35th WEDC International Conference, Loughborough, UK, 2011

THE FUTURE OF WATER, SANITATION AND HYGIENE:
INNOVATION, ADAPTATION AND ENGAGEMENT IN A CHANGING WORLD

Methodological lessons and findings from an impact evaluation of a WASH project in Indonesia

C. Sijbesma, B. Sikoki, W. Suriastini & M. Ponsonby, Netherlands

REFEREED PAPER 1096

Internationally, there is growing emphasis on evaluations that aim not only at output and performance data but also show the project's impacts. This paper reports on the experiences, findings and lessons from the impact evaluation of the Second Water and Sanitation for Low Income Communities Project (WSLIC-2) in the Ministry of Health of Indonesia. During project formulation an impact study was planned using the strongest model of a double difference evaluation (differences between before and after situations and practices in project and matched control villages). In this paper, the team that designed and implemented the study reports on the problems and how they were addressed, followed by a brief summary of the findings. It concludes that the classical double difference study is becoming ever harder to practice in development research and suggests ways to address the challenge.

The WSLIC-2 project

From 2000 to 2010, the Government of Indonesia carried out the Second Water and Sanitation for Low Income Communities Project (WSLIC-2). The project was implemented at field level by the project districts and villages with technical support and facilitation from district based teams of Indonesian consultants and facilitators (technical, community development and health disciplines) and at national level by a predominantly Indonesian team established within the Ministry of Health.The International Development Association (IDA) provided a loan for the project supported by a grant from the Australian Agency for International Development (AusAID) for technical assistance and capacity building. The project was carried out in 37 districts in eight provinces: East Java, West Nusa Tenggara (NTT), West Sumatra, South Sumatra, Bangka-Belitung Islands, West Sulawesi, South Sulawesi, and West Java. The province of Bangka-Belitung was not included in the impact studies, because this small island on the east coast of Sumatra had only 2% of the project villages. The targeted coverage under the project budget was 2000 villages after an implementation period of five years. The actual coverage was 2,294 villages by June 2010.

The project goal was to improve the water, sanitation and hygiene conditions and practices and the health status, productivity and quality of life of poor communities in under-served rural villages in the selected project provinces. This was to be done by enabling the selected villages to improve their domestic water supply services, sanitation and hygiene in a sustainable manner and in such a way that safe water and individual household toilets near the homes would become accessible for all. The approach empowered village women and men to identify their problems and plan, implement and sustain village water services and sanitation and hygiene programs solutions with the help of participatory methods. Local schools were assisted to install water supplies, toilets and handwashing facilities and promote good hygiene habits in school. At the end of the project period, WSLIC-2 had attained the following outputs:

- In total, 2,294 villages provided minimally 80% of their population with improved water services;
- All were trained to plan and implement their own sanitation program aiming at ending open defecation and achieving access to and use of safe and sanitary household toilets;

- The project had built the institutional capacities of gender-balanced Village Water, Sanitation and Hygiene (WASH) committees to manage the water supplies, continue village sanitation and hygiene promotion and serve equitably the interests of all households, including the poor ones; and
- Some 6,164 village schools obtained varying combinations of a school water supply, toilets and handwashing facilities and hygiene promotion activities.

The impact evaluation

The growing focus on the effectiveness of aid, has led to heightened interest in evaluating impact (Jones, 2009). International development banks such as the Asian Development Bank (2006) and the World Bank (n.d.) and bilateral agencies, such as IOB, the Dutch independent agency for evaluating bilateral cooperation, see quasi-experimental design as the most appropriate scientific model to attribute impacts to project interventions. Others, such as the American Evaluation Association and the European Evaluation Society (EES) contest that the most rigorous way to assess impacts is through randomised controlled trials (RCTs) and support multi-method approaches (Barahona, 2010).

The project design included an impact study in the project formulation to assess the impacts on health, especially for children under five and school children, productivity (in terms of cost of water) and quality of life, especially of mothers and children in poor households. A baseline study was carried out in 2003 in a sample of 60 intended project villages and 60 matched control villages. This study was to be replicated at the mid term[i] and at the end of the project intervention. The argument for including an impact study was that the project's monitoring system could measure progress in outputs and immediate outcomes, such as increased access to improved water supply and sanitation, including for poor households. However, it could not evaluate changes in conditions, practices, knowledge in the households, nor any impacts on socio-economic conditions and health. Monitoring could also not attribute such impacts to the project rather than to other, external factors, such as economic growth or the more general national programmes for the promotion of better hygiene and health. The model chosen for the impact study was the afore-mentioned quasi-experimental design of a before and after study in a random sample of project villages and matched control villages. The before and after survey for households and schools was complemented by several other studies as depicted in Figure 1.

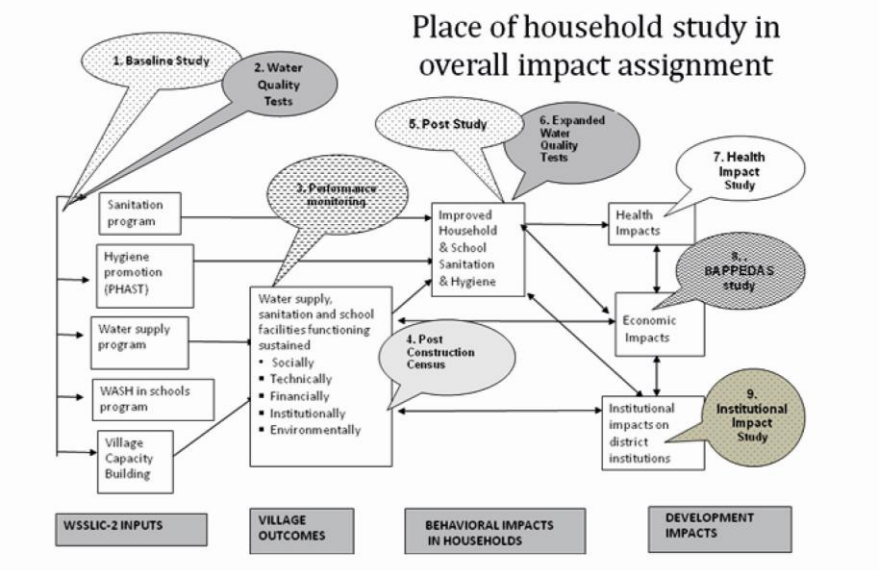

Figure 1. The design of the impact assignment

Source: Sijbesma (2010)

The impact study itself consisted of:

1. A baseline and post study in households and schools;
2. A study of the chemical and bacteriological water quality of the sources used for the improved services and the indicative bacteriological water quality in one household with optimal and another with worst drinking water chain management (storage, treatment, post-treatment storage, drawing);
3. Linkage of household findings to quality of community management to see if better or worse management at village level was associated with better or worse household impacts;
4. An institutional study of selected districts, sub-districts and village governments to understand the roles of higher level support; and
5. A health statistics study to see if good performance in access and use was reflected in local health statistics.

Methodological challenges

For the baseline study the consultants had drawn a random sample of 10 districts from Batch 1 project provinces (four provinces). The sample was proportional to the number of project districts per province and to the type of terrain (hilly and coastal flat), as the latter would influence the type of water technology. In each district they had randomly chosen 15 project villages from the project long list, and 15 matched control villages without project intervention[ii]. The household sample consisted of 310 households per province with mothers and children under the age of 5, and schoolchildren aged 9-11. Assuming a 30% incidence of acute diarrhea among the target population and used a significance criterion of 5% (that is, a risk of 5% of wrongly ascribing found differences to the project) and a statistical power of 80% (that is, an 80% chance that differences will be correctly picked up), the researchers arrived at a total sample size of 2480 households, or 310 project households and 310 control households per province. The actual number of households depended on the size of *the* actual project and control villages, but was about 20 per village.

During the preparation of the post study, a number of methodological issues emerged that seriously challenged the evaluation model and reduced the likelihood of finding valid, reliable and significant differences which could be ascribed to the impact of the project rather than to chance (Survey Meter, 2010; Sijbesma, 2010). They can be summarized as follows:

1. The baseline study was only done in Batch 1 (East Java, West Nusa Tenggara, West Sumatra and South Sumatra). No baseline study was done in Batch 2 (West Sulawesi, South Sulawesi and West Java) and so there were no 'before' data for project and control villages in these provinces.
2. The number of baseline study villages was selected in proportion to the number of project villages in the long list. However, under influence of the on-site village need assessment and feasibility studies a number of control villages became project villages and a smaller number of project villages changed over to the control list. As a result the baseline data base no longer represented the real baseline and control villages.
3. Since the start of WSLIC-2, two other large projects, KDP[iii] and UPP[iv] (both of which were continued under PNPM[v]) had also undertaken some water projects in the area;
4. Household sampling was proportional to the total number of households with children under five and between 9 and 11 in each village, but until a fixed number of households (310) for each province was achieved, however the project was not distributed equally over these provinces;
5. In some cases there was confusion about the identity of project and control villages. In some districts some sub-districts and villages were split up after the baseline survey was implemented. In West Sumatra villages were named differently in the baseline survey and the project records because of different village and hamlet naming systems. There were a number of gaps in the survey instruments, such as water use conditions and practices were not assessed separately for the dry and the wet season. Toilet use and hand washing practices were not assessed for the different household members.
6. No objective criteria were used to score toilet hygiene – faeces and mud could both be scored as 'unclean'. Presence of water/water and soap in places of food handling such as kitchens was not observed.
7. The study used much longer periods of recall for WASH-related diseases than advised by the Ministry of Health, e.g. three months for child diarrhoea. Internationally, long recalls have also been rejected because of their poor reliability.
8. The practices and conditions were not analyzed according to socio-economic levels of the households (poor and non-poor).

To address these issues, the following actions were taken in the design of the post study:

1. The post study was carried out in the same villages of Batch 1. For Batch 2, mothers and school children were asked to recall conditions and practices before the project wherever this was possible. For questions such as incidence of diseases the period before the project was obviously too long ago for a reliable recall, so these were replaced by questions on the perceived change in incidence of WASH-related diseases. In the control villages we could use two widely known local events as reference for the 'before' dates: the tsunami of 2004 and the Yogya earthquake of 2006.
2. The baseline data set for Batch 1 was reorganized according to the actual distribution of project and control villages. This resulted in a sample of 69 project villages and 41 control villages. From Batch 2, 17 project and 17 control villages were sampled. Overall, the sample was 3% of all Batch 1 villages and 4% and 5% of the Batch 2 project villages in West Java and West and South Sulawesi respectively.
3. With a lot of effort, the consolidated lists of water project villages were obtained from KDP/PNPM and UPP-WASH. No overlap was found between these projects and the control villages for WSLIC-2, but an extra interview with a key informant was included to check on other WASH projects in control villages.
4. The same type and number of households were sampled in proportion to their presence in the post study villages. Under the new distribution, this resulted in new sample sizes as depicted in Table 1 below.

Table 1. Composition of the study samples

	Baseline study		Post study	
	# Mothers	**# Children 9-11**	**# Mothers**	**# Children 9-11**
Project households	1572*	510*	2009, of which 1975 with children below 5	819 (402 boys, 417 girls)
Control households	737*	225*	1921, of which 1875 with children below 5	744 (396 boys, 348 girls)
TOTAL	2309	735	3850	1563

Source: SurveyMETER, 2010

5. The errors in the identification of baseline villages were corrected and each village was identified in the post study by its unique Badan Pusat Statistik (BPS) code which the central or national level Statistics Office gives to each village in Indonesia.
6. Questions in the questionnaire were improved and recall questions added where the baseline data were missing or lacked rigor. For some adequate baseline data we still asked recall data and compared the outcomes. We found that recall of before situations gave similar results to the baseline data. For livelihood impacts we used nine scenario scales that ranged from a typical bad situation via slightly better conditions and an acceptable average, or 'benchmark', to very good and ideal situations. The mothers scored their current and past situation on these scales.
7. Each interview was followed by structured observations using objective indicators such as [absence of] faecal stains in toilets, numbers of flies, and presence of water and soap for hand washing, and findings were correlated between reported handwashing practices, knowledge of critical handwashing occasions through multiple choice questions and observed presence of water/water and soap.
8. Disease incidence recall was reduced to two weeks for infant diarrhoea and one month for other WASH-related diseases in line with the National Socioeconomic Survey (SUSENAS).
9. All data were analysed for poor and non-poor. In the project villages, WSLIC-2 had used participatory methods to agree on the local poverty indicators and monitor access for the poor. Because this data was not available in the control villages, we ran the data base three times to see which general indicator gave a comparable result. The Median Per Capita Income proved to be the most similar divider between poor and less poor in the control households. In addition, special focus group discussions were held with poor women and men to assess the degree and impacts of their participation and their satisfaction levels.

Implications for the quality of the data

While some problems could thus be solved, two choices from the baseline study could not be adjusted. The first remaining challenge was the imbalance in the composition of the baseline sample, with a much larger number of project villages than control villages, and relatively worse conditions in the former than in the

latter. The relative under-representation of villages in East Java means a relatively larger representativeness of poorer villages from Sumatra and NTT. The overall findings for Batch 1 may thus be somewhat worse than if the sampling had been more proportional. However, since WSLIC-2 was meant especially for low-income villages, the deviation can be considered acceptable.

The imbalanced sample of a larger number of project households in a wider set of project villages may have caused a larger variation in conditions and practices in the project sample than in the smaller control sample. A theoretically lower homogeneity may have reduced the chance of significance in differences in comparison with the findings for a more homogeneous sample. In the latter case (that is, with better matched samples) we might have found a greater number and/or strength of significant differences between the net changes in WSLIC and non-WSLIC households.

That project and control villages were not fully comparable is consequence of the reality of a development program. While in a pure research project it is possible to allocate an intervention to some randomly chosen participants and not to equally randomly chosen others, in a water project that aims at serving the worst one cannot draw lots which village is to get the project and which one will be bypassed.

However, the double-difference design of the impact study does circumvent this problem. The reason is that the impact study does not compare worse situations in project villages with better situations in control villages. Instead, it measures the difference-in-difference, or net impact between the two groups, as set out in Table 2.

Table 2. Method of analysis: difference in difference

Sample	Before	After	Difference
Project	P1	P2	p=P2-P1
Control	C1	C2	c=C2-C1
Difference in difference or net impact			p-c

Source: SurveyMETER, 2010

The second challenge was the decision made during the project design to target the impact assessment only on households with mothers of children under five and school age children aged 9-11. Because these baseline households did no longer have the same composition several years later, new households with the same composition were indentified, sampled and interviewed. Recall questions to these households did thus not always relate to situations in which the household composition corresponded with that of the targeted types of households. The possible difference in household composition for the recall questions may have meant that the reported conditions may have been different for households without and with children under five or schoolchildren aged 9-11. However, this is very unlikely, since at that time there were no specific sanitation or hygiene programs for these two groups (young mothers and schoolchildren).

From the above it may be clear that the challenges encountered reduced considerably the chances to find statistically significant differences between project and control households. In consequence the implementers of the post study remained in suspense until the very last moment whether the large investment of time and money would show statistical impacts. The tension was increased when the extra check on other interventions showed that while two other large projects (KDP/PNPM and UPP-WASH) had no overlap with WSLIC-2, *all* control villages had successfully lobbied for interventions from the districts or local NGOs or undertaken WASH improvements themselves. While it is of course very positive that WSLIC-2 stimulated non-selected villages to undertake their own WASH projects, it reduced the chance that impacts from WSLIC-2 could be statistically proven at the end of the project period.

Summary of the study findings

Fortunately enough statistical evidence was found to show that part of the goals of the WSLIC-2, improved water, sanitation and hygiene conditions and practices, and through them a better health status, productivity and quality of life of poor communities in underserved areas could be met. There were also valuable lessons on what could *not* be achieved, and why. The main outcomes were as follows:

On water use:

- A greater net shift occurred in project households to using improved water supply sources in the wet and dry season, although the current average use is just at or below what is internationally seen as a critical mass of safe water sources (75%-80% use), which is one of the conditions for an impact on public health. The project now had an average of 74% in the dry and 77% in the wet season.
- There was a significantly greater shift to house connections. As they were financed by the users themselves, the access was larger for the better-off (57%) than the poor (36%). This was not better in the project villages than in the control villages in spite of the project being pro-poor.
- There was a significant decrease in water collection distance, especially for the poor.
- There was also a significant impact on enough water for all domestic needs of women and men.
- The impact was further significant for time- and energy savings for the poor and the better-off, with a stronger impact for the better-off because of the house connections.

On sanitation:

- A significant project impact was found on access to household sanitation for both the poor and the better off, but only after the project rules for toilet loan were made more pro-poor in 2005 and CLTS was adopted for promotion in 10 districts in 2006 and project-wide in 2007. There was also a significantly greater access of poor households to toilet loans after 2006. However, most toilets (85%) were fully self-financed. Average coverage levels (62% in project sample, 70% in control sample) are still below the standard of 75% to 80% for a 'critical mass' to contribute to health, and still far from access for all.
- The shift was significant for toilets with water seals; which prevent flies' spread faeces-related diseases. However, most had direct pits and mothers said they would either abandon filled-up toilets and face the investment costs for new ones, or practice unsanitary emptying and end-disposal, since no sanitary services are available from the private sector. Only one-third said they would bury the contents.
- No statistical difference was found in observed faecal contamination of toilets, but there was a significant project impact on observed presence of cleaning brushes in toilets, which was equally good for poor and better-off households, and in lower observed numbers of flies in kitchens and toilets of project households.
- We also found a significant net reported shift away from open defecation by project households for all members (adult women and men, girls and boys). The percentage of households with a toilet but not using it were equal to that of households without a toilet, but who shared the toilet of a relative or neighbour or used a public facility. However, average reported use of toilets by the different household members was never above 60%, still too low for a critical mass.

On handwashing:

- Mothers reported a significant shift from washing hands with water only to sometimes with soap. There was also a significant net increase of mothers, girls and boys reporting to now wash hands with soap after defecation and mothers, both poor and better-off, who reportedly did so now before handling food;
- Batch 2 (with stronger sanitation and hygiene inputs) showed a significant net gain for soap in toilets (observed now, reported for earlier). Soap presence was still modest: 34% for toilets and 41% for kitchens.
- The project households had significantly better choices of critical times for washing hands with soap. Knowledge on good practice before food handling and infant feeding was significantly positive correlated with soap observed in kitchens, while knowledge on washing hands with soap after defecation and cleaning infants' bottoms correlated significantly with soap in bathrooms and toilets. However, correct knowledge scores were fairly low. In general, management of the chains of infant faeces and infant feeding were gaps in the mothers' knowledge and reported practice.

On drinking water treatment in the home:

- While boiling remained the reportedly most used method for treating drinking water at home, the project brought a significant increase in other methods (ceramic filters, solar disinfection and chlorination products from local shops). This is important because only 21% of the improved water supply sources tested met the national standards for e-coli. Chemical and physical water standards were better. When good water chain management in the home was practiced, the percentage of samples that met the national e-coli standard increased to 28%. Under poor home management only 19% met the standard. This data was only indicative, since only one good and one poor management household was sampled per village.

On hygiene promotion:
- The project had a significant greater impact on access to hygiene promotion, for poor and better-off mothers and poor fathers.
- Mothers said to be satisfied with the promotion, but the promotion methods had not improved and they would like better adjustment to their needs, more time given to promotion and more interesting methods.

On WASH-in-schools:
- The project had a significant impact on the presence of sanitary school toilets (96%), separate toilets for girls in Batch 1. However only 15% of toilets were separate for girls, while almost one in three girls in age group 9-11 said to have started menstruation.
- There was also a significant impact on handwashing stations. They were observed in 80% of project schools, but only 28% of the schoolchildren interviewed said they always had soap to wash hands.
- There was no significant impact on school water supply. Here control area schools improved to the same degree. Hygiene promotion in schools had also not improved significantly.

On health and quality of life:
- Mothers in the project sample reported a significantly higher reduction in child diarrhoea than mothers in the control group. However, the statistical study did not confirm this, probably due to problems in diagnosis and recording *and* interventions in the control group.
- Scale scoring of post- and pre-situations showed significant impacts on workload reduction, water availability for drinking and cooking, hygiene and domestic productivity, and a reduction in poor sanitation and its related social problems such as privacy and safety for women. There were also significant positive impacts for reduced WASH-related tensions between the sexes over water collection.

On village management and good governance:
- Poor women noted a significant difference on their knowing the local WASH management and contacting them when needed.
- Better-off women scored significantly more often on now having roles in village WASH management and on transparency and accountability: the village WASH leadership informed them better than before the WSLIC-2 project, including on financial matters.

Lessons for future impact studies

The impact study of the WSLIC-2 project shows how field realities can interfere with the quasi-experimental design that a part of the development agencies prefer for impact studies. The study showed that both the growing number of WASH projects and the decentralization process that enables villages to undertake their own projects made it harder to find control villages that remained without intervention over a longer period of time. Easier communication through better transport, mobile phones and mass media also made it hard to isolate control villages from project developments, as indicated by the finding that 1% of toilet loans were taken by households in the control sample!

It is however unrealistic to demand from control villages to forego opportunities for improvement so that foreign tax payers can get statistical evidence of the impacts of allocations to development cooperation. Good evidence is nevertheless important to show to both national governments and donors that the project has been worthwhile. *If* a double difference statistical impact study is chosen, a first lesson is to carry out the baseline and post study per batch over a shorter period, so that control villages for tan earlier batch have the possibility to become project villages in the next batch. This would fit especially in ongoing district-wide programs. A related lesson is that baseline studies should be done after the actual project allocation and not on the basis of long-lists.

When using questionnaires, scientific rigor is crucial. Especially questions into promoted good practice and knowledge, such as on toilet use and handwashing produce inflated outcomes. This study used multiple choice questions cross-checked with observed conditions for the reported practice and triangulated the answers of mothers and older children to detect and correct inflated figures.

In the present study all data entry was done in the villages. Because the view was held that the villagers have a right to the outcomes that their inputs produced, the study teams gave the village leadership direct feedback on the main local outcomes, e.g. the quality of the water sources and the key household management practice. A participatory assessment tool with all good practices in the chain noted ‘above the

line' and the bad ones below proved a useful tool. A still more participatory baseline study could enhance the learning function of the data and provide local cadres and village WASH leaders with a basis and commitment for planning and monitoring improvements.

Acknowledgements

The authors would like to extend thanks to the Ministry of Health, Directorate General for Disease Control and Environmental Health for their permission to share the findings and lessons from the impact study, and to the Australian Government and IDA for financing the study.

References

ADB (2006) *Impact evaluation: Methodological and operational issues*. Asian Development Bank: Manila.

Barahona, C. (2010) *Randomised Control Trials for the Impact Evaluation of Development Initiatives: A Statistician's Point of View*. ILAC Working Paper 13, Rome, Italy: Institutional Learning and Change Initiative.

Jones, H. (2009) The *'gold standard' is not a silver bullet for evaluation.ODI Opinion* Paper no. 127. Overseas Development Institute: London.

Sijbesma, C.A. (2010) *Impact Evaluation of WSLIC-2: End of Assignment Report*. IRC International Water and Sanitation Centre: The Hague.

SurveyMETER (2010) *WSLIC-2: Post-implementation household survey and impact study*, Jakarta, 7 October 2010

World Bank (n.d.) *The Development IMpact Evaluation (DIME) Initiative*. Project Document. World Bank: Washington, D.C

Note/s

[1] Villages included in WSLIC-2 were selected from a long list of applicants on the following criteria: (a) low levels of access to water and sanitation; (b) high incidence of diarrhoea; and (c) high poverty levels. Therefore it is likely that project villages had worse WS&S situations and higher incidence of diarrhea.
[2] The mid-term study was not implemented for various reasons including delay of the baseline study.
[3] Kecamatan (subdistrict) Development Program
[4] Urban Poverty Project
[5] National Community Empowerment Program (Program Nasional Pemberdayaan Masyrakat).

Contact details

Christine Sijbesma
IRC International Water Supply and Sanitation Centre
POB 82327
2508 EH The Hague, The Netherlands
Tel: +31 70 304 4011
Email: sijbesma@irc.nl
www.irc.nl

Bondan Sikoki
Jl.Babadan Gang Kartika VII
No.99,Banguntapan,Bantul
Yogyakarta 55198, Indonesia
Tel: +62 811253956
Email: bsikoki@gmail.com