Fostering the use of quasi-experimental designs for evaluating public health interventions: Insights from an mHealth project in Malawi

By Jean Christophe Fotso[1], Amanda Lea Robinson[2], A. Camielle Noordam[3], Jessica Crawford[4]

1. Introduction

As governments and international organizations strive to meet the health related Millennium Development Goals (MDGs), the rapidly expanding use of mobile phones to improve health and health systems (mHealth) presents an unprecedented opportunity to increase access to health care and improve health outcomes (mHealth Alliance, 2012a; Al-Shorbaji & Geissbuhler, 2012; WHO, 2011). Previous research has demonstrated that mobile phones can be a successful tool in behavior change for achieving positive health outcomes such as smoking cessation, weight loss, diet and physical activity, treatment adherence, and disease management (Free et al, 2013a; 2013b) and for improving health worker and health system performance (Gurman et al., 2012).

The declining cost of mobile phones, growth in subscriptions, and rapid advances in technology have triggered an explosion of mHealth pilot projects and programs in low- and middle-income countries (LMIC) since the mid-2000s (Qiang et al., 2011). A recent report shows that between 2005 and 2010 mobile telephone subscriptions grew almost three-fold in sub-Saharan Africa and more than six-fold in South Asia (World Bank & ITU, 2012). Not surprising, the second World Health Organization (WHO) global survey on mobile technologies reported that 83% of the 112 participating countries had at least one mHealth initiative, and of these, three quarters reported four or more mHealth initiatives (WHO, 2011).

For maternal, newborn and child health (MNCH), mHealth can support and strengthen existing efforts along the continuum of care. Existing mHealth interventions for MNCH in LMICs tend to fall under three broad objectives: Stimulating demand for healthy behaviors among pregnant women and mothers; strengthening human resource capacity for health care delivery; and transforming health system capacity (mHealth Alliance, 2012b; Lemaire, 2011). Demand generation activities in particular are increasingly prioritizing information services that increase women's awareness of health issues and encourage women to use available health resources; complementary patient-support services addressing the management of health issues; and patient communication services that directly connect women to peer networks or expert resources in the community such as skilled birth attendants (Noordam et al., 2011; Gurman et al., 2012; Tamrat & Kachnowski, 2011). Phones are also used to improve point-of-care decision

[1]Concern Worldwide US, New York, USA
[2]Department of Political Science, The Ohio State University, USA
[3]Maastricht University, Maastricht, The Netherlands
[4]VillageReach, Balaka, Malawi

making by community- and facility-based health workers (Mitchell et al., 2013; Tamrat & Kachnowski, 2012), and in connecting patients to health facilities and strengthening referral systems (Crawford et al., 2014). Moreover, there is an increasing interest in using mobile phones to empower women and improve social dynamics that impact access to quality health services for women (mHealth Alliance, 2012b).

How robust is the mHealth evidence base?
While mHealth interventions are gaining traction in almost all parts of the world, the evidence base to support this trend is relatively nascent (Tomlinson et al., 2013; Leon et al., 2012; mHealth Alliance, 2012b). Evidence of mHealth effectiveness is primarily limited to studies in developed countries or for chronic disease prevention and management and life-style changes and robust analyses of mHealth for MNCH are still lacking (mHealth Alliance, 2013). A recent review on the effectiveness of mHealth to improve service delivery noted that out the 42 controlled trials identified, none was of high quality or implemented in LMICs. The authors concluded that robust evaluations are needed to ascertain the effectiveness of the interventions and provide evidence for scale up (Free et al., 2013a). The same authors conducted a systematic review of the effectiveness of mHealth to improve health behaviors or disease management and found that of the 75 trials identified, only 3 were conducted in LMICs (Free et al., 2013b). A review of publications on mHealth behavior change communications found that the key findings from almost all 16 qualifying articles were from less solid bases like summative evaluations, formative research or process evaluations (Gurman et al., 2012). A recent review of maternal health focused mHealth project evaluations revealed a scarcity of quantitative assessments with only 4 out of 34 articles and reports identified including a quantitative design (Tamrat & Kachnowski, 2011). More generally, many of the studies published on mHealth suffer weakness ranging from lack of intervention and control groups, absence of baseline, and small sample sizes to being primarily descriptive in nature (mHealth Alliance, 2013; Tamrat & Kachnowski, 2011; Noordam et al., 2011). Overall, it appears that a major attribute of mHealth projects is a general failure to frame and answer critical evaluation questions with the needed rigor.

In addition, concerns are raised on the limited number of mHealth projects that are brought to scale, as a result of our limited understanding of the likely uptake, best strategies for engagement, recurring challenges and the linkage of these interventions to broader existing structures (Noordam et al., 2014; Lemaire, 2011).

Why is impact evaluation needed?
It is well known that public health programs might appear promising, yet fail to generate the expected impacts or benefits. With a proliferation of mHealth pilot projects in LMICs (Qiang et al., 2011), there is an urgent need for substantial evidence documenting the benefits of these interventions. Robust impact evaluations that can attribute changes in outcomes to a particular mHealth project and that fill gaps in understanding what works and what doesn't work are

important to promote accountability in the allocation of resources and to maximize the potential of mHealth to improve MNCH outcomes (Victora et al., 2011).

Against this backdrop, the aim of this paper is three-fold. First, we describe the design and key quantitative findings of an evaluation of a MNCH mHealth project in Malawi. Second, models for causation in quasi-experimental designs are discussed, comparing the intention-to-treat (ITT) principle and the effect of treatment on the treated (TOT). Third, we discuss potential sources of self-selection bias in quasi-experimental evaluations in relation to the level of randomization and the assumption of time invariant unobserved heterogeneity.

2. Data & Methods

2.1. The MNCH mHealth Project in Malawi

The MNCH mHealth project called Chipatala cha pa foni (CCPF) – meaning health center by phone – aimed to increase knowledge and use of home-based and facility-based MNCH services. To achieve these objectives, the intervention offered a toll-free case management hotline and an automated and personalized mobile messaging service. Village volunteers, trained and provided with phones, mobilized users in the intervention sites and ensured access to services to those without phones (IKI, 2011). The project was implemented between July 2011 and June 2013 in Balaka District in the southern region of Malawi, an area with some of the poorest MNCH indicators in the country (Crawford, 2011).

The toll-free case management hotline provided protocol-based health information, advice and referrals, allowing mothers/caregivers of under-five children who may not be able to access an in-person consultation at a health center to connect to trained personnel based at the District Hospital. While women using a personal phone could dial the hotline number directly, those without access to phone had to go to the nearest community volunteer's house and use the phone to connect privately to the hotline center (Innovations, 2012). On the other hand, the automated and personalized mobile messaging service for pregnant women, guardians of children under one year of age and women of child-bearing age offered the opportunity for registered clients to receive weekly messages on appropriate care seeking and health practices. The messages were delivered to a user's phone or could be retrieved by voicemail through calling the toll-free number and responding to voice prompts provided by the system. Registration of women for this service was usually done during the first hotline call or during antenatal (ANC) visits.

The project was implemented in the catchment areas of four health centers in the district namely, Phimbi, Chiyendausiku, Kalembo and Mbera, the only health centers that met the selection criteria set the program and research teams (presence of electricity and at least two maternity nurses, and cell phone coverage). The district hospital and Christian Health Association of Malawi (CHAM) facilities were not considered for the intervention, as their

clients are likely to have different socioeconomic and health profiles (e.g. services are not free at CHAM) (IKI, 2013).

2.2. Evaluation design and source of data

A two-arm quasi-experimental, pre-post design was used to quantify the impact of the intervention on knowledge and use of home-based and facility-based care for mothers and children. The neighboring Ntcheu District was selected to serve as control, as it is more likely to exhibit similar MNCH outcomes as Balaka district. With the same eligibility criteria as for the intervention area, the two health centers of Bwanje and Kasinje were selected (IKI, 2013). Clearly, randomization of subjects or groups was neither practical nor feasible. In this context, a quasi-experimental design - use of intervention and comparison groups, but nonrandom assignment to the groups – is shown to approximate the randomized experiment (Duflo et al. 2008), with differences between the intervention and control sites accounted for in the analyses (Heckman, 2005; Meyer, 1995). The core of the assessment were baseline and endline cross-sectional population-based surveys of mothers aged 15-49 who had children under 5 years of age, as well as pregnant women and caregivers of children under 5 years of age, conducted in June-July 2011 and April-May 2013, respectively. Qualitative methods were also used, but they are beyond the scope of this paper.

Villages were the primary sampling units. At baseline, GIS information and maps from the health centers were used to define the catchment areas of the health facilities and to create a comprehensive list of villages with information on each village's total population, estimated number of women of child-bearing age and estimated total of children less than five years of age. Using a systematic sampling approach, villages were randomly selected to be included in the study such that villages in each health center catchment area had the same probability of being selected into the sample regardless of their population size. All households with eligible women (aged 15-49 or caretaker of a child under 5 years) were selected and questionnaires administered to all eligible respondents (IKI, 2011). The number of survey respondents is shown in Table 1. A total of 6,453 households (2,810 at baseline and 3,643 at endline) were successfully visited, yielding a total of 6,693 women aged 15-49 and 6,846 children under the age of five years.

[Table 1 about here]

2.3. Outcome and control variables

This paper analyzes four aggregate outcome measures:

- Use of home-based practices for maternal health, derived from the following indicators: 1) Used a bed net during pregnancy; and 2) Breastfed child within one hour of birth - Both questions were asked of women who had given live birth in the last 18 months, about their last birth.
- Use of home-based practices for child health, derived from 1) Child breastfed exclusively until six months of age; 2) Under-five child slept under a bed net the previous night; and 3) Under-

five child sick with diarrhea in the previous two weeks received oral rehydration salt - The last questions was asked for children who had experienced diarrhea in the past two weeks, while the two others were asked for all children.

- Use of facility-based services for maternal health, aggregated from 1) Received the correct dosage of the tetanus toxoid (TT) vaccine during pregnancy; 2) Received a Vitamin A dose during last pregnancy; 3) Received the recommended 4 antenatal care (ANC) consultations; 4) Started ANC in first trimester; 5) Gave birth under the supervision of a skilled birth attendant; and 6) Received one postnatal care (PNC) check-up within two days of birth - All questions were asked of women who had given live birth in the last 18 months, about their last birth.

- Use of facility-based services for child health, derived from 1) Child was fully immunized by first birthday; 2) Child with symptoms of acute respiratory infections (ARI) in the previous two weeks sought care from health facility; and 3) Child with fever in the previous two weeks sought care from health facility – the three questions were asked of children between 12 and 24 months of age, children with symptoms of ARI in the past two weeks, and children with fever in the past two weeks, respectively.

The indicators were aggregated into aggregate measures using the following protocol (Kling et al., 2007): First, recode each indicator in a way that a higher value always indicates a better outcome (i.e., behavior encouraged by the intervention). Second, impute missing values on individual indicators at the treatment assignment group mean. Third, compute the average across the set of indicators that apply to a given respondent. For example, if a woman has not yet given birth, facility-based indicators #5 (Gave birth under the supervision of a skilled birth attendant) and # 6 (Received one PNC check-up within two days of birth) will be excluded from the average.

Variables at the community level (mean distance to the health center); at the household level (household wealth, number of under-five children, and ethnicity and religion of the household head); at the woman level (access to phone, education, marital status and age); and at the child level (age and sex) are controlled for in the multivariate models. The household wealth variable is constructed on the overall sample from household characteristics (presence of electricity and type of drinking water, toilets, wall, roof and floor) and household possessions of durable goods (e.g. bicycle, TV, fridge, watch), using principal component analysis (Jolliffe, 2002). The variable was further recoded as a dichotomous variable using the median value as the cut-off point (low 50%, high 50%). The specification of these variables is detailed in Table 2.

2.4. Methods of analysis

We assess the impact of the intervention using the difference-in-difference (DID) - the most widely used method for impact evaluation in the context of quasi-experimental designs (Heckman, 2005; Meyer, 1995) - in three steps of increasing complexity. **First**, we estimate the simple DID for a given outcome Y as follows:

$$DID(Y) = \left(\overline{Y}_{IE} - \overline{Y}_{CE}\right) - \left(\overline{Y}_{IB} - \overline{Y}_{CB}\right) \qquad (1)$$

where $\overline{Y}_{IE}$ and $\overline{Y}_{CE}$ represent the average outcome at endline in the intervention site and control area, respectively, and $\overline{Y}_{IB}$ and $\overline{Y}_{CB}$ the average outcome at base line in the intervention site and control area, respectively. The DID estimator allows for unobserved heterogeneity between the intervention and the control sites, but assumes this unobserved heterogeneity is time invariant; so the potential bias cancels out through differencing (Bertrand et al., 2004; Rubin, 1974). Since CCPF was offered but not compulsory, this estimate, referred to as average treatment effect (ATE) is to be interpreted as intention-to-treat (ITT) effect – as it entails comparing the intervention and control areas without regard for the actual use of CCPF.

The DID estimate in (1) can also be calculated within a regression framework as follows:

$$Y_{ivt} = \beta_0 + \beta_1 T_v + \beta_2 P_t + \beta_3 (T * P)_{vt} + \varepsilon_{ivt} \qquad (2)$$

where Y_{ivt} is the outcome measure for woman/child i, in village v, at time t. T_v is a dummy variable taking the value 1 for individuals in treatment areas and 0 for individuals in control areas, P_t is a dummy variable taking the value 0 for the baseline data and 1 for the endline data, and ε_{ivt} is the idiosyncratic error, clustered by village. The difference-in-difference estimator of interest is the coefficient β_3 of the interaction between T_v and P_t.

Next, we estimate the adjusted effect using regression-based DID and controlling for possible confounders according to the following formulae:

$$Y_{ivt} = \beta_0 + \beta_1 T_v + \beta_2 P_t + \beta_3 (T * P)_{vt} + \mathbf{W}'_{ivt}\Phi + \mathbf{X}'_v\Pi + \varepsilon_{ivt} \qquad (3)$$

where $\mathbf{W}_{ivt}$ is a vector of controls at the household and individual levels, and $\mathbf{X}_v$ is a vector of village-level controls.

Finally, to assess the impact of the intervention on the women who actually used the services, we estimate the treatment effect on the treated (TOT) which, in contrast to the ITT (or ATE), compares the individuals who used the services in both the intervention and the control areas, to those who did not. The method uses instrumental variable analyses to construct a proper counterfactual – the women who would have used the services in control communities had they been offered (Angrist et al., 1996).

3. Evaluation Results
3.1. Sample characteristics
A combined (baseline and endline) total of 4,230 women were interviewed in the intervention area, compared to 2,463 in the control site. The corresponding figures for under-fives were

4,406 and 2,440, respectively. Tables 2A and 2B compare the distribution of women and children in the intervention and control areas. The distribution of women by education and age is similar in both areas, with about three women in four having a primary level of education, and approximately 42% in their 20s. Women's marital status, the mean distance to the health center, and the number of under-five children in the household, are almost similarly distributed across the intervention and the control sites. Table 2A also shows that the proportion of women from wealthier households is almost comparable across the two areas - household wealth was constructed on the pooled sample, and could allow the comparison of the socioeconomic status across the two sites. By contrast, women's access to phone was higher in intervention area (35%) than in control communities (23%). The variables displaying the largest distributional differences in the women sample are ethnicity, and to a lesser degree, religion. Finally, the distribution of children by sex and age in Table 2B is similar across the intervention and the control groups.

[Table 2A & 2B about here]

At endline, awareness of the services offered was high, with nearly 77% of respondents in invention communities reporting that they had heard about the hotline services, as presented in Table 3. In contrast, only a third of them were aware of the mobile messaging services, all of whom had also heard of the hotline. Use of services was relatively low among women who were aware of the intervention, at around 24% for the hotline services and 23% for the messaging services. Based on the total sample of women, these proportions stand at 18% and 7.5%, respectively.

[Table 3 about here]

3.2. Effects of the intervention on home-based and facility-based MNCH care

Table 4 presents two estimates of the impact of CCPF the four aggregate maternal and child health outcomes of interest: the intention-to-treat (ITT) effect (which compares the intervention and the control areas regardless of the use of services), and the treatment effect on the treated (TOT) (which compares the individuals who used the services offered by CCPF to those who did not). The sub-population corresponding to each outcome is also described in the table.

Average treatment effect: Unadjusted results

The results of the unadjusted DID of the intention-to-treat (ITT) model are presented in the first column of Table 4. They show a strong, negative average treatment effect of the project on facility-based care for child health (-0.172; p<0.01), which results from a decrease in the intervention area, and an increase in the control site (not shown). Likewise, there is a negative and statistically significant effect on home-based care for children (-0.059; p<0.05), though of lower magnitude, compared to the effect on facility-based care for child health. Unlike for facility-based care, the negative effect is a result of a steeper increase in the control area compared to the intervention site (not shown). The results in Table 4 also show that the intervention did not have any significant impact on maternal health, the small, positive effect

on facility-based care failing the reach statistical significance, and the effect on home-based care appearing negligible.

[Table 4 about here]

Average treatment effect: Adjusted results

The second column of Table 4 presents the DID of the intention-to-treat (ITT) estimates controlling for community, household, women and child-level variables. As can be seen, the inclusion of controls did not result in major changes in the magnitude, direction and statistical significance of the effects. While the size of the effect on home-based care for child health increased slightly (from -0.059 to -0.071), the level of significance did not change. The impact of the three other outcomes remained largely unchanged.

Treatment effects on the treated

Besides the ITT effect and owing to the low uptake of the services (only 18% of women in the intervention communities used CCPF – and the associated dilution of the treatment effect), it is critical to evaluate the impact of the intervention on women who actually used it services. The results in the third column of Table 4 reveal a strikingly different feature of the treatment effects on treated (TOT) compared to intention effects to treat (ITT), in terms of magnitude, direction and statistical significance. We note are a large, positive and statistically significant TOT effect of the intervention on the aggregate home-care for child health (+0.603; p<0.01), in opposition to the negative ITT effect described above. There is a large, positive TOT effect on home-based care for maternal health (+0.479; p<0.01). CCPF also recorded a positive, significant effect on facility-based care for maternal health among women who used the services (+0.239; p<0.05). The negative ITT on facility-based child health care is amplified among women who used CCPF (TOT effect of -0.499; p<0.01).

4. Summary and Discussion

To measure the impact of the mHealth CCPF project on maternal and child health, we used a DID estimator to deal with unobserved differences between control and intervention communities, and an instrumental variable analysis to address the fact that use of the intervention was voluntary, introducing self-selection bias into the study (Angrist et al., 1996; Bertrand et al., 2004).

Our analyses show a large, positive effect of CCPF on the utilization of home-based care for child health among those who used the services offered, in contrast to the modest, negative average treatment effect observed on the same outcome. The project also resulted in large, positive TOT effects on home-based and facility-based care for maternal health on the one hand, and virtually no ITT effect on either of the two maternal health outcomes. Finally, there is a large, negative ITT effect of the intervention on facility-based care for child health, an effect which is amplified among women who used the services offered by CCPF. Another study showed that this negative effect is driven solely by the use of facility-care among children who had fever in the two weeks preceding the survey, and concluded that CPPF contributed to

strengthening the home-to-facility continuum of care, by reducing the unnecessary visits to health facilities for conditions that can be adequately managed through home-based care (Fotso, Bellhouse and Jezman, Draft). These overall positive findings correlate with another study's findings indicating that all types of messaging (SMS or voice) led to high levels of satisfaction, comprehension, and new information learned (Crawford et al., 2014). The merits and limitations of the TOT and ITT approaches are discussed below.

Differences between control and intervention in quasi-experimental designs

The difference in difference component of our design helped us deal with pre-treatment differences between control and intervention communities. Indeed, though the two communities were adjacent to each other, there were notable disparities in a few health outcomes prior to the intervention (IKI, 2012). The key assumption allowing difference in differences to identify a causal effect is that in the absence of the intervention, the unobserved differences between intervention and control communities would remain the same over time (parallel trends assumption).

In this study, the parallel trend assumption was likely to be violated in two different ways. First, when control communities were much worse than treatment communities at baseline, as they were for the home-based behavior for child health, the parallel trends assumption may not hold owing to the fact the indicators vary between 0% and 100%. For example, at baseline children were significantly more likely to have slept under a bed net the previous night in intervention communities (82%) compared to control communities (71%). By endline, 93% of children in control communities were sleeping under a bed net, an improvement of 22 percentage points; by the parallel trend assumption then, intervention communities would have also improved by 22 percentage points to 104%. Second, there may have been important other changes between baseline and endline in the control communities that invalidate the parallel trends assumption. In particular, other MNCH interventions may have been targeted at control communities at the same time that CCPF was implemented in intervention communities (IKI, 2013). In fact, the data indicate that new MNCH programs were introduced at a higher rate in control communities compared to intervention communities (IKI, 2013), suggesting that the average treatment effects reported may be underestimates.

Non-compliance and self-selection bias

The uptake of different components of the intervention was low, varying between 7% and 18%. As a result, the ITT estimates are not likely to provide a good indication of the impact of the intervention. Indeed when compliance is low, the TOT effects may be more relevant than the ITT estimates. The TOT method attempts adjust for two self-selection biases: a) there are individuals in the control areas who would not be able/willing to use the services even if they were offered; and b) only a subgroup of the individuals assigned to the intervention area actually used the services, *and* use of the services was non-random (Angrist et al., 1996).

The substantial differences between the ITT and TOT estimates observed in this study are driven by the fact that there was fairly low uptake in intervention areas and that self-selection into the use of services in intervention communities introduced substantial bias. Once the non-compliance is accounted for, and the likelihood of service uptake modelled, the TOT estimate provides an unbiased estimate of the impact of CCPF services, furthering our understanding of what the impact would be if there were full compliance.

5. Conclusion

This study has contributed to the evidence base on the effectiveness of mHealth to improve MNCH outcomes. The level of analysis provided by the datasets generated in this evaluation provided invaluable insight for the program in terms of understanding the impact on the population and among those who used the service. Furthermore, this paper demonstrates that rigorous quasi-experimental evaluation designs can be successfully applied to mHealth pilot projects, helping to understand what works and what does not (Victora et al., 2011

References

Al-Shorbaji N & Geissbuhler A. 2012. Establishing an evidence base for e-health: the proof is in the pudding. Bulletin of the World Health Organization 90:322-322A.

Angrist JD, Imbens GW, Rubin DB. 1996. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 91(434):444-455.

Bertrand M, Duflo E, and Mullainathan S. 2004. How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics* 119 (1): 249–75.

Crawford J, Larsen-Cooper E, Jezman Z, Cunningham SC, Bancroft E. 2014. SMS versus voice messaging to deliver MNCH communication in rural Malawi: assessment of delivery success and user experience. *Global Health: Science & Practice*, 2014.

Duflo E, Glennerster R and Kremer M. 2008. Using Randomization in Development Economics Research: A Toolkit." In *Handbook of Development Economics*, vol. 4, ed. T. Paul Schultz and John Strauss, 3895–962. Amsterdam: North-Holland.

Free C, Phillips G, Galli L, Watson L, Galli L, Felix L, et al. 2013a. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. *PLoS Med* 10(1):e1001362.

Free C, Phillips G, Watson L, Galli L, Felix L, et al. 2013b. The Effectiveness of Mobile-Health Technologies to Improve Health Care Service Delivery Processes: A Systematic Review and Meta-Analysis. *PLoS Med* 10(1):e1001363.

Gurman TA, Rubin SE, Roess AA. 2012. Effectiveness of mHealth Behavior Change Communication Interventions in Developing Countries: A Systematic Review of the Literature. Journal of Health Communication 17: 82-104.

Heckman J. 2005. The Scientific model of causality. *Sociological Methodology* 35(1): 1-97

Innovations for MNCH. 2012. Bridging Women and Children to Better Health Care: Maternal, Newborn and Child Health (MNCH) Access through Mobile Technology in Balaka District, Malawi. Concern Worldwide, 2012.

Invest in Knowledge (IKI). 2011. Baseline Report for an Evaluation of the Information and Communications Technology to Improve Health Services for Mothers and Children Project. Malawi, 2011.

Invest in Knowledge (IKI). 2013. Evaluation of The Information and Communications Technology to Improve Health Services for Mothers and Children Project. Malawi, 2013.

Jolliffe IT. 2002. Principal Component Analysis. Series: Springer Series in Statistics, 2nd ed., Springer, NY

Kling JR, Leibman JB, Lawrence KF. 2007. Experimental Analysis of Neighborhood Effects. *Econometrica* 75(1):83-119.

Lemaire J. 2011. Scaling up Mobile Health: Elements Necessary for the Successful Scale up of mHealth in Developing Countries. Geneva: Advanced Development for Africa.

Leon N, Schneider H, Daviaud E. 2012. Applying a framework for assessing the health system challenges to scaling up mHealth in South Africa. *BMC Medical Informatics and Decision Making,* 12:123.

Meyer BD. 1995. Natural and quasi-experiments in economics. *Journal of Business & Economic Statistics* 13(2): 151-161.

mHealth Alliance. 2013. mHealth and MNCH: State of the Evidence - Trends, Gaps, Stakeholder Needs, and Opportunities for Future Research on the Use of Mobile Technology to Improve Maternal, Newborn, and Child Health.

mHealth Alliance. 2012a. Leveraging Mobile Technologies to Promote Maternal & Newborn Health: The Current Landscape & Opportunities for Advancement in Low-Resource Settings.

mHealth Alliance. 2012b. Leveraging Mobile Technologies for Maternal, Newborn & child Health: A Framework for Engagement.

Mitchell M, Hedt-Gauthier BL, Msellemu D, Nkaka M and Lesh N. 2013. Using electronic technology to improve clinical care – results from a before-after cluster trial to evaluate assessment and classification of sick children according to Integrated Management of Childhood Illnesses (IMCI) protocol in Tanzania. *BMC Medical Informatics and Decision Making*, 13:95.

Noordam AC, Kuepper BM, Stekelenburg J, Milen A. 2011. Improvement of maternal health services through the use of mobile phones. Tropical Medicine & International Health 16(5):622–626.

Noordam AC, George A, Sharkey AB, Jafarli A, Bakshi SS, Kim JC. 2014. Assessing scale up of mHealth innovations based on intervention complexity: Two case studies of child health programmes in Malawi and Zambia. *Journal of Health Communication*: Forthcoming

Qiang CZ, Yamamichi M, Hausman V, Altman D. 2011. Mobile applications for the health sector. Washington DC: World Bank.

Rubin. 1974. Estimating causal effect of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 66(5): 688-701.

Tamrat T, Kachnowski S. 2012. Special delivery: An analysis of mHealth in maternal and newborn health programs and their outcomes around the world. *Maternal and Child Health Journal* 16(5):1092-101

Tomlinson M, Rotheram-Borus MJ, Swartz L, Tsai AC. Scaling Up mHealth: Where Is the Evidence? PLoS Med. 2013 Feb 12;10(2):e1001382.

Victora CG, Black RE, Boerma J, Bryce J. 2011. Measuring impact in the Millennium Development Goal era and beyond: a new approach to large-scale effectiveness evaluations. The Lancet 377(9759):85–95.

WHO (World Health Organization). 2011. mHealth: new horizons for health through mobile technologies: second global survey on eHealth. Geneva, 2011.

World Bank & ITU (International Telecommunication Union). 2012. The Little Data Book on Information and Communication Technology. Washington DC, 2012.

Table 1. Number of survey respondents by population group at baseline and at endline

	Baseline survey			End-line survey			Total
	Control	Intervention	Total	Control	Intervention	Total	
Health facility catchment areas	2	4	6	Same as at baseline			6
Villages[1]	10	8	18	39	38	77	NA
Households[2]	1,112	1,698	2,810	1,284	2,359	3,643	6,453
Women 15-49[2]	1,119	1,721	2,840	1,344	2,509	3,853	6,693
Under-5 children[2]	1,365	2,220	3,585	1,075	2,186	3,261	6,846

[1]Random selection of villages (primary sampling units) in the catchment areas of all six qualified health centers in the intervention (4 health centers) and control (2 health centers) sites

[2]All households in the selected villages, and all women and under-fives in those households

Table 3. Awareness and use of the services among women of child bearing age, at endline

	Intervention area		Control area	
	%	N	%	N
Awareness and use of the hotline services				
Heard about the services	76.9	2,509	2.8	1,344
Used the services	18.3	2,509	0.1	1,344
Awareness and use of the Mobile messaging services				
Heard about the services	33.3	2,509	0.4	1,344
Used the services	7.5	2,509	0.0	1,344

Table 2A. Distribution of mothers/caretakers of children under 5 and pregnant women

	Intervention	Control
	%	%
Community level covariates		
Mean distance to the health center (km)	4.98 (3.44)	5.52 (4.13)
Household level covariates		
Wealth		
Poor (Lowest 50%)	54.3	57.8
Rich (Highest 50%)	45.8	42.2
Number of under-5 chidren		
0	18.2	19.3
1	56.1	59.6
2+	25.7	21.0
Ethnicity of the head of household		
Lomwe	21.5	5.6
Ngoni	40.2	6.1
Yao	19.7	80.3
Others	18.5	8.1
Religion of the head of household		
Catholic	18.8	16.2
Other Christian	40.9	69.9
Muslim	37.2	4.6
Other/No Religion	3.1	9.2
Woman-level covariates		
Access to a phone		
No	65.5	77.4
Yes	34.5	22.6
Education		
None	13.8	15.2
Primary	73.1	75.0
Secondary+	13.1	9.9
Marital status		
Not in union	22.3	16.7
In union	77.7	83.3
Age in years		
<20	19.6	19.3
20-29	40.9	43.6
30+	39.6	37.2
N	**4,230**	**2,463**

Table 2B. Percentage distribution of children under 5

	Intervention area	Control area
	%	%
Age in months		
<12	23.6	25.0
12-23	20.7	22.7
24+	55.7	52.3
Sex		
Male	49.5	49.6
Female	50.5	50.5
N	**4,406**	**2,440**

Table 4. Effects of the intervention on aggregate home-based and facility-based MNCH outcomes

	Intention-to-treat (ITT) effect)		Treatment on the treated effect (TOT)	Sub-sample	
	Simple DID[1]	Adjusted[2] DID		N	Description
Use of home-based care for maternal health	-0.006	-0.012	0.479***	2,813	Women 15-49 who had a live birth in last 18 months
Use of home-based care for child health	-0.059**	-0.071**	0.603***	6,846	All children under five
Use of facility-based care for maternal health	0.092	0.085	0.239**	2,813	Women 15-49 who had a live birth in last 18 months
Use of facility-based care for child health	-0.172***	-0.171***	-0.499***	4,068	Children who are 12-23, or had ARI or fever in last 2 weeks

[1]Difference in difference

[2]Controlling for variables at community, household, woman and child levels (in Tables 2A and 2B)

Statistical significance: *p<0.10, **p<0.05, ***p<0.01