of works and outputs for the function and other relevant design information.

Objective

3. To provide an independent and rigorous monitoring, verification and evaluation function for the WASH Results Programme.

Recipient

4. The recipient of this service will be the DFID WASH Policy Team.

Background

5. The Department for International Development (DFID) manages the UK's aid to poor countries and works to get rid of extreme poverty. DFID is working to reach the Millennium Development Goals (MDGs), the international targets agreed by the United Nations (UN) to halve world poverty by 2015.
6. Progress on Water, Sanitation and Hygiene (WASH) provision is critical to the achievement of these targets. Millennium Development Goal (MDG) 7c specifically relates to sustainable access to safe drinking water and basic sanitation by 2015.
7. The UK Government has committed to scale up our results on Water, Sanitation and Hygiene to support 60 million people gain access to sustainable WASH services before the end of December 2015. To achieve these results, DFID has commissioned a portfolio of programmes, including a WASH Results Programme managed by the WASH Policy Team in Human Development Department of DFID's Policy Division.
8. The WASH Results Programme is a results oriented fund which provides funding to organisations capable of delivering water, sanitation and hygiene interventions for people who currently do not have access to clean water or safe sanitation. WASH Results Programme projects must demonstrate real, positive, sustainable changes to the lives of poor people. DFID will award an expected three contracts[98] to organisations based on payments linked to output results (payment by results).
9. WASH interventions supported by DFID must be aimed at providing poor people with access to improved drinking water supplies, access and use of basic sanitation and the adoption of behaviours that reduce the health risks caused by poor hygiene, including hand washing at critical times. This support should enable approximately 4.5 million people to receive access to sustainable WASH services by December 2015. The WASH Results Programme will be managed within DFID with contracts reporting to an identified fund manager within DFID. Full

[98] As per the TORs for the main programme, 3 contracts are expected (two WASH contracts each to reach 1 million people, and one sanitation and hygiene contract to reach 2.5 million people). However DFID reserves the right to award more contracts (up to a maximum of five) or fewer contracts, as outlined in the TORs.

details of the WASH Results Programme will be available on DFID's website www.gov.uk/DFID

10. DFID' approach[99] to payment by results involves three key elements:
    i. Disbursements tied to the achievement of clearly specified results: payment for outcomes such as completion of education, rather than payment for inputs such as provision of textbooks;
    ii. Recipient discretion – the recipient has space to decide how results are achieved; and
    iii. Robust verification of results as the trigger for disbursement.

    These relate directly to this results programme: i. the TORs for the overall programme set the results expected; ii, organisational discretion with the successful contract holder on the approach to reaching the required results, and iii. robust verification, which for this programme includes strong systems in the contracted organisations for achieving the results, and the services covered in these terms of reference to independently audit the quality of the contracted organisations results reporting.
11. The monitoring and verification of the results achieved through the WASH Results programme will be critical to the credibility and robustness of the data used to track and report on progress. To support the DFID Fund Manager in ensuring robust monitoring, verification and evaluation systems and data are in place, we are now looking to appoint a Monitoring, Verification and Evaluation (MVE) Service Provider to run for the lifetime of the WASH Results Programme over the next 5 years.

Scope of Work and Requirements

12. Design and implement a system of third-party results verification (including working with WASH Results Programme suppliers to define measurement methods and monitoring, verification and reporting requirements for payment) for those parts of projects operating under a Payment by Results framework. This will be used to verify results and outputs to inform payments. The MVE provider will audit contractors' results reporting within the contract terms related to payment by results.
13. The MVE Service Provider will develop a monitoring framework to assess progress and performance of individual projects funded under the WASH Results programme. Where this indicates that corrective action is required, the MVE Service Provider will inform both DFID and the implementing partner and facilitate discussion of corrective actions required. The MVE Services Provider will also assess and audit the monitoring systems and results reporting within each project and make recommendations for improvement where required. Assist DFID with monitoring contracts under this Results Programme. These monitoring tasks will include:
    - early monitoring of activities undertaken in the inception phase of the main results contracts
    - Un-announced visits to project work.
    - Quality assurance of contract holders results returns
    - Facilitation of annual review missions with DFID involvement.
14. Develop one or more impact evaluations on key issues agreed during the inception with DFID, but potentially covering one or more of the following: approaches to service delivery at scale, efficiency and effectiveness across different contexts, sustainability and Payment by

[99] https://www.gov.uk/government/publications/payment-by-results

Results.

15. The MVE Service Provider has knowledge and understanding of:
    - monitoring and evaluation of development programmes using both quantitative and qualitative methods;
    - monitoring and evaluation of WASH programmes, including testing of WASH outcomes;
    - social research management;
    - Payment by Results programmes
    - WASH sector governance and political economy analysis.
    - management of impact evaluations; and
    - evaluations in the context of major donor interventions, ideally focused outside of government.
16. In addition the MVE Service Provider is expected to:
    - establish a good working relationship with the DFID Fund Manager and implementing partners;
    - support the DFID Fund Manager to establish appropriate monthly reporting mechanisms;
    - provide technical support on the use of Payment by Results;
17. The services under the monitoring and verification are primarily to support DFID's role in holding results contract holders to account for delivery within their contracts. Timeliness of work under the monitoring and verification is therefore critical. The evaluation component will also provide findings to inform annual reviews of the programme, to allow for learning-feedback to inform progress under the project and permit course correction.
18. Where possible the evaluation questions, identified in the inception phase, should follow the OECD DAC Criteria for Evaluating Development Assistance[100] including the headings of relevance, effectiveness, efficiency, impact and sustainability.

Outputs

19. The MVE Service Provider is responsible for drafting a Monitoring, Verification and Evaluation Framework for approval by DFID within the first 3 months. This inception report should contain:
    - a monitoring, verification and evaluation strategy and implementation plan;
    - a risk management plan;
    - a quality assurance plan;
    - an outline of proposed methods for assessing core indicators;
    - an outline of the proposed approach to assessing project-specific additional indicators;
    - an outline of the proposed approach to assessing programme performance;
    - an outline of proposed evaluation questions (based on identification of what is feasible within given constraints), design and methods; and

[100] http://www.oecd.org/dac/evaluation/daccriteriaforevaluatingdevelopmentassistance.htm

- a strategy for disseminating information to key stakeholder and partners.

20. Once the inception report it is approved, it is expected that the MVE Service Provider will be responsible for delivering the outputs outlined below.

21. Support technical review of contractors proposed work, including assessment of the approaches proposed, review of contractors approach to sustainability, and checks on environment appraisals.

22. Tracking progress to ensure robust measurements of performance at the project and programme level including:

    - auditing partner approaches to monitoring and carrying out reviews, field-work and spot checks to provide assurance of robust performance management processes;
    - establishing data requirements to inform value for money and sustainability metrics. Value for money metrics will include key unit costs (eg. cost per latrine, cost per person of providing improved water supply) and also measures of effectiveness (measures of behaviour change such as the percentage of people reported as benefitting from hygiene interventions that are hand washing with soap and at critical times or percentage of beneficiaries that continue to use improve sanitation facilities). DFID has provided further information[101] on understanding value for money at the economy, efficiency and effectiveness levels. Sustainability metrics will include system functionality, coverage of lifetime costs and other key factors likely to promote sustainability such as institutional capacity and environmental factors. Programme monitoring data and additional spot checks will inform these metrics.
    - allowing comparison between programmes and to inform future DFID and wider sector programming;
    - quality assuring project progress reports, with a focus on ensuring robust tracking of results and performance based on agreed milestones and targets and challenging data and conclusions if necessary;
    - notifying DFID of progress with projects at least quarterly and monitoring, managing and advising DFID on risks to programme delivery and mitigation measures;
    - aggregating data and preparing programme progress reports to provide overall assessments of programme performance against milestones and targets in the WASH Results Programme logframe; and
    - aggregating end of project data and preparing a programme completion report.

23. The Monitoring, Verification and Evaluation Service Provider will also be responsible for verifying outputs, by auditing the reports of service providers, to inform results-based payments. This will include:

    - Working with contract holders to establish results baselines
    - Undertaking assessments of organisations capacity to report on results
    - Advising organisations' on improvements to process for results reporting.
    - Provide comments (by desk review) on the quarterly reports from service providers.
    - Provide a full audit, annually, of contract holders' results reporting, including visiting projects.

[101] https://www.gov.uk/government/publications/dfids-approach-to-value-for-money-vfm

24. In early 2016 the DFID Fund Manager will arrange an event at which the WASH Results Programme projects will be able to demonstrate the main results, findings and lessons from programme funded activities. The MVE services provider will support this by preparing a document and a presentation which summarises the results to-date and the lessons learned.

Evaluation Requirements

25. Development of a detailed strategy and implementation plan for evaluation including evaluation design and methods and timings for collection of baseline data.

26. Work with DFID to select, design and administer one or more depth evaluations at programme level and on a select number of project interventions and thematic areas. These decisions will be based on relevance to the overall objectives of the WASH Results Programme, potential for wider DFID and global lesson learning and the potential to fill key knowledge gaps and feasibility and cost of collecting data.

27. Implementation of the evaluation framework agreed on with DFID including:

    - tracking whether assumptions set out in the Theory of Change and logframe hold;
    - evaluating innovative elements of projects (including results-based finance);
    - developing the evidence base on value for money metrics, comparing suppliers and approaches across contexts;
    - identifying factors which have enhanced or impeded the sustainability of WASH interventions;
    - undertaking case studies of WASH projects at scale, to support the evaluations.
    - addressing knowledge gaps (e.g. effective approaches to behaviour change, community monitoring and payment by results); and
    - assessing how the wider environment has enabled or impeded achievement of programme objectives and identifying implications for programming.

28. Dissemination of lessons to inform WASH Results Programme evolution and wider DFID and global programming including:

    - with the DFID Fund manager, disseminating lessons learned and reporting those to DFID to agree actions for the WASH Results Programme;
    - ensure lessons are drawn out on what works in WASH, triangulated with other evidence, and reported to DFID. These lessons may be both immediate and used to inform future WASH Results Programme evolution or longer term and used to inform future DFID or others' interventions;
    - ensure lessons are drawn out on the costs and benefits of the Results Programme approach compared to other approaches including DFID bilateral aid and other DFID operations;
    - linking with the WASH Operational Research Programme to share monitoring information and evaluation findings and to co-ordinate analysis and dissemination;
    - design an innovative strategy to disseminate lessons through a variety of mediums and engage key partners and stakeholder in lesson learning on implementation and good practice from the WASH Results Programme, ensuring this reaches a wide audience (including project implementing partners, national governments, DFID country offices, the private sector and civil society).

Constraints and Dependencies

29. The MVE Services Provider will be expected to provide its own overseas duty of care in relation to its employees and other personnel it retains and logistical arrangements. If deemed necessary DFID may need to be convinced that systems and procedures that it has in place are adequate if traveling to conflict affected countries.

30. Clear separation between the team implementing the monitoring and verification elements of these services and the team implementing evaluation will be required to ensure independence of the evaluation elements.

Reporting and Monitoring and Evaluation of the MVE Services Provider

31. Key Performance Indicators (KPIs) will be agreed between the MVE Services Provider and DFID during the Inception phase. These will ensure that the management of the contract is undertaken as transparently as possible and to ensure that there is clarity of roles and responsibilities between the DFID WASH Results Programme team and the MVE Services Provider. The MVE Services Provider will need to demonstrate to DFID, at quarterly intervals its performance against these KPI's.

32. DFID will evaluate the performance of the MVE Services Provider throughout the life of the programme and at least twice yearly including as part of DFID's standard Annual Review of the programme. The MVE Services Provider will be expected to submit progress reports and lessons presented written and orally to DFID twice annually in-line with DFID's programme cycle as outlined in the requirements section of this ToR. It is expected that the MVE Services Provider take a proactive approach to notifying DFID of any matters which may require immediate attention.

33. The Monitoring, Verification and Evaluation components will report directly to DFID. To ensure that the evaluation (which would include evaluation of DFID's role in fund management of the programme) is rigorous, all evaluations will be made public and subject to the scrutiny which DFID operates within, which includes parliament and the Independent Commission on Aid Impact.

Timeframe and Contractual Arrangements

34. The contract for the MVE Services Provider will be awarded from May 2014 to November 2018. The contract is designed to end after financing is dispersed to allow a final evaluation of projects to be completed. The inception phase will be for a period of 6 months. DFID reserves the right to extend the contract for a further 2 years, on basis of continued need, and availability of funding.

35. The WASH Results Programme will run for 4.5 years (2014 – 2018).

36. DFID will issue a contract for the full programme duration however there will be a formal break clause in the contract at the end of the inception period. Progression to the implementation phase will be dependent on strong performance by the SP during the inception period and delivery of all inception outputs, including a revised proposal for implementation period. We expect costs for implementation to remain in line with what was indicated in your original proposal (any increase in costs will be subject to agreement with DFID), with costs such as fee rates fixed for contract duration. DFID reserves the right to terminate the contract after the inception phase if it cannot reach agreement with the SP on the activities, staffing, budget and timelines for the implementation phase.

37. DFID reserves the right to scale back or discontinue this programme at any point (in line with our Terms and Conditions) if it is not achieving the results anticipated. Conversely, we may also scale up the programme should it prove to be having a strong impact and has the potential to yield better results.

DFID coordination and management

38. The DFID WASH Fund Manager will have the day-to-day oversight and management of the MVE Services Provider. The DFID WASH Fund Manager will monitor operational and financial progress drawing on inputs from the WASH Team staff and will raise any issues that require attention to DFID senior management and Ministers as necessary.
39. The DFID Evaluation Department (EvD) support the DFID Fund Manager by providing strategic advice as required and ensuring that evaluation and monitoring activity undertaken by the MVE Services Provider aligns with wider DFID activity. The DFID Fund Manager team will work alongside the MVE Services Provider to consider what input is required, by whom and at what times to ensure technical advice is on hand at the right times.
40. There will be regular meetings between the DFID Fund Manager, staff from the WASH Policy team and the MVE Services Provider.

H.1 Clarifications to the Terms of Reference

This section clarifies our understanding of the terms of reference (TOR) for this evaluation. The TOR specify developing "one or more impact evaluations on key issues agreed during the inception with DFID, but potentially covering one or more of the following key issues agreed during the inception with DFID: approaches to service delivery at scale, efficiency and effectiveness across different contexts, sustainability and Payment by Results" (see above for the full TOR).

We propose to address a selection of these issues through activities across two interlinked workstreams – namely the **Programme evaluation Workstream (PEW) and the RCT Research Study.** Based on the list outlined in the TOR, these will jointly gather evidence on 'Payment by Results', 'efficiency and effectiveness across different contexts' and 'sustainability'.

With regard to the evaluation purpose

Item 17 of the TOR state that the evaluation component "will […] provide findings to inform annual reviews of the programme, to allow for learning-feedback to inform progress under the project and permit course correction". Item 27 also highlights that the evaluation framework would "address knowledge gaps (e.g. effective approaches to behaviour change, community monitoring and payment by results)". We therefore understand the evaluation to be primarily undertaken for learning purposes as opposed to accountability purposes. As a result, our evaluation design favours depth over breadth: We propose a case study approach to understand the mechanisms operating under certain contexts, as opposed to an evaluation design which is able to verify programme outcomes in every programme country.

Nonetheless, for practical reasons, the dissemination of the evaluation findings (i) to the Suppliers, (ii) to DFID, and (iii) to a wider policy audience will be organised under the Learning and Dissemination component managed by ITAD. This choice was made because the verification team is already planning regular feedback and learning sessions with DFID and the Suppliers. Instead of duplicating efforts, the evaluation team will participate at these meetings to disseminate the findings of the evaluation.

With regard to the evaluation scope

Item 26 of the Terms of Reference outlined that the evaluation team would select, design and administer one or more depth evaluations at programme level and on a select number of project interventions and thematic areas. As a result, we propose both in-depth evaluation activities through country case-studies, as well as broader evaluation activities at programme-level.

With regard to the RCT Research Study

The OECD-DAC criteria situate 'impact' as the causal effect of the programme on indicators of interest. The programme's effect can relate to medium-term 'outcomes' (e.g. use of water or sanitation services) or to longer-term 'impacts' (e.g. improved health and welfare as a result of the use of water or sanitation services). In both cases we would refer to these as "impact" evaluations.

While the delivery of programme outputs should generally lie within the control of Suppliers, outcomes and impacts lie progressively further from their control. The achievement of desirable **outcomes** involves the use of services and the practice of new behaviours. There is a level of intent and choice on the part of service users and those practising improved hygiene behaviours which Suppliers cannot fully guarantee – although they can and should influence those choices.

Beneficial **impacts** on health may require concurrent changes in nutrition, in the cleanliness of the living environment, and in the behaviours of others (e.g. community-wide compliance with latrine use and good hygiene practice), amongst other factors. Similarly, impacts on educational attainment, income and quality of life only follow if numerous other conditions are fulfilled. These mostly lie outside the sphere of influence of WASH implementing agencies. Better WASH outputs and outcomes are necessary but not sufficient conditions for the achievement of beneficial impacts.

In short, there are four main reasons why our evaluation will not focus on attributing any changes in longer-term health or non-health 'impacts' to the WASH Results Programme:

1. Impact indicators (especially those related to health and quality of life) are difficult and expensive to measure – undertaking such measurement would not represent value for money;
2. We currently possess inadequate theories of change and explanatory capability to say definitively why certain impacts have or have not materialised – we may find that diarrhoea morbidity in under-fives has reduced by a certain percentage, but we cannot explain why it was not reduced by more than this; or we find no change in some other indicator, but can only speculate as to why;
3. We already know that outcome-level results in WASH are some of the pre-conditions for beneficial impacts – we would learn nothing new; and
4. The use of better WASH services and the practice of better hygiene are worthwhile results in themselves, and they should be available to all.

As a result, in consultation with DFID, we have proposed the "impact evaluation" – referred to as RCT research study in this document – to focus our evaluation on medium-term 'outcomes', which would focus on the following aspects set out in the TOR:

- *"identifying factors which have enhanced or impeded the sustainability of WASH interventions"*
- *"assessing how the wider operating environment has enabled or impeded achievement of programme objectives and identifying implications for programming".*

The likelihood of health and non-health impacts being achieved by the WASH Results Programme will nonetheless be explored in a qualitative manner, to better understand the prerequisites which favour health impacts, and to understand if there have been any unintended positive or negative impacts on service users as a result of the WASH Results Programme.

With regard to the Programme Evaluation Workstream (PEW)

The programme evaluation will focus on two of the aspects listed in the TOR:

- *"Efficiency and effectiveness across different contexts: develop the evidence base on value for money metrics, comparing Suppliers and approaches across contexts"; and*

- *"Evaluate innovative elements of projects (i.e. the results-based finance modality)".*

We understand the former to be part of the tracking of VFM metrics, listed under *verification* activities in the TOR. However, the TOR assume that programme monitoring data and additional spot checks will be sufficient to inform these VFM metrics. As this is not the case, VFM analysis will primarily fall under the responsibility of the *evaluation* team, instead of the verification team. Due to the additional cost of complementing the insufficient monitoring data, this VFM analysis will be undertaken in case studies, as opposed to in every programme country.

With regard to the latter, we will examine the role of the PBR modality in delivering outcomes through a theory-based approach.

With regard to the evaluation questions

The TOR did not outline any evaluation questions, proposing that these, where possible "should follow the OECD DAC Criteria for Evaluating Development Assistance including the headings of relevance, effectiveness, efficiency, impact and sustainability". As a result, we have drafted our own high-level evaluation questions along the DAC criteria, which were signed off by DFID in September 2014.

H.2 Identification of research questions

In discussing how the thematic focus of this research study was developed it is worth noting that the focus of the RCT component of the evaluation changed during the 'exploratory phase' (post-inception and pre-implementation) of the evaluation. Initially the RCT design centred on the relationship between the role of water points in ensuring sustained sanitation outcomes. This section discusses how the initial focus was defined, why the water points RCT was rejected, and how the revised thematic focus was established.

For the thematic focus on the research study, we considered several options. The choice of proposed research is the result of a careful process of interaction between stakeholders, striking the balance of meeting four key aims:

i) The research questions should be relevant to the sector and contribute to the current status of the literature, specifically, they should speak to one or more of the key knowledge gaps identified in the DFID Evidence Paper: WASH (2013);

ii) The research questions should link to the grant;

iii) The design should be rigorous so as to be able to make a significant contribution;

iv) The design should fit the implementing partners' timeline and should imply only minimal adjustments to their implementation approach.

An assessment of the overall PbR was early on decided could not be a focus on the RCT study for one because the establishment of a control group in which no overall PbR activities would be implemented (including output phase) was deemed infeasible and also did not fit with timings of discussions.

When discussions started, evaluation of both output and outcome phase activities was considered, but overall keeping a focus on understanding the effectiveness on achieving sustainable outcomes.

Two important challenges were faced at the design stage. For one, the majority of outcome phase activities were undefined. This was the case since consortium partners were very engaged in the output phase and had limited time to think about the outcome phase. In discussing the design, we hence had to strike a balance between identifying research areas that met the above mentioned criteria, and working with limited information. The second constraint was that output phase activities had already started, which posed challenges in identifying areas in which certain activities had not yet taken place in order to introduce the randomization of activities to treatment and control.

Despite these significant challenges, detailed discussions amongst all stakeholders led to the identification of five possible research questions. These identified questions are as follows:

1. The role of water points in achieving sustained sanitation outcomes
2. The role of demonstration latrines in inducing uptake and moving up the sanitation ladder
3. Sanitation marketing and its role in supporting uptake and moving up the sanitation ladder
4. The role of Community Resource Persons (CRPs)in achieving sustainable sanitation
5. The role of follow-up activities in achieving sustainable sanitation

Table 47 below summarizes how each of these research areas fared in terms of the four key objectives, following the table is a brief discussion of the key considerations surrounding each area.

Table 47: Comparison of the research themes again the study's feasibility and aims

	I	ii	iii	Iv	V
	The role of water points in achieving sustained sanitation	**The role of demonstration latrines in inducing uptake and moving up the**	**Sanitation marketing and its role in supporting uptake and moving up the**	**The role of Community Resource Persons (CRPs) in achieving sustainable sanitation**	**The role of follow-up activities in achieving sustainable sanitation**

			sanitation ladder	**sanitation ladder**		
Key gaps the research speaks to* **		D, E, F	D, E	D, E, G	D, E, F	D,E,F,G
Link to sustainability		Strong	Weak	Weak	Strong	Strong
Design		Rigorous	Rigorous	Weak	Rigorous	Rigorous
Fit with implementation:	**Timeline**	Good	Tight	Weak	Weak	Good
	Approach	Good	Good	Unclear	Weak	Moderate

****Key Gaps identified in "DFID Evidence Paper: WASH" (2013):**

A. Relative effectiveness and efficacy of sanitation interventions on health
B. Environmental transmission of excreta-related infections and the role of different pathways
C. The effect of enteric exposures on multiple health and developmental outcomes
D. Non-health impacts associated with poor WASH, especially for the long-term
E. How vulnerable populations are affected and how they can be reached
F. Behaviour change for hygiene and sanitation, especially at scale
G. Costs and cost-effectiveness across social and physical settings

(1) The role of water in achieving sustained sanitation outcomes. These water points were planned as an activity in the outcome phase. The provision of water points in achieving sustainable outcomes has not been rigorously studied and this area can be addressed through a rigorous design while at the same time fitting the implementer's timeline, and not altering their current approach disturbingly. Initially this was viewed as the strongest option and a RCT design developed around this option. However the developed design was rejected by DFID and the suppliers in favour of placing a larger focus on general outcome phase activities that have a stronger focus on sanitation and hygiene behaviour.

(2) Examining the role of demonstration latrines in inducing uptake and moving up the sanitation ladder. Demonstration toilets are provided by WaterAid during the output phase. Although this had already started at the time of RCT discussions, there would have been enough areas were the provision was still outstanding. As is reflected in the secondary research questions in the current design, this question was a subject of important interest to the suppliers (Plan/WaterAid). The design could be rigorous and the fit with implementation was straightforward. However, the concern was that the research question is tangential as the focus is on uptake rather than sustainability. Moreover, the baseline survey would have to be conducted around April/May, which made it infeasible given the proposal timeline.

(3) Research on **sanitation marketing was generally felt as extremely relevant**. However, these activities were part of the output phase and hence already ongoing. More importantly, the intervention activities were implemented at a high geographical level (district), which makes the identification of any type of convincing comparison group a close to impossible task. Furthermore, there is currently an RCT on this subject currently in progress in Nigeria, bringing out the difficulties in evaluating a sanitation marketing intervention and possibly reducing the impact and relevance of pursuing this question.

(4) The main concern relating to examining t**he role of Community Resource Persons (CRPs) in achieving sustainable sanitation** was that it was felt impossible to think of a design which would not lead to difficulties for the implementing agencies in the field and hence this research are was

dropped early on. Suppliers and their partners were (understandably) quick to reject any research design that may result in them missing their output targets and consequently not get paid.

(5) Initially, focusing on t**he role of follow-up activities in achieving sustainable sanitation** was rejected by the supplier's partners over concerns that the design of this option would adversely affect their outcomes and consequently introduce a risk of non-payment through lower achievement of targets[102]. At a much later date it transpired that the suppliers did not share these concerns, furthermore DFID ensured that this would be taken into account when payment decisions were made, and consequently this option was re-visited and considered superior to the 'water point'.

WRP theory of change

[102] It should be noted that this decision was made during a scoping visit to Pakistan in Jan 2015. At that point in time the outcome indicators for the programme were still poorly defined.

Figure 14 below presents a modified version of the WASH Results Programme's TOC. This was developed under the programme evalution following a critical assessment of the original TOC, detailed review of the project design documentation and discussion with DFID. We have included elements that were not explicitly articulated in the original TOC, but were set out in in other DFID documentation (in particular the Business Case). The revised version of the TOC was discussed together with DFID in February 2015 and revised accordingly.
Table 48 below sets out all the explicit assumptions that underpin the theory of change. It is not an exhaustive list of all the assumptions, which are *implied* by the TOC (but not stated in any documentation). It includes assumptions stated in the Business Case (the item number is indicated) as well as DFID's evidence papers. The assumptions relating to PBR are not included in the table, as these are discussed separately in Table 48.

The TOC forms the basis for the WASH Results Programme's evaluation, with the evaluation questions developed to test the key assumptions that underpin it (explicitly and implicitly). It is important to note that the programme-wide TOC remains generic in nature, without a detailed explanation of how specific service providers' interventions will achieve the programme outcomes. This is a logical consequence of the PBR modality of the programme that allows the service provider flexibility in how best to achieve outputs and outcomes given the implementation context.

Table 48: Assumptions stated to be underlying DFID's TOC

Inputs to Processes	Processes to Outputs	Outputs to Outcomes	Outcomes to Impact
Sufficient engagement, level of expertise, continuity and speed of response by DFID advisor (#53) MVE contract enables "prompt corrective action" (#124) Sufficient user demand for services (and capacity of the community to manage improved services) (evidence papers).	Any construction and capacity development activities are appropriate and of high quality Vulnerable groups are included in Programme (#23) Households make labour and capital available (#37) Private sector has capacity to construct and rehabilitate water points (#37) Communities are able to operate and maintain the improved water supply (#37) Local Government Authorities have the capacity to manage inputs and deliver outputs (#37) Hydrogeological, hydrological and water resources management assessments have been undertaken (#68) Climate risk assessments undertaken (#78, 79, 80). Sustainability assessments undertaken (#82) Effective targeting of poor/vulnerable groups (#117-121)	Households change their behaviours as a result of sanitation and hygiene campaigns (#37). Life-cycle costs and responsibilities for their payment have been included (#15). Community participation and capacity (#15, #16). No impact of time-bound PBR targets on sustainability (#117-121)	A critical mass of households change their behaviours, resulting in health and non-health benefits at community level (evidence papers). No unsustainable abstraction of water or pollution from sanitation (#68). No use of contaminated water sources (#68). No contamination of water after collection from a 'safe' source (evidence papers).

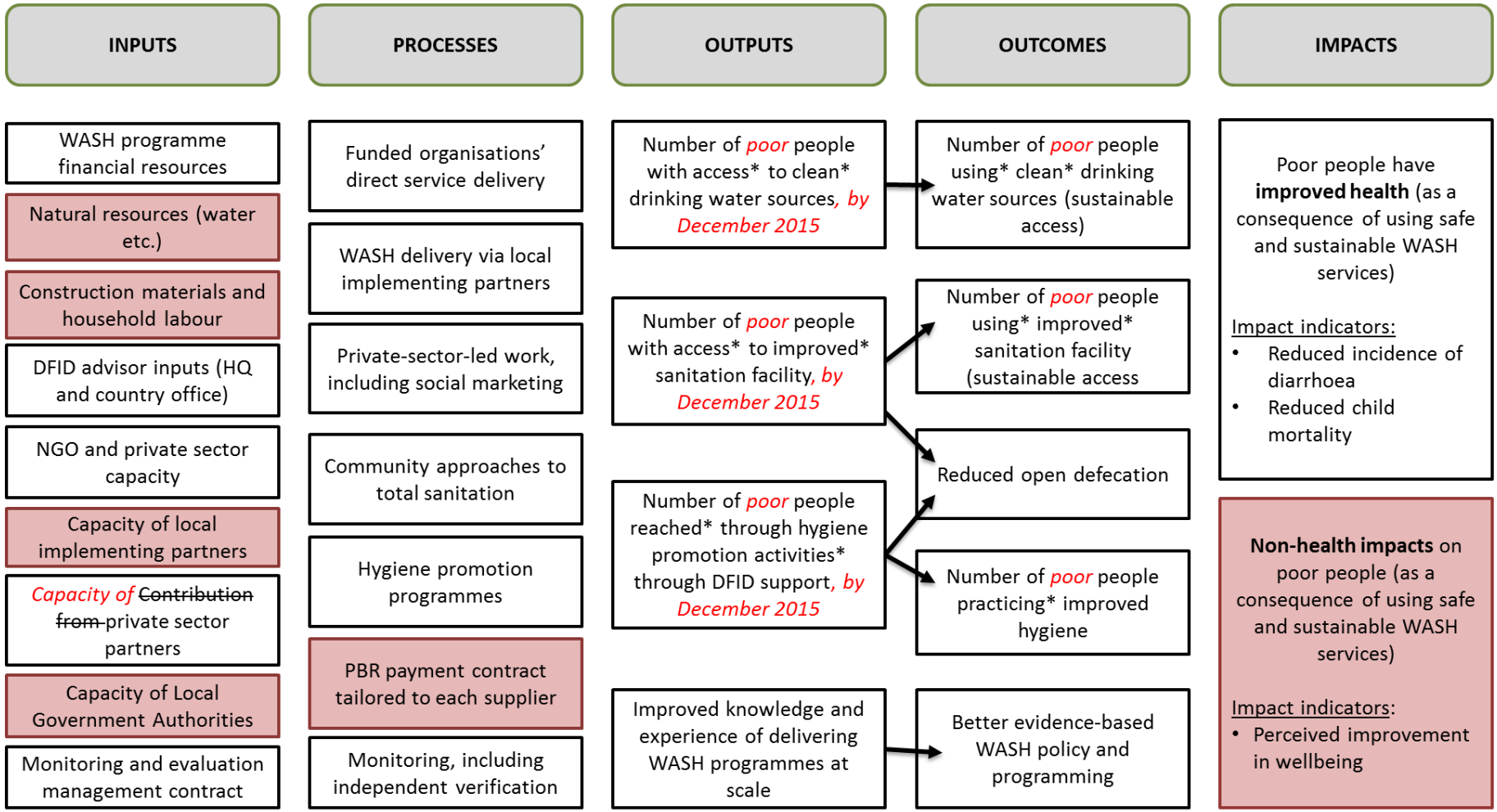
Figure 14: The WASH Results Programme Theory of Change

Boxes shaded in red represent items which were not articulated in the original TOC. Red text in italics represent wording used in DFID documentation but also not articulated in the original TOC.

This impact assessment is being carried out by e-Pact. Britta Augsburg (britta_a@ifs.org.uk) from the Institute of Fiscal Studies is the workstream leader for the randomised control trial. Zach White (zach.white@opml.oc.uk) from Oxford Policy Management is the project manager of the evaluation. Julia Larkin is the evaluation Team Leader.

The contact point for the client is Stephen Lindley-Jones (stephen-jones@dfid.gov.uk). The client reference number for the project is PO6507.

e-Pact
6 St Aldates Courtyard
38 St Aldates
Oxford OX1 1BN
United Kingdom
Registered in England: 3122495
Tel +44 (0) 1865 207300
Fax +44 (0) 1865 207301
Email admin@opml.co.uk
Website www.opml.co.uk



(*) These items were defined differently by each supplier consortia, as articulated in their M&E framework