4. Advances in impact evaluation

Kara Hanson
London School of Hygiene and Tropical Medicine, United Kingdom of Great Britain and Northern Ireland

Although there is a rich body of literature on health programme evaluation, the work that focuses on system-level interventions is smaller. However, recent years have seen a growth of interest in understanding the ‘impact’ of development interventions, including health system interventions, in order to guide development practice and investments using evidence about ‘what works’ and an understanding of why it works (Evaluation Gap Working Group, 2006). New bodies have been established to promote and finance impact evaluations, such as the International Initiative for Impact Evaluation (3IE) (http://www.3ieimpact.org), and bilateral donors and other funders have given renewed emphasis to strengthening their approaches to evaluation and their capacity to use this evidence in their decision-making. At the same time, influenced by trends within social programme evaluation in higher-income countries (Harrison, 2001), there is an emerging interest in critical realist approaches to evaluation (for example see FEMhealth: http://www.abdn.ac.uk/femhealth). Such approaches consider the question: What works for whom in what circumstances? All approaches to impact evaluation, thus, aim to explain health policy and systems changes and interventions.

Rigour in impact evaluation

There are different meanings of ‘impact’ in the general evaluation literature, but in the contemporary literature, impact is understood to refer to a causal mechanism – the change in an outcome that is caused by a particular programme. This focus on causal mechanisms has meant that a lot of attention is paid to methods for arriving at an unbiased measure of the change that is due to the programme or intervention. A starting point to measure such impacts is to consider what would have happened without the intervention – known as the ‘counterfactual’ – in order to be able to attribute the observed change to the intervention under study. Methodological development in this field has focused to a substantial degree on different approaches to establishing this counterfactual, and on how best to minimize different forms of selection bias.

This body of work also recognizes the importance of external validity – the extent to which findings can be generalized to other settings. This requires understanding the causal mechanism, looking more closely at its causal pathway and testing the validity of assumptions that are made about the route between intervention and impact, in order to assess whether those assumptions are likely to hold in other contexts. It also means paying careful attention to the implementation setting and how this mediates the effects of the intervention.

Two main types of study design are currently used within impact evaluations:

- **Experimental design**: This involves a random assignment of the programme to an intervention group and a control group, with the effect that potential unobserved confounding factors are also randomly distributed between the two groups, minimizing risks of bias.

- **Quasi-experimental designs**: These can involve ‘natural experiments’ which take advantage of a policy or other change that generates an appropriate control group. Study designs then compare groups or areas with and without the intervention; make before-and-after comparisons; adopt ‘difference-in-difference’ approaches (before and after with a control group); or take advantage of a phased implementation that provides variation in the duration of exposure to the programme. Another approach is to use matching methods (such as propensity score matching) in a cross-sectional design to create a control group that is matched on as many observable factors as possible.

Health system interventions have some particular features that influence the choice of evaluation approach. First, they often work through complex causal pathways and are particularly influenced by features of the policy and implementation context. Recent guidance on the
Another feature of evaluation designs for health system interventions is that it is often difficult to use a ‘control group’ to establish the counterfactual because, for example, a policy change takes place at national level (the ‘small n’ (sample size) problem). For instance, changes in regulatory or health financing systems often occur across a whole country at one time so there is no other unit to use as a comparison group.

For both reasons – complexity and the need for alternative approaches to establish the counterfactual – it seems appropriate to recommend that to enhance their rigour all evaluations of health system interventions should be based on a strong programme theory (White, 2009).

Indeed, theory-based evaluation approaches represent a third form of study design for impact evaluation. These approaches are based on an explicit programme theory that sets out the links between inputs, outputs and impacts and tests these causal links using a mix of qualitative and quantitative methods. Realist evaluation, meanwhile, focuses attention on the links between context, mechanisms of change and outcomes, given its interest in how the intervention leads to which effects, under what circumstances (Pawson & Tilley, 1997). It requires that middle range theory, the analysts’ initial ideas about these links is developed prior to, and then tested through, the evaluation. Realist evaluation tends to rely on mixed-methods, with greater use of qualitative methods than other impact evaluations, and adopts approaches to generalization which rely more on analytic, rather than statistical, generalization. Its rigour is then safeguarded by the adoption of approaches common in case-study practice (see section on the case study approach).
References


Overview of selected papers

The papers in this section were chosen because they address system-level interventions and reflect a broad range of approaches to impact evaluation.

- Björkman and Svensson (2009) use a randomized study design to evaluate the impact of a report-card approach to improving community accountability. This paper was selected because of its focus on a novel health system intervention and its use of an experimental design to measure impact.

- Macinko et al. (2007) examine a large-scale health system intervention (a national community-based primary care programme in Brazil) using a quasi-experimental design which takes advantage of the gradual expansion of the programme to generate an internal control group to measure impact.

- Marchal, Dedzo & Kegels (2010) use realist evaluation methods to examine the impact of a particular human resource management approach within one hospital in Ghana. It looks at the link between organizational practices and performance, has strong theoretical underpinnings, and uses exclusively qualitative methods to explore the causal links between management practice and behaviour within the organization.

- Wang et al. (2009) look at the impact on health status of a community-based health insurance scheme in China, in which increased financial risk protection was accompanied by service innovations including more selective purchasing, changes to the provider payment mechanism, and changes to the prescription system. They both adopt a quasi-experimental approach (before-and-after with a control group) and employ propensity score matching to construct a comparison group.
References for selected papers


Reproduced in the print version only, by permission of Wiley. Copyright Wiley, 2009.
POWER TO THE PEOPLE: EVIDENCE FROM A RANDOMIZED FIELD EXPERIMENT ON COMMUNITY-BASED MONITORING IN UGANDA

MARTINA BJÖRKMAN AND JAKOB SVENSSON

This paper presents a randomized field experiment on community-based monitoring of public primary health care providers in Uganda. Through two rounds of village meetings, localized nongovernmental organizations encouraged communities to be more involved with the state of health service provision and strengthened their capacity to hold their local health providers to account for performance. A year after the intervention, treatment communities are more involved in monitoring the provider, and the health workers appear to exert higher effort to serve the community. We document large increases in utilization and improved health outcomes—reduced child mortality and increased child weight—that compare favorably to some of the more successful community-based intervention trials reported in the medical literature.

I. INTRODUCTION

Approximately eleven million children under five years die each year and almost half of these deaths occur in sub-Saharan Africa. More than half of these children will die of diseases (e.g., diarrhea, pneumonia, malaria, measles, and neonatal disorders) that could easily have been prevented or treated if the children had had access to a small set of proven, inexpensive services (Black, Morris, and Bryce 2003; Jones et al. 2003).

Why are these services not provided? Anecdotal, and recently more systematic, evidence points to one possible reason—ineffective systems of monitoring and weak accountability

"This project is a collaborative exercise involving many people. Foremost, we are deeply indebted to Frances Nsonzi and Ritva Reinikka for their contributions at all stages of the project. We would also like to acknowledge the important contributions of Gibwa Kajubi, Abel Ojoo, Anthony Wasswa, James Kanyesigye, Carolyn Winter, Ivo Njosa, Omiat Omongin, Mary Bitekerezo, and the field and data staff with whom we have worked over the years. We thank the Uganda Ministry of Health, Planning Division, the World Bank’s Country Office in Uganda, and the Social Development Department, the World Bank, for their cooperation. We are grateful for comments and suggestions by Paul Gertler, Esther Duflo, Abhijit Banerjee, and seminar and conference participants at Stanford, Berkeley, LSE, Oxford, IGIER, MIT, World Bank, NTNÚ, Namur, UPF, CEPR/EUDN conference in Paris, and BREAD & CESifo conference in Venice. We also thank three anonymous referees and the editor, Lawrence Katz, for very constructive suggestions. Financial support from the Bank-Netherlands Partnership Program, the World Bank Research Committee, the World Bank Africa Region division, and the Swedish International Development Agency, Department for Research Cooperation is gratefully acknowledged. Björkman also thanks Jan Wallander’s and Tom Hedelius’ Research Foundation for funding.

© 2009 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology.
The Quarterly Journal of Economics, May 2009
relationships.¹ This paper focuses on one of these accountability relationships, citizen-clients’ ability to hold providers accountable, using primary health care provision in rural Uganda as a testing ground.

To examine whether community-based monitoring works, we designed and conducted a randomized field experiment in fifty communities from nine districts in Uganda. In the experiment, local nongovernmental organizations (NGOs) facilitated village and staff meetings in which members of the communities discussed baseline information on the status of health service delivery relative to other providers and the government standard. Community members were also encouraged to develop a plan identifying key problems and steps the providers should take to improve health service provision. The primary objective of the intervention was to initiate a process of community-based monitoring that was then up to the community to sustain and lead.

The community-based monitoring project increased the quality and quantity of primary health care provision. A year after the first round of meetings, we found a significant difference in the weight of infants—0.14 z-score increase—and a markedly lower number of deaths among children under five—33 percent reduction in under-5 mortality—in the treatment communities. Utilization for general outpatient services was 20 percent higher in the treatment compared to the control facilities and the overall effect across a set of utilization measures is large and significantly positive. Treatment practices, including immunization of children, waiting time, examination procedures, and absenteeism, improved significantly in the treatment communities, thus suggesting that the changes in quality and quantity of health care provision are due to behavioral changes of the staff. We find evidence that the treatment communities became more engaged and began to monitor the health unit more extensively. Using variation in treatment intensity across districts we show that there is a significant relationship between the degree of community monitoring and health utilization and health outcomes, consistent with the community-based monitoring mechanism.

Community-based, randomized, controlled field trials have been used extensively in medical research to evaluate the

effectiveness of various health interventions (see footnote 14). Our paper is related but differs in one important dimension. Whereas the medical field trials address the question of impact of a biological agent or treatment practice when the health workers competently carry out their tasks, we focus on how to ensure that the health workers actually carry out their tasks and the impact that may have on health utilization and health outcomes.

This paper also relates to a small literature on improving governance and public service delivery through community participation. Olken (2007) finds minor effects of an intervention aimed at increasing community participation in the monitoring of corruption in Indonesia. Our work differs in several ways. First, the intervention we evaluate was structured in a way to reduce the risk of elite capture. Second, unlike corruption, which is not easily observable, the information discussed in the meetings was basic facts on utilization and quality of services based on the community’s own experience. Finally, the intervention sought to address two constraints highlighted in the literature on community monitoring: lack of relevant information and inadequate participation. Banerjee, Deaton, and Dufo (2004) evaluate a project in Rajasthan in India where a member of the community was paid to check whether the nurse-midwife assigned to the health center was present at the center. The intervention had no impact on attendance and the authors speculate that a key reason for this is that the individual community member did not manage to use his or her information on absenteeism to invoke community participation. Here, on the contrary, we explicitly try to address the participation constraint by involving a large number of community members and encouraging them to jointly develop a monitoring plan.

Finally, the paper links to a growing empirical literature on the relationship between information dissemination and accountability (Besley and Burgess 2002; Strömberg 2004; Ferraz and Finan 2008). In this paper, however, we focus on mechanisms through which citizens can make providers, rather than politicians, accountable. Thus, we do not study the design or allocation of public resources across communities, but rather how these resources are utilized. Second, we use microdata from households and clinics rather than disaggregated national accounts data. Finally, we identify impact using an experimental design.

The next section describes the institutional environment. The community-based monitoring intervention is described in
Section III. Section IV lays out the evaluation design and the results are presented in Section V. Section VI concludes. Details about the experiment and additional results are reported in the Online Supplemental Appendix.

II. INSTITUTIONAL SETTING

Uganda, like many newly independent countries in Africa, had a functioning health care system in the early 1960s. The 1970s and 1980s saw the collapse of government services as the country underwent political upheaval. Health indicators fell dramatically during this period until peace was restored in the late 1980s. Since then, the government has been implementing major infrastructure rehabilitation programs in the public health sector.

The health sector in Uganda is composed of four types of facilities: hospitals, health centers, dispensaries, and aid posts or subdispensaries. These facilities can be government-operated and -owned, private for-profit, or private not-for-profit. The impact evaluation focuses on public dispensaries. Dispensaries are in the lowest tier of the health system where a professional interaction between users and providers takes place. Most dispensaries are rural. According to the government health sector strategic plan, the standard for dispensaries includes preventive, promotional, outpatient care, maternity, general ward, and laboratory services (Republic of Uganda 2000). As of 2001, public health services are free. In our sample, on average, a dispensary was staffed by an in-charge or clinical officer (a trained medical worker), two nurses, and three nursing aids or other assistants.

The health sector in Uganda is decentralized, and a number of actors are responsible for supervision and control of the dispensaries. At the lowest tier, the Health Unit Management Committee (HUMC) is supposed to be the main link between the community and the facility. Each dispensary has an HUMC, which consists of both health workers and nonpolitical representatives from the community. The HUMC should monitor the day-to-day running of the facility but it has no authority to sanction workers. The next level in the institutional hierarchy is the health subdistrict. The health subdistrict monitors funds, drugs, and service delivery at the dispensary. Supervision meetings by the health subdistrict are supposed to appear quarterly but, in practice, monitoring is infrequent. The health subdistrict has the authority to reprimand,
but not dismiss, staff for indiscipline. Thus in severe cases of indiscipline, the errand will be referred to the chief administrative officer of the district and the District Service Commission, which are the appointing authorities for the district. They have the authority to suspend or dismiss staff.

Various local NGOs, so-called community-based organizations (CBOs), focusing primarily on health education, are also active in the sector.

III. EXPERIMENTAL DESIGN AND DATA

III.A. Overview

In response to perceived weak health care delivery at the primary level, a pilot project (citizen report cards) aimed at enhancing community involvement and monitoring in the delivery of primary health care was initiated in 2004. The project was designed by staff from Stockholm University and the World Bank, and implemented in cooperation with a number of Ugandan practitioners and eighteen community-based organizations.

The main objective of the intervention was to strengthen providers’ accountability to citizen-clients by initiating a process, using trained local actors (CBOs) as facilitators, which the communities themselves could manage and sustain.

Based on a small but rigorous empirical literature on community participation and oversight, and extensive piloting in the field, our conjecture was that lack of relevant information on the status of service delivery and the community’s entitlements, and failure to agree on, or coordinate expectations of, what is reasonable to demand from the provider, were holding back initiatives to pressure and monitor the provider. Although individual community members have private information—for example, they know whether their own child has died and whether the health workers did anything to help them—they typically do not have any information on aggregate outcomes, such as how many children in their community did not survive beyond the age of 5 or where citizens, on average, seek care, or what the community can expect in terms of quality and quantity of service provision (Khemani 2006). Partly as a response to this information problem, and partly because monitoring a public facility is a public good that may be subject to serious free-rider problems, few people actively participate in monitoring their service providers. Relaxing
these two constraints was therefore the main objective of the intervention.

The key behavioral change induced by more extensive community-based monitoring was expected to be increased effort by the health unit staff to serve the community. In Uganda, as in many other developing countries, health workers have few pecuniary incentives to exert high effort. Public money does not follow patients, and hiring, salaries, and promotions are largely determined by seniority and educational qualifications—not by how well the staff performs. An individual worker may of course still put in high effort if shirking deviates from her ideal choice (Akerlof and Kranton 2005). The effort choice may also be influenced by social rewards from community members or social sanctions against shirking workers. Social rewards and sanctions are key instruments available to the community to boost the health worker’s effort.

III.B. Experimental Design

The experiment involved fifty public dispensaries, and health care users in the corresponding catchment areas, in nine districts covering all four regions in Uganda. All project facilities were located in rural areas. We define a facility’s catchment area, or the community, as the five-kilometer radius around the facility.² A community in our sample has, on average, 2,500 households residing within the five-kilometer radius of the clinic, of which 350 live within a one-kilometer radius. For the experimental design, the facilities were first stratified by location (districts) and then by population size. From each group, half of the units were randomly assigned to the treatment group and the remaining 25 units were assigned to the control group.

III.C. Data

Data collection was governed by two objectives. First, data were required to assess how the community at large views the quality and efficacy of service delivery. We also wanted to contrast the citizens’ view with that of the health workers. Second, data were required to evaluate impact. To meet these objectives, two surveys were implemented: a survey of the fifty providers and

² Dispensaries are designed to serve households in a catchment area roughly corresponding to the five-kilometer radius around the facility (Republic of Uganda 2000).
COMMUNITY-BASED MONITORING IN UGANDA

a survey of users. Both surveys were implemented prior to the intervention (data from these surveys formed the basis for the intervention) and one year after the project had been initiated.

A quantitative service delivery survey was used to collect data from the providers. Because agents in the service delivery system may have a strong incentive to misreport key data, the data were obtained directly from the records kept by facilities for their own need (i.e., daily patient registers, stock cards, etc.) rather than from administrative records. The former, often available in a highly disaggregate format, were considered to suffer the least from any incentive problems in record keeping. Data were also collected through visual checks by enumerators.

The household survey collected data on both households’ health outcomes and health facility performance as experienced by the household. A stratified random sample of households within the catchment area of each facility was surveyed. In total, roughly 5,000 households were surveyed in each round. To the extent that it was possible, patient records (i.e., patient exercise books and immunization cards) supported the household’s response. The postintervention household survey also included a shorter module on health outcomes. Specifically, data on under-5 mortality were collected and we measured the weight of all infants in the surveyed households.

III.D. Intervention

A smaller subset of the findings from the preintervention surveys, including utilization, quality of services, and comparisons vis-à-vis other health facilities, were assembled in report cards. Each treatment facility and its community had a unique report card, translated into the main language spoken in the community, summarizing the key findings from the surveys conducted in their area.

The process of disseminating the report card information, and encouraging participation, was initiated through a series of meetings: a community meeting, a staff meeting, and an interface meeting. Staff from various local NGOs (CBOs) acted as facilitators in

3. The sampling strategy for the baseline household survey was designed to generate representative information on the core users’ variables in each community (such as the proportion of patients being examined with equipment). In total, 88% of the households surveyed in the baseline survey were resurveyed in the ex-post survey. The households that could not be surveyed were replaced.
these meetings. A time line of the intervention is depicted in Figure I.

The community meeting was a two-afternoon event with approximately 100 invited participants from the community. To avoid elite capture, the invited participants consisted of a selection of representatives from different spectra of society (i.e., young, old, disabled, women, mothers, leaders). The facilitators mobilized the village members by cooperating with village council representatives in the catchment area. Invited participants were asked to spread the word about the meeting and, in the end, a large number of uninvited participants also attended the meeting. More than 150 participants per day attended a typical village meeting.

In the community meeting, the facilitators used a variety of participatory methods to disseminate the information in the report cards and encouraged community members to develop a shared view on how to improve service delivery and monitor the provider. Information on patients’ rights and entitlements was also discussed. The participants were divided into focus groups so that also more marginalized groups such as women and youth could raise their voices and discuss issues specific to their group.

4. The eighteen participating CBOs had been active in 64% of the treatment communities and half of the control communities prior to the intervention. A handful of them covered more than one treatment community. The CBOs were primarily focused on health, including issues of health education and HIV/AIDS prevention, although other objectives such as agricultural development, women’s empowerment, support of orphans and vulnerable children, and peace-building initiatives, were also common. The CBO facilitators were trained for seven days in data interpretation and dissemination, utilization of the participatory methodology, and conflict resolution and management. Various other CBOs also operate in the project communities.
At the end of the meeting, the community’s suggestions for improvements, and how to reach them without additional resources, were summarized in an action plan. The action plan contained information on health issues/services that had been identified by the community as the most important to address, how these issues could be addressed, and how the community could monitor improvements (or lack thereof). Although the issues raised in the action plans differed across communities, a common set of concerns included high rates of absenteeism, long waiting time, weak attention of health staff, and differential treatment.

The health facility meeting was a one-afternoon event held at the facility with all staff present. In the meeting, the facilitators contrasted the information on service provision as reported by the provider with the findings from the household survey.

An interface meeting with members from the community, chosen in the community meeting, and health workers followed the community and health facility meetings. During the interface meeting, the community representatives and the health workers discussed suggestions for improvements. The participants discussed their rights and responsibilities as patients or medical staff. The outcome was a shared action plan, or a contract, outlining the community’s and the service provider’s agreement on what needs to be done, and how, when, and by whom. The “community contract” also identified how the community could monitor the agreements and a time plan. Because the problems that were raised in the community meetings constituted the core issues discussed during the interface meetings, the community contract was in many respects similar to the community’s action plan.

The three separate meetings aimed at kick-starting the process of community monitoring. Thus, after the initial meetings the communities were themselves in charge of establishing ways of monitoring the provider. After a period of six months, the communities and health facilities were revisited. The CBOs facilitated a one-afternoon community meeting and a one-afternoon interface meeting with the aim of tracking the implementation of the community contract. Health facility staff and community members jointly discussed suggestions for sustaining or improving progress, or in the case of no improvements, why so.5

5. Details on the report cards and the participatory methods used, as well as an example of an action plan, are provided in the Online Supplemental Appendix.
IV. Evaluation Design and Expected Outcomes

IVA. Outcomes

The main outcome of interest is whether the intervention increased the quantity and quality of health care provision and hence resulted in improved health outcomes. We are also interested in evaluating changes in all steps in the accountability chain: Did the treatment communities become more involved in monitoring the health workers? Did the intervention change the health workers’ behavior?

As a robustness test we also assess alternative explanations. One concern is spillovers. Another concern is that the intervention did not only (or primarily) increase the extent of community monitoring, but had an impact on other agents in the service delivery chain, such as the health subdistrict. The intervention could also have affected the health workers’ behavior directly, or affected it through the actions of the CBOs, rather than through more intense community-based monitoring as we hypothesize. Although this would not invalidate the causal effect of the intervention, it would, of course, affect the interpretation. Therefore, these alternative hypotheses are also subject to a battery of tests.

IVB. Statistical Framework

To assess the causal effect of the intervention we estimate

\[ y_{ijd} = \alpha + \beta T_{jd} + X_{jd}\pi + \theta_d + \varepsilon_{ijd}, \]

where \( y_{ijd} \) is the outcome of household \( i \) (when applicable) in community/health facility \( j \) in district \( d \), \( T_{jd} \) is an indicator variable for assignment to treatment, and \( \varepsilon_{ijd} \) is an error term. Equation (1) also includes a vector, \( X \), of preintervention facility-specific covariates and district fixed effects (\( \theta_d \)). Because of random assignment, \( T \) should be orthogonal to \( X \), and the consistency of \( \beta \) does not depend on the inclusion of \( X \) in the model. The regression adjustment is used to improve estimation precision and to account for stratification and chance differences between groups.

6. The baseline covariates included are number of villages in the catchment area, number of days without electricity in the past month, indicator variable for whether the facility has a separate maternity unit, distance to nearest public health provider, number of staff with less than advanced A-level education, indicator variable for whether the staff could safely drink from the water source, and average monthly supply of quinine.
in the distribution of pre-random assignment (Kling, Liebman, and Katz 2007).

We report the results of estimating equation (1) with $X$ and $\theta$ excluded in the Online Supplemental Appendix. For a subset of variables we can also stack the pre- and postdata and explore the difference-in-differences in outcomes; that is, we estimate

$y_{ijt} = \gamma_{POST} + \beta_{DD}(T_j \ast POST) + \mu_j + \varepsilon_{ijt}$, (2)

where POST is an indicator variable for the postintervention period, $\mu_j$ is a facility/community specific fixed effect, and $\beta_{DD}$ is the difference-in-differences estimate (program impact).

For some outcomes we have several outcome measures. To form judgment about the impact of the intervention on a family of $K$ related outcomes, we follow Kling et al. (2004) and estimate a seemingly unrelated regression system,

$Y = [I_K \otimes (TX)]\theta + \nu$, (3)

where $I_K$ is a $K$ by $K$ identity matrix. We then derive average standardized treatment effects, $\tilde{\beta} = 1/K \sum_{k=1}^{K} \hat{\beta}_k / \hat{\sigma}_k$, where $\hat{\beta}_k$ and $\hat{\sigma}_k$ are the point estimate and standard error, respectively, for each effect (see Duflo, Glennerster, and Kremer [2007]). The point estimate, standard error, and $p$-value for $\tilde{\beta}$ are based on the parameters, $\hat{\beta}_k$ and $\hat{\sigma}_k$, jointly estimated as elements of $\theta$ in (3).

V. RESULTS

V.A. Preintervention Differences

The treatment and the control group were similar on most characteristics prior to the intervention. Average standardized pretreatment effects are estimated for each family of outcomes (utilization, utilization pattern, quality, catchment area statistics, health facility characteristics, citizen perceptions, supply of resources, and user charges) using preintervention data. As shown in Table I, we cannot reject the null hypotheses of no difference between the treatment and the control group.\footnote{It is a subset of variables because the postintervention surveys collected information on more variables and outcomes.}

\footnote{We report the test of difference in means across control and treatment groups for each individual variable in the Online Supplemental Appendix.}
TABLE I
Pretreatment Facility and Catchment Area Characteristics and Average Standardized Effects

<table>
<thead>
<tr>
<th>Variables</th>
<th>Treatment group</th>
<th>Control group</th>
<th>Difference</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Key characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outpatient care</td>
<td>593 (75)</td>
<td>675 (57)</td>
<td>−82 (94)</td>
<td></td>
</tr>
<tr>
<td>Delivery</td>
<td>10.3 (2.2)</td>
<td>7.5 (1.4)</td>
<td>2.8 (2.6)</td>
<td></td>
</tr>
<tr>
<td>No. of households in catchment area</td>
<td>2,140 (185)</td>
<td>2,224 (204)</td>
<td>−84.4 (276)</td>
<td></td>
</tr>
<tr>
<td>No. of households per village</td>
<td>93.9 (5.27)</td>
<td>95.3 (6.32)</td>
<td>−1.42 (8.23)</td>
<td></td>
</tr>
<tr>
<td>Drank safely today</td>
<td>0.40 (0.10)</td>
<td>0.32 (0.10)</td>
<td>0.08 (0.14)</td>
<td></td>
</tr>
<tr>
<td>No. of days without electricity in past month</td>
<td>18.3 (2.95)</td>
<td>20.4 (2.90)</td>
<td>−2.12 (4.14)</td>
<td></td>
</tr>
<tr>
<td><strong>Average standardized pretreatment effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utilization</td>
<td>0.11 (0.77)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utilization pattern</td>
<td>−0.48 (0.33)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality measures</td>
<td>−0.35 (0.84)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Catchment area statistics</td>
<td>0.11 (0.66)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health facility characteristics</td>
<td>0.14 (0.31)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Citizen perceptions</td>
<td>0.37 (0.67)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supply of drugs</td>
<td>0.73 (0.83)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>User charges</td>
<td>−0.65 (0.63)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes. Key characteristics are catchment area/health facility averages for treatment and control group and differences in averages. Robust standard errors in parentheses. Description of variables: Outpatient care is average number of patients visiting the facility per month for outpatient care. Delivery is average number of deliveries at the facility per month. Number of households in catchment area and number of households per village are based on census data and Uganda Bureau of Statistics maps. Drank safely today is an indicator variable for whether the health facility staff at the time of the preintervention survey could safely drink from the water source. Number of days without electricity in the month prior to preintervention survey is measured out of 31 days. Average standardized pretreatment effects are derived by estimating equation (3) on each family of outcomes. Utilization summarizes outpatients and deliveries. Utilization pattern summarizes the seven measures in Supplemental Appendix Table A.I, reversing sign of traditional healer and self-treatment. Quality measures summarize the two measures in Table A.I, reversing sign of waiting time. Catchment area statistics summarize the four measures in Table A.I. Health facility characteristics summarize the eight measures in Table A.I and drank safely today and days without electricity, reversing sign of days without electricity and distance to nearest local council. Citizen perceptions summarize the four measures in Table A.I. Supply of drugs summarizes the five measures in Table A.I. User charges summarize the four measures in Table A.I, reversing all signs. The $\chi^2$ test-statistic on the joint hypothesis that all average standardized effects are 0 is 4.70 with p-values = .79.
V.B. Processes

The initial phase of the project, that is, the three separate meetings, followed a predesign structure. A parallel system whereby a member of the survey team originating from the district participated as part of the CBO team also confirmed that the initial phase of the intervention was properly implemented. After these initial meetings, it was up to the community to sustain and lead the process. In this section we study whether the treatment communities became more involved in monitoring the providers.

To avoid influencing local initiatives, we did not have external agents visiting the communities and could therefore not document all actions taken by the communities in response to the intervention. Still, we have some information on how processes in the community have changed. Specifically, the CBOs submitted reports on what type of changes they observed in the treatment communities and we also surveyed the local councils in the treatment communities. We use facility and household survey data to corroborate these reports.

According to the CBO reports and the local council survey, the community-based monitoring process that followed the first set of meetings was a joint effort mainly managed by the local councils, HUMC, and community members. A typical village in the treatment group had, on average, six local council meetings in 2005. In those meetings, 89% of the villages discussed issues concerning the project health facility. The main subject of discussion in the villages concerned the community contract or parts of it, such as behavior of the staff.

The CBOs reported that concerns raised by the village members were carried forward by the local council to the facility or the HUMC. However, although the HUMC is an entity that should play an important role in monitoring the provider, it was in many cases viewed as being ineffective. As a result, mismanaged HUMCs were dissolved and new members elected. These claims are confirmed in the survey data: more than one-third of the HUMCs in the treatment communities were dissolved and new members were elected or received following the intervention, whereas we observed no dissolved HUMCs in the control communities. Further, the CBOs report that the community, or individual members, also monitored the health workers during visits to the clinic, when they rewarded and questioned issues in the community contract that had or had not been addressed, suggesting a
more systematic use of nonpecuniary rewards. Monitoring tools such as suggestion boxes, numbered waiting cards, and duty rosters were also reported to be put in place in several treatment facilities.

In Table II, we formally look at the program impact on these monitoring tools. We use data collected through visual checks by enumerators during the postintervention facility survey. As shown in columns (1) and (2), one year into the project, treatment facilities are significantly more likely to have suggestion boxes (no control facility had these, but 36% of the treatment facilities did) and numbered waiting cards (only one control facility had one, but 20% of the treatment facilities did). Columns (3) and (4) show that a higher share of the treatment facilities also posted information on free services and patients’ rights and obligations. The enumerators could visually confirm that 70% of the treatment facilities had at least one of these monitoring tools, whereas only 4 of 25 control clinics had at least one of them. The difference is statistically significant (Online Supplemental Appendix, Table A.II). Column (5) reports the average standardized effect of the monitoring tools. The estimate is significantly different from zero at the 1% level.

The results based on household data mirror the findings reported in columns (1)–(5). The performance of the staff is more often discussed in local council meetings in the treatment communities, shown in column (6), and community members in the treatment group are, on average, better informed about the HUMC’s roles and responsibilities, as reported in column (7). Combining the evidence from the CBO reports and the household survey data thus suggests that both the “quantity” of discussions about the project facility and the subject, from general to specific discussions about the community contract, changed in response to the intervention.

V.C. Treatment Practices

The qualitative evidence from the CBOs and, to the extent that we can measure them, the findings reported in Table II suggest that the treatment communities became more involved in monitoring the provider. Did the intervention also affect the health workers’ behavior and performance? We turn to this next.

We start by looking at examination procedures. The estimate based on equation (2) with the dependent variable being
### TABLE II
**PROGRAM IMPACT ON MONITORING AND INFORMATION**

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Suggestion box (1)</th>
<th>Numbered waiting cards (2)</th>
<th>Poster informing free services (3)</th>
<th>Poster on patients' rights (4)</th>
<th>Average standardized effect (5)</th>
<th>Discuss facility in LC meetings (6)</th>
<th>Received information about HUMC (7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specification</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Program impact</td>
<td>0.32*** (0.08)</td>
<td>0.16* (0.09)</td>
<td>0.27*** (0.09)</td>
<td>0.14 (0.10)</td>
<td>2.55*** (0.55)</td>
<td>0.13*** (0.03)</td>
<td>0.04*** (0.01)</td>
</tr>
<tr>
<td>Mean control group</td>
<td>0</td>
<td>0.04</td>
<td>0.12</td>
<td>0.12</td>
<td>—</td>
<td>0.33</td>
<td>0.08</td>
</tr>
<tr>
<td>Observations</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>3,119</td>
<td>4,996</td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** Robust standard errors in parentheses. Disturbance terms are clustered by catchment areas in columns (6)–(7). Point estimates, standard errors, and average standardized effect, columns (1)–(5), are derived from equation (3). Program impact measures the coefficient on the assignment to treatment indicator. Outcome measures in columns (1)–(4) are based on data collected through visual checks by the enumerators during the postintervention facility survey. Outcome measures in columns (6) and (7) are from the postintervention household survey. The estimated equations all include district fixed effects and the following baseline covariates: number of villages in catchment area, number of days without electricity in the past month, indicator variable for whether the facility has a separate maternity unit, distance to nearest public health provider, number of staff with less than advanced A-level education, indicator variable for whether the staff could safely drink from the water source, and average monthly supply of quinine. Specification (1) indicator variable for whether the health facility has a suggestion box for complaints and recommendations; (2) indicator variable for whether the facility has numbered waiting cards for its patients; (3) indicator variable for whether the facility has a poster informing about free health services; (4) indicator variable for whether the facility has a poster on patients' rights and obligations; (5) average standardized effect of the estimates in columns (1)–(4); (6) indicator variable for whether the household discussed the functioning of the health facility at a local council meeting during the past year; (7) indicator variable for whether the household has received information about the Health Unit Management Committee's (HUMC's) roles and responsibilities.

*Significant at 10%.
**Significant at 5%.
***Significant at 1%.
## TABLE III  
**PROGRAM IMPACT ON TREATMENT PRACTICES AND MANAGEMENT**

<table>
<thead>
<tr>
<th>Spec.</th>
<th>Dep. variable</th>
<th>Model</th>
<th>Program impact</th>
<th>Mean control</th>
<th>Obs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>Equipment used</td>
<td>DD</td>
<td>0.08**</td>
<td>-0.07***</td>
<td>0.41</td>
</tr>
<tr>
<td>(2)</td>
<td>Equipment used</td>
<td>OLS</td>
<td>0.01</td>
<td>(0.02)</td>
<td>0.41</td>
</tr>
<tr>
<td>(3)</td>
<td>Waiting time</td>
<td>DD</td>
<td>-12.3**</td>
<td>131</td>
<td>6,602</td>
</tr>
<tr>
<td>(4)</td>
<td>Waiting time</td>
<td>OLS</td>
<td>-5.16</td>
<td>(5.51)</td>
<td>131</td>
</tr>
<tr>
<td>(5)</td>
<td>Absence rate</td>
<td>OLS</td>
<td>-0.13**</td>
<td>(0.06)</td>
<td>0.47</td>
</tr>
<tr>
<td>(6)</td>
<td>Management of clinic</td>
<td>OLS</td>
<td>1.20***</td>
<td>(0.33)</td>
<td>-0.49</td>
</tr>
<tr>
<td>(7)</td>
<td>Health information</td>
<td>OLS</td>
<td>0.07***</td>
<td>(0.02)</td>
<td>0.32</td>
</tr>
<tr>
<td>(8)</td>
<td>Importance of family planning</td>
<td>OLS</td>
<td>0.06***</td>
<td>(0.02)</td>
<td>0.31</td>
</tr>
<tr>
<td>(9)</td>
<td>Stockouts</td>
<td>OLS</td>
<td>-0.15**</td>
<td>(0.07)</td>
<td>0.50</td>
</tr>
</tbody>
</table>

Notes. Each row is based on a separate regression. The DD model is from equation (2). The OLS model is from equation (1) with district fixed effects and baseline covariates as listed in Table II. Robust standard errors, clustered by catchment areas, are in columns (1)–(4) and (7)–(8), in parentheses. Program impact measures the coefficient on the assignment to treatment indicator in the OLS models and the assignment to treatment indicator interacted with an indicator variable for 2005 in the DD models. Specifications: (1) and (2) indicator variable for whether the staff used any equipment during examination when the patient visited the health facility; (3) and (4) difference between the time the citizen left the facility and the time the citizen arrived at the facility, minus the examination time; (5) ratio of workers not physically present at the time of the postintervention survey to the number of workers employed preintervention (see text for details); (6) first component from a principal components analysis of the variables Condition of the floors of the health clinic, Condition of the walls, Condition of furniture, and Smell of the facility, where each condition is ranked from 1 (dirty) to 3 (clean) by the enumerators; (7) indicator variable for whether the household has received information about the importance of visiting the health facility and the danger of self-treatment; (8) indicator variable for whether the household has received information about family planning; (9) share of months in 2005 in which stock cards indicated no availability of drugs (see text for details).

* Significantly different from zero at 90% confidence level.
** Significantly different from zero at 95% confidence level.
*** Significantly different from zero at 99% confidence level.

An indicator variable for whether any equipment, for instance, a thermometer, was used during examination is shown in the first row in Table III. Fifty percent (41) of the patients in the treatment (control) community reported that equipment was used the last time the respondent (or the respondent’s child) visited the project clinic. The difference-in-differences estimate, a 20% increase, is highly significant. The cross-sectional estimate in row (2), based on equation (1), is less precisely estimated.
In row (3) we report the result with an alternative measure of staff performance—the waiting time—defined as the difference between the time the user left the facility and the time the user arrived at the facility, subtracting the examination time. On average, the waiting time was 131 minutes in the control facilities and 119 in the treatment facilities. The estimate based on equation (1), shown in column (4), is less precisely estimated.

The results on absenteeism are shown in row (3). The point estimate suggests a substantial treatment effect. On average, the absence rate, defined as the ratio of workers not physically present at the time of the postintervention survey to the number of workers on the list of employees as reported in the preintervention survey, is 13 percentage points lower in the treatment facilities. Thus, in response to the intervention, health workers are more likely to be at work.

Enumerators also visually checked the condition of the health clinics, that is, whether floors and walls were clean, the condition of the furniture, and the smell of the facility. We combine these variables through principal components analysis into a summary score. Treatment clinics appear to have put more effort into keeping the clinic in decent condition in response to the intervention. The point estimate, reported in row (6), implies a 0.56 standard deviation improvement in the summary score in the treatment compared to the control facilities.

According to the government health sector strategic plan, preventive care is one of the core tasks for health providers at the primary level. A significantly larger share of households in the treatment communities have received information about the dangers of self-treatment, reported in row (7), and the importance of family planning, reported in row (8). The difference is 7 and 6 percentage points, respectively.

There is no systematic difference in the supply of drugs between the treatment and control groups (see Section V.F). However, as shown in row (9), stockouts of drugs are occurring at a higher frequency in the control facilities even though, as reported

---

9. The postintervention survey was not announced in advance. At the start of the survey, the enumerators physically verified the provider’s presence. A worker was counted as absent if, at the time of the visit, he or she was not in the clinic. Staff reported to be on outreach were omitted from the absence calculation. Four observations were dropped because the total number of workers verified to be present or reported to be on outreach exceeded the total number of workers on the preintervention staff list. Assuming instead no absenteeism in these four facilities yields a point estimate (standard error) of −0.20 (0.065).
below, the control facilities treat significantly fewer patients. These findings suggest that more drugs leaked from health facilities in the control group.\textsuperscript{10}

The findings on immunization of children under five are reported in Table IV. We have information on how many times (doses) in total each child has received polio, DPT, BCG, and measles vaccines and vitamin A supplements. On the basis of the recommended immunization plan, we create indicator variables taking the value of 1 if child $i$ of cohort (age) $j$ had received the required dose(s) of measles, DPT, BCG, and polio vaccines, respectively, and 0 otherwise.\textsuperscript{11} We then estimate (3), for each age group, and calculate average standardized effects.

The average standardized effects are significantly positive for the younger cohorts. Looking at individual effects (Online Supplemental Appendix Table A.IV), there are significant positive differences between households in the treatment and control community for all five vaccines, although not for all cohorts. For example, twice as many newborns in the treatment group have received vitamin A supplements, 46% more newborns have received the first dose of BCG vaccine, and 42% more newborns have received the first dose of polio vaccine as compared to the control group.

V.D. Utilization

To the extent we can measure it, the evidence presented so far suggests that treatment communities began to monitor the health unit more extensively in response to the intervention and that the health workers improved the provision of health services. We now turn to the question of whether the intervention also resulted in improved quantity and quality of care.

Cross-sectional estimates based on equation (3) are given in Table V, Panel A. For outpatients and deliveries, we have

\textsuperscript{10} The dependent variable is the share of months in 2005 in which stock cards indicated no availability of drugs, averaged over erythromycin, mebendazole, and septrin. We find no significant difference between treatment and control clinics for chloroquine—the least expensive of the drugs on which we have data. Not all clinics had accurate stock cards and these clinics were therefore omitted.

\textsuperscript{11} According to the Uganda National Expanded Program on Immunization, each child in Uganda is supposed to be immunized against measles (one dose at nine months and two doses in case of an epidemic); DPT (three doses at six, ten, and fourteen weeks); BCG (one dose at birth or during the first contact with a health facility); and polio (three doses, or four if delivery takes place at the facility, at six, ten, and fourteen weeks). Because measles vaccination should not be given at birth, we exclude immunization against measles in the plan for infants under twelve months.
### TABLE IV

**Program Impact on Immunization**

<table>
<thead>
<tr>
<th>Group Specification</th>
<th>Newborn</th>
<th>Under 1 year</th>
<th>1 year old</th>
<th>2 years old</th>
<th>3 years old</th>
<th>4 years old</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td>Average standardized effect</td>
<td>1.30*</td>
<td>1.44**</td>
<td>1.24**</td>
<td>0.72</td>
<td>2.01***</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>(0.70)</td>
<td>(0.72)</td>
<td>(0.63)</td>
<td>(0.58)</td>
<td>(0.67)</td>
<td>(0.80)</td>
</tr>
<tr>
<td>Observations</td>
<td>173</td>
<td>929</td>
<td>940</td>
<td>951</td>
<td>1,110</td>
<td>526</td>
</tr>
</tbody>
</table>

Notes: Average standardized effects are derived from equation (3) with the dependent variables being indicator variables for whether the child has received at least one dose of measles, DPT, BCG, and polio vaccines and vitamin A supplement, respectively (see text for details), and with district fixed effects and baseline covariates listed in Table II included. Robust standard errors clustered by catchment areas in parentheses. Groups: (1) Children under 3 months; (2) Children 0–12 months; (3) Children 13–24 months; (4) Children 25–36 months; (5) Children 37–48 months; (6) Children 49–60 months.

*Significant at 10% level.

**Significant at 5% level.

***Significant at 1% level.
### TABLE V

**Program Impact on Utilization/Coverage**

<table>
<thead>
<tr>
<th>Dep. variable</th>
<th>Outpatients</th>
<th>Delivery</th>
<th>Antenatal</th>
<th>Family planning</th>
<th>Average std effect</th>
<th>Use of project facility</th>
<th>Use of self-treatment/traditional healers</th>
<th>Average std effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: Cross-sectional data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Program impact</td>
<td>130.2**</td>
<td>5.3**</td>
<td>15.0</td>
<td>3.4</td>
<td>1.75***</td>
<td>0.026</td>
<td>-0.014</td>
<td>1.43*</td>
</tr>
<tr>
<td>(60.8)</td>
<td>(2.1)</td>
<td>(11.2)</td>
<td>(3.2)</td>
<td>(0.63)</td>
<td>(0.016)</td>
<td>(0.011)</td>
<td>(0.87)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>B: Panel data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Program impact</td>
<td>189.1***</td>
<td>3.48*</td>
<td>78.9</td>
<td>15.2</td>
<td>2.30***</td>
<td>0.031</td>
<td>-0.046**</td>
<td>1.96**</td>
</tr>
<tr>
<td>(67.2)</td>
<td>(1.96)</td>
<td></td>
<td>(8.9)</td>
<td>(0.69)</td>
<td>(0.017)</td>
<td>(0.021)</td>
<td>(0.89)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Mean control group 2005</td>
<td>661</td>
<td>9.2</td>
<td>78.9</td>
<td>15.2</td>
<td></td>
<td>0.24</td>
<td>0.36</td>
<td></td>
</tr>
</tbody>
</table>

Notes. Panel A reports program impact estimates from cross-sectional models with district fixed effects and baseline covariates as listed in Table II, with robust standard errors in parentheses. Panel B reports program impact estimates from difference-in-difference models with robust standard errors clustered by facility in parentheses. Point estimates, standard errors, and average standardized effects in specifications (1)–(4) and (9)–(10) are derived from equation (3). Program impact measures the coefficient on the assignment to treatment indicator in the OLS models and the assignment to treatment indicator interacted with an indicator variable for 2005 in the DD models. Specifications: first column is average number of patients visiting the facility per month for outpatient care; second column is average number of deliveries at the facility per month; third column is average number of antenatal visits at the facility per month; fourth column is average number of family planning visits at the facility per month; fifth column is average standardized effect of estimates in specifications (1)–(4) and (9)–(10), respectively; sixth column is the share of visits to the project facility of all health visits, averaged over catchment area; seventh column is the share of visits to traditional healers and self-treatment of all health visits, averaged over catchment area; eighth column is average standardized effect of estimates in specifications (6)–(7) and (12)–(14), respectively, reversing the sign of use of self-treatment/traditional healers.

*Significant at 10% level.
**Significant at 5% level.
***Significant at 1% level.
preintervention data and can also estimate difference-in-differences models, shown in Panel B, and value-added models, shown in Table A.V in the Online Supplemental Appendix.\footnote{12}

One year into the program, utilization (for general outpatient services) is 20\% higher in the treatment facilities as shown in specification (1). For the difference-in-differences and the value-added models (reported in specification (9) in Table V and specification (ix) in Table A.V), the coefficients on the treatment indicator are larger both in absolute magnitude and relative to their standard errors. Thus, controlling for baseline outcomes, $y_{jt-1}$, improves the precision of the treatment effect, which is to be expected given the persistent nature of the outcome variable. The difference in the number of deliveries, shown in specification (2), albeit starting from a low level, is 58\% and is fairly precisely estimated. There are also positive differences in the number of patients seeking antenatal care (19\% increase) and family planning (22\% increase), although these estimates are not individually significantly different from zero. The average standardized effect, reported in specification (5), however, is highly significant.

The last three columns in Table V, Panels A and B, report changes in utilization patterns based on household data. We collected data on where each household member sought care during 2005 in case of illness that required treatment and collapsed this information by community. There is an 11\%–13\% increase, specifications (6) and (12), in the use of the project facility in treatment as compared to the control group—a result consistent with that reported in specification (1) using facility records.

Households in the treatment community also reduced the number of visits to traditional healers and the extent of self-treatment, specifications (7) and (13), but there are no statistically significant differences across the two groups in the use of other providers (not reported). Thus, as summarized in the average standardized treatment effects, specifications (8) and (9), households in the treatment communities switched from traditional healers and self-treatment to the project facility in response to the intervention.

\footnote{12. The value-added specification is $y_{jt} = \alpha_{VA} + \beta_{VA}T_j + \lambda y_{jt-1} + \epsilon_{jt}$.}
V.E. Health Outcomes

We collected data on births, pregnancies, and deaths of children under five years in 2005. We also measured the weight of all infants (i.e., under age 18 months) and children (between ages 18 and 36 months) in the surveyed households.

Health outcomes could have improved for several reasons. As noted in the Introduction, access to a small set of proven, inexpensive services could, worldwide, have prevented more than half of all deaths of children under age 5. For a country with an epidemiological profile as in Uganda, the estimate of preventable deaths is 73 percent (Jones et al. 2003). In the community monitoring project specifically, increased utilization and having patients switch from self-treatment and traditional healers to seek care at the treatment facility could have an effect. Holding utilization constant, better service quality, increased immunization, and more extensive use of preventive care could also have resulted in improved health status.

As a reference point we review the set of health interventions feasible for delivery at high coverage in low-income settings with sufficient evidence of effect on reducing mortality from the major causes of under-5 deaths (Jones et al. 2003). We focus on community-based, randomized, controlled field trials that bear some resemblance (because they are community-based) to our project. Several of these field trials document reductions in under-5 mortality rates of 30%–50% one to two years into the project. There is, however, a fundamental difference between the

13. This is likely to be a conservative number because only medical interventions for which cause-specific evidence of effect was available were included in the estimation. For example, increased birth spacing, which has been estimated to reduce under-5 mortality by 19 percent in India, was not considered. Several perinatal and neonatal health interventions that could be implemented in low-income countries were not included either (Darmstadt et al. 2005).

14. For example, a project in Tigray, Ethiopia, in which coordinators, supported by a team of supervisors, were trained to teach mothers to recognize symptoms of malaria in their children and provide antimalarials, reduced under-5 mortality by 40% (Kidane and Morrow 2000). Bang et al. (1999) document a 30% reduction in under-5 mortality from an intervention that included mass education about childhood pneumonia and case management of pneumonia by trained village health workers—a result similar to the meta-analysis estimate by Sazawal and Black (2003). Bang et al. (1999) evaluate a project in which trained village health workers, assisted by birth attendants and supervisory visits, provided home-based neonatal care, including treatment of sepsis. Two years into the project, they document a reduction in infant mortality by nearly 50 percent. Rahmathullah et al. (2003) assess the impact of a community-based project in two rural districts of Tamil Nadu, India, where newborn infants in the treatment group were allocated oral vitamin A after delivery. The intervention resulted in a 22% reduction in total mortality at age 6 months. Manandhar et al. (2004) evaluate a project in which a
interventions discussed in footnote 14 and our work. The medical field trials study the impact of a biological agent or treatment practice in a community setting when the community health workers and medical personnel competently carry out their tasks. In the experiment we consider, on the contrary, no new health interventions were introduced and the supply of health inputs was unchanged. Instead, we focused on incentivizing health workers to carry out their tasks through strengthened local accountability.

 Estimates for births and pregnancies are given in Table VI, columns (1) and (2). To the extent that the intervention had an effect on fertility, for example, through increased use of family planning services, it would primarily affect the incidence of pregnancies in 2005, given the forty-week period between conception to birth. The incidence of births is not significantly different across treatment and control groups. However, the treatment groups had 10% fewer incidences of pregnancies in 2005.

 Column (3) shows the treatment effect on under-5 mortality. The point estimate suggests a substantial treatment effect. The average under-5 mortality rate in the control group is 144, close to the official figure of 133 for 2005 (UNICEF 2006). In the treatment group, the under-5 mortality rate is 97, which is a 33% reduction in under-5 mortality. The difference is significant (and somewhat larger in absolute magnitude) when controlling for district fixed effects as reported in column (3). Although the effect is large, it is worth emphasizing that the 90% confidence interval of our estimate also includes much lower effects (90% CI: 8%–64% reduction in under-5 mortality rate). With a total of approximately 55,000 households residing in the treatment communities, the treatment effect corresponds to approximately 550 averted under-5 deaths in the treatment group in 2005.

 facilitator convened nine women’s group meetings every month in the Makwanpur district in Nepal in which perinatal problems were identified and strategies to address them formulated. Two years into the project they document a 30% reduction in neonatal mortality. Rahman et al. (1982) evaluate the impact of immunization of women with tetanus injections during pregnancy in rural Bangladesh. The intervention reduced neonatal mortality by 45%. Mungo and Neuvians (1986) evaluate a project in rural Tanzania in which trained village health workers visited families at their homes every six to eight weeks, giving health education on recognition and prevention of acute respiratory infections, treating children with pneumonia with antibiotics or referring them to the next higher level of care. Within a two-year period, they document a 27% reduction in under-5 mortality—a reduction slightly lower than that found in a similar study in rural Bangladesh (Fauveau et al. 1992).

 15. The under-5 mortality rate is the sum of the death rates for each cohort (age groups 0–1, 1–2, 2–3, 3–4, and 4–5) per community in 2005, expressed per thousand live births.
**TABLE VI**

**Program Impact on Health Outcomes**

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Births (1)</th>
<th>Pregnancies (2)</th>
<th>U5MR (3)</th>
<th>Child death (4)</th>
<th>Weight-for-age z-scores (5)</th>
<th>Mean control group 2005</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specification:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Program impact</td>
<td>−0.016</td>
<td>−0.03**</td>
<td>−49.9*</td>
<td>0.14**</td>
<td>0.14**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.014)</td>
<td>(26.9)</td>
<td>(0.07)</td>
<td>(0.07)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child age (log)</td>
<td></td>
<td></td>
<td></td>
<td>−1.27***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.07)</td>
<td>(0.09)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td></td>
<td>0.27***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.09)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Program impact ×</td>
<td></td>
<td></td>
<td></td>
<td>−0.026**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>year of birth 2005</td>
<td></td>
<td></td>
<td></td>
<td>(0.013)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Program impact ×</td>
<td></td>
<td></td>
<td></td>
<td>−0.019**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>year of birth 2004</td>
<td></td>
<td></td>
<td></td>
<td>(0.008)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Program impact ×</td>
<td></td>
<td></td>
<td></td>
<td>0.003</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>year of birth 2003</td>
<td></td>
<td></td>
<td></td>
<td>(0.009)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Program impact ×</td>
<td></td>
<td></td>
<td></td>
<td>0.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>year of birth 2002</td>
<td></td>
<td></td>
<td></td>
<td>(0.006)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Program impact ×</td>
<td></td>
<td></td>
<td></td>
<td>0.002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>year of birth 2001</td>
<td></td>
<td></td>
<td></td>
<td>(0.006)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean control group 2005</td>
<td>0.21</td>
<td>0.29</td>
<td>144</td>
<td>0.029</td>
<td>−0.71</td>
<td>−0.71</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>4,996</td>
<td>4,996</td>
<td>50</td>
<td>5,094</td>
<td>1,135</td>
<td>1,135</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Estimates from equation (1) with district fixed effects and baseline covariates as listed in Table II included. Specification (4) also includes a full set of year-of-birth indicators. Robust standard errors in parentheses (3), clustered by catchment area (1)–(2), (4)–(6). Program impact measures the coefficient on the assignment to treatment indicator.

Specifications: (1) Number of births in the household in 2005; (2) indicator variable for whether any women in the household are or were pregnant in 2005; (3) USMR is under-5 mortality rate in the community expressed per 1,000 live births (see text for details); (4) indicator variable for child death in 2005; (5)–(6) weight-for-age z-scores for children under 18 months excluding observations with recorded weight above the 90th percentile in the growth chart reported in Cortinovis et al. (1997).

*Significant at 10% level.
**Significant at 5% level.
***Significant at 1% level.
Column (4) shows the age range of the mortality effects. We have information on the birth year of all children (under age 5) alive at the beginning of 2005 and the birth year of all deceased children in 2005. Using these data we estimate (1), replacing the treatment indicator with a full set of year-of-birth indicators and year-of-birth-by-treatment interactions. We can then address the question: Conditional on having a child of age $x$ at the end of 2004, or a child born in 2005, what is the probability that the child died in 2005? As evident, children younger than two years old drive the reduction in under-5 mortality. The point estimate for the youngest cohort, for example, implies a 35% reduction in the likelihood of death of a child born in 2005 in the treatment compared to the control group.

The program impact on the weight of infants is reported in columns (5) and (6). On the basis of weight-for-age z-scores, Ugandan infants have values of weight far lower than the international reference of the U.S. National Center for Health Statistics of the Centers for Disease Control and Prevention (CDC) and the gap increases for older infants, consistent with the findings in Cortinovis et al. (1997). The difference in means of z scores of infants between the treatment and the control group is reported in column (5): The estimated effect (difference) is 0.14 in weight-for-age. Figure II plots the distribution of z scores for the treatment and control groups. The difference in measured weight is most apparent for underweight children. This is consistent with a positive treatment effect arising from improved access and quality of health care, rather than a general increase in nutritional status, because underweight status causes a decrease in immune and nonimmune host defenses and, as a consequence, underweight children are at a higher risk of suffering from infectious diseases or severe complications of infectious diseases, and therefore in higher demand of health care. In column (6) of Table VI, we add controls for age and gender. The results remain qualitatively unchanged.

The treatment effect is quantitatively important. For this purpose, the baseline proportion of infants in each risk category

16. The z-score is a normally distributed measure of growth defined as the difference between the weight of an individual and the median value of weight for the reference population (2000 CDC Growth Reference in the United States) for the same age, divided by the standard deviation of the reference population. We exclude $z$-scores $>4.5$ as implausible and omit observations with a recorded weight above the 90th percentile in the growth chart reported in Cortinovis et al. (1997). Because weight is measured by trained enumerators, the reporting error is likely due to misreported age of the child. The coefficient estimate (standard error) on the treatment indicator is 0.16 (0.09) when including these outliers.
Distributions of Weight-for-Age z-Scores for Treatment and Control Groups

Weight-for-age z-scores for children under 18 months excluding observations with recorded weight above the 90th percentile in the growth chart reported in Cortinovis et al. (1997). Sample size is 1,135 children. Solid line depicts the distribution for the treatment group and dashed line the distribution for the control group. Vertical solid line denotes mean in treatment group; dashed line denotes mean in control group.

(severe, \(< -3\) z-scores; moderate, \(-3 \leq z\)-scores \(< -2\); mild, \(-2 \leq z\)-scores \(< -1\)) in the control group was calculated. Applying the shift in the weight-for-age distribution (adding 0.14 z-score) with the odds ratio for each category—children who are mildly (moderately) [severely] underweight have about a twofold (fivefold) [eightfold] higher risk of death from infectious disease (Jones et al. 2003)—the reduction in average risk of mortality is estimated to be approximately 7 percent.17

V.F. Getting Inside the Box and Robustness Tests

The findings of large treatment effects on our proxies of community-based monitoring and outcomes are consistent with the community-based monitoring mechanism, but the findings do

17. To put this into perspective, a review of controlled trials designed to improve the intake of complementary food for children ages six months to five years showed a mean increase of 0.35 z-score (Jones et al. 2003). Jones and colleagues argue that this is one of the most effective preventive interventions feasible for delivery at high coverage in a low-income setting.
not rule out other explanations. In this section we assess a number of these alternative hypotheses.

To examine the plausibility of community-based monitoring as a key mechanism for the health utilization and health outcomes treatment effects, we follow the methodology used by Kling, Liebman, and Katz (2007). Specifically, we test whether the differences between treatment and control in outcomes across districts are larger in districts with large treatment-control differences in monitoring and information outcomes. This relationship is summarized by the parameter \( \delta \), the coefficient on the summary index of monitoring and information, in the outcome equation

\[
y_j = \delta M_j + X_j \pi + \epsilon_j.
\]

The summary index of monitoring \( M \) in (4) is the first component from a principal components analysis of the six monitoring and information variables in Table II. We examine two outcome measures \( y_j \), under-5 mortality and number of outpatients.

Following Kling, Liebman, and Katz (2007), we estimate (4) by two-stage least squares (2SLS), using a full set of district-by-treatment interactions as the excluded instruments for the monitoring index \( M \), while controlling for district fixed effects. The IV estimation of (4) will be consistent if \( M \) is the mediating factor between treatment and outcomes.

The IV approach is depicted graphically in Figure III. There is a consistent pattern across districts and groups that larger differences in monitoring (relative to the district mean) are associated with larger differences in outcomes—a result in line with the community-based monitoring mechanism.

Estimates based on equation (4) are given in Table VII. The first two columns show 2SLS estimates of \( \delta \) with district-by-treatment interactions as excluded instruments for the the monitoring index \( M \). To increase precision, we control for baseline outcomes \( y_{jt-1} \), when data allow it (i.e., for number of outpatients treated). The estimates are large in absolute terms and precisely estimated.

18. If \( X \) contains only district indicators, the 2SLS estimate of \( \delta \) using the district-by-treatment interactions instruments is the slope of the line fit through a scatterplot of the outcome and monitoring index means for the treatment and control groups in each of the nine districts, normalized so that each district has mean 0 (Kling, Liebman, and Katz 2007). We plot the average values by group (treatment and control) for each district for \( y \) and \( M \) expressed in standard deviation units relative to the control group overall standard deviation for each variable.
Figures III

Differences in Treatment-Control in Outcomes and Monitoring across Districts

Partial regression plots. The community monitoring index, outpatients, and under-5 mortality rate in the community (all three variables are described in the main text) are expressed in standard deviation units relative to the control group overall standard deviation for each variable. The points are the average values by group (treatment and control) for each district, normalized so that each district has mean 0. The line passes through the origin with the slope from the 2SLS estimation of equation (4) of the outcome on community monitoring and district indicators, using district-by-treatment interactions as instrumental variables. T (C) denotes treatment (control) group.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Community monitoring index</td>
<td>0.77***</td>
<td>-0.43*</td>
<td>0.86*</td>
<td>-0.43</td>
<td>0.77**</td>
<td>-0.54*</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.22)</td>
<td>(0.25)</td>
<td>(0.53)</td>
<td>(0.82)</td>
<td>(0.21)</td>
<td>(0.30)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Staff’s knowledge about</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>patients’ rights</td>
<td>-0.01</td>
<td>0.47</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.28)</td>
<td>(0.29)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Program impact</td>
<td>-0.12</td>
<td>0.01</td>
<td></td>
<td></td>
<td>190.5**</td>
<td>-41.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.66)</td>
<td>(0.88)</td>
<td></td>
<td></td>
<td>(92.6)</td>
<td>(45.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CBO presence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-8.3</td>
<td>-21.0</td>
<td></td>
<td></td>
<td>(69.4)</td>
<td>(37.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Program impact × CBO presence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-127.9</td>
<td>-4.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(126.1)</td>
<td>(58.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-test on program impact</td>
<td>6.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-test on CBO presence</td>
<td>0.37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-test on program impact × CBO presence</td>
<td>1.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes. Columns (1)–(4) report 2SLS estimates from equation (4) with district-by-treatment interactions as the excluded instruments and district fixed effects and outpatients \(_t−1\) (in specifications (1) and (3)) as controls. The variables in columns (1)–(4) are expressed in standard deviation units relative to the control group overall standard deviation for each variable. Robust standard errors are in parentheses. Program impact measures the coefficient on the assignment to treatment indicator. F-test statistics (with \(p\)-values in parentheses) on the excluded instruments Community monitoring and Staff’s knowledge about patients’ rights are 15.9 (0.00) and 7.23 (0.00), respectively. Point estimates and standard errors in columns (5)–(6) and columns (7)–(8), respectively, are jointly estimated from equation (3). Explanatory variables: Community monitoring is the first component from a principal components analysis of the six monitoring and information proxies presented in Table II. Staff’s knowledge about patients’ rights is a measure of the in-charge’s knowledge about patients’ rights and obligations (see text for details). CBO presence is an indicator variable for whether a participating CBO had been operating in the community before the intervention. F-test on program impact (CBO presence) | Program impact × CBO presence is the test statistic, with \(p\)-values in parentheses, on the test that the coefficients on program impact (CBO presence) | Program impact × CBO presence are jointly 0 in columns (5)–(6) and (7)–(8), respectively.

*Significant at 10% level.
**Significant at 5% level.
***Significant at 1% level.
A stricter test of whether the extent of the program impact varies with the size of the community monitoring impact is to add a treatment dummy (an overall treatment effect regardless of the community monitoring impact) to the IV regressions in equation (4). The community monitoring index is then identified by cross-district variation in changes in community monitoring by treatment from the district-by-treatment interactions as the excluded instruments, with the main effect for treatment no longer excluded; the results are reported in columns (3) and (4) of Table VII. Comparing the results without and with controls for treatment is quite similar for both outpatients and under-5 mortality, while the coefficients on the treatment indicator have the wrong sign and are small relative to their standard errors, providing some evidence that community monitoring had the primary effects on outcomes as opposed to other effects induced by the intervention.

To examine the hypothesis that differences in monitoring are driving the results as opposed to the supply-driven hypothesis that health workers, once being informed that their effort deviates from what is expected (in the health facility staff meeting), decided to exert greater effort in serving the community, we augment specification (4) with a measure of the staff's knowledge about patients' rights and obligations. \(^{19}\) This model thus has two endogenous variables. If large treatment effects on outcomes across districts are associated with differences in staff knowledge about patients' rights rather than more intense community monitoring, this would be evidence against the community-based monitoring hypothesis. As reported in columns (5) and (6), the coefficients on community monitoring remain largely unaffected, and the coefficients on staff knowledge are insignificant and with the wrong signs, providing additional evidence, albeit not conclusive, that the demand-driven mechanism is more important than the supply-driven mechanism.

The CBOs played an integral role in the intervention as facilitators of the meetings. However, it is possible that these CBOs had a role (as educators or activists, for example) beyond the described treatment itself. There is no definitive way to sort out the

\(^{19}\) The in-charge was asked to list patients' rights and obligations according to the Ministry of Health's plan for basic health service delivery. Patients' rights were discussed in the interface meeting. Each correct answer (out of five) was given a score of 0.2, and so this test score ranges from 0 to 1. We also examined other measures of staff engagement, including number of staff meetings in 2005 and if the in-charge had initiated training of staff on proper conduct. The results using these alternative proxies mirror those reported in Table VII.
role of community-based monitoring from the possible roles of the CBOs, but because around 60 percent of the CBOs that took part in the intervention had been operating in the communities before the intervention, and several of them also had activities in the control areas, we can investigate whether the outcomes are correlated with preintervention CBO activity. This would be the case if the CBOs that participated in the experiment, and that had been present in the communities prior to intervention, had a direct impact on health outcomes (through various preventive activities, for example) or indirectly by being more involved in monitoring the provider. The number of outpatients treated per month, shown in column (7), and the under-5 mortality rate, shown in column (8), are not significantly different in communities where the CBOs had been active prior to the intervention. We have also examined whether the treatment effect varies conditional on observable CBO characteristics or actions. For example, CBOs that are located (have an office) in the community might, everything else equal, be in a better position to monitor the health provider. Moreover, at ten of the treatment sites, the CBOs reported that they regularly visited the clinic. If the CBOs, rather than the community, were pushing the service providers into action, presumably the effect would be more pronounced at sites where the CBO actually visited the clinic regularly. However, the treatment effects are independent of whether the office of the CBO is located within a five-kilometer radius of the health facility or if the CBO reported that it regularly visited the clinic.

Given that within each district there are both treatment and control units, one concern with the evaluation design is the possibility of spillovers from one catchment area to another. In practice, there are reasons to believe spillovers will not be a serious concern. The average (and median) distance between the treatment and control facility is thirty kilometers, and in a rural setting, it is unclear to what extent information about improvements in

20. Given the small sample size, we test whether the distribution of outcomes in the subsample \( T = 1 & CBO \text{ located in community } = 1 \) is the same as in the subsample \( T = 1 & CBO \text{ located in community } = 0 \), and whether the distribution of outcomes in the subsample \( T = 1 & CBO \text{ regularly carries out monitoring visits to the facility } = 1 \) is the same as in the subsample \( T = 1 & CBO \text{ regularly carries out monitoring visits to the facility } = 0 \), using the Wilcoxon rank-sum test. The test statistics (with p-values in parentheses) are 0.88 (.38) and –1.10 (.27) for outpatients and 0.31 (.76) and –0.03 (.98) for under-5 mortality rate. We get similar results if we enrich equation (1) with an interaction term \( T \times CBO \text{ characteristic} \). The estimates of the interaction term are not statistically different from 0 in any of the specifications.
treatment facilities has spread to control communities. Still, the possibility of spillovers is a concern. Following Miguel and Kremer (2004), and taking advantage of the variation in distance to the nearest treatment clinic induced by randomization, we estimate spillovers from treatment to control groups by enriching \( \mathbf{X} \) in equation (1) to include an indicator variable for whether the control clinic is within ten kilometers of the nearest treatment clinic. The results are presented in the Online Supplemental Appendix (for utilization, delivery, and child death). We do not find evidence in favor of the spillover hypothesis.

Another concern is if the district or subdistrict management changed its behavior or support in response to the intervention. For example, the health subdistrict or local government may have provided additional funding or other support to the treatment facilities. The results in Table A.VIII in the Online Supplemental Appendix do not provide any evidence of this being the case. The treatment facilities did not receive more drugs or funding from the subdistrict or district as compared to the control facilities during 2005.

Upper-level authorities could also have increased their supervision of treatment facilities in response to the intervention. As shown in Online Supplemental Appendix Table A.IX, however, supervision of providers by upper-level government authorities remained low in both the treatment and the control group. As a complement we also assessed sanctions. Only a handful of staff were dismissed or transferred in 2005 and there is no systematic pattern that distinguishes treatment from control facilities. There is also no difference between treatment and control facilities in the number of staff that voluntarily left the facility during 2005 (Table A.IX).

VI. DISCUSSION

Based on a small but rigorous empirical literature on community participation and oversight, and extensive piloting in the field, our conjecture was that lack of relevant information and failure to agree on, or coordinate expectations of, what is reasonable to demand from the provider were holding back individual and group action to pressure and monitor the provider. We designed an intervention aimed at relaxing these constraints. Through two rounds of community meetings, local NGOs initiated a process
aimed at energizing the community and agreeing on actions to improve service provision.

We document large increases in utilization and improved health outcomes that compare favorably to some of the more successful community-based intervention trials reported in the medical literature. However, whereas medical field trials address the question of impact of a biological agent or treatment practice when the health workers do what they are supposed to do, we focus on a mechanism to ensure that health workers exert effort to serve the community.

The project was implemented in nine districts in Uganda with an estimated catchment population of approximately 55,000 households. In this dimension, therefore, the project has already shown that it can be brought to scale. However, the literature on how to enhance local accountability and participation is still in its infancy. And although the results in the paper suggest that community monitoring can play an important role in improving service delivery when traditional top-down supervision is ineffective, there are still a number of outstanding questions. For example, we know little about long-term effects and cross-sector externalities. It may also be the case that combining bottom-up monitoring with a reformed top-down approach could yield even better results. Before scaling up, it is also important to subject the project to a cost-benefit analysis. This would require putting a value on the improvements we have documented. To provide a flavor of such a cost-benefit analysis, consider the findings on averting the death of a child under five. A back-of-the-envelope calculation suggests that the intervention, including the cost for collecting data for the report cards (the main cost item), at $3 per household in the catchment areas or $160,000 in total, only judged on the cost per death averted, must be considered to be fairly cost-effective. The estimated cost of averting the death of a child under five is around $300, which should be compared to the estimate that the average cost per child life saved through the combined and integrated delivery of 23 interventions shown to reduce mortality from the major causes of death in children younger than 5 years is $887 (Bryce et al. 2003).

As argued in a recent *Lancet* article, a systematic program of research to answer questions about how best to deliver health (child survival) interventions is urgently needed (Bryce et al. 2003). In this paper we have focused on a mechanism that has been highlighted, but not examined, in the literature—a mechanism of
accountability enabling (poor) people to scrutinize whether those in authority have fulfilled their health responsibilities. Future research should address long-term effects, identify which mechanisms or combination of mechanisms that are important, and study the extent to which the results generalize to other social sectors.

IGIER, BOCONI UNIVERSITY, AND CEPR
INSTITUTE FOR INTERNATIONAL ECONOMIC STUDIES, STOCKHOLM UNIVERSITY, NHH, AND CEPR

REFERENCES


Going to scale with community-based primary care: An analysis of the family health program and infant mortality in Brazil, 1999–2004

James Macinko\textsuperscript{a,}\textsuperscript{*}, Maria de Fátima Marinho de Souza\textsuperscript{b}, Frederico C. Guanais\textsuperscript{c}, Celso Cardoso da Silva Simões\textsuperscript{d}

\textsuperscript{a}University of Pennsylvania, RWJF Health and Society Scholars program, 3641 Locust Walk, Philadelphia, PA 19104, USA
\textsuperscript{b}Secretariat of Health Surveillance, Ministry of Health, Brasília, Brazil
\textsuperscript{c}Ministry of Social Development and the Fight against Hunger, Brasília, Brazil
\textsuperscript{d}Institute of Geography and Statistics (IBGE), Rio de Janeiro, Brazil

Available online 8 August 2007

Abstract

This article assesses the effects of an integrated community-based primary care program (Brazil’s Family Health Program, known as the PSF) on microregional variations in infant mortality (IMR), neonatal mortality, and post-neonatal mortality rates from 1999 to 2004. The study utilized a pooled cross-sectional ecological analysis using panel data from Brazilian microregions, and controlled for measures of physicians and hospital beds per 1000 population, Hepatitis B coverage, the proportion of women without prenatal care and with no formal education, low birth weight births, population size, and poverty rates. The data covered all the 557 Brazilian microregions over a 6-year period (1999–2004).

Results show that IMR declined about 13 percent from 1999 to 2004, while Family Health Program coverage increased from an average of about 14 to nearly 60 percent. Controlling for other health determinants, a 10 percent increase in Family Health Program coverage was associated with a 0.45 percent decrease in IMR, a 0.6 percent decline in post-neonatal mortality, and a 1 percent decline in diarrhea mortality ($p < 0.05$). PSF program coverage was not associated with neonatal mortality rates. Lessons learned from the Brazilian experience may be helpful as other countries consider adopting community-based primary care approaches.

\textsuperscript{*}Corresponding author. Tel.: +1 2129985592.
E-mail addresses: jmj5@nyu.edu (J. Macinko), mfmsouza@usp.br (M. de Fátima Marinho de Souza).

Introduction

There is renewed worldwide interest in primary health care and its potential for improving human health (Pan American Health Organization, 2007).

But there have been few peer-reviewed studies that assess the effectiveness of national primary health care strategies on improving population health in the developing world. This study attempts to fill this gap by analyzing the effects of the Brazilian Family Health Program (Programa Saúde da Família or PSF in Portuguese) on child health. Child health outcomes are particularly important to examine because of the higher-than-expected rates of infant mortality.
and child mortality in Brazil as compared to other countries of similar gross national income, the PSF’s emphasis on improving infant and child health, and the possibility of observing rapid changes in these outcomes over a relatively short period of time.

The PSF is the main approach to provide primary care services within Brazil’s national health system, known as the Sistema Único de Saúde or SUS (Almeida & Pêgo, 2002). The PSF has its roots in the community health agents program begun in the state of Ceará in the early 1990s (Cufino Svitone, Garfield, Vasconcelos, & Araujo Craveiro, 2000). Since it was adopted as a national strategy in 1994, the program had grown by 2007 to encompass 26,730 community-based teams responsible for providing care to about 85 million people, making it one of the world’s largest systems of community-based primary care (Brazilian Ministry of Health Department of Primary Care, 2007).

The PSF is a decentralized approach to providing core primary care functions, including first-contact access for each new health need, comprehensive and person-focused care over the lifecourse, coordination of care between different providers and types of health services, and family and community-oriented health promotion activities (Ministry of Health of Brazil, 2003). These functions are achieved through the program’s organization (municipalities manage the program with national supervision and each PSF team is assigned to a geographical area with responsibility for enrolling and monitoring the health status of about 3500 people), its financing (services are delivered free of charge, are financed on a capitation basis, and municipalities have incentives for increasing the number of neighborhoods with access to the program), and delivery mechanisms (multidisciplinary teams are composed of, at minimum, a physician and nurse who deliver clinic-based care and most teams include community health workers who make regular home visits and perform community-based health promotion activities) (Ministry of Health of Brazil, 2003). Family Health Program teams in many areas also include dental and social work professionals.

Despite the ambitious scope of this undertaking there have been only a few evaluations of the program (Conill, 2002; Escorel et al., 2002; Ministério da Saúde, 2004; Serra, 2005; Viana & Pierantoni, 2002) although several more are underway. To date, only one peer-reviewed article has assessed the relationship between PSF coverage and changes in health outcomes at the national level (Macinko, Guanais, & Marinho de Souza, 2006). The present article expands and strengthens earlier work by employing local-level analyses, examining several different outcomes, and by assessing the effects of the rapid expansions in PSF coverage over the past few years.

Methods

This study follows a quasi-experimental design since each municipality in Brazil adopted the PSF at different times and coverage in each municipality grew at different rates. To take advantage of this heterogeneity, we use a pooled, cross-sectional, time series approach to assess the relationships between dependent and independent variables over a 6-year period. This technique pools together 6 years (1999–2004) of cross-sections (composed of all 557 Brazilian microregions for each year) for a maximum sample size of 3342 observations. The approach provides an estimate of the health effects of program expansion by testing the association between differences in coverage in each microregion with differences in infant mortality outcomes, while controlling for potential confounders (Hsiao, 2003).

In order to strengthen the study design, we analyze two types of outcomes. Based on previous literature, we hypothesize that the PSF will have a strong association with outcomes most sensitive to primary care: post-neonatal mortality (deaths of children from 30 days to 1 year per 1000 live births) and deaths from diarrheal diseases (deaths of children under 1 year from diarrhea per 1000 live births) (Caldeira, França, & Goulart, 2001; Caldeira, França, Perpetuo, & Goulart, 2005). It should have a modest impact on IMR (all deaths of children under 1 year per 1000 live births in the same year) that will depend on the proportion of IMR that is composed of post-neonatal mortality (Moore, Castillo, Richardson, & Reid, 2003). We hypothesize that there should be little or no relation between PSF coverage and neonatal mortality rates (deaths of children within their first month of life per 1000 live births), since these outcomes are most sensitive to care provided primarily by specialist and hospital services outside the scope of the PSF (Lansky, França, & Leal Md, 2002).

The unit of analysis is the microregion. Each of the 557 microregions contains several of Brazil’s 5564 municipalities that have been grouped together to be geographically contiguous and homogeneous.
in terms of demography, agriculture, and transportation. Microregions represent smaller units of analysis and thus capture greater variation than would analysis of the 27 Brazilian states. Microregions also have a larger population size than individual municipalities, thus allowing for more stable mortality estimates over time.

Data on PSF coverage, health resources, and outcomes are from the Brazilian Ministry of Health (Ministry of Health of Brazil, 2007). In this study, we use official estimates of IMR that have been adjusted for underreporting of child deaths (Rede Interagencial de informações para a saúde (RIPSA), 2002; Szwarcwald, Leal Medo, de Andrade, & Souza, 2002). All other outcomes (neonatal, post-neonatal, and diarrhea mortality rates) are constructed directly from observed counts.

Independent variables known to influence infant mortality include poverty (proportion of the population in the lowest income quintile), women’s health and development (proportion of women over 15 with no formal schooling, and proportion of women with no prenatal care), child health (proportion of children with Hepatitis B immunizations, low birth weight defined as percent of births under 2500 g), and health services (physicians and hospital beds per 1000) (Moore et al., 2003; Wang, 2003). Data on these variables are based on population surveys conducted by the Brazilian Institute of Geography and Statistics (IBGE) and developed for state-level representativity by the Institute of Applied Economic Research (IPEA) (Brazilian Institute of Geography and Statistics, 2005; Instituto de Pesquisa Econômica Aplicada (IPEA), May 2005).

Some independent variable data were missing for some years. Missing data were imputed using non-linear interpolation methods that modeled within-municipal changes as a function of prior values at the municipal level and contemporaneous values at the state level (Allison, 2002; Guainais, 2006). All values were then summed up to the microregional level.

Statistical analyses

The study uses a fixed-effects specification in order to correct for serial correlation of repeated measures and to control for time-invariant unobserved or unobservable microregional characteristics. An alternative approach, the random effects model, was rejected due to results of the Hausman test \( p < .0001 \) that tested correlation between the regressors and error terms. (Wooldridge, 2002) All analyses were conducted using Stata 9 software and use robust standard errors to correct for heteroskedasticity (Statacorp, 2005).

Advantages of the fixed-effects model over cross-sectional analyses include the fact that it is able to establish temporal ordering between exposures and outcomes and it can control for unmeasured time-invariant characteristics of the microregion (such as geography, historical disadvantages, urban/rural location, and local cultural practices) that might influence health outcomes (Hsiao, 2003). One disadvantage of the fixed-effects approach is that the results obtained are conditional on the data used to estimate them; that is, results cannot be generalized to other years or microregions not included in the study (Hsiao, 2003).

In order to compare how variables changed over time, we calculate the mean values and standard deviations for 1999 and 2004 and the percent change during this time. Differences in mean values between time periods were assessed using \( t \)-tests. Regression analyses are presented as a series of nested models. The \( F \)-test is used to assess whether the inclusion of an additional set of independent variables improved regression models. In order to compare the magnitude of the effects of the main explanatory variables on the outcomes, we calculated their marginal effects. This statistic represents the percent change in the outcome given a one-percent change in the independent variable, when all other values are set at their mean (Greene, 2003).

We also assessed several pathways by which the PSF might influence IMR. Primary care access is associated with lower post-neonatal mortality and fewer deaths from diarrhea (UNICEF, 2002). In order to test potential mechanisms of the health effects of PSF expansion we developed a set of dummy variables representing microregions in the highest 75th percentile of under-five deaths from both of these conditions (called “high diarrhea deaths” and “high postneonatal deaths,” respectively). We then created interaction terms between these binary variables and PSF coverage to test if the PSF effect was higher in those microregions where a greater share of infant and child mortality was amenable to primary care. Other interactions of the PSF term (with physicians per 1000 population and Hepatitis B coverage) were not significant and therefore not included in the final models.

Because there are great differences in health and economic development between the poorer north
Results

Table 1 presents descriptive statistics. Between 1999 and 2004 some measures of infant mortality declined: IMR was reduced by 13 percent, post-neonatal mortality by 16 percent and diarrhea-specific mortality by 44 percent. However, neonatal mortality increased by 5 percent and the percentage of births that were low birth weight increased 10 percent. By 2004, the PSF covered about 60 percent of the population in the microregions, ranging from a low of 6 percent to over 100 percent for the top 90th percentile. Access to some forms of healthcare appeared to increase: Physician availability increased by 87 percent, Hepatitis B coverage increased by 20 percent, and access to prenatal care increased by 50 percent. Hospital beds per 1000 declined slightly. Average population size for microregions increased by nearly a quarter and most of this increase occurred in large metropolitan areas. The proportion of the population in the lowest income quintile increased slightly, while the proportion of mothers with no education declined by nearly one-third from 1999 levels.

Table 2 presents the results of the fixed effects analyses. Model 1 shows the bivariate relationship between PSF and IMR: the larger the proportion of the state’s population served by the PSF, the lower the expected infant mortality rate. Model 2 adds health system covariates to model 1. PSF coverage remains significant and negatively associated with IMR. In terms of covariates, physician supply and Hepatitis B coverage were negatively associated with IMR, while hospital beds were positively associated with it. The F-test is statistically significant, suggesting that addition of these covariates improves the explanatory power of model 2 over model 1.

Model 3 adds a set of social and economic variables. Population size was negatively associated with IMR, suggesting that IMR is lower in microregions with larger populations. Both the proportion of women with no formal education and the proportion of the population in the lowest income quintile were positively associated with IMR. The PSF coefficient remains significant and negative (although slightly reduced in magnitude),
and socioeconomic variables remain stable. Based on the results of the F-test, Model 3 is considered superior to the previous models.

Model 4 includes additional maternal and child health indicators. The proportion of women with no prenatal care is positively associated with IMR while the percentage of births that are low weight is negatively associated with IMR. The PSF variable is lightly reduced in magnitude, but remains similar in direction and statistical significance. Results of the F-test indicate that Model 4 is superior to any previous models. The $R^2$ value suggests that the model explains up to 73 percent of the within-microregion variation in IMR from 1999 to 2004.

Model 5 further explores the relationship between PSF and IMR by including interaction terms between PSF coverage and microregions with high proportionate mortality from diarrhea and high total mortality from diarrhea and post-neonatal mortality.

### Table 2
Fixed effects regression models of infant mortality rates* for the microregions of Brazil, 1999–2004

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family health program (% population covered)</td>
<td>$-0.057^{**}$</td>
<td>$-0.042^{**}$</td>
<td>$-0.038^{**}$</td>
<td>$-0.030^{**}$</td>
<td>$-0.028^{**}$</td>
<td>$-$</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.003)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>$-$</td>
</tr>
<tr>
<td>Family health program (Coverage quartile 2)</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-0.715^{**}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.066)</td>
</tr>
<tr>
<td>Family health program (Coverage quartile 3)</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-1.346^{**}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.105)</td>
</tr>
<tr>
<td>Family health program (Coverage quartile 4)</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-2.188^{**}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.147)</td>
</tr>
<tr>
<td>Physicians (per 1000 population)</td>
<td>$-0.385^{**}$</td>
<td>$-0.333^{**}$</td>
<td>$-0.296^{**}$</td>
<td>$-0.299^{**}$</td>
<td>$-0.353^{**}$</td>
<td>$-$</td>
</tr>
<tr>
<td></td>
<td>(0.139)</td>
<td>(0.128)</td>
<td>(0.113)</td>
<td>(0.112)</td>
<td>(0.117)</td>
<td></td>
</tr>
<tr>
<td>Hospital beds (per 1000 population)</td>
<td>$0.478^{**}$</td>
<td>$0.343^{**}$</td>
<td>$0.350^{**}$</td>
<td>$0.366^{**}$</td>
<td>$0.365^{**}$</td>
<td>$-$</td>
</tr>
<tr>
<td></td>
<td>(0.063)</td>
<td>(0.056)</td>
<td>(0.055)</td>
<td>(0.055)</td>
<td>(0.056)</td>
<td></td>
</tr>
<tr>
<td>Hepatitis B vaccination (% children covered)</td>
<td>$-0.013^{**}$</td>
<td>$-0.014^{**}$</td>
<td>$-0.012^{**}$</td>
<td>$-0.011^{**}$</td>
<td>$-0.012^{**}$</td>
<td>$-$</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td></td>
</tr>
<tr>
<td>Population (1000s)</td>
<td>$-$</td>
<td>$-0.012^{**}$</td>
<td>$-0.009^{**}$</td>
<td>$-0.009^{**}$</td>
<td>$-0.009^{**}$</td>
<td>$-$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td></td>
</tr>
<tr>
<td>Population in poverty (% in lowest income quintile)</td>
<td>$0.035^{**}$</td>
<td>$0.027^{**}$</td>
<td>$0.025^{**}$</td>
<td>$0.025^{**}$</td>
<td>$0.025^{**}$</td>
<td>$-$</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.007)</td>
<td>(0.007)</td>
<td>(0.007)</td>
<td>(0.007)</td>
<td></td>
</tr>
<tr>
<td>No formal education (% of all mothers)</td>
<td>$0.158^{**}$</td>
<td>$0.098^{**}$</td>
<td>$0.097^{**}$</td>
<td>$0.104^{**}$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.015)</td>
<td>(0.014)</td>
<td>(0.015)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No prenatal care (% of all mothers)</td>
<td>$-$</td>
<td>$-$</td>
<td>$0.160^{**}$</td>
<td>$0.149^{**}$</td>
<td>$0.160^{**}$</td>
<td>$-$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.016)</td>
<td>(0.015)</td>
<td>(0.015)</td>
<td></td>
</tr>
<tr>
<td>LBW births (% of all births)</td>
<td>$-$</td>
<td>$-$</td>
<td>$-0.214^{**}$</td>
<td>$-0.217^{**}$</td>
<td>$-0.213^{**}$</td>
<td>$-$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.028)</td>
<td>(0.028)</td>
<td>(0.028)</td>
<td></td>
</tr>
<tr>
<td>High diarrhea deaths (75th percentile)</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$0.452^{**}$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.096)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interaction: (PSF* high diarrhea deaths)</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-0.007^{**}$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.002)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High post-neonatal mortality (75th percentile)</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$0.333^{**}$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.092)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interaction: PSF* high post-neonatal mortality</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-0.003^{**}$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.001)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>$30.200^{**}$</td>
<td>$30.340^{**}$</td>
<td>$32.530^{**}$</td>
<td>$32.531^{**}$</td>
<td>$32.277^{**}$</td>
<td>$32.059^{**}$</td>
</tr>
<tr>
<td></td>
<td>(0.059)</td>
<td>(0.393)</td>
<td>(0.574)</td>
<td>(0.516)</td>
<td>(0.512)</td>
<td>(0.526)</td>
</tr>
<tr>
<td>Observations</td>
<td>3342</td>
<td>3337</td>
<td>3337</td>
<td>3337</td>
<td>3337</td>
<td>3337</td>
</tr>
<tr>
<td>Number of microregions</td>
<td>557</td>
<td>557</td>
<td>557</td>
<td>557</td>
<td>557</td>
<td>557</td>
</tr>
<tr>
<td>$R^2$-squared (within)</td>
<td>0.567</td>
<td>0.625</td>
<td>0.688</td>
<td>0.733</td>
<td>0.725</td>
<td>0.74</td>
</tr>
<tr>
<td>$F$-test (model 2 v 1)</td>
<td>$71.49^{**}$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td>$F$-test (model 3 v 2)</td>
<td>$-$</td>
<td>$55.52^{**}$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td>$F$-test (model 4 v 3)</td>
<td>$-$</td>
<td>$-$</td>
<td>$78.25^{**}$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses. Microregion fixed effects not shown.

*Significant at the 0.05 level.
**Significant at the 0.01 level.

*Infant mortality rate expressed as per 1000 live births and adjusted for underreporting of infant deaths in some municipalities.
post-neonatal mortality. The coefficients for high diarrhea mortality and for high post-neonatal mortality are positive and significant, suggesting that IMR is higher in those microregions with very high levels of diarrhea-related and post-neonatal deaths. The interaction variable for PSF*diarrhea and PSF*post-neonatal deaths is significant and negative, suggesting that increases in PSF coverage have a particularly strong impact on lowering IMR by reducing diarrhea and post-neonatal deaths in areas where these rates are high.

Model 6 tests a transformation of the PSF variables to reflect quartiles of coverage. The results show that as PSF coverage increases, the magnitude of the regression coefficient likewise increases, suggesting a dose–response relationship.

Table 3 presents results for neonatal, post-neonatal, and diarrhea mortality rates. All covariates are the same as in the full model (Model 4 from Table 2). Family health program coverage was not associated with neonatal mortality, although it was negatively associated with both post-neonatal and diarrhea mortality rates.

Table 4 presents analyses stratified by geographic region. The main finding is that the PSF has a consistently significant negative association with IMR in each region. Covariates are generally similar to the full sample analysis, although in the regional analyses poverty is significant only for the north region, physicians are not significant in the southeast, and low birth weight is not significant for the south.

Table 5 presents the marginal effects of the main explanatory variables included in the final model (Model 4 in Table 2). Marginal effects have been multiplied by 10 to give a measure of the percent change in infant mortality associated with a 10 percent increase in the independent variable. Controlling for all other covariates, a ten percent increase in PSF coverage was associated, on average, with a 0.45 percent decrease in IMR, a 0.6 percent decrease in post-neonatal mortality, and a 1 percent decrease in diarrhea-related mortality. The largest contributor to reductions in all outcomes was the size of the microregion’s population, suggesting an important urban advantage. For mortality from diarrhea, a ten percent increase in Hepatitis B coverage was associated with a 3.7 percent decline. Most other covariates had marginal effects near or less than that of PSF coverage.

Discussion

The analyses presented here suggest that PSF coverage is independently associated with better primary care-sensitive child health outcomes, including IMR, post-neonatal mortality, and deaths from diarrhea. As hypothesized, PSF coverage was not associated with neonatal mortality, which is strongly influenced by the availability and quality of care during and post-delivery, special care for low birth weight babies, and some aspects of prenatal care (Martines et al., 2005).

### Table 3

<table>
<thead>
<tr>
<th>Variable</th>
<th>Neonatal mortality rate</th>
<th>Post-neonatal mortality rate</th>
<th>Diarrhea mortality rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family health program (% covered)</td>
<td>$-0.004$ (0.003)</td>
<td>$-0.022**$ (0.006)</td>
<td>$-0.012*$ (0.006)</td>
</tr>
<tr>
<td>Physicians (per 1000 population)</td>
<td>$-0.008$ (0.069)</td>
<td>$-0.264$ (0.14)</td>
<td>$-0.216$ (0.182)</td>
</tr>
<tr>
<td>Hospital beds (per 1000 population)</td>
<td>$0.084$ (0.103)</td>
<td>$-0.297$ (0.225)</td>
<td>$-0.176$ (0.217)</td>
</tr>
<tr>
<td>Hepatitis B coverage (% of children covered)</td>
<td>$0.016**$ (0.002)</td>
<td>$0.006$ (0.005)</td>
<td>$0.020**$ (0.006)</td>
</tr>
<tr>
<td>Population (1000s)</td>
<td>$-0.003$ (0.002)</td>
<td>$-0.015**$ (0.003)</td>
<td>$-0.007**$ (0.002)</td>
</tr>
<tr>
<td>Population in poorest income quintile (%)</td>
<td>$0.011$ (0.007)</td>
<td>$0.035**$ (0.017)</td>
<td>$0.017$ (0.011)</td>
</tr>
<tr>
<td>Mothers with no formal education (%)</td>
<td>$-0.024$ (0.024)</td>
<td>$0.027$ (0.082)</td>
<td>$0.164**$ (0.056)</td>
</tr>
<tr>
<td>Mothers with no prenatal care (%)</td>
<td>$0.046$ (0.026)</td>
<td>$0.279**$ (0.066)</td>
<td>$-0.066$ (0.052)</td>
</tr>
<tr>
<td>LBW births (% of all births)</td>
<td>$0.192**$ (0.063)</td>
<td>$0.214$ (0.156)</td>
<td>$-0.104$ (0.124)</td>
</tr>
<tr>
<td>Constant</td>
<td>$2.919**$ (0.77)</td>
<td>$18.256**$ (1.775)</td>
<td>$9.969**$ (1.479)</td>
</tr>
<tr>
<td>Observations</td>
<td>3336</td>
<td>3336</td>
<td>3228</td>
</tr>
<tr>
<td>Number of microregions</td>
<td>556</td>
<td>556</td>
<td>538</td>
</tr>
<tr>
<td>R-squared (within)</td>
<td>0.335</td>
<td>0.545</td>
<td>0.407</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses; microregion fixed effects not shown.

* $p<0.05$; ** $p<0.01$.

*All rates expressed as per 1000 live births and are based observed counts that have not been adjusted for underreporting of infant deaths in some municipalities.
**Table 4**

Determinants of infant mortality rate by region, 1999–2004

<table>
<thead>
<tr>
<th>Variable</th>
<th>North</th>
<th>Northeast</th>
<th>Southeast</th>
<th>South</th>
<th>Central-west</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coverage of family health program (%)</td>
<td>-0.037** (0.006)</td>
<td>-0.023** (0.003)</td>
<td>-0.038** (0.003)</td>
<td>-0.014** (0.002)</td>
<td>-0.013** (0.002)</td>
</tr>
<tr>
<td>Physicians (per 1000 population)</td>
<td>-1.401** (0.232)</td>
<td>-1.140** (0.111)</td>
<td>-0.048 (0.055)</td>
<td>-0.603** (0.065)</td>
<td>-0.853** (0.091)</td>
</tr>
<tr>
<td>Hospital beds (per 1000 population)</td>
<td>0.338** (0.113)</td>
<td>0.222* (0.09)</td>
<td>0.371** (0.07)</td>
<td>0.433** (0.124)</td>
<td>-0.255** (0.059)</td>
</tr>
<tr>
<td>Hepatitis B immunization (% of children covered)</td>
<td>-0.016** (0.003)</td>
<td>-0.011** (0.002)</td>
<td>-0.006* (0.002)</td>
<td>0.003* (0.001)</td>
<td>-0.007* (0.003)</td>
</tr>
<tr>
<td>Population (1000s)</td>
<td>-0.006* (0.003)</td>
<td>-0.010** (0.003)</td>
<td>-0.005** (0.001)</td>
<td>-0.004 (0.004)</td>
<td>-0.011** (0.001)</td>
</tr>
<tr>
<td>Population in poorest income quintile (%)</td>
<td>0.046* (0.022)</td>
<td>-0.01 (0.036)</td>
<td>0.017 (0.009)</td>
<td>-0.01 (0.017)</td>
<td>-0.005 (0.005)</td>
</tr>
<tr>
<td>Mothers with no formal education (%)</td>
<td>0.060** (0.02)</td>
<td>0.147** (0.025)</td>
<td>0.062* (0.03)</td>
<td>0.071* (0.03)</td>
<td>0.088** (0.026)</td>
</tr>
<tr>
<td>Mothers with no prenatal care (%)</td>
<td>0.065** (0.023)</td>
<td>0.106** (0.021)</td>
<td>0.329** (0.04)</td>
<td>0.243** (0.048)</td>
<td>0.125** (0.043)</td>
</tr>
<tr>
<td>LBW births (% of all births)</td>
<td>-0.106* (0.052)</td>
<td>-0.194** (0.055)</td>
<td>-0.212** (0.033)</td>
<td>-0.066 (0.035)</td>
<td>-0.107** (0.029)</td>
</tr>
<tr>
<td>Constant</td>
<td>31.888** (0.670)</td>
<td>47.913** (0.845)</td>
<td>22.874** (0.614)</td>
<td>19.754** (1.098)</td>
<td>26.871** (0.544)</td>
</tr>
<tr>
<td>Observations</td>
<td>378</td>
<td>1122</td>
<td>957</td>
<td>564</td>
<td>312</td>
</tr>
<tr>
<td>Number of microregions</td>
<td>63</td>
<td>187</td>
<td>160</td>
<td>94</td>
<td>52</td>
</tr>
<tr>
<td>$R^2$-squared (within)</td>
<td>0.76</td>
<td>0.84</td>
<td>0.74</td>
<td>0.76</td>
<td>0.90</td>
</tr>
</tbody>
</table>

**Mean values for selected variables (1999–2004)**

<table>
<thead>
<tr>
<th>Variable</th>
<th>North</th>
<th>Northeast</th>
<th>Southeast</th>
<th>South</th>
<th>Central-west</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infant mortality rate (IMR)</td>
<td>27.12</td>
<td>42.37</td>
<td>19.18</td>
<td>17.81</td>
<td>21.47</td>
</tr>
<tr>
<td>PSF coverage (%)</td>
<td>30.82</td>
<td>53.73</td>
<td>32.30</td>
<td>34.44</td>
<td>47.44</td>
</tr>
<tr>
<td>Physicians</td>
<td>1.22</td>
<td>1.83</td>
<td>3.34</td>
<td>2.67</td>
<td>2.16</td>
</tr>
<tr>
<td>Hospital beds</td>
<td>2.00</td>
<td>2.15</td>
<td>3.29</td>
<td>3.17</td>
<td>3.56</td>
</tr>
<tr>
<td>Hepatitis B coverage</td>
<td>80.69</td>
<td>85.49</td>
<td>96.98</td>
<td>97.02</td>
<td>92.41</td>
</tr>
<tr>
<td>Mothers with no prenatal care</td>
<td>7.48</td>
<td>10.24</td>
<td>1.97</td>
<td>1.77</td>
<td>2.34</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses; microregion fixed effects not shown.

* $p < 0.05$; ** $p < 0.01$.

Infant mortality rate expressed as per 1000 live births and adjusted for underreporting of infant deaths in some municipalities.

**Table 5**

Marginal effects* by outcome, Brazilian microregions 1999–2004

<table>
<thead>
<tr>
<th>Variable</th>
<th>Infant mortality rate</th>
<th>Post-neonatal mortality rateb</th>
<th>Diarrhea mortalityb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coverage of family health program (%)</td>
<td>-0.447** (-0.506, -0.387)</td>
<td>-0.591** (-0.909, -0.273)</td>
<td>-1.034* (-2.030, -0.037)</td>
</tr>
<tr>
<td>Physicians per 1000 population</td>
<td>-0.251** (-0.439, -0.064)</td>
<td>-0.401* (-0.816, 0.015)</td>
<td>-1.088 (-2.880, 0.704)</td>
</tr>
<tr>
<td>Hospital beds per 1000 population</td>
<td>0.348** (0.242, 0.455)</td>
<td>-0.526 (-1.310, 0.257)</td>
<td>-1.038 (-3.536, 1.459)</td>
</tr>
<tr>
<td>Hepatitis B coverage (%)</td>
<td>-0.376** (-0.452, -0.300)</td>
<td>0.359 (-0.252, 0.971)</td>
<td>-3.770* (-5.951, -1.589)</td>
</tr>
<tr>
<td>Population (1000s)</td>
<td>-1.048* (-1.340, -0.755)</td>
<td>-2.873** (-4.154, -1.593)</td>
<td>-4.715* (-7.439, -1.991)</td>
</tr>
<tr>
<td>Population in poorest income quintile (%)</td>
<td>0.213** (0.102, 0.323)</td>
<td>0.497* (0.035, 0.958)</td>
<td>0.801 (-0.229, 1.831)</td>
</tr>
<tr>
<td>Mothers with no formal education (%)</td>
<td>0.190** (0.134, 0.245)</td>
<td>0.093 (-0.459, 0.644)</td>
<td>1.879** (-0.628, 3.130)</td>
</tr>
<tr>
<td>Mothers with no prenatal care (%)</td>
<td>0.252* (0.204, 0.300)</td>
<td>0.784** (0.420, 1.147)</td>
<td>-0.619 (-1.570, 0.332)</td>
</tr>
<tr>
<td>LBW births (%) (%)</td>
<td>-0.546* (-0.686, -0.406)</td>
<td>0.973 (-0.414, 2.360)</td>
<td>-1.572 (-5.245, 2.101)</td>
</tr>
</tbody>
</table>

Robust 95% confidence intervals in parentheses; microregion fixed effects not shown.

* $p < 0.05$; ** $p < 0.01$.

*Marginal effects represent percent change in the outcome associated with a 10 percent change in the independent variable. All marginal effects were calculated in terms of elasticities evaluated at the means of all other independent variables.

**Rates expressed as per 1000 live births and are based on observed counts that have not been adjusted for underreporting of infant deaths in some municipalities.

Our results are consistent with evidence of potential mechanisms through which the PSF might work to lower primary care-sensitive infant mortality. For example, higher PSF coverage has been found to be associated with higher population rates of breastfeeding, oral rehydration therapy, immunizations, and treatment of respiratory and other infections—interventions that address the leading...
causes of post-neonatal mortality (Emond, Pollock, Da Costa, Maranhão, & Macedo, 2002; Escorel et al., 2002; Shi et al., 2004; Starfield, 1985).

The magnitude of the PSF effect was significant, albeit of lesser magnitude than observed in previous studies. This is likely to be due to the fact that IMR has experienced a dramatic decline throughout Brazil as a function of a range of interventions, including PSF coverage, improved water and sanitation, and better women’s health and development (Macinko et al., 2006). Moreover, as noted above, as IMR declines a greater proportion of infant deaths tend to happen within the first month of life due to conditions that are less amenable to primary care. Neonatal mortality has been linked to increased preterm and low birth weight births and has become a more significant contributor to IMR in Brazil as post-neonatal mortality declined (Barros et al., 2005; Caldeira et al., 2001).

There were also important regional differences in the effects of PSF coverage. In the region-stratified analyses, the effect of the PSF program was reduced for the more developed southern regions where IMR has been lower relative to the north and northeast. The apparent protective effect of population size may represent either an urban advantage or the fact that since 1998, PSF expansion has focused on municipalities with populations greater than 100,000 people.

Physician supply was also associated with lower infant mortality: a finding that is consistent with other studies (Anand & Barnighausen, 2004). Sensitivity tests using nurses per 1000 instead of physicians found similar results, although both variables could not be included in the analyses due to their high correlation ($\rho = 0.74; p<0.001$). This result suggests that the PSF has made progress in expanding primary care physician supply in under-served regions (such as the northeast) (Ministério da Saúde, 2004). This argument is supported by the observation that the physician supply effect was significant in all regions except the southeast where there has historically been less of a physician deficit than in other regions and where most physicians are specialists (rather than family practitioners or other primary care providers).

Not surprisingly, measures such as poverty, female illiteracy, lack of prenatal care, and low levels of Hepatitis B immunization were all found to be associated with higher mortality. Although earlier studies found no relationship between immunization rates and IMR (Macinko et al., 2006), they used measures of all immunization schedules which are already over 90 percent in most states. Hepatitis B vaccination is a more recent initiative and coverage varies substantially between microregions, making it potentially a more sensitive indicator of primary care access.

Availability of hospital beds was positively associated with outcomes—a result that was not expected. One possible explanation is that in recent years hospitals may have experienced declines in accessibility, quality, or both. This hypothesis is partially supported by the results in Table 3 which show that hospital beds were not associated with neonatal mortality, the outcome that should be most highly correlated with indicators of hospital care. Lansky, França, and Kawachi (2007) suggest that there is considerable variation in hospital quality and this variation is associated with elevated perinatal mortality from potentially avoidable conditions such as intrapartum asphyxia. Potentially avoidable infant mortality was found to be especially high for normal birth weight babies born in government-contracted private hospitals in large urban areas, which were found to have lower quality care (Lansky et al., 2007).

Finally, low birth weight births were found to be negatively associated with IMR in this study. This “low birth weight paradox” has been observed elsewhere and may be explained by the fact that low birth weight infants from population groups in which LBW is most frequent often have a lower risk of death than low birth weight infants from the general population (Hernandez-Diaz, Schisterman, & Hernan, 2006). Our ecological analysis might be more prone to picking up this phenomenon than would an individual-level study. Removing LBW from the analyses does not significantly change any of our conclusions.

Limitations

This is an ecologic study, so it is not possible to test whether the reductions in IMR and other outcomes occurred within families that actually visited the Family Health Program. Ideally, we would conduct a multi-level analysis but there are currently no nationally representative data on individual PSF users and non-users. Nevertheless, there is evidence that improving PSF coverage leads to improvements in determinants of child health. For example, PSF clients regularly receive health education about breastfeeding, use of oral rehydration...
therapy, immunization, and infant growth monitoring (Emond et al., 2002; Escorel et al., 2002). In a study of several large urban centers, more than three-quarters of PSF clients interviewed believed that child health services were of good quality and that the PSF was responsible for improvements in the health of the neighborhood and their family (Escorel et al., 2002). There is also evidence to suggest that the PSF program decreases financial barriers to access (Goldbaum, Gianini, Novaes, & Cesar, 2005). Finally, other studies have confirmed that in areas where the PSF or similar programs have been implemented, infant mortality has actually declined (Cufino Svitone et al., 2000; Macinko et al., 2006; Serra, 2005).

Ecological analyses are vulnerable to omitted variable problems. That is, there could be some latent, unmeasured variable confounding the apparent relationship between PSF and IMR. In this case, the existence of such a variable is unlikely given that we employed a comprehensive model of health determinants, included fixed effects to control for time-invariant unobserved characteristics of microregions, and tested several pathways and alternative explanations. The high R-squared values of the main regression models suggest that they explain a large proportion of the variation in infant mortality.

Finally, conclusions about outcomes based on unadjusted rates (post-neonatal, neonatal, and diarrhea deaths) need to be interpreted with caution since there is evidence of undercounting of child mortality in Brazil. Note that this undercounting has improved in recent years, so each year’s data should be closer to real values. In this study, adjusted IMR values corresponded with observed IMR rates 85 percent of the time with an average difference of 4.7 deaths/1000 live births. Most of this variation was in the Northeast region of the country (60 percent agreement in the northeast, 84 percent agreement in the north, 90 percent agreement in the central-west, 99 percent agreement in the south and southeast). In sensitivity tests that excluded the 982 (out of 3337) data points with outcome data that was one or more standard deviation above or below the adjusted IMR rates for any year, there was no change in the main conclusions of the relationship between PSF coverage and IMR, neonatal mortality, or post-neonatal mortality. However, several covariates did become non-significant as did the relationship between PSF coverage and diarrhea mortality. This may be due to the fact that the microregions excluded due to poor quality data were also those with the highest rates of diarrhea deaths and underscores the importance of using adjusted rates when available.

**Conclusions**

The study has shown that expanding coverage of a community-based primary care program, hand-in-hand with other socioeconomic developments, was consistently associated with reductions in primary care-sensitive measures of infant mortality. Despite the consistency of these findings, several issues need to be addressed in order to assess the program’s overall effectiveness and potential relevance to other countries.

First, there is little data on the contribution of the PSF to health inequalities within Brazil. This study provides some evidence that due to its expansion in the north and northeast regions of the country, the PSF may have contributed to reducing interregional inequalities in primary care-sensitive infant mortality. But within regions, expanded PSF coverage has not always occurred in the most deprived municipalities (Morsch, Chavannes, van den Akker, Sa, & Dinant, 2001). In order to maximize the equity-enhancing potential of the program, national efforts should be directed at encouraging adoption of the program in the poorest municipalities. Within municipalities, program expansion should be encouraged within the most underserved neighborhoods. Such a strategy is likely to improve equity in outcomes since the greatest impact is likely to occur where infant mortality is still the highest, especially once outcomes have already improved for higher income groups (Victora, Vaughan, Barros, Silva, & Tomasi, 2000).

Second, financial incentives for municipalities to adopt the program are currently linked to increasing population coverage, but there are few systematic monitoring and evaluation processes in place to assess municipal or service-level performance. Surveys show that clients are generally satisfied with the quality of care delivered, but sustaining this level of satisfaction will be a critical challenge in maintaining popular and political support for the program (Trad et al., 2002). New initiatives have been proposed that would provide financial incentives for municipalities that reach or exceed certain health targets as a means to enhance access and quality of care. For these reasons, a major challenge will be to develop and use systems to monitor and
improve the quality of care delivered in order to maximize the potential health gains of this innovative approach to integrated primary care delivery.

Third, there is little data available on the cost-effectiveness of the PSF. In 2005, Federal government transfers to municipalities totaled $5.7 billion Brazilian Reais (approximately $US 2.6 billion), which represents about $US 14 per person covered. This figure does not include the municipal contribution (which varies from zero to nearly 100 percent). Thus we estimate that the true costs of the program may be as much as $US 30 per capita. While this is still a modest amount, there is, as yet, no national data to compare how well this program performs vis-a-vis the status quo. Such information will become increasingly important to mobilize the additional political and financial capital needed to reach the rest of the Brazilian population not currently covered and then to maintain adequate coverage in light of Brazil’s rapid epidemiologic and demographic transition.

Fourth, Brazil has a large supply of health workers, which might make it different from most other developing countries. So far, the PSF strategy has been successful in hiring more than 26,000 physicians and nurses and over 220,000 community health workers. In principle, any trained health professional may apply to the program and competitive salaries have made it an increasingly attractive option. On-going training in primary care is an additional benefit of PSF affiliation, but there is not yet enough known on how well this in-service training prepares formerly specialty-trained physicians to function as primary care providers. As the program continues to expand, health authorities will need to develop longer-term plans for maintaining and expanding the health workforce, with particular attention to improving the stability of physician contracting mechanisms, and more concerted efforts to enhance provider skills in community-based primary care.

Finally, because it serves as part of the Brazilian national health system, the PSF is vulnerable to health system level factors that could undermine its potential impact, such as access to pharmaceuticals; the quality and supply of needed specialty, diagnostic, or hospital care; or the availability, training, and salaries of health workers (Chiesa & Batista, 2004; Franco, Bastos, & Alves, 2005). Key challenges as the program moves forward include ensuring coordination between different types of health services and vertically focused disease control programs, improving quality of care, and maximizing community-based health promotion. Taken together these actions may help to assure that the PSF becomes more than just another program, but fulfills its promise as a central organizing feature of a more accessible, effective, and equitable national health system.

Acknowledgments

This study was partially supported by the Brazilian Ministry of Health. Frederico Guanais was supported by the National Council for Research and Development (CNPq). The conclusions presented in this paper represent the opinion of the authors alone.

References


References


A realist evaluation of the management of a well-performing regional hospital in Ghana
Bruno Marchal1*, McDamien Dedzo2, Guy Kegels3

Abstract
Background: Realist evaluation offers an interesting approach to evaluation of interventions in complex settings, but has been little applied in health care. We report on a realist case study of a well performing hospital in Ghana and show how such a realist evaluation design can help to overcome the limited external validity of a traditional case study.

Methods: We developed a realist evaluation framework for hypothesis formulation, data collection, data analysis and synthesis of the findings. Focusing on the role of human resource management in hospital performance, we formulated our hypothesis around the high commitment management concept. Mixed methods were used in data collection, including individual and group interviews, observations and document reviews.

Results: We found that the human resource management approach (the actual intervention) included induction of new staff, training and personal development, good communication and information sharing, and decentralised decision-making. We identified 3 additional practices: ensuring optimal physical working conditions, access to top managers and managers’ involvement on the work floor. Teamwork, recognition and trust emerged as key elements of the organisational climate. Interviewees reported high levels of organisational commitment. The analysis unearthed perceived organisational support and reciprocity as underlying mechanisms that link the management practices with commitment. Methodologically, we found that realist evaluation can be fruitfully used to develop detailed case studies that analyse how management interventions work and in which conditions. Analysing the links between intervention, mechanism and outcome increases the explaining power, while identification of essential context elements improves the usefulness of the findings for decision-makers in other settings (external validity). We also identified a number of practical difficulties and priorities for further methodological development.

Conclusion: This case suggests that a well-balanced HRM bundle can stimulate organisational commitment of health workers. Such practices can be implemented even with narrow decision spaces. Realist evaluation provides an appropriate approach to increase the usefulness of case studies to managers and policymakers.
recent realist synthesis of the effect of human resource management interventions on health worker performance in LMIC found that very few studies provide adequate information on the assumptions, the context and the underlying mechanisms of these interventions [7]. The same applies to a review of the effect of HRM policies on supply, distribution, efficient use and performance of health workers [8]. Rowe and colleagues came to similar conclusions [9]. Fourth, few of these studies have been carried out in LMIC [8].

All this notwithstanding, policies and management strategies are still imported from other settings into health services of LMIC without a blink of the eye. The surge of performance based financing (PBF) provides a good example. PBF is being introduced at different levels of the health system [10-13] and in a wide variety of countries, including Nicaragua [13], Cambodia [14], Rwanda [15,16], Zambia [17], Sri Lanka, Ghana, Zimbabwe, Thailand and India [18]. The evidence base, however, is very narrow [19]. Most PBF studies were found to lack controls and to neglect the analysis of confounding factors [6], which reduces the validity of the attribution of the reported effects to the intervention. Furthermore, very few studies offer indications of the conditions in which these approaches are working (see [20] for an example of a study that does).

In part, the methodological weakness of the health workforce management research resides in insufficiently rigorous studies. Some problems also stem from the widespread use of the case study. Indeed, although organisational studies is a domain marked by a lack of consensus on ontology and epistemology [21] and the consequent lack of consensus on methodology, the case study is a common research design for a number of reasons. First, it allows exploring a “phenomenon within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident” [22], and thus suits well the open systems-nature of human organisations. Second, it enables investigation of organisational behaviour as it happens in its natural setting [23]. Case studies are also useful in dynamic and complex situations where multiple, interacting variables may act upon intervention and outcome [24,25]. It is well suited to research on HRM [26]. Finally, Hartley argues that case studies can help in probing and developing theory [27].

Since the publication of *Experimental and quasi-experimental designs for research* by Campbell & Stanley [28], the major limitation held against the case study design is its limited external validity, or the weak potential to generalise findings from one case to another. Other authors raise its limited attribution power: case studies are good at analysing the intervening processes or documenting evolution in time, but weak at demonstrating the causal links between intervention and outcome [29]. Much of this critique has its origins in quantitative criteria of validity, according to which case studies are based on too small numbers of cases and on non-randomised case selection, thus leading to problems of representation and inference [25]. It is exactly here that its adherents claim that theory-based methodologies can make a difference.

During the 1980s, Chen and Rossi developed the theory-based evaluation approach as an answer to policy and programme evaluation approaches that remained limited to before-after and input-output designs or that focused narrowly on methodological issues (method-driven evaluation) [30,31]. The theories of change approach [32] and realist evaluation (RE) [33] are among the most recent applications of theory-based evaluation. As we will discuss in detail below, both approaches aim at opening the black box between intervention and outcome.

For organisational research, realist evaluation seems to offer a number of advantages. It promises, first, to increase the external validity of case studies. Building upon existing knowledge, RE analyses why change occurs, or why not, and in which conditions. It aims at providing information that allows decision-makers to judge whether the lessons learnt could be applied elsewhere [34]. Repeated case studies lead to more refined middle range theories that offer increasingly refined information of context conditions, thereby increasing generalisability of such case studies [21,27,29] and improving our understanding of causal processes [35]. Second, based on its generative perspective on causality, it seeks to explain change by referring to the actors who change a situation under influence of particular external events (such as an intervention) and under specific conditions [33]. Accepting the role of actors in change (agency), realist evaluation also considers structural and institutional features to exist independently of the actors and researchers. If human action is embedded within a wider range of social processes and structures, then causal mechanisms reside in social relations and context as much as in individuals. As a consequence of this ontological perspective, evaluators need to unearth the social layers in order to understand the root causes of the problem at hand [36] and to find the mechanism that explains the outcomes of the intervention [33]. In short, Pawson & Tilley argue that realist evaluation indicates ‘what works in which conditions for whom’, rather than merely answering the question ‘does it work?’. Realist evaluation is thus well suited to assessment of interventions in complex situations, which most organisational research is all about.

While the merits of theory-driven and realist evaluation have been amply discussed in journals on
evaluation (see for instance [36-41] and [42-44], there is little documented experience in the domain of health service organisation and public health, notable exceptions being [45] and [46]. This scarcity of realist studies could be interpreted as a sign of the limited academic credibility of theory-driven evaluation in general: ‘objectivist’ arguments overrule ‘subjectivist’ research [47]. Other reasons may be practical in nature: carrying out a full-blown theory-driven evaluation is resource- and time intensive [48]. The need of assessing the underlying theory in addition to the efficacy/outcome evaluation adds to the burden [41].

In this paper, we examine whether and how a realist evaluation design can be applied in research of well performing hospitals. We present the case of Central Regional Hospital (CRH) in Cape Coast, Ghana and discuss how we applied this method, from the stage of hypothesis formulation to the synthesis of the results. This case study is part of a longitudinal study on the links between management and performance in well-performing hospitals. We describe the latter as hospitals that ensure equitable access to high quality care and that provide such services in an efficient manner. We choose CRH both because it won the award for the best hospital of the Ghana Health Service in 2004 and on the basis of previous research.

The objective of the study was to analyse the management approach at CRH. We formulated the following research questions: (1) What is the management team’s vision on its role?; (2) Which management practices are being carried out?; (3) What is the organisational climate? (defined by Takeuchi et al. as the perceptions of employees regarding how the management approach is practiced and implemented in their organisation [49]); (4) What are the results?; (5) What are the underlying mechanisms explaining the effect of the management practices?

Methods

Principles of realist evaluation

Drawing inspiration from [34,50,51], we structured our study in 4 steps: the formulation of the Middle Range Theory, the design of the study, the data analysis and synthesis, and presentation of the results. We briefly introduce these steps from a theoretical point of view, and then describe how we developed each step in practice.

A realist evaluation research starts from a middle range theory (MRT), which is understood as “theory [y] that lie [s] between the minor but necessary working hypotheses (...) and the all-inclusive systematic efforts to develop a unified theory that will explain all the observed uniformities of social behavior, social organization and social change” [32] p. 39). In essence, this MRT states how the intervention leads to which effect in which conditions. Lipsey & Pollard identify different mechanisms to develop this MRT [53]. It can be formulated on the basis of existing theory and past experience.

If the latter is not available, exploratory on-site research can be done to unearth the models used implicitly by the actors to make sense of the intervention - what Pawson & Tilley call ‘folk theories’ [33]. Through individual interviews or group discussions, the key elements of the problem or intervention, the expected outcomes and potential moderating factors are to be identified [50], p. 196. Additional information may be derived from programme or policy documents. Cause mapping or concept mapping can be used in this process [54]. Ideally, the resulting MRT is then compared with existing knowledge. A literature review identifies studies reporting other causal chains, moderating factors or unintended outcomes, allowing a plausibility check of the preliminary MRT. The result is then again discussed with the stakeholders and results in the middle range theory that will be tested. Byng constructed the middle range theory on the basis of a literature review, a description of the intervention and discussions with facilitators involved in the programmes in question [40].

Regarding designs and research methods, realist evaluation is neutral [33]: the hypothesis as expressed by the MRT is guiding the choice of data that should be collected and the methods and tools to do so. Most theory-driven evaluations in healthcare used the case study design and combine both quantitative and qualitative methods.

Pawson & Tilley call the working hypotheses that emerge during the analysis phase ‘Context-Mechanism-Outcome configurations’ (CMOC) [33]. Realist evaluators describe not only the intervention and its outcome, but also the context and the underlying mechanism. They seek to establish patterns or regularities that explain outcomes of interventions. In practice, the data from interview transcripts, document analysis and observation are coded with codes drawn from the initial MRT (See [40] for a practical example). Similar to other analysis methods, subsequent rounds of analysis lead to a refined set of themes, categories and codes. The emerging findings are compiled as conjectural CMOCs, which indicate how the intervention led to particular outcomes in which context and by which mechanism. Their fit with the data is checked to ensure internal validity. The retained CMOCs are then compared with the MRT, which in turn is modified if necessary [55]. In some studies, the resulting ‘new’ MRT was discussed with key actors in order to validate it. A new study then further refines the MRT and this cyclical process leads to accumulation of better insights in how particular interventions work, in which conditions and how [33,34].
In order to be useful in decision-making, the synthesis should present the combinations of attributes required for an intervention to be effective, a presentation of the various alternative explanations, an indication of the potential of transferability by showing the links with existing knowledge, and an indication of the preliminary nature of the findings [56].

**Formulation of our MRT**

We formulated our preliminary MRT on the basis of an explorative study at CRH. During that study, interviewees indicated the importance of trust between health workers and their management, and the high levels of commitment of staff to the hospital. We also found arguments that pointed to the importance of a contingency approach to management of health workers: effective managers implement management practices that have a good fit with the nature of their workforce, the tasks of the organisation and its environment.

A second source of inspiration was our literature review of human resource management and hospital performance, which led us to high commitment management (HICOM). We retained this concept because its comprehensive approach to management fitted well with our initial analysis. The central attribute of HICOM is the combination of several complementary practices (e.g. good selection of staff, providing training on a needs basis and individual mentoring) in what is called ‘bundles’. Through their research in the industrial, commercial and service sectors, Pfeffer & Veiga identified a bundle of 7 elements, which they claim is universally valid [57]: providing employment security, ensuring comparatively high compensation contingent on organisational performance, instituting training and development, putting in place selective hiring, instituting self-managed teams and decentralisation, reduction of status differences and information sharing. Organisational commitment was identified as an outcome of such HRM practices [58] and has been shown to contribute to higher organisational performance. Such balanced bundles of management practices lead to better organisational performance [59-61]. We described elsewhere the key elements of high commitment management in health care organisations [62]. Some of the mechanisms that link HICOM to better performance include positive psychological links between managers and staff, organisational commitment and trust.

We drew another element from the work of Cameron & Quinn on organisational culture [63], which points to the importance of the coherence between the vision of the managers on their role, the practices they choose to implement, and the perception of their employees of these practices. Good fit between these would contribute to better organisational performance.

A final element is the notion of ‘decision space’. This concept was developed by Bossert [64] to describe the margins of freedom of health service managers at the operational level. His framework analyses how decentralisation policies affect the management practice at operational level. We retained adequate decision spaces as a potentially important context factor and a potential condition for HICOM to be possible.

It should be noted that there is considerable debate about the outcomes of HRM, and even more about the methods to demonstrate these. In general, we would describe the proximal outcomes of human resource management in terms of three categories: improved staff availability, improved staff attitudes and affects (commitment, job satisfaction) and better staff behaviour (in terms of higher task performance and organisational citizenship behaviour, and lower absenteeism). We selected organisational commitment and trust as proximal outcomes of human resource management, because our literature review pointed out that these outputs are often found to explain the effect of HICOM.

Combining all these elements with the findings of our first exploration visit, we formulated the MRT as follows:

“Hospital managers of well-performing hospitals deploy organisational structures that allow decentralisation and self-managed teams and stimulate delegation of decision-making, good flows of information and transparency. Their HRM bundles combine employment security, adequate compensation and training. This results in strong organisational commitment and trust. Conditions include competent leaders with an explicit vision, relatively large decision-making spaces and adequate resources.”

**Study design and data collection tools**

As will be clear at this point, we used the case study design as the basis. We collected both qualitative and quantitative data through document review of GHS and hospital records and reports, focusing on hospital HRM policies, and staffing levels and skill mix data.

In-depth interviews with all 6 members of the hospital management team (HMT) explored their management vision and practices. We based the HRM part of the interview guide on the 7 elements set of Pfeffer & Veiga [57](see some questions in Additional file 1). It must be noted that the interview guides only served as a guide to structure the interview when necessary, not as a questionnaire list that must be applied similarly in all interviews. In line with the concern that most studies focus on managers and ignore the perceptions of employees [65], we also explored the perceptions of staff regarding the management approach (the organisational climate).
In order to cover a wide range of views of different cadres, we made a purposive selection of staff. We identified the main cadres and within these, we randomly selected candidates for the interviews. This resulted in individual in-depth interviews of 3 nurses, 1 midwife, 1 doctor, 1 radiographer, 1 physiotherapist, 2 laboratory technologists, 1 clerical officer and 1 ward assistant. We also carried out 3 group discussions with heads of units, nurses, and paramedical staff (orderlies, clerical officers and account staff). Opportunistic non-participant observations were made of management meetings, ward procedures and OPD clinics.

We also developed a data collection form that focused on numbers of different cadres of staff (stocks) and on movement of personnel in or out of the hospital (flows in terms of transfer in/out, deceased staff, dismissed staff, absconded staff, retired staff).

During the preparation phase, a self-assessment of ethical issues, based on the working paper “Notes regarding ethical guidelines for health services research”, of the Department of Public Health, Institute of Tropical Medicine was done. This covered the following issues: Minimal risk to participants; Invitation, information and informed consent; Feedback to interviewees and staff. We sought and obtained a written informed consent from all interviewees. Measures were taken to safeguard confidentiality and anonymity. All interviews were recorded and transcribed verbatim.

Data analysis
We used NVivo 2.0 software for data management and analysis. The initial coding was based on a preliminary list of codes inspired by the MRT and on additional ideas that emerged during the fieldwork.

In a second round of analysis, some themes and patterns emerged (see below). In order to structure these as CMO configurations, we found it useful to borrow categories from theory-driven evaluation [66]. We described the intervention (in this case the HRM practices) in terms of content and application, and the intended and actual outcomes. We drew on our interviews and observations to differentiate (proclaimed) vision (what the team wants), the discourse (what they say) and the actual practices (what they do). We described the organisational climate, defined as “the atmosphere that employees perceive is created in their organisation by practices, procedures and rewards” [67]. In order to indicate how the intervention works, we analysed both the context and the intervening mechanisms, and attempted to identify the essential conditions.

To assess the intensity of the implementation of the practices, we developed an analytical framework based on the paper by Richardson & Thompson [59]. These authors questioned the research tools used in HRM surveys, which in their opinion often lack assessment of the intensity of application and coverage of the HRM practices. We selected coverage, intensity, internal fit and external fit as dimensions. ‘Coverage’ is understood as the degree to which the elements of the HRM bundle are applied to all cadres. ‘Intensity’ looks at the intensity of application. ‘Internal fit’ examines the synergistic and/or counterbalancing effect of the different elements. ‘External fit’ examines the appropriateness of the bundle for the cadre and organisation in question.

Reporting of findings
The preliminary results were discussed with the management team of the hospital, and the final analysis subsequently refined. A research report was sent to the commissioner of the study, a policy brief posted on the web and the findings were presented at the 2008 Geneva Health Forum.

Results
In this section, we present both primary findings and results from the analysis of the qualitative data in terms of the management vision, the actual management practices and the organisational climate. These sections correspond with the research questions presented above and are drawn from a ‘thick’ description of the case, or a detailed account of what the interviewees said, what we observed and what we learned from our document review.

The management vision
A first element we analysed was the views of the management team members of their own role in the hospital and on how they should manage the personnel. During the interviews, the management team members did not use words like ‘bundle’ or ‘high commitment management’, but they nonetheless expressed a clear view of the hospital’s roles and of how the health workers should be managed accordingly. Key terms include striving for excellence, offering services to all, attention for their personnel and sound financial management. This vision is transmitted through what they say during staff meetings or write in the mission statement and the annual reports.

This vision is well shared: not only do the director, the financial manager, the nursing manager and the non-medical administrator maintain the same discourse, also interviewees from the operational staff expressed this vision clearly, from nurses to cleaners.

“Their vision is that, they want this place to be a first class hospital. Their aim is to save life, so that is their main focus. And whatever they want to do so that life is saved, to me is their agenda.” (Non-medical worker, group discussion Non-medical staff)
The actual human resource management - what the management does

Based on the analysis of our interviews, observations and collected documents, we found that the actual set of practices at CRH includes more and different elements than Pfeffer and Veiga [57] listed. These authors list of seven elements includes:

- putting in place selective hiring
- providing employment security
- ensuring comparatively high compensation contingent on organisational performance
- instituting training and development
- deploying self-managed teams and decentralisation
- reduction of status differences
- information sharing

We found that selective hiring took place at the start-up of the new hospital in 1998, when the medical and para-medical staffs were almost handpicked from the pool of health workers in the region. At the time of the study (2005), however, the Ghana Health Service (GHS) regulations allowed only local recruitment of labourers and administrative staff.

The employment security offered by the GHS to its appointed staff was an often-mentioned reason why interviewees prefer employment in the GHS rather than the private sector.

At the time of the study, setting compensation levels was not within the decision space of the HMT. Only financial incentives for night duties and expatriate doctors could be given. Remuneration was not linked to actual performance. Just prior to the study, health sector strikes led to the Additional Duty Hours Allowance (ADHA) policy, which significantly improved the purchasing power of the health workers - the ADHA initially constituted a mark-up of 100-250% to the salary of a doctor and of lesser proportions for other health workers.

Training and personal development was found to be an important part of the HRM package. A full-time in-service training coordinator was appointed and a budget allocated to organise continued medical education activities, including clinical meetings, mortality meetings, seminars and conferences. Staffs were actively stimulated to follow external courses, even during working hours and personnel from all cadres actually did.

We found decentralised decision-making to be a central feature. The different units enjoyed a moderate level of autonomy in terms of decision-making and objective setting. Considerable decision-making authority over a number of domains, including the highly sensitive distribution of ADHA funds, was delegated to committees composed of different cadres of staff. The management team members argued that such decision-making structure would foster active participation of staff in decisions that affect the hospital.

In this decentralised decision-making structure, we found that teamwork is understood as ‘working all together, all engaged, all involved’. In the daily practice of curing and caring, teamwork was most visible at operational unit level. Deliberate efforts were made to include cleaners, sweepers and auxiliary staff in decision-making.

The nursing cadre decided to introduce an all-white uniform instead of the colour-coded uniforms. Interviewed nurses indicated this reduction of status differences as an important policy and perceived it as a sign of respect by management. In contrast, reduction of status differences between the management team and the operational staff seemed not a concern, neither for management, nor for the staff.

Information sharing was one of the most striking features. Formal communication channels were in place at all levels, including regular unit and ward meetings, heads of unit meetings and top management meetings. These were complemented by the committees mentioned above. General quarterly meetings (staff “durbars”), open to all staff, offered a voice even to the hierarchically lowest cadre. Observation showed that durbars effectively contributed to low-threshold, two-way communication.

Additional practices

We also found that the HMT developed HRM practices not included in Pfeffer & Veiga’s set: they made substantial efforts to ensure good physical working conditions, ensured good accessibility of the top managers and stressed hands-on involvement of managers and staff socialisation.

Major attention was given to creating optimal working conditions. The interviewees pointed to the good communication system in the hospital, the promptness of repairs, the general cleanliness of building and compound, the availability of air conditioning in virtually all rooms and the good amenities for patients. Other elements of the physical environment that were appreciated include the subsidised staff canteen, the internet café, the staff bus and the staff library. This points to the leverage of improving the working conditions. In Ghana, this may be a management intervention that increases not only the effectiveness of health workers, but also their job satisfaction.

Top managers are accessible for all staff. As in most Ghanaian hospitals, we found a clear hierarchy, whereby superiors should never be bypassed. Hierarchy was strong in the nursing and administrative cadres. However, interviewees mentioned the possibility to see the director or nursing manager in person when problems
could not be solved with their direct supervisor. Our observations showed that staff members of any cadre effectively made use of this open door policy.

Management stays involved at the operational level. Interviewees reported that the nursing managers were regularly helping out staff in the wards during their twice-daily supervision rounds, while the director was still involved in clinical work. The interviews show that this was a deliberate management strategy: the top management aimed at boosting staff morale by actually working with them and by leading by example. We also found that the heads of unit steered this process by inviting senior managers and heads of other departments to their unit meetings in case of cross-border problems.

At the time of the study, socialisation of staff was a central element at CRH. Newcomers were given a formal induction course and rotated for a few weeks through different units before being posted to their first station. Both close supervision and peer pressure contributed to maintenance of the standards of work. Interviews show that unit heads would identify staff not following the procedures and correct such behaviour through tutoring.

Intensity of implementation

We analysed the actual implementation of the HRM practices with the framework we presented under Section ‘Study design and data collection tools’ and which was based on the paper by Richardson & Thompson [60]. First, our observations and interviews show that the elements of the HRM bundle are applied to all cadres (good coverage). The intensity of application was variable. The management team, indeed, adapts its practices in response to emerging priorities. For example, when confronted with problems of permanence of doctors at the emergency department, a custom-made incentive package was put in place. This unequal approach was not contested because all staff recognised the role of doctors in the performance of the hospital.

Second, it seems the management team reached a good internal fit of the bundle (good degree of synergy between elements of the bundle). There were no practices that cancelled each other out, except perhaps for the emphasis on training. This had the unintended effect of enabling staff to leave CRH for better posts. Most other elements have mutually reinforcing effects: (1) information sharing, recognition and participative decision-making; and (2) bottom-up access to management and managers getting involved in the wards.

Finally, the external fit of a HRM bundle is the fit of the management practices with the core activities of the hospital (caring and curing) and with the mission of the organisation (providing accessible quality care). The HRM practices stimulate good professional practice by nurses, midwives and doctors by providing adequate autonomy to the operational units regarding their daily activities, while ensuring coordination between these units. The management is also perceived to provide effective support, information and resources (see below). As such, the bundle fits well to the task and mission of the hospital and to the professional values.

The organisational climate: the management practices as perceived by the staff

Four themes emerged in the analysis of the perceptions of the operational staff of the HMT’s actions: teamwork, strong perceptions of support by the management team, recognition and trust. As we will discuss below, these themes point to mechanisms that help explain how the management strategies worked.

Teamwork stimulates staff from all cadres to be involved in care

The interviews indicated a strongly shared feeling among staff members that team work matters: they maintain that quality of care can only improve if all types of staff are involved.

“In some places, nobody gets close to the Nurse Manager and it is like she only decides what she wants at the place. (…) But here, everybody is important. We see everybody’s job as important aspect of the health care delivery system, so we include everybody in the care.” (IO 1, Unit head, Ind. interview)

Junior staff members pointed out the ‘free’ relations with their superiors.

“We are all free in our units. My head always comes round to see what is going on over here. If something is not in the right place, he will show you to do this or that. So, always the heads are helping us, so we also feel free to work with them.” (Non-medical staff, GD Non-medical staff)

‘Free relations’ strengthen the collaboration between operational staff and their heads of units, but also with the top managers. Interviewees similarly mentioned the easy communication between the middle line staff and the HMT.

“I would say there is good relationship both formally and informally. We communicate by memos, but as soon as I came, I can just walk straight to Director and tell him: ‘This is the problem’, and we just brainstorm to see how the problem can be solved.” (IM5, HMT member, Ind. interview)

Perceptions of support by the management team

Interviewed staff members often mentioned that they feel supported by the HMT. First, interviewees
expressed the feeling that the HMT is effectively solving problems. Unit meetings or ward conferences are a good example of how formal meetings can prevent or solve coordination problems.

“The ward conference is very good. The accountant is there, the pharmacist is there, the lab man is there, everybody is there. The meetings or presentations are not for fault finding. We pick issues from there and we make our corrections or cover loopholes.” (Head of unit 1, GD Unit heads)

Informal and non-structured opportunities exist, too, and are used to good effect. Interviewees pointed out how open relationships and good access to top managers allows them to take a problem to the ‘next level’.

“As a unit head, if I think that something is not going on well, my demands are not being met, I can approach the director and we sit down and talk about it. (...) You are free to enter his office anytime to discuss your problem, especially when you think things are not going on well” (Head of unit 4, GD Unit heads)

Staff members appreciated not only the possibility to discuss work-related problems with their superiors, but also the attention given by the latter to their professional development. This also applies to members of the hospital management team.

“He [the director] made every opportunity for my career advancement. He is always looking out, listening and trying to help where he can, to see how he can help people to progress. So, when you have someone doing that for you, at least you also have to return the same to him.” (IMS, hospital management team member, Ind. interview)

**Strong perception of recognition**

The interviewees expressed strong feelings of recognition by the management team. They explained how a range of practices, from a word of appreciation to tangible rewards expresses the appreciation of the HMT for their work.

“At the end of the year, every staff here is given a token. Sometimes, something in the form of food, money, a get-together, occasionally words of motivation, a tap on your shoulder, meeting you and finding out how is it, how is the work going on. This serves as motivation.” (IO7, Head of unit, Ind. interview)

Interestingly, several interviewees mentioned the initial staff selection, when the hospital was started up, as a key event, not only because it helped set standards, but also because of its strong undertone of recognition.

“To start with, I can surely say that, the standard that was set right from the inception of the hospital has made such a mark. Because immediately when this hospital was instituted, we were to come for an interview. So, a high standard was set (...) and they see if you have the call to work. On that note, in coming out to publish the names of those to come here, it is like Government releasing a white paper. By that time, you feel as if you are in heaven. (...) With that alone, that standard was set and everybody was expected to give of his best.” (Nurse, GD Nurses)

**Perceptions of trust**

We explored the issue of trust, which we found to be an important element in the explorative study, by asking staff how they would rate the levels of trust at CRH and how they believe trust is generated. The interviewees indicated fair levels of trust both amongst staff and between management and staff.

“In the whole hospital, there is some trust, but I don’t think it is 100%. May be it is between 80% - 95%.” (Unit head 3, GD Unit heads)

Asked how management practices influence the levels of trust, they pointed out the importance of meetings during which information is exchanged, the willingness of managers to discuss decisions and the resulting perception of transparency.

“At least, we have management meetings and after that, management meets the unit heads and tells them what the institution wants to do, the programmes they have embarked on. They discuss with the unit heads and if somebody does not understand something, management explains it. The unit heads are supposed to go down and explain to their subordinates. And when we have staff durbars, these things are also brought up. So, transparency is there, we can understand things. Anything they want to do is explained to workers.” (Unit head 3, GD Unit heads)

These consultations and opportunities to discuss important issues contribute to perceived fairness of the decisions. Interviewees said that less rumour mongering and suspicion arise when people are informed why certain measures are implemented and others not.

“At the end of the day, like we had our last year’s meeting after we presented our reports, management too presented their report, their financial report,
what they got and how they spent their expenditure and those things. So, there will be no room to think that somebody is cheating on you, or management is hiding certain things from us. So, we know what is happening; you don’t need to or there is no room for suspicion. (…) I think it is a fair deal between management and staff.” (IO2, Midwife, Ind. interview)

Another source of trust is the effective support staffs receive from their superiors in case of problems.

“Even the Director himself came here three days ago. So, what he said, he has done it. That is why I say I trust him.” (Non-medical worker, GD Non-medical staff)

“The trust comes from the urgent action taken when there is a problem. If there is any problem on discipline for example, an ad-hoc committee is set up and within days, the matter is settled.” (IM4, HMT member, Ind. interview)

Analysis

After categorising, and thus making sense, of the primary data in the form of CMO Configurations, a realist evaluation seeks to examine the link between these findings and the middle range theory it set out to examine. In practice, we searched for potential causal pathways between the management practices and the apparent outcomes of commitment and trust. To do so, we summarised the above findings and then searched for CMO configurations.

A summary of the intervention and its outcomes

Our interviews and document review show that the Hospital Management Team identified good hospital performance as the intended distant outcome of its management practices and a motivated and well-performing workforce as the proximal outcome. As mentioned above, the scope of this study did not allow examining the association between management practice and hospital performance, and we focus on the effect of these practices on organisational commitment and trust, the proximal outcomes we retained on the basis of our preliminary theory-building.

The actual intervention can be summarised as a combination of HRM practices: socialisation of (new) staff, training and personal development, good communication and information sharing between different levels of the organisation, and decentralised decision making to the level of ward and department teams. We also found important additional management practices: the creation of good working conditions, the good accessibility of top managers, and the active involvement of the manager on the work floor.

Regarding the process of implementation, we noted a good coherence between the HRM practices and the management team’s vision. Indeed, in line with their vision, the management team motivates the staff through different interventions: remuneration, effective support and recognition. The HRM practices are reinforcing each other (good internal fit). The bundle is well adapted to the different cadres of a healthcare organisation and its mission (good external fit). It is applied similarly to all cadres (good coverage). The intensity is variable, but this poses no problems for the staff.

Realist evaluation improves external validity of a case study by describing the implementation context. During the study, we found several potentially important elements in the context of Central Regional Hospital. First, as testified by the brain drain, Ghana has a well-trained health workforce from which the GHS (and thus CRH) can draw personnel. Its medical and paramedical cadres display a high degree of professionalism, and there is a general culture of professionalism in the GHS. Second, reasonably good resource availability in terms of hospital funding and management capacity allows investing in the workforce. Indeed, commitment eliciting management practices are costly, especially in management time and in terms of training costs.

We found that the outcomes of the HR management bundle at CRH included trust, commitment and strong perceptions of recognition and of support by management, which result in a positive organisational climate.

CMO configurations

During the later phases of the analysis, we found that the management practices can be grouped according to their key mechanism and this led to the description of two parallel CMO configurations, each with their own outcome.

The first CMO can be summarised as ‘keeping up standards of excellence through organisational culture’. The hospital had a head start: staff members were selected on professional and motivational grounds by the management team. This lengthy selection procedure gave the staff a feeling of belonging to an elite corps of health professionals and reinforced their professional identity. The management team used this opportunity to initiate a culture of high standards of professional excellence. They set up an induction programme for new staff, and much attention was paid to teamwork and supervision. This reflects findings of Schein [68]: such practices serve as strong embedding mechanisms of the organisational culture. There was equally much attention for a clear role distribution and for task monitoring. In summary, both ‘hard’ and ‘soft’ management practices are balanced in the bundle. The former include general rules and procedures, task distribution for clinical and administrative staff and monitoring of task performance; the latter include induction courses, peer pressure mechanisms and training/personal development.
opportunities. All this reinforced the initial capital of professional excellence. Availability of a pool of professional health workers is an important context element, and may be essential for such a bundle to work.

The second CMO configuration can be summarised as follows: a hospital management team can attain higher organisational commitment if it strengthens positive reciprocity relationships that are based on social exchange, even in hospitals with limited HRM decision spaces. Key practices in this set include creating open access to managers for all staff and grass-root involvement of managers at operational level. This reinforces open relationships and contributes to solving operational problems and conflict resolution. In turn, it stimulates the feeling of perceived organisational support. Eisenberger and colleagues describe this as the beliefs and perceptions of employees regarding the support provided and the commitment demonstrated by the organisation in their staff [69]. Employees interpret decisions and actions of their managers and their trust-worthiness in terms of the commitment of managers to their staff. At CRH, the leadership and management style is indeed perceived to be effective (in meeting its promises and in ensuring adequate physical working conditions) and supportive, even on the personal level. Ultimately, such practices stimulate reciprocity and as a result, organisational commitment. This in turn contributes to organisational performance [70].

Availability of well-trained health workers and adequate funding seem intuitively to be essential context elements in both CMO configurations.

**The new MRT**

Our analysis identified two CMOCs that indicate causal pathways between sets of HRM practices and HRM outcomes, and we modified the MRT accordingly:

“The management of a well-performing hospital deploys organisational structures that allow decentralisation and self-managed teams and stimulates delegation of decision-making, good flows of information and transparency. In the management of health workers, they implement a balanced bundle of management practices that includes both clear goal setting, role distribution and task monitoring (hard HRM) and training, support and recognition (soft HRM). Based on the mechanism of perceived organisational support and reciprocity, such combinations lead to a positive organisational climate that includes recognition, respect, commitment and trust. If these are taken up into the organisational culture and newcomers are inducted into the OC, enduring effects of such practices can be expected.

Conditions for such management practices to work include competent leaders with an explicit vision, a minimum of resources and conducive institutional arrangements, including effective decentralisation and appropriate decision spaces (although the latter can be narrow for HRM).”

**Discussion**

On the basis of this one study, we cannot yet draw firm policy recommendations. Nevertheless, it offers interesting insights in health workforce management and in the use of realist evaluation.

**Lessons for policy and practice**

First, we found a proof of concept for HICOM in resource-poor health services. Second, our study found variant practices compared with the bundle described by Pfeffer and Veiga, which supports the findings of Richardson & Thompson [60] and Marchington & Guruli [71]. Third, this case reinforces the point that in management of health workers, we need to apply coherent bundles of practices, and not focus on singular interventions. In HRM, the quality of management practices counts more than the quantity. It is not the actual number of practices, but rather the process by which these practices are put in place that is related with positive staff attitudes like commitment, job satisfaction and procedural justice [65]. This is in line with conclusions of other studies in other sectors [26,72].

Regarding the mechanisms, our findings relate to the analysis of Evans & Davis [73], who situate the underlying mechanisms of high commitment management at the level of the internal social structure of the organisation. Such practices improve knowledge, skills and abilities, but they exert also major effects at the level of relationships. Weak ties are strengthened [74], reciprocity is established and maintained [75] and shared mental models contribute to a strong organisational culture. This in turn affects behaviour of staff and improves organisational efficiency and flexibility, and ultimately, organisational performance. The evidence of the impact of such reciprocity relations or of organisational commitment on organisational performance is not strong, and further research should investigate whether and how high commitment leads to better performance in healthcare organisations.

We found that the decision spaces managers require to develop a responsive HRM approach may be smaller than is often thought. At the time of study, the decision spaces of regional hospital managers in Ghana were quite limited concerning HRM. As important as the formal decision space is its actual utilisation. At CRH, the team exploited its decision spaces well to create its own...
way of management within the defined institutional arrangements of a ‘regional hospital’ (e.g. by using committees and delegation of decision-making power).

Finally, a balanced management approach is costly, especially in management time. It requires reasonable financial resources and a management capability to deal not only with administration but also with the less tangible issues of relationships, organisational culture and motivation of staff.

Future research should establish what other HRM approaches lead to high commitment, under which conditions HICOM works, and how it can be stimulated. This last question deserves attention. Health services in many LMIC are both ill equipped and not sufficiently supported to implement a HRM approach that differs from a mere administrative approach. In the first place, the managers of health services are mostly medical doctors. Human resource management is not an element of the medical education curriculum. Even if they received additional public health or management training, the curriculum mostly equates HRM to personnel administration and this hardly prepares future health service managers for responsive management.

Methodological lessons
In this case study, we used a realist evaluation approach because we consider health care organisations to be essentially social entities. Pawson argues that realist evaluation is well suited to investigate change in such social systems [34]. Its focus on the generative causality that underlies interventions, stimulates the analysis of how the intervention works and in which context conditions.

This results in more detailed conclusions that indicate how the intervention was carried out, which effect it had and how it worked. It also offers insights in the context elements. Such theory building helps to overcome the limits of traditional case studies, and specifically their low external validity and low power to explain change [42]. However, appealing as it is, realist evaluation poses a number of challenges for the researcher.

The attribution paradox
Perhaps the most critical issue is the attribution paradox. Because of its ontological and epistemological basis, realist evaluation is quite fit to assess complexity [76,77] and may contribute most in research of exactly such topics. However, research of complex problems needs to confront multi-causality. In complex systems, the behaviour of people and organisations alike is determined by many interlinked factors. Health professionals act under influence of their professional norms, social pressure, management interventions, and not least, their intrinsic motivation. Assessing the exact contribution of a set of management practices to overall organisational performance may therefore be virtually impossible.

What realist evaluation can do is to stimulate the researcher to describe a detailed picture of the causal web that includes the multiple determinants and to categorise these as intervention, underlying mechanism or essential context factor. In our case, we have arguments to say that both commitment-eliciting management and personnel administration are required, but we cannot (yet) indicate which among these two sets is the most important in which setting.

The conclusion may be that one needs to accept that the kind of evidence provided by realist evaluation can never be put in the same categories of evidence produced by controlled experimental methods, not only because of its perspective on causality, but also because of the complexity of the subjects on which it will be applied.

The MRT fallacy
While any researcher adopts specific reference frameworks during her research, realist evaluation asks researchers to make these frameworks explicit in the form of a MRT. This implies a risk of developing a tunnel vision: the researcher may remain blind for the unexpected factors and alternative explanations. This risk can be reduced by the plausibility check during the development of the initial MRT, triangulation of findings, analysis by multiple researchers and discussion with stakeholders and peers.

The MRT fallacy also operates at the stage of analysis and of dissemination. During analysis, we did several rounds of plausibility checks, because we kept finding alternative explanations in disciplines such as organisational psychology, organisational theory and sociology. The CMOCs and resulting MRTs are indeed most often just one way of explaining the findings. A middle-range theory can indeed never cover all possible explanations of change [34]. In Pawson and Tilley’s view, a realist evaluator does not strive at nor pretend to provide the ultimate evidence that the intervention works. Rather, she aims at enlightening the decision-maker, a process of utilisation of research that may be the most frequent in case of social science [78]. In such cases, a pragmatic position should be taken, whereby one tries to refine the middle range theory as much as practically possible, with the explicit aim of providing options for improvement rather than reaching a perfect understanding of the intervention as such [56,79].

The CMO dilemma
As we mentioned, the CMO configuration is a powerful model to go beyond the classic case study, as it forces the researcher to go beyond description. However, a true application of realist evaluation requires not only a systematic description of the intervention in terms of intervention, outcome, context and mechanisms. Also the generative causal relationships between these
elements need to be assessed. In our analysis, this proved difficult at several levels.

The first important hurdle is the differentiation of the effect of the context from that of the intervention. This feeds the attribution paradox: is the outcome the result of the intervention - and to which degree - or are there context elements that explain the change in outcome - and to which degree? Furthermore, some context elements can be expected to moderate the relation between intervention and outcome, and in some cases, the outcome of an intervention will influence its context (initiating feedback). Regarding our case, probably more attention needs to be given to the role of professionalism. Professional values can steer providers’ behaviour to an important degree and could partially explain the behaviour of certain staff, irrespectively of the management strategies. Most likely, we may find that the observed management strategies are in a close fit with professional behaviour traditions.

Secondly, the realist researcher seeks to describe the mechanism that is triggered by the intervention and that leads to the outcome. Confusion may result from the finding that some context elements are essential for the outcome: is this context element then part of the mechanism? We clarified this issue by considering context elements as actors or factors that are external to the intervention - that are present or occurring even if the intervention does not lead to an outcome -, but which nevertheless may have an influence on the outcome. The mechanism is the causal pathway that explains how the intervention leads to an observed outcome in a particular context. In other words, the intervention leads to an outcome in specific contexts if it triggers certain mechanisms. If the mechanism is found to be context-dependent, which in health services may often be the case, essential context elements can be identified. In our case, the professionalism of the staff selected to work at CRH is a context element, the decision to introduce an induction training was a management decision, and the effect of building an organisational culture was a mechanism.

The efficiency question

By its very nature, RE may yield information that is particularly useful for policymakers. However, by its same nature, a RE needs considerable expertise and ample time and resources, because of its comprehensive scope. Indeed, besides the efficacy/outcome evaluation, also the underlying theory and the context must be accounted for [41,80]. In our case, work at CRH started in 2004 with an exploratory visit, and much analysis went on after the second visit in 2005 and the third visit in 2007. Such timelines may still be acceptable in case of non-urgent issues, but far less in case of high-interest policy issues.

Conclusions

Realistic evaluation offers a comprehensive approach to assessment of interventions in complex situations that can go beyond the simple efficacy question. We developed a realist case study that unravelled the management practices put in place by a hospital management team in Ghana. This study shows that it is possible to implement high commitment management practices in LMIC and that these are perceived to be relevant by the health workers. We found that through a well-balanced bundle of HRM practices, management teams can stimulate organisational commitment and an organisational culture of excellence. At CRH, the HRM bundle included sound administrative management. Reciprocity and perceived organisational support emerged as an important underlying mechanism. In applying the realist methodology, we also encountered a number of pitfalls and paradoxes. Only through further practical applications will we find out how these can be overcome.

Acknowledgements

We would like to thank all the staff of Central Regional Hospital (Ghana) for their warm welcome and collaboration during this study. This work was funded through the framework agreement (Razmalkaard) between the Belgian Directorate-General for Development Cooperation and the Institute of Tropical Medicine, Antwerp. The sponsors had no role in the study or in the writing of the paper.

Author details

1Department of Public Health, Institute of Tropical Medicine-Antwerp, Nationalestraat 155, B-2000 Antwerp, Belgium. 2Volta Regional Health Directorate, PO Box HP 72, Ho, Volta Region, Ghana. 3Department of Public Health, Institute of Tropical Medicine-Antwerp, Nationalestraat 155, B-2000 Antwerp, Belgium.

Authors’ contributions

All three authors contributed to the original design and analysis. BM and MD carried out the data collection. BM and MD analysed the data. BM, MD and GK contributed to the discussion section and to writing the manuscript. BM edited the final draft. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 15 July 2009 Accepted: 25 January 2010 Published: 25 January 2010

References


Pre-publication history
The pre-publication history for this paper can be accessed here: http://www.biomedcentral.com/1472-6963/10/24/prepub

Cite this article as: Marchal et al. A realist evaluation of the management of a well-performing regional hospital in Ghana. BMC Health Services Research 2010 10:24.

Submit your next manuscript to BioMed Central and take full advantage of:

• Convenient online submission
• Thorough peer review
• No space constraints or color figure charges
• Immediate publication on acceptance
• Inclusion in PubMed, CAS, Scopus and Google Scholar
• Research which is freely available for redistribution