

**HOUSEHOLD AND PROVIDER BEHAVIOR IN THE HEALTH SECTOR IN
AFRICA: WHAT HAS BEEN LEARNED FROM PROGRAM EVALUATIONS?**

By

Peter Glick
Cornell University

June 2009

Framework Paper prepared for the African Economic Research Consortium
Project on Health, Economic Growth and Poverty Reduction in Africa

TABLE OF CONTENTS

1. INTRODUCTION	1
2. CASUAL INFERENCES AND PROGRAM EVALUATION	4
2.1 <i>Inference in Randomized Controlled Trials</i>	4
2.2 <i>Alternatives to Randomization</i>	5
2.3 <i>Debates over Randomization</i>	9
2.3.1 Threats to external validity	9
2.3.2 Threats to internal validity.....	12
2.3.3 Other critiques	13
3. HOUSEHOLD AND PROVIDER HEALTH BEHAVIOR	15
3.1 <i>Demand Side Behavior: Households</i>	15
3.1.1 Price.....	15
3.1.2 Bamako once again: price increases with quality improvements.....	19
3.1.3 Other evidence for household response to service quality	22
3.1.4 Externalities and social learning.....	24
3.2 <i>Supply Side Behavior</i>	26
3.2.1 Accountability and community monitoring.....	27
3.2.2 Pay for performance incentives	29
3.2.3 Contracting with private providers	31
3.2.4 Private sector and positive prices	33
4. EVIDENCE ON CHILD HEALTH INTERVENTIONS	35
4.1 <i>Water, Sanitation, and Child Diarrheal Illness</i>	35
4.2 <i>Bednets for Malaria</i>	38
4.3 <i>Community-based Programs</i>	42
5. SUMMARY AND CONCLUSIONS.....	46
REFERENCES	49

1. INTRODUCTION

Africa has experienced a number of dramatic successes in health in the decades since independence, particularly with respect to child mortality and morbidity and the eradication or reduction of specific diseases. Successes include vaccination programs against measles and other childhood illnesses, the expansion of oral rehydration therapy for treating infant diarrheal diseases, and the near-elimination of river blindness and Guinea worm (Levine 2007). At least until the 1980s, under-five mortality fell sharply in Africa. However, Africa remains behind most of the developing world in child and other health indicators, and progress in child health has slowed considerably in the last two decades. From 1990-2006, the share of malnourished children under five fell slightly, from 32 to 28% in Subs-Saharan Africa (United Nations 2008). Under-five mortality showed some progress, falling from 184 to 157 per 1000 in the period, but this improvement is smaller than in other regions.

Adding to these remaining challenges is the enormous burden of HIV/AIDS, which affects Africa more severely than any other region and in some countries has put tremendous strain on health care systems. While African economies overall have performed significantly better over the last decade than in the previous two, and donor contributions for health have increased sharply (though very disproportionately for HIV/AIDS), governments still must face these challenges with severely limited financial, infrastructural, and human resources. Policymakers therefore need to know which kinds of interventions can have the strongest public health impacts. An additional challenge is to insure that programs reach the poorest segments of the population. In most African countries, as in much of the developing world, sizable gaps exist between poor and wealthy citizens in terms of both coverage of basic health services and indicators of health status (Gwatkin et al. 2007; Boerma et al. 2008; Castro-Leal et al. 1999).

In the area of child health, clinical studies have shown the efficacy of many low cost interventions for improving child survival, such as oral rehydration therapy to treat diarrhea, insecticide-treated bednets for malaria prevention, hand washing with soap for diarrhea prevention, breastfeeding, vaccinations, and vitamin and food supplementation (Bhutta et al. 2008). With efficacy established, attention has increasingly turned to insuring effectiveness, meaning that an intervention has significant health impacts at the population level. Effectiveness is a function of service delivery, uptake, and quality, and achieving it requires overcoming constraints affecting both the demand (consumer) and supply sides. Public health spending will fail to improve health outcomes when households choose not to use services or lack access to them, or when the services are of poor quality. Therefore the success of policies depends to a very significant extent on the behavior of two sets of actors: households or health care consumers, and providers of services.¹

¹ Beyond household and provider behavior are broader issues relating to the effectiveness of health system organization and financing. These are obviously important in explaining the success or failure of policy, and constraints in these areas will be noted at times in this paper.

A great deal of recent research, much of it employing randomized controlled trials, has considered household responses to interventions that alter the costs or availability of health services or products such as bednets or drinking water disinfectant or that promote better health behaviors (such as hand washing or safe sex). On the supply side, quality and performance issues have come to the fore in research and policy, both through the collection and analysis of comprehensive facility level data and through analysis of provider incentives and (somewhat controversially) the potential benefits of contracting with private providers or expanding the role of the private sector in other ways.

This paper reports on what has been learned about these issues in the African context through (primarily) evaluations of health interventions. The focus, as in many surveys of research, is broadly speaking on ‘what works’ (or does not work). However, as the previous discussion implies, the concern is not primarily about whether certain treatments improve health in a clinical sense, that is, efficacy. Rather, almost all of the program evaluations and other research considered here involve strategies to improve the delivery, uptake, and effective use of services or products, that is, program effectiveness. Central to the success of these strategies, as indicated, are behavioral responses of households or providers. To clarify the distinction with an example, Cohen and Dupas (2008) use a randomized controlled design to examine the effects of variation in the price of insecticide-treated bednets on the demand for the nets and their actual utilization by women attending antenatal clinics in Kenya. The focus of this study is not on whether such bednets, if used appropriately, can prevent malaria in children; efficacy had already been established in a number of clinical trials.²

The remainder of this paper is structured as follows. Section 2 discusses methodological issues in estimating the impacts of health programs. Substantial discussion is given over to randomized controlled trials (RCTs), or policy experiments, which have been rapidly gaining ground in health policy research in developing countries. However, the research discussed in this paper draws on a much broader array of evidence than RCTs alone. Despite clear advantages in terms of obtaining causal inferences of program impacts, experiments have limitations, and a lively and at times sharp debate is taking place over how prominent—or dominant—a role they should play in research and policy evaluation (Duflo and Kremer 2008; Deaton 2009; Heckman and Urzua 2009, Imbens 2009). These limits are discussed and alternative approaches to evaluation are presented.

Sections 3 and 4 consider lessons learned from empirical research. The focus is on African experiences but with findings drawn from other contexts as well. In Section 3 the discussion is organized thematically rather than by specific illnesses. First, the subject of take-up of interventions (household demand behavior) is considered. There is a particularly large empirical microeconomic literature on this topic, recently significantly enhanced by experimental studies, that considers the role of price, quality, information, and (to a lesser extent) behavioral and health externalities that impact program effectiveness. For example, recent experiments have provided substantial insight into price responsiveness on the part of consumers—which has been found to quite high—and have thus contributed significantly to debate over whether health goods and services should be positively priced (the influential

² That said, the distinction is not always so clear, and many of the studies reviewed in this paper consider efficacy as well as effectiveness.

view of social marketing) or free. Next the discussion turns to the behavior of providers of health goods and services. Here the microeconomic empirical literature (that is to say, experimental or other statistical evaluation of policies) is considerably thinner. However, drawing on this literature and other evidence, this subsection considers the evidence for the effectiveness of approaches such as community monitoring of public health care providers, performance-based pay, and contracting with private providers for service provision. Also discussed is the problem of insuring adequate supply when a health good or service is provided free to much of the population and consequently incentives to private providers and distribution networks are weak.

Section 4 considers evidence for several specific interventions to reduce child morbidity and mortality, namely treated bednets for malaria and water and sanitation interventions to reduce diarrheal diseases. This is not a comprehensive overview of interventions to address major childhood illness (for example, acute respiratory infection and micronutrient deficiency are not discussed), but the existence of a fairly extensive recent evaluation literature, using RCTs and other methods, makes these two topics particularly attractive.³ This is followed by consideration of community-based interventions for child health which encompass the full spectrum of child illness but are linked by their reliance on low cost non-professional health workers to expand the contact of households with the health care system. This approach merits particular attention because it is directly focused on increasing the delivery of services where uptake is low, a common theme of much of the literature this paper will discuss.

Section 5 summarizes the main findings from the literature and concludes with several observations on methodology and evaluation.

³ Nor, of course, does it cover many other pressing health needs in Africa, above all HIV/AIDS but also including maternal morbidity and mortality and inadequate curative care for adults as well as children. However, a number of evaluations of programs to address these problems are discussed in Section 3. In addition, a prior framework paper (Glick 2008) discusses HIV/AIDS interventions in Africa in detail.

2. CASUAL INFERENCES AND PROGRAM EVALUATION

2.1 Inference in Randomized Controlled Trials

Since most of the evidence discussed in this paper takes the form of program evaluation of one sort or another, this section describes basic concepts of, and techniques for estimating, the causal impacts of programs. The discussion is not at all comprehensive, as many thorough guides are available (Duflo, Glennerster, and Kremer 2008; Todd 2008). We begin with a description of randomized controlled trials (RCTs) and cover the main points of contention in the debate over RCTs. Descriptions of alternatives to randomization follow.

The purpose of evaluation is to make causal inferences about the effects of programs on outcomes of interest. Defining Y as the outcome (child nutritional status, curative health care consultations, etc.), we are interested, at least as the first order of business, in finding the mean program or treatment effect in the population or in a target population, represented as follows:

$$E[Y_i^T - Y_i^C] \tag{1}$$

where T and C superscripts denote treatment and non-treatment states, respectively, and i denotes the individual; i could equally refer to a group of individuals such as a school class or community. Y_i^C is the no-treatment counterfactual and cannot be observed for i if i actually receives the treatment; we can only observe outcomes for some other group of individuals that is not treated. Thus the measurement of impacts involves comparing outcomes for a group that receives the treatment (denoted by T) with outcomes for a group that does not (denoted by C):

$$E[Y_i^T|T] - E[Y_i^C|C] \tag{2}$$

Subtracting this from both sides of (1) yields:

$$E[Y_i^T|T] - E[Y_i^C|T] + E[Y_i^C|T] - E[Y_i^C|C] \tag{3}$$

The first two terms show the expected value of the effect of treatment on the treated and is (usually) the quantity of interest in the evaluation. The last two terms show the expectation of the difference in untreated outcomes of the treated group and the untreated group. If the two groups are statistically equivalent, this expectation is zero and thus the comparison in (2) yields the effect of treatment on the treated. If the two groups differ in ways that affect outcome Y this expectation is nonzero, and the comparison yields a biased estimate. If as in an experiment individuals can be randomized into treatment and control groups, the first case holds. In the absence of randomization, one cannot assume an absence of correlation of treatment status T with other factors affecting outcomes, implying a non-zero value of $E[Y_i^C|T] - E[Y_i^C|C]$ and a biased estimate of the effect of treatment on the treated. Without random assignment, in other words, selection into the treatment group is likely to be endogenous to potential outcomes. This means that other, non-experimental approaches are needed to make causal inferences.

It is useful also to present the same idea in a regression framework:

$$Y_i = \alpha + \beta T_i + e_i \quad (4)$$

Where T_i is a dummy variable equaling 1 if i belongs to the treatment group and zero otherwise. If treatment status is associated with the level of Y_i other than through treatment itself (i.e., if outcomes in the absence of the program would differ for the treatment and no treatment groups), then T_i and the error term are correlated and the estimate of β is a biased measure of the treatment effect. Randomization ensures that T_i is independent of the error term.

In standard clinical settings, or in field experiments that approach this set-up, individuals are randomized to treatment and control groups; the former gets the intervention and the latter does not. The difference in mean outcomes for the two groups is the effect of treatment on the treated as described above. In other cases—very common in RCTs of health interventions in developing countries—randomization is at the level of communities (or other units such as schools) rather than individuals. For example, the intervention (HIV testing, an iron supplementation program) might be introduced in a community health center, with all in the local target population be eligible to participate. Eligibility, however, does not imply participation: not all adults will decide to get tested for HIV, not all parents will accept iron tablets for their children. The difference in mean outcomes for treatment and control communities, with means calculated over both participants and non-participants, is an estimate of ‘intention to treat’. In many cases this is more important from a public health perspective than the effect of treatment on the treated, since it measures the overall impact of the intervention on the target population.

The great benefit of randomization, if carried out appropriately, is that it provides truly causal inference of the effect of the treatment on the study population—what is known as ‘internal validity’. Whether these results hold in other contexts—whether this refers to other areas in the same country or other countries—is another matter. This issue, known as ‘external validity’ is at the heart of the debate over randomization in development research and is taken up in some detail below.

2.2 Alternatives to Randomization

Randomization is very often not possible logistically, politically, or ethically (one usually would not deny life-saving treatment to a control group). There are several alternative, quasi-experimental, approaches to understanding the effects of programs.

Matching techniques construct comparison groups based on the assumption that relevant differences among treated and non-treated individuals or communities can be completely captured by measured (observable) characteristics. That is, conditional on these controls X , expected outcomes would be the same for both groups in the absence of treatment:

$$E[Y_i^C|T,X] - E[Y_i^C|C,X] = 0 \quad (5)$$

The widely used propensity score matching (PSM) technique involves matching treated and non-treated units according to an index score, which is the predicted probability of treatment or program participation, derived in most applications from a probit or logit model including a series of covariates. Techniques for the matching itself range from simply pairing each individual treatment observation with the most similar (based on propensity score) non-treatment observation, at one extreme, to matching each participant to a weighted average of all non-treated observations, at the other (Jalan and Ravallion 2003; Heckman et al. 1997). The assumption represented by Equation (5), that unobserved factors do not correlate with both treatment and outcomes, may be a strong one, and it cannot be directly tested. Its validity will depend on the context, as well as on how comprehensive is the information on the characteristics X at the disposal of the researcher.

One reason PSM is used so often is that it is generally far less demanding of data and planning than experiments. Whereas RCTs typically involve collecting baseline and follow-up data (in addition to randomizing program assignment), many studies using PSM make do with a single cross section household survey with information on the outcome of interest and an indicator for participation in a program or use of a particular service such as vaccinations or antenatal care. This information is readily available, along with variables to serve as controls, in many standard household surveys such as Demographic and Health Surveys (DHS). However, combining PSM with baseline and follow-up surveys is highly advantageous as this can considerably reduce the potential from bias due to unobservable characteristics, as discussed below.

In principle, PSM can be readily used to assess impacts of programs that are already scaled up. Indeed, when working with existing household surveys like the DHS, this is the only type of program that would be adequately captured in such data, given that the samples are nation-wide rather than focused on particular communities or areas as would be the case with purposive data collection to evaluate smaller pilot projects. This is an advantage: by assessing impacts of projects that are expanded to national scale, this approach may not suffer from external validity problems that potentially compromise results of smaller scale experiments. At the same time, they it lacks RCTs' strong claim to internal validity.

Another non-experimental approach is regression discontinuity design. If eligibility for a program is defined on a threshold value of some indicator (e.g., household income for a training program), comparison of individuals just below and just above the threshold can provide an estimate of exogenous program impacts. There are not very many applications of this to health, at least in the African context. It requires that a program be provided on the basis of some easily measured continuous indicator, such as income or age.⁴

⁴ One health-related application is the study by Duflo (2003), who estimates the effect of pension receipt by grandmothers and grandfathers on height-for-age of girls and boys in South Africa. The analysis exploits the existence of age thresholds for eligibility (60 for women, 65 for men).

The intertemporal analogue of this approach, the ‘interrupted time-series’ design, involves looking for breaks in trends occurring at the time of program implementation. Whereas regression discontinuity designs are usually associated with individual-level interventions, the interrupted time series approach is particularly useful for interventions implemented for larger units, including region- or nation-wide programs. For these it will typically be difficult or impossible to identify an appropriate control or comparison group. Interrupted time series designs, which strictly speaking do not require comparison groups, will likely be the best option for evaluation in these cases. However, interrupted time series designs generally require several years of observation before and (optimally) after a program is introduced.

Another standard quasi-experimental approach is difference-in-difference (D-D) or its close regression analogue, group or individual fixed effects. It assumes that while treated and non-treated groups may differ in levels of potential outcomes (since treatment is not randomly assigned), the trend in outcomes for the two groups would be the same in the absence of the treatment. Using the same notation as above (though here C refers to ‘comparison’ rather than control group) and denoting the pre-intervention period with subscript 0 and the post-intervention period with subscript 1, the difference in difference estimates of program impact is:

$$E[Y_1^T|T] - E[Y_0^C|T] - E[Y_1^C|C] - E[Y_0^C|C] \quad (6)$$

which is the difference over time in outcomes for the treatment group minus that for the comparison group. Provided that trends (and any shocks) are the same for both groups, Equation (6) provides an unbiased estimate of the treatment effect. Typically, in applications the treatment and controls are defined as areas (villages, regions) receiving and not receiving an intervention. Unlike matching methods and like randomized designs, D-D does control for differences in unobservables across treatment and control groups—provided these unobservable influence only levels, not changes, in outcomes. This assumption cannot usually be tested.⁵ If, in addition, one matched individuals or villages from the control and treatment groups using propensity scores before applying the difference-in-difference, bias due to non-random assignment would be further reduced to the extent that it is also due to observables (Heckman et al. 1997, 1998; for an example, see Lokshin and Yemtsov 2005).

Many applications of D-D to evaluation of nutrition and health interventions will be discussed below, even if they are not explicitly described as such by the study researchers (they might be referred to as ‘observational designs with a comparison group’; they are D-D if data from multiple points in time are used to compare trends in the two groups). For example, a number of studies compare trends in child mortality or malaria in districts with a program and districts

⁵ Sometimes it can be tested indirectly. One way (as in Duflo 2001) is to compare trends in the outcome of direct interest in the two areas over a period prior to the period of study (and prior to the policies). If these do not differ statistically, the areas are comparable in terms of these movements for the study period as well on the assumption that the trend itself is stable over periods. The other approach is to focus on the period directly under consideration and consider trends in outcomes that would not *a priori* be affected by the policy in question but would nonetheless similarly reflect local preferences or shocks. For example, changes in health outcomes of older children should be correlated with that of younger children but not be directly affected by programs directed at infant or under-five mortality.

without it (where assignment not randomized). Often the trends in these outcomes are derived from regular surveys or sentinel site data rather than surveys carried out specifically for the evaluation.

Elaborations of this simple D-D framework allow a number of interesting applications to program evaluation. It is sometimes possible to assess not simply the effect of participation in or exposure to a program, but of the duration of exposure. For example, for a child health intervention introduced over (say) the prior decade, different cohorts of children would have been exposed to the program for different lengths of time, providing continuous variation in exposure. Since many indicators of child health (vaccination, birthweight) are collected retrospectively in surveys, the necessary variation in outcomes and exposure can sometimes be obtained from a single cross-section survey, when matched with information from other sources on the timing of the introduction of the program by area. Related applications have used household or sibling fixed effects. Here the variation in exposure is across children of different ages in the same family, thereby allowing control for fixed family, not just community, unobservable factors.⁶ As with PSM techniques, these applications of D-D allow estimation of program impacts for programs that are scaled up, provided there is variation by area in the presence (or else in the timing of introduction) of the program.

Finally, the traditional means by which economists deal with endogeneity is the use of instrumental variables (IV). This involves finding exogenous variables that affect T in Equation (4) but do not affect the outcome variable Y ; that is, conditional on included covariates X , the correlation of T_i and e_i is zero. Typically, the endogenous variable T is regressed on the instrumental variable(s) and other controls X in a first stage, and the predicted value of T generated from this regression is used in Equation (4). The validity of the IV approach hinges on whether the instruments are truly excludable from the second stage Equation (4). This is often hard to justify *a priori*, as there is usually some story that can be told in which the instrument would have a direct effect on the outcome Y . This is particularly a problem when trying to estimate the effect of treatment on the treated in non-experimental data: one needs a variable that affects individual or household participation in a program but is not correlated with the outcome that the program is intended to affect.

However, IV can more readily be used to estimate the effect of treatment on the treated in *experimental* data when the design of the experiment was intention to treat. In the example above of iron supplementation for children, the mean comparison of children's iron status and other health outcomes in treatment and control communities indicates the effect of the program on the target population of children—including those who participate in the supplement intervention and those who do not. In contrast, the health impact of actually receiving the iron supplement, the effect of treatment on the treated, cannot be measured simply by comparing participants to non-participants. Because receiving the supplement is voluntary, it has to be regarded as potentially endogenous to the outcome. On the other hand, an individual's *eligibility* for the treatment, determined by being a resident of an intervention community, is exogenous because communities were randomized to get the program. Hence

⁶Applications of D-D exploiting variation in exposure across cohorts include Frankenberg, Suriastini, and Thomas (2005); Duflo (2001); Pitt, Rosenzweig, and Gibbons (1993); Alderman, Hoddinott, and Kinsey (2006); Parker, Todd, and Wolpin (2006).

eligibility satisfies the exclusion requirement, and of course, it should affect T (treatment or participation in the program). Therefore the indicator for eligibility is a valid instrument for treatment and thus can be used to obtain an unbiased impact of the effect of treatment on health outcomes.

In non-experimental contexts, while as noted it is usually difficult to apply IV methods to estimate the effect of treatment on the treated, the possibilities are richer for estimating intention to treat impacts. Here one must deal with the endogeneity of program placement with respect to the outcomes of interest, hence find variables that can predict program placement but are not directly related to the outcomes. The idea is to find geographical, economic, or political factors that affect program placement. For example, Angeles, Guilkey and Mroz (1998) estimate the effect of availability of family planning programs on fertility in Tunisia, instrumenting program presence using area characteristics such as infrastructure (e.g., roads, other health care) and political ties to the central government. The assumption that these factors affect placement of family planning programs but are uncorrelated with fertility outcomes, however, seems questionable.

Several more recent IV studies deal with endogenous program placement using information on the *relative* characteristics of different areas. They are informed by the notion that governments make allocation decisions based both on the mean levels of various indicators (poverty, ethnic makeup, enrollment rates, environmental factors) in different provinces or other administrative units *and*—given that public resources are limited—on how the different areas compare or rank along these dimensions. This implies that, that for community or region j , the characteristics of other areas k , or how j ranks nationally (or subnationally) with regard to them, can be used as instruments for program allocations or timing. That is, the characteristics of *other* areas should not directly affect behavior *in j* conditional on these allocations. This approach is promising but as of yet there are few applications to health policy evaluation (see Pörtner, Beegle, and Christiaensen (2007) and Pitt and Menon (2007) for applications to family planning interventions).

2.3 Debates over Randomization

Long considered the gold standard for clinical research in medicine, in just the past decade or so RCTs have made considerable inroads into developing country research on a much broader set of health interventions. Indeed, randomization, forcefully advocated by researchers associated with the Poverty Action Lab at MIT among others, may be said to have gained a position of dominance in the field. The World Bank has to a significant extent embraced this view, and some donor organizations such as the U.S. Millennium Challenge Account explicitly favor randomized trials to evaluate their projects. As RCTs have gained ground, they have also attracted criticisms. In what follows I discuss the key issues that have been raised.

2.3.1 Threats to external validity

The most important criticism of RCTs is that they do not assure external validity, or generalizability of the experimental results—to other populations or, relatedly, to an actual

scaled up policy that attempts to reach the full target population in the country in question. First, the study population for an RCT is usually small and homogenous relative to the target population overall; mean response in the population may be different for that reason. This would also occur if study participants respond positively to the idea of being part of a new and innovative project. Second, once a project is scaled up, the quality of implementation may differ significantly from the initial experimental study. For example, it is very common for researchers conducting RCTs to work with Non-Governmental Organizations (NGOs), with the latter implementing the intervention for the experiment. In these cases it is not hard to imagine that the experiment uses a more dedicated, competent, or better-supervised staff than would a scaled up program using government (e.g., ministry of health) resources and personnel. Implementation quality would also weaken as an expanded program begins to run into human, infrastructural, and other resource constraints that would not have been important for a smaller program.⁷ There are many analyses of how the success of public health programs has been hindered by difficulties associated with implementation on a large scale. A comprehensive look at this issue was provided by the recent multi-country evaluation of the Integrated Management of Childhood Illnesses strategy discussed below in Section 4.3 (Bryce et al. 2005; Victora et al. 2005).

One response to this is that the NGO itself, with infusion of the necessary resources from donors, can implement the program on an expanded scale. Indeed, because NGOs and private providers are thought to suffer less from incentive and monitoring problems than the public sector, many have proposed relying on partnerships of government with NGOs (or with the private sector) to deliver basic health services (Leonard and Leonard 2004; Loevinsohn and Harding 2005). And of course, NGOs already play a large role in health delivery in many African countries. While if feasible this might (as a side effect, one could say) partially⁸ address the external validity problem, having non-governmental actors essentially replace the public sector in health care delivery has potentially important implications for the development of state capacity in the social sectors. This issue is discussed below in Section 3.2.3. An alternative is to try to conduct the experiment on a larger scale, say district rather than village level, and using the agency (e.g., the health ministry) that realistically will be the one carrying out the scaled up program.

Another response to the external validity criticism is that an experiment conducted in one setting should be repeated in many other environments to understand whether the findings are

⁷ These external validity issues should be kept in mind when considering studies that directly compare experimental and non-experimental approaches--studies that are often used to highlight the superiority of the former. Using data collected for experiments, these analysis (reviewed in Glazerman, Levy, and Myers 2002; and Cook, Shadish, and Wong 2006) are able to compare approaches with the knowledge that the experimental method provides the 'true'—that is, internally valid—result. Regression discontinuity designs seem to work well when the rules regarding thresholds are strictly observed (Cook, Shadish, and Wong 2006; Buddlemeyer and Skoufias 2003). As for the more common propensity score approaches, results are mixed, but there is a suggestion that the method can work well if the number of control variables used is sufficiently large (e.g., Diaz and Handa 2006). These are all comparisons of internal validity, however. One might want to know instead whether, for example, propensity score matching combined with difference-in-difference techniques on data from a large scale, government-administered program provide more or less accurate information on actual impact than an RCT on a smaller study sample. This question is impractical, if not impossible, to answer.

⁸ NGOs themselves would likely suffer some of the same problems in maintaining implementation quality when scaling up.

universally applicable or under what conditions they might be applicable (Duflo and Kremer 2008). Critics argue that this would not happen because researchers, especially academic ones, would have little interest in simply replicating previous studies in different populations (Rodrik 2008, Deaton 2009). This may be true, but it is not an argument against repeated experiments but rather an argument in favor of public subsidies for them, since while researchers may not perceive individual benefits, there is a public benefit. A more substantive critique of the ‘repeated experiments’ argument is that while conducting similar experiments in different contexts (meaning, typically, countries) indicates whether the experimental results are robust to change in contexts, it does not address the scaling up issues just described.

It is also often pointed out that small scale experiments will usually not be able to incorporate general equilibrium effects of an intervention. A classic non-health example is a training program that has a large impact on the earnings of participants in a small pilot study but has smaller effects if implemented broadly, because it increases the supply of individuals in the labor market with such training. In contrast, large scale (and possibly, longer term) non-experimental assessments of actual policies will, whatever their other shortcomings, be in a better position to capture these effects. In health, a number of important general equilibrium effects may be generated by non-pecuniary externalities. Some individuals will change their health behaviors because others have done so; some interventions will reduce illness prevalence in the non-participating population (from having fewer mosquitoes carrying malaria, an environment with less open defecation, fewer individuals with HIV infection, etc.). These impacts in turn may have further reverberations on behavior and health within the community. In many cases an appropriately designed experiment can capture these effects. The most obvious way is to randomize the intervention at the community level, where community is defined to be large enough to encompass the externalities.

The distinction between internal and external validity is frequently equated to that between an intervention’s ‘efficacy’ and its ‘effectiveness’, terms arising in the literature on medical interventions. In that context efficacy refers to whether a drug or therapy has been shown to work in ideal clinical (experimental) settings where patients and treatment are homogeneous, and compliance is perfect or nearly so. Effectiveness refers to the potential benefits—possibly smaller—of the drug in real world health care settings where these conditions do not hold. For the types of health interventions considered in this review, effectiveness would refer to the health impacts in the population of a program once it has been scaled up and is being delivered under routine conditions. In child health, early efficacy studies of interventions such as oral rehydration therapy and micronutrient supplementation led to overly optimistic expectations that childhood mortality in poor countries could be reduced cheaply and quickly. These expectations failed to take into account problems of weak health system infrastructure and human resources, lack of provider incentives, and lack of demand on the part of households arising from inadequate information or other sources—in other words, the range of obstacles to translating efficacy into effectiveness (Bryce et al. 2003, 2006). Since the efficacy studies were usually small scale RCTs, this corresponds to the standard external validity critique of experiments. In a variation on this argument, Easterly (2009) complains that the links between many interventions and health are often so obvious that RCTs are hardly necessary to show them, and by extension, are often not very useful in health policy research in developing countries.

These complaints against RCTs are not entirely valid, because the efficacy-efficiency distinction is not the same as the internal-external validity distinction. RCTs in health in developing countries are often most valuable precisely because they go beyond purely clinical assessments to address behavioral issues that can determine whether interventions are able to affect public health outcomes. As described in later sections of this paper, experiments have been used to learn what determines households' demands for specific health services or products (price incentives, promotion and education, externalities in uptake via social learning or imitation), as well as, on the supply side, to understand the role of provider incentives and monitoring mechanisms on the performance of health workers. All of these address issues related to modes of service delivery, uptake, hence ultimately, effectiveness. On the other hand, it remains the case that even behaviorally-oriented experiments still potentially suffer from external validity problems, because the behavior of households and of providers (and more broadly the quality of implementation) may be different under scaled up delivery of the interventions being evaluated.

2.3.2 Threats to internal validity

RCTs derive their power from being able to make valid causal statements about the study population, if not the target population overall. Further, by virtue of the randomization, it can do so with minimal statistical or econometric assumptions, unlike methods (difference-in-difference, instrumental variables, matching) that attempt to infer causality using non-experimental data. However, even the claims to internal validity are subject to qualification, as argued by Deaton in a recent paper (2009). First, there are several standard and well known threats, including contamination whereby some individuals in the control group (or living in control communities) are able to get access to the treatment, and the opposite problem whereby some in the treatment group do not participate. These clearly bias the differences in means of assigned treatment and control groups as a measure of the effect of the intervention. To some extent an appropriate study design can forestall these problems, for example, by insuring that in a community-randomized setup the treatment and control villages are not near each other. Post-trial statistical methods can also be employed, namely using the assignment to the treatment group as an instrumental variable to predict actual participation in a two stage procedure. One can also place upper and lower bounds on the estimates in the face of non-compliance or contamination (Manski 1990; Imbens 2009). If these problems are relatively limited, the bounded estimates still allow useful inference about causal effects.

A more subtle point made by Deaton is that experimenters frequently resort to more traditional econometric approaches, with attendant assumptions and judgments, thereby giving up the simplicity and nonambiguity of a pure experiment. Randomization, when valid, allows the researcher to answer the relatively narrow question of whether the treatment leads to a difference in the mean of an outcome variable for the treated group. Analysts typically are interested in much more than this, in particular, whether an intervention affects different subgroups differently (poor vs. non-poor, very ill vs. less ill, educated vs. uneducated). However, this opens up the possibility of data mining to find and present some results that are statistically significant. Researchers often also, as noted, estimate treatment effects in regression form including both the treatment indicator and control covariates. The latter will control (at least in part) for a randomization that is imperfect, and further, may lead to greater

statistical precision in the estimate of the treatment—an important benefit for many experiments where sample size is not large (see Duflo, Glennerster, and Kremer 2008). But this practice also is open to data mining (in the selection of control variables), and further, the model with covariates will yield a biased estimate of the mean treatment effect if this effect is heterogeneous (see the exposition in Deaton 2009).

The general point applies as well to the practice just noted of using treatment assignment as an instrumental variable for actual receipt of treatment. The IV estimate measures the mean effect of treatment on those individuals who change treatment status as a result of the instrument ('compliers', in this case, those for whom eligibility induces participation in the program). This is the Local Average Treatment Effect or LATE (Imbens and Angrist 1994). If the response to treatment of compliers is different from that of the target population overall, LATE will not provide an estimate of the mean effect of treatment for this population. This is a potentially important issue in interpretation since the mean effect on the target population for an intervention is usually of primary interest. How important a concern this is depends on the expected level of heterogeneity in treatment effects, which cannot be known. This is no more or less of a problem than for IV estimation on non-experimental data, but like those approaches it entails supplemental assumptions or judgments, thereby losing the simplicity and clarity of a pure experiment.

However, while experimenters frequently do resort to assumptions and judgments that are similar in kind to those they criticize in non-experimental studies, the problem of reliance on questionable assumptions is also a matter of degree, and here RCTs are generally in a better position. For example, IV estimation on non-experimental data relies on a maintained untestable assumption about one or more instruments: that these variables affect the variable of interest (the treatment or program) but not the outcome of interest (effect of the treatment). In many—even the majority—of cases, these assumptions are questionable, even for a natural experiment, and in any case, they cannot be tested.⁹ In contrast, an experiment—provided the treatment assignment is truly random—at least has a fully credible exogenous variable, and potential instrument, to start with, even if additional assumptions need to be made in the analysis. Similarly, the essential assumptions of propensity score matching and difference-in-difference techniques are ones that an RCT does not have to make. It should be kept in mind that all this refers only to internal validity. The problems with external validity noted above—and the relative benefits of experimental vs. non-experimental data in terms of external validity—remain.

2.3.3. Other critiques

A different concern with RCTs, voiced by Deaton and others (e.g., Heckman and Smith 1995), is the 'black box' critique: experiments prove that an intervention 'works' (at least in the study population) but do not show why. RCTs, in other words, are not structural, and often are inadequately informed by theory in the sense that empirical analysis should involve testing theories of economic behavior. However, this is a problem primarily with the way

⁹ Exogeneity tests like the Hausman test and the Sargan-Hansen test rely on having at least one valid instrument as a maintained assumption.

many experiments have been carried out rather than an inherent shortcoming. Many researchers are designing experiments (or in some cases, using them *ex post*) to address behavioral or theoretical questions in microeconomics. Attanasio, Meghir, and Santiago (2005) use the exogenous program assignment in Mexico's Progressa conditional cash transfer program to identify dynamic structural models of education choice. The experimental study of farmers and fertilizer use in Kenya by Duflo, Kremer, and Robinson (2006), discussed below, was designed to explain the role of social learning in the adoption of new technologies. Several other experimental studies are directly informed by new insights from behavioral economics, including analyses of savings behaviors (Ashraf, Karlan, and Yin 2006; Duflo, Kremer, and Robinson 2006) and smoking cessation (Giné, Karlan, and Zinman 2009), and empirical tests of the effects of sunk costs on use of a purchased product (Ashraf, Berry, and Shapiro 2008; Cohen and Dupas 2008). Analyses like these can provide insight not just into whether an intervention 'works', but why and how it works (or not). At the same time, it must be recognized that there will often be a conflict between the objectives of policymakers and donors, on the one hand, and researchers on the other. The former (especially, policymakers in a given country) are likely to be primarily concerned with what works in a specific context, while the latter will also want to investigate more general economic questions which may involve more complex and expensive designs.

Finally, as more and more emphasis is given to conducting experiments, an emerging concern is that many important policy questions in development that are not amenable to randomization will be given short shrift by researchers and funding agencies. Often the distinction is drawn between interventions within the health and education sectors on the one hand and broader policy changes that are essential for development on the other; the latter generally fall outside the feasible scope of experiments (Rodrik 2008). So, too, will many interventions in sectors such as infrastructure and finance. Since we focus here only on interventions within the health sector, this argument is less of a concern. Still, many policies in the health sector have not been or cannot be addressed using randomized trials. These include major overhauls of health sector financing and organization as well as national mobilization strategies for behavior change. For these types of large scale changes (for the most part not discussed in this paper), observational evidence on changes at the country level or cross-country comparisons are usually required.¹⁰

¹⁰ Still, RCTs or sophisticated non-experimental approaches can and have been used to evaluate fairly systemic changes on the supply side, including the introduction of provider pay for performance schemes and contracting with private providers. These are discussed in subsequent sections.

3. HOUSEHOLD AND PROVIDER HEALTH BEHAVIOR

3.1 *Demand Side Behavior: Households*

This section examines different aspects of the problem of the delivery and uptake of health services, drawing on African evidence to the extent possible. It considers first the extensive evidence on how households respond to the cost of health services. It next considers the role of service quality, with and without concomitant changes in cost. It concludes with a discussion of the relatively less researched question of externalities in household health and demand behavior occurring through social learning and other mechanisms.

3.1.1 Price

How households respond to the cost of health services is an important issue for policy. A consensus exists that for many people—the poor, and possibly also those with inadequate knowledge of the benefit of health care—and for many types of services, health care should be subsidized. This reflects a common view that health status is inherently valuable and/or that it generates significant positive externalities. The question is whether it is good policy to require households to contribute some fraction of the costs of providing these services via user fees, or if the services should be free—or even have a negative price via a direct subsidy to households to use a service.¹¹ The main rationale for user fees (“cost sharing”) is that the revenues thus provided can be used to improve quality or increase the supply of drugs, leading to higher utilization and better health outcomes despite the higher costs to households. In the late 1980s, when many African countries were undergoing structural adjustment and governments were severely resource-constrained, the idea of cost-sharing for basic health care services gained considerable appeal. It was codified in the Bamako Initiative arising out of a meeting of African Health ministers in 1987, which also stressed the potential benefits of decentralization and community control of the revenues generated by fees. Cost-sharing in health care was strongly pushed by the World Bank and other international agencies during the period of the late 1980s through the 1990s, and most African countries adopted the practice.

Other rationales exist for positive pricing of health services or goods such as mosquito bednets and water purification products. Charging something plausibly will reduce waste by screening out those who do not really value or need the service or goods. Somewhat in contradiction to this hypothesis, it is also argued that fees or positive prices will increase demand and (for health products) use, because people associate a positive price with higher quality. With regard specifically to products such as bednets, water purification, and condoms, it is argued that in psychological terms, having already paid for something makes people more committed to using it. This ‘sunk cost fallacy’ is a major part of the rationale for NGOs and other organizations charging for these products (see e.g., Population Services International, 2003).

¹¹ It should be kept in mind that private costs to households can be significant even when services are nominally free, and include direct payments for medicines and transportation (and possibly, informal payments to providers) and indirect costs in terms of work time lost to travel for care.

Focusing first on the quality improvement rationale, the effect of cost sharing will depend on the responsiveness of demand to price, in terms of both utilization and intensity (and how this varies by income level), and the responsiveness of demand to quality. It also will depend how the supply of the good or service, and its quality, responds to price. The fairly limited evidence at the time of the Bamako initiative suggested that households' price elasticity of demand for basic health services in developing countries was low (Heller 1982, Akin et al. 1984), leading to some confidence that the negative effect of user fees on demand would be small. However, subsequent research suggests that price elasticities are in fact quite large.

First, the introduction of cost-sharing in a number of African countries after Bamako was followed by large declines in health service utilization, especially among more vulnerable groups (Blas and Limbambala 2001; Frankish 1986; Hussein and Mujinja 1997; Kipp et al. 2001; Meuwissen 2002; Mwabu, Mwanzia, and Liambila 1995; Waddington and Enyimayew, 1989; Waddington and Enyimayew 1990). Exemption measures to protect the poor from the burden of fees were often ineffective (Gilson, Russell, and Buse 1995; McPake, Hanson, and Mills 1992; Willis and Leighton 1995). Most of these analyses consist of simple pre-post comparisons of facility uptake before and after cost-sharing went into effect. In most cases, the full range of policy changes envisioned in the Bamako Initiative were not implemented in that price increases were not accompanied by effective community control of health care and the use of fee revenues to improve drug supply or quality of care. This means that these findings essentially show (subject to caveats about not having comparison groups to control for other factors that may have been changing) the implications of raising costs alone.

In large part, based on these outcomes, cost sharing has since fallen out of favor in Africa and among donors, including the U.S. government.¹² The (for the most part) more recent experiences of countries that eliminated user fees—for primary education as well as basic health care—provide a further source of evidence of price responsiveness of the demand for services. These policy shifts were followed by large (often extremely large) increases in utilization. In Uganda, where user fees in public facilities for basic health services had been instituted in 1993, there were major increases in utilization following their removal in 2001. Data from 78 health facilities in 10 districts, collected 8 months before and 12 months after the fees were eliminated, reported new visits increasing by 53% on average and repeat visits by 24% (but much more for children) (Burnham et al. 2004). The gains appear to have been largest among the poor and were sustained over time (Yates 2004; Deininger and Mpuga 2004). While the Uganda experience suggests large price elasticities, it is not a valid natural experiment, since resources allocated to the health sector also increased substantially at the same time, presumably affecting access and/or the quality of care.

Nevertheless, experiences were qualitatively similar in other countries where fees (mostly for schooling) were removed and where complementary changes in resources did not always occur. In the last 15 years, in Uganda, Tanzania, Malawi, Madagascar, and Kenya, such

¹² In a reversal of the previous stance of the U.S. Government, in 2001, the foreign appropriations bill report required the U.S. Congress to oppose any World Bank, IMF, or other multilateral development bank loan which included user fees for basic health or education services.

policies resulted in sudden and very large surges in enrollments. In some cases, these changes occurred in the context of other moves to promote education (as in Uganda) or recovery from economic and political crisis (as in Madagascar); in others (such as Kenya), they did not. Taken together, they suggest that households in poor countries are highly sensitive to the costs of services.

Next we come to evidence from randomized experiments. A significant number of experiments involving pricing of services (primarily in health and education) have now been carried out in developing countries. As Holla and Kremer (2009) note in their review of these studies, a very consistent finding is that small changes in cost leads to large reductions in take-up. This was also true specifically for studies of health interventions in African contexts.

For example, as part of the evaluation of a school-based deworming project in Western Kenya (Miguel and Kremer 2004), a subset of the schools was randomly selected to charge fees for the drugs (U.S. \$0.30 per family). Uptake in these schools was only 19 percent compared to 75 percent in the free treatment schools. Nor did the imposition of a positive price lead to relatively larger uptake among those in greatest need, i.e., those most likely test positive for intestinal helminth (worm) infection.

Two randomized studies of insecticide-treated mosquito bednets, one in Kenya (Cohen and Dupas 2008) and the other in Uganda (Hoffmann 2008), also found uptake to be very sensitive to cost. In the first study, when antenatal clinics offered free nets to women, take-up was virtually 100 percent, but declined sharply as price was varied upward from zero in other randomly selected clinics. Take-up was 61% lower when the cost was equivalent to U.S. \$0.60. Following the methodology of Ashraf, Berry, and Shapiro (2008), Cohen and Dupas use a two stage approach that allows them to distinguish screening effects from sunk cost effects.¹³ The presence of both effects is rejected. Regarding screening, charging for the nets did not increase uptake among those in greatest need (as measured by anemia, a sign of malaria). Regarding sunk costs, variation in price among those who had already decided to purchase nets did not affect the probability that a net is actually hung in the home. Hoffman (2008) also reports that free nets are as likely to be used as those that are paid for. Furthermore, free nets were more likely to be used for young children, for whom the need is greatest.

An experiment in Lusaka, Zambia (Ashraf, Berry, and Shapiro 2008) considered the role of pricing in a door to door promotion to get households to use a chlorine-based water disinfectant. As with the studies just cited, charging higher prices led to a large progressive reduction in take-up. Also as in those studies, targeting to those most in need (in terms of poor initial water quality) did not improve when households were charged for the product. On the

¹³ In the first stage, women visiting antenatal clinics were offered nets at varying prices, including zero, with the price variation at the clinic level. This variation is used to estimate the effect of price on take-up and how this effect varies by individual characteristics (including need as proxied by anemia), hence to test the screening hypothesis. In the second stage, discounts of varying amounts off the initial clinic price were offered to women who had agreed to buy at the clinic price. This (random) variation in price among committed buyers allows a test of whether sunk costs are important to actual usage. That is, among women who had agreed to buy the nets at the same initial price, the sunk cost hypothesis implies that use will be lower among those whose realized (second stage) price is lower.

other hand, there was some evidence that charging increased the propensity for actual use (the sunk cost hypothesis), as measured by random checks of chlorine in households' drinking water.

Policies can induce uptake not just by reducing prices or setting them to zero, but also by paying people (in cash or in kind) to use a service—that is, the price can be negative. In an experimental study in rural Udaipur, India (Banerjee et. al. 2008a) once-monthly mobile 'camps' offering immunizations for young children were introduced in randomly selected villages. In one treatment arm, the camps also offered women an incentive in the form of a kilo of lentils (and several metal food plates) for completed immunizations. Though the incentive was small (the lentils were worth about US \$1.00), it led to a more than doubling of the fraction of children who were fully immunized relative to villages with the immunization camps but without the incentive.

An African example is an experiment involving HIV testing in Malawi (Thornton 2008). Nurses visited households and administered free HIV tests, and the results could be obtained at temporary voluntary counseling and testing centers at various distances from the household. Some individuals were additionally offered vouchers, the values of which were varied randomly, that were redeemable upon retrieving the results at the centers. The decision to get the results was very sensitive to the availability and amount of the voucher. Those receiving a voucher were twice as likely to get their test results, and the probability of getting results increased about nine percentage points for every additional dollar offered.

The idea of providing cash incentives for using social services—conditional cash transfers (CCT)—is growing in popularity in development policy following the success of the *Progresa* program (now *Oportunidades*) in Mexico and several similar programs in Latin America. These programs, unlike the Malawi experiment, combine demand side incentives to use health and education services with the objective of providing income support for poor households. Typically, cash payments are made to households (or specifically, mothers) regular conditional on regular school attendance and health visits for children (and sometimes, pregnant women or even all adults). Health-related outcomes from these programs (usually assessed using RCTs) have been generally positive, especially with regard to increased visits to health centers. In Mexico, Brazil, and Nicaragua, CCT programs led to more frequent nutrition monitoring, higher immunization rates, and higher child growth (Gertler 2004; Maluccio and Flores 2005; Rawlings and Rubio 2005). The *Progresa* program also reduced child morbidity. The nutritional supplement component of these programs was less successful. In Africa, CCT programs have been slower to take hold. There are some constraints to operating CCT programs in African contexts. In particular, since these are programs of regular monetary transfers to the poor, the resources involved are significant, as are the administrative burdens. Still, a number of pilots are under discussion or being implemented in Africa, for example in Tanzania and Uganda.

The experimental studies just described, and others in education, confirm the earlier non-experimental evidence that household demand is very sensitive to the costs of services. Further, since many of the experiments randomly allocate individuals or communities to varying price levels (including no price), it is possible to determine whether a specific

importance attaches to having a positive price—i.e., where there is a discontinuity in the demand curve at zero. A survey of the experimental evidence, including the African studies just discussed, suggests that this is the case: a larger reduction in use occurs when going from zero to some small price than from equivalent increases at non-zero prices (see Holla and Kremer 2009). This contradicts the often expressed belief, which informs the perspective of social marketing, that it is important to charge at least something so that people feel the product is valuable and will use it.

Why price elasticities are so high is an interesting question. Most households in these study contexts are quite poor, but the changes in cost that were seen lead to such large changes in uptake are still often just a tiny fraction of their annual incomes. Presumably, rational individuals would weigh this cost against the potentially large gains to individual or family health. It is possible that, despite the information that is often provided along with the offer of the good or service, households do not fully understand the benefits, so they are not willing to sacrifice even a small amount to obtain them. If this is the case, education efforts need to be strengthened. Another possibility is that people discount the future extremely highly, so that even prices that imply only small reductions in current consumption strongly reduce demand for goods or services that could substantially raise welfare in the future. Modern behavioral economics supplies a more plausible variant of this that is stressed by Holla and Kremer (2009): individuals may have time-inconsistent preferences, meaning that while they do not heavily discount the future, discount rates are much greater in the short run than the long run. This would explain why individuals appear to respond so much more strongly to short term costs than long term benefits.

So far we have been discussing evidence of the effect of price changes while holding (by and large) other things constant. Two important ‘other things’ are: changes in service quality that also affect demand, and changes on the supply side, which also can be expected to respond to changes in price. The next subsection discusses evidence for the effects on demand for changes in quality. Supply side responses to prices are discussed in Section 3.2.

3.1.2 Bamako once again: price increases with quality improvements

The evidence discussed above indicates that imposing user fees can have large negative effects on the utilization of health services, especially among the poor. While this constitutes a strong warning that cost-sharing policies can have detrimental impacts, it should be recalled they these are estimates of price effects only. In contrast, the Bamako Initiative and proponents of cost sharing envisioned the use of fee revenues to improve service quality, usually via community co-management of health care (including access to and control over local fee revenues). There is evidence that when quality does improve, households will be willing to pay more for services.

A study by Litvack and Bodart (1993) for Cameroon took advantage of the staggered rollout of the change to community financing to compare three health centers where the system was introduced to two others where it had not yet been (though the selection of health centers/communities into the two groups was not randomly assigned). Baseline and follow-up household surveys in the catchment areas of each facility indicated that the probability of

using the health center increased significantly for people in the treatment areas compared to those in the control areas. This is attributed to the greater availability of drugs in the health centers charging fees. Further, utilization in the ‘treatment’ areas households rose the most for poorer households.¹⁴

Other case studies are, like Litvak and Bodart (1993), non-experimental, but also often lack comparison groups and instead rely on simple pre- and post comparisons of utilization and other outcomes using health facility information or household surveys (as is also the case for much of the counter-evidence cited above for reductions in uptake after the introduction of cost sharing). An analysis in Niger (Diop, Yasbeck, and Bitran 1995) of a pilot test of alternative community financing schemes (which did have a comparison group) examined the effects on uptake of two strategies relative to standard free service: having centers impose relatively high user fees, and having smaller fees while also imposing a small tax on all residents to support the health system. Utilization increased among the poorest 25% of the population following the introduction of community financing schemes compared to centers where free services were offered, where utilization actually fell over the period. The increases in demand were even larger where fees were kept low and the health care tax was imposed. As in the Cameroon case, the gains in utilization were attributed to quality improvements financed by the revenue from fees (or taxes).

Kipp et al. (2001), using a simple pre-post test design, studied the impacts of the introduction of community-controlled cost sharing in Uganda, whereby user fees were used to finance incentive payments to health workers (it is not clear if or how these were based on performance) and possibly also to improve drug supply. Results on utilization for outpatient services were mixed: it fell in urban and semi-urban areas (possibly reflecting a shift to private providers) but increased in more remote rural areas. Overall, there was a decline in utilization. This is consistent with the substantial increase in health care utilization captured by national household surveys in Uganda following the later removal of user fees in 2001, as already noted.

Soucat and Wagstaff (2001) contrast the experiences of Benin and Burkina Faso, two West African countries that introduced community financing via user fees in different manners. The two countries are similar in terms of income level, demography, and disease types and prevalence. In Benin, prices were set differently for different services: they were higher (though still moderate) for most curative care services and lower for preventative care such as immunizations, antenatal care, as well as for other services for children for which price elasticities were thought to be high. National household surveys in 1988 and 1993 from Benin show very large increases in antenatal care, immunizations, and even curative care. The 1993 utilization levels were substantially higher than in Burkina Faso, where fee levels were not differentiated and overall, were higher than in Benin. Unfortunately, the changes in Benin are not compared to changes in Burkina Faso over the same period. Assuming however that the

¹⁴ It should be noted that this study, like many early community level experimental and non-experimental evaluations, had relatively few communities in each study arm and did not adjust standard errors of the estimates for the clustering of observations in communities. Since there were only five communities in the study, this adjustment would likely have substantially reduced the precision of the estimates, affecting the inferences that were possible from the data. See Hayes and Bennett (1999).

utilization differences in 1993 in the two countries are indeed policy-driven, this probably was due in part to how effective each country was in allocating fee revenues to quality improvements and to insure that the costs were kept low for the poor. For example, Ridde (2003), studying one district in Burkina Faso, reports that local management committees were not effective in releasing funds to ensure access by the poor; in this district, the introduction of fees led to reductions in utilization.

The Benin policy is consistent with the perspective on user fees for primary health care laid out in Filmer, Hammer, and Pritchett (2000, 2002), which has formed the basis for recommendations by the World Bank and others. Subsidies should be high (fees low or zero) for interventions with significant positive externalities such as treatment of infectious diseases, immunizations, health education, and preventative care.¹⁵ Subsidies should also be high for services for which the elasticity of demand for public provision is high—possibly because there are few private substitutes—so that reductions in price yield large increases in demand, hence health impact. In contrast, subsidies should be low (fees higher) for less essential but common types of (mostly curative) care, for which the *net* impact of public subsidies on health outcomes is relatively low—in part because public provision to a large extent merely crowds out private markets that are easily able to offer these services.

These are efficiency arguments and are well grounded in the theoretical framework of public economics. The equity implications of this pricing scheme are not clear. A key concern would be that the poor would be less likely to access private providers for services where fees were set high; their net (from any source) utilization of these services would fall disproportionately when prices at public centers were increased. On the other hand, as Filmer et al. (2002) point out, high subsidies for interventions to prevent communicable diseases, vaccination, and education are likely to disproportionately benefit the poor who, for example, currently suffer more than the well off from exposure to most communicable diseases. Rigorous evaluations of this approach in African contexts are needed, with particular attention paid to equity impacts.¹⁶

The evidence discussed in this section suggests that the negative effects of higher prices, even among the poor, can be offset, or more than offset, by concomitant improvements in service quality. This was the thinking behind the Bamako Initiative. However, it is not at all clear that the community co-management of funds to insure improved local facility quality or drug availability, which was an essential component of the Bamako Initiative, can be implemented effectively in all contexts. The same applies to measures to insure that the poor have lower cost access. The negative case studies cited earlier suggest that this will be difficult in many contexts. Such outcomes are not surprising in view of the experiences in Africa with decentralization (of which local control of revenues is one form), which has at best a mixed record in terms of impacts on poverty and service delivery (Von Braun and Groote 2002; Crook 2003; Jütting, Corsi, and Stockmayer 2005). Obviously, strategies for cost sharing

¹⁵ Preventative care is characterized by market failure because it is usually less lucrative for private care providers to provide than curative treatment, and further may be under demanded because of lack of health knowledge.

¹⁶ A policy experiment currently being conducted by the World Bank in Rwanda, combining differential service pricing with provider performance incentives, may shed light on this issue.

need to take into account local capacity for managing these funds, and if this is lacking, employ more centralized approaches for providing resources to improve quality in health facilities. It is clear that as a stand-alone policy, fees for basic health services in poor environment will likely have detrimental impacts on uptake.

3.1.3 Other evidence for household response to service quality

The varying experiences with cost-recovery with and without quality improvements suggests that consumers in poor countries are aware of and respond to at least some dimensions of health care quality. Recognition of the importance of service quality, and of evident shortfalls in this dimension in resource-stretched public health sectors, has led to increasing research focus on quality—its measurement, its determinants, and its effects on demand. Some evidence on policies to improve health care quality will be discussed later. On the demand side, a number of studies have estimated the effects of provider characteristics on utilization and health outcomes in African contexts using household survey data matched to information on local health facilities. Interpretation of the results from these studies must take into account several possible sources of bias. Health service quality may be positively correlated with unobserved health practices, knowledge, or preferences in the community that lead to better health outcomes or higher demand for care independently of quality. This would bias upward the estimated impacts of quality. Alternatively, the authorities may choose to direct more resources to areas where these factors are in short supply and health is poor, implying a downward bias. Further, some indicators of quality, particularly supplies of drugs and other items, may reflect not just supply characteristics but demand as well. Where utilization of public facilities is high, for example, the stocks of many drugs may frequently be depleted, so OLS estimates will underestimate the magnitude of any positive effects of drug availability on demand. In addition, where health programs are effective (for example, in preventing child diarrhea), demand for care might appear low simply because people are ill less frequently or for shorter periods of time.

That said, the cross-section evidence does suggest that households respond positively to health care quality. Surveying a range of studies, including several from African settings (Ghana, Kenya, and Nigeria), Alderman and Lavy (1996) note, in particular, the evidence of a positive response of the demand for care to drug availability. In Côte d'Ivoire, Thomas, Lavy, and Strauss (1996) find large positive effects of drug availability measures on child heights. These findings are consistent with the experimental study of Cameroon noted earlier (Litvack and Bodart 1993), showing higher utilization relative to controls in clinics that imposed fees and used the revenues to increase drug supplies. It should be noted, however, that the apparent robustness across studies of the findings for drug availability compared with other dimensions of provider quality could reflect the presence of greater measurement error in the latter. For example, because of often very high absenteeism (Chaudhury et al. 2006), the numbers of actually available professional staff (doctors, nurses, technicians) may be severely mismeasured if surveys rely on official employee rosters rather than recording actual attendance. Thomas, Lavy, and Strauss (1996), using data on both official and actual staff availability, report that in some cases using the latter, more accurate, measure substantially altered the magnitudes or even the signs of the effects on child heights.

More recent efforts to measure health service quality using facility surveys have not only been careful to record absenteeism of personnel, they have also significantly expanded the scope of quality-related questions that researchers can address. An example of this type of instrument is the World Bank's Quantitative Service Delivery Survey (QSDS) which has been implemented in several African countries. The QSDS includes detailed information on supplies and other inputs, facility condition and amenities, and organizational structure and management, including policies used to reward good staff performance or punish bad performance. Further, while earlier surveys typically relied principally on measures of structural quality (number and qualifications of staff, drugs availability, equipment, etc), a number of more recent facility surveys also consider indicators of process quality, referring to the care practices of the practitioners in a facility (see Donabedian 1980). While some information on process quality can be garnered from interviews with clients, accurate assessments of the quality of treatment can only be made by medically trained personnel. This has been done via direct observation of practitioners treating patients, or by using structured interviews in which the interviewer asks practitioners how they would treat patients with given symptoms (the 'vignettes' approach) (Leonard and Masatu 2005).

Mariko (2003) examines the choice among providers in Bamako, Mali, utilizing matched household and facility information, the latter having data on both standard structural attributes and process quality (the latter represented by an index of 'good' consultation practice based on direct observation). Among structural factors, drug supply increases the probability a facility is chosen as seen in many earlier studies. The process measure of good consultation practice is also associated with higher demand. It is also found that leaving out the process measures leads, via omitted variable bias, to misleading estimates for the effects of some structural attributes, namely number of staff.

Using detailed facility data from a district of Tanzania that included information on patients, Leonard, Mliga, and Mariam (2002) examine the role of quality in the health care seeking behavior of individuals, focusing on the pattern of 'bypass' whereby individuals will seek care at a facility other than the one closest to them. Both structural and process quality information was collected, the latter including direct clinical observation of practitioners. Patients generally choose providers with higher consultation quality as well as those who give more appropriate drug prescriptions. Further, health care customers seem to seek higher quality care for those illnesses where this is especially important, though there are exceptions to the pattern. Glick (2009), using matched facility and household data from Madagascar, finds that consumer satisfaction and perceptions of quality reflect both objective structural measures of quality and process measures. Similarly, Leonard (2008) finds patient satisfaction to be a positive function of an index of several aspects of practitioner behavior in Tanzania.¹⁷

¹⁷ Glick, however, finds that while patient satisfaction with consultation increases with several measures of process quality that patients can easily observe (such as whether they were given a physical exam), these measures are uncorrelated with quality recorded in direct clinical observation at the same facilities (though not necessarily with the same practitioners). The latter, which involve making correct diagnoses and giving appropriate prescriptions, are not easily observed by patients.

These analyses suggest that households at least to some extent are able to recognize health care quality, and that their demand behavior is affected by quality. Hence these studies are in accord with earlier demand modeling using less comprehensive facility data. To my knowledge, data on quality including process indicators of appropriate clinical practice have not been used in the evaluations (randomized or not) of specific health care reform interventions or programs in Africa. Such analyses would be very valuable, particularly for policies that affect provider incentives. This is because the process measures discussed above are not merely a function of provider competence and training but also provider motivation and effort (see Leonard, Masatu, and Vialou 2007) which will be affected by incentives. It is important to know these impacts of policies, not just the effects on the number of patients of different types seen—indeed it is possible that there is a conflict between improvements in the quantity and quality of services delivered, on the one hand, and their quality, on the other.

3.1.4 Externalities and social learning

The level of uptake and public health impacts of many interventions may depend strongly on externalities. This is especially the case for communicable diseases and can involve both prevention and treatment behavior. For example, a household's use of bednets to prevent mosquito bites will benefit not just those in the household but also others living nearby via an overall reduction in the probability of malarial infection. Similarly, use of condoms by an individual with multiple sex partners reduces not just that person's chances of contracting HIV, but the chances of his partners and their other partners. On the treatment side, an individual's therapy for TB, or sexually transmitted infections such as syphilis, reduces the likelihood of infection for others. Where positive externalities such as these are significant, there is a strong rationale for public subsidization of these activities.

The externalities associated with health interventions need not be positive, however. In the condom example, if some people are convinced to use condoms, others may be less inclined to do so (or to protect themselves in other ways such as having fewer sexual partners), since use by others lowers their own infection risk via reductions in overall HIV or STI prevalence. This principal applies to treatment or prevention for other types of infections that can be passed on the others (e.g., intestinal worms, malaria). In these examples, the externalities are negative in the sense that health protecting behavior by one individual leads to less of such behavior by others. It should be kept in mind, however, that these negative *behavioral* responses are themselves the product of positive *health* (or epidemiological) externalities that reduce risks for all.

Clearly, it is important for evaluations to insure that measured outcomes incorporate positive or negative spillovers in order to understand the true impacts of a program. As noted earlier, externalities are captured implicitly in many existing program evaluations, in particular community-randomized designs in which outcomes are measured at the community level (e.g., village level uptake of a program, or communicable disease prevalence or incidence among direct participants and non-participants). Indeed, community-level designs are often recommended precisely because they include local spillover effects.¹⁸ On the other hand, such

¹⁸ Such designs essentially use an intention to treat framework implemented at the community level.

studies usually do not provide direct insight into the extent of the externalities or how they operate, which if known could be used to design policies to maximize the positive impacts on public health.¹⁹ This requires more complex study designs. So far there are only a few examples of such analysis in Africa (or elsewhere).

An important study is Miguel and Kremer's (2004) randomized evaluation of a project in rural Kenya to provide school-based mass treatment with de-worming drugs. The project was designed to improve children's health, school attendance, and academic performance by reducing intestinal worm infections. These are spread from one child to another through several means, including open defecation in the bush or through contact with infected fresh water when bathing or fishing. Since the randomization was at the school level, spillover effects from children who comply (take the drugs) to those who do not are captured in comparisons of school level mean outcomes of treatment and control schools. Further, given the modes of transmission of intestinal worms, externalities may extend to children living nearby who attend other schools. Miguel and Kremer measure these impacts using variation in the local density of treatment school pupils. That is, the effects on children in neighboring schools will depend on the density of treated schools in a broader area and on the number of children assigned the de-worming drugs in the treated schools. Both variables are plausibly exogenous because of the random school assignment. The program was found to reduce school absenteeism in treatment schools by one-quarter, though there were no measurable effects on test scores. Further, attendance impacts on children in neighboring schools were large, indicating positive externalities of treatment. Ignoring the latter would have substantially underestimated the overall benefits of de-worming, and would have provided misleading estimates of cost-effectiveness relative to other strategies to increase school participation.

A very different example of modeling treatment externalities is the study by Thirumurthy, Graff Zivin, and Goldstein (2008) of the impacts on children of having a parent receiving antiretroviral (ARV) therapy for HIV/AIDS, also in Kenya. The treatment significantly extends the life of the recipient and in a prior analysis was shown to have strong positive effects on labor supply (Graff Zivin, Thirumurthy, and Goldstein 2006). Random assignment of ARVs to some individuals and not others is not possible ethically, so a non-experimental approach was required. The researchers have baseline data (at treatment inception) both from households with ARV patients and from other randomly selected households, so they can compare changes in nutritional status and school attendance of children in ARV-patient households with other children, a difference-in-difference approach. Relative to the comparison group, children's weekly hours of school attendance increased by over 20 percent within six months after treatment was initiated for an adult household member. Large gains in young children's short-term nutritional status (weight-for-height) were also seen.

Individual participation in health interventions can also generate externalities via processes of social learning or peer effects. Either through direct interactions with participants or simple

¹⁹ If randomization is at the cluster (community, school, etc.) level, the random assignment allows for identification of the mean cluster level effect only. It does not permit a decomposition of that effect into direct impacts on participants and indirect effects on non-participants.

observation, others may also decide to participate in a new program or to change their health behaviors in similar ways. In this way aggregate demand impacts would be larger than individual level impacts. This process has long been studied as a factor in the spread of modern contraception (Montgomery and Casterline 1996; Behrman, Kohler, and Watkins 2002) but has much wider potential applicability. There do not seem to be examples for Africa of health-related program evaluations that have tried to explicitly model these pathways.²⁰ An example from elsewhere of this type of analysis is the study by Dickinson and Pattanayak (2007) of social interactions and the adoption of latrines in Orissa, India, discussed below. The concept of social learning has also informed evaluations of several non-health interventions in African contexts, in particular related to farmer adoption of new technologies for agricultural production. These have used both non-experimental approaches to identification (Conley and Udry 2001) and experimental designs (Duflo, Kremer, and Robinson (2006).

The Kenya experiment described in Duflo, Kremer, and Robinson is an example of how a well thought out study design can solve the key identification problem in estimating causal peer effects. In a simple OLS framework, estimates of individual adoption behavior as a function of (say) the mean adoption rate among members of an individual's social network will be subject to serious potential biases. The estimates may pick up the effect of common unobserved traits in the group affecting the behavior of all members, or suffer from reverse causality due to the influence of the individual's behavior on that of others. In the Kenya experiment a randomly selected a group of farmers was provided with fertilizer and hybrid seeds sufficient for small demonstration plots on their farms and was also given instruction in their use. Surveys in successive seasons recorded their subsequent adoption (or not) of the fertilizer, and also that of other persons listed as 'contacts' by these farmers. Since the selection of the recipient farmers was random, the design also effectively randomly selected non-recipients (the contacts) to have someone they know receive the intervention. The behavior of the latter could be compared to randomly selected farmers in control villages that did not receive the intervention, hence did not have contact with someone getting the fertilizer/seed/training intervention.²¹ This approach can be transferred to the analysis of health related interventions that are designed to change knowledge or behavior. Learning about externalities and social learning in health is an important area for future research.

3.2 Supply Side Behavior

The bulk of empirical and microeconomic modeling of behavior related to health in developing countries, including evaluations of programs, has focused on demand side behavior and health outcomes for individuals. However, empirical analysis, and in particular program evaluation, has also increasingly focused on analyzing the impacts of policies designed to affect the supply of services—both their quantity and quality. This focus (in both

²⁰Research has been done on the role of social networks in changing health-related behavior outside the context of program evaluation, especially with respect to HIV/AIDS (e.g., Bühler and Kohler 2003).

²¹ In the event, while the intervention did have a small positive effect on treatment farmer's probability of using the fertilizer, knowing a treatment farmer had no effect on other's uptake—i.e., social learning did not take place, or at least, did not affect other's behavior.

policy and research) reflects the awareness of problems of poor quality in health care, and in particular, of principal-agent issues leading to incentives problems. In this section, I discuss evidence from Africa for several types of policies to improve health service delivery by changing provider behavior. Of course, this is only one determinant of service quality. There are many other constraints to increasing the effectiveness of the health sector, involving, for example, levels of resources or donor support, available human and physical infrastructure, and the overall organization and management of the health sector. The limited focus in this section reflects our emphasis on microeconomic behavior and related evidence. While the strategies discussed below are not directed at the broader constraints just enumerated, they do (potentially) make them less binding by increasing the efficiency of existing human and physical resources in the health sector.

3.2.1 Accountability and community monitoring

Providers of public health services in many poor countries are insufficiently accountable to the public they should be serving, creating a lack of incentives to provide better care (World Bank 2004). Evidence of poor performance is not hard to find. A key, unambiguous indicator is the rate of absenteeism. A series of recent surveys in developing countries (described in Chaudhury et al. 2006) reveal high—sometimes incredibly high—levels of absenteeism for health professionals as well as teachers in public facilities. For health workers the average rate of absenteeism for seven countries was 35 percent. Absenteeism tends to be higher for health workers than teachers, likely because they can more easily divert their skills and time to private practice. Anecdotal evidence of disrespectful behavior of health center staff toward clients also suggests a lack of accountability. Obviously, even if health personnel are on hand to perform their jobs, the quality of the care they provide is constrained by the level of skills they possess. Still, as noted, performance also reflects motivation and effort, hence incentives and accountability potentially play a role here as well.

Monitoring problems exist in all public health sectors, but in developing countries standard institutional procedures for monitoring clearly often function very poorly. With standard mechanisms failing, interest has grown in developing approaches whereby the service beneficiaries themselves—health care clients—are able to monitor provider performance and possibly, enforce behavior. As in other areas, much of the impetus for this comes from the World Bank (the position was laid out in the 2004 World Development Report “Making Services Work for Poor People”). However, it is part of the broader move in policy discussion and practice toward greater decentralization of governance and service delivery and toward increasing emphasis on ‘participatory’ decision making. Some early precedents support the notion that community oversight may insure effective performance of health providers. In Kerala, India, which famously achieved dramatic improvements in health despite being one of the poorest states in India, citizens were instilled with a strong sense of entitlement to health care, as described in Caldwell (1986). Community political leaders would lead protests if primary clinics were unmanned, and there was the threat of sanctions from the community for non-performance.

The advantage of monitoring by beneficiaries is that they should have better information on local provider performance than external evaluators in the health sector bureaucracy, as well

as presumably having a greater stake in improving quality. On the other hand, if beneficiary groups are not granted actual power to enforce changes ('direct control') and instead are only able to report what they observe to central authorities ('indirect control'), impacts may be limited. Also, there is a collective action problem in that serving on monitoring committees imposes a burden on specific individuals while the benefits are felt by the community overall (Banerjee and Duflo 2006). This participatory approach relies on some level of civic engagement.

Community monitoring is being tried in a number of countries in Africa and elsewhere, typically using surveys of users ('Citizen Report Card') to provide information on provider performance. As yet, there are only a handful of completed rigorous evaluations, counting interventions both in health and education. They provide a mixed picture. In health care, Björkman and Svensson (2009) report on a randomized evaluation of community-based monitoring in Uganda. In 25 randomly selected communities, local NGOs organized meetings between community members and public health care personnel to discuss the quality of care provided, using surveys (citizen report cards) that had been administered to measure citizen perceptions of performance. In terms of organization, district Health Unit Management Committees (HUMCs), which were already in existence but barely functional, were given a central role. In 25 control communities, similar surveys were given, but there was no organized public discussion. In the intervention communities both provider attendance and quality of service measures (including waiting times and perceived quality of care) improved. Health outcomes also improved: immunization rates rose and child mortality fell in the treatment communities. The results were especially noteworthy as the HUMCs' power was limited to indirect control over health centers as defined above.

Less success with community monitoring of health care was found in an experiment in India involving government health clinics in Udaipur district, Rajasthan (Banerjee, Deaton and Duflo, 2004). In treatment communities, a member of the community was paid to check once a week, on unannounced days, whether the auxiliary nurse-midwife assigned to the health subcenter was present at work. External monitoring by the survey personnel confirmed that the monitoring was accurate. How this information was used by the community (what type of sanctions or rewards for attendance performance were employed) was left up to the community. Weekly monitoring of treatment clinics was found to have no effect on absence rates in local clinics, which remained very high for both treatment and control clinics (44 and 42 percent, respectively, verified externally).

An experiment with community monitoring in education, elsewhere in India, also did not lead to improved service delivery, measured again (in part) by absence rates. (Banerjee et al. 2008b). In 195 randomly selected villages in the state of Uttar Pradesh, an education NGO facilitated information gathering about school quality and learning (using 'report cards' on reading achievement of the children) and ways to improve it. The intervention made use of Village Education Committees (VEC) that like the HUAC in Uganda already existed but were largely inactive at baseline. The VEC were able to exercise indirect control via monitoring the performance of the schools and reporting problems to higher authorities and, in principle, also had some direct control, over community-based teachers and some school resources. At follow-up, there was no improvement in community participation, teacher presence, or

learning outcomes in public schools in the treatment villages relative to 85 randomly selected control villages where no mobilization took place. In contrast, a different intervention (in a different arm of the randomized study), which trained volunteers to teach children to read, attracted substantial community participation and had a large impact on reading skills.

Hence only one of the three programs evaluated, the Uganda health care monitoring, delivered the hoped for benefits of beneficiary control in terms of changes in service quality and outcomes. Banerjee et al. compare their findings in India to the more successful program in Uganda, both of which had similar designs, including the use of previously existing but largely inactive local oversight committees. They note among other factors that the external promoter (the participating NGO) was less actively involved in the India case, and the political power of the teachers in India was likely stronger than the health care providers in Uganda, making them less subject to the will or sanctions of the community. Some more general factors that may inhibit the success of community monitoring may be noted. These include the fact that the sanctions or rewards available to committees themselves are typically limited; most importantly, hiring and firing decisions usually remain with higher health or school authorities, who choose how, if at all, to respond to the committee's reports on performance. Nor will committees be effective if participation of community members is inadequate, as seemed to be the case with the India education experiment. This in turn may reflect low expectations that reports made to the authorities will have any effect, as well as the collective action problem noted above. Incentives to participate will also be weak if either the demand for formal health care or education is low to begin with, or if individuals are conditioned by experience to have very low expectations of quality (Banerjee and Duflo 2006).

If this limited body of evidence is any guide, programs to improve accountability via beneficiary monitoring and control are not a 'magic bullet' for solving incentive problems in public health care. They may work in certain contexts and not others, based on a range of factors that include the strength of local institutions, traditions of collective action, and the political power of different actors. For example, in Kerala, strong local participation and effective community control was related to the politicization of the poor, fostered by the State government, which created a sense of health care as a right of all (Caldwell 1986). The variation in outcomes seen for community monitoring is quite different from the studies of price responsiveness reviewed in Section 3.1.1, which showed a reassuring consistency (high responsiveness of poor people to the cost of services) across many contexts. For some interventions designed to change individual behavior, such as promotion of hand-washing, there were also fairly consistent results across a number of studies. The lesson we may draw from this is that when interventions are complex and involve a range of institutions, not just individual incentives and behavior, context will matter a great deal.

3.3.2 Pay for performance incentives

A different, supply- rather than demand-side based, approach to the incentive problem in health care is for pay to providers to be based on performance. There are few evaluations of this framework, which has gained currency relatively recently in discussions of developing country health care. Rwanda is one country that instituted output-based payments relatively

early on. Meessen, Kashala, and Musango (2007) consider the effects on the performance of clinics in Kabuare district of a switch from a standard contract with providers (including fixed salaries and fixed annual bonuses) to performance-based contracts. The latter replaced the fixed bonuses with payments to providers (meaning the facilities, not individual health workers) based on the number of key services performed, with different rates for different services. These payments were shared among staff at the facility according to a pre-determined system based on experience and other factors. A simple before-after comparison (comparing outputs in 2001 and 2003) suggests very large increases in outputs among the 15 clinics studied: the average increase in the output index (with services weighted by their specific pay rates) was 80%. Costs, reflecting the pay-for-performance scheme, increased by about 30%. In considering these findings it should be kept in mind that there was no comparison group of clinics to control for trends.²²

The types of health care providers involved in the pay-for-performance scheme in Kabuare is noteworthy. Although an international NGO managed both the financing and contracting, the providers themselves were non-profit centers run alternatively or in combination by the government, churches, and the community, and overseen by the health ministry. Therefore, pay for performance is compatible with the direct public provision of health services, provided that institutional factors do not rule out such a payment mechanism.

Several potential problems with this strategy should be noted. The first is the questions of how to monitor performance. Only aspects of care that can relatively easily be observed by the contracting agency (for example, the numbers of clients seen for different services) can feasibly be used in rating performance. Related to this are several perverse incentive problems noted by the authors of the Kabuare study. There is an incentive to substitute unremunerated or less well remunerated services for better remunerated services, even if the latter are not warranted medically or are services which the center lacks the resources or capability to perform. There is also an incentive (if output is measured by number of patients treated) to substitute quantity for quality of care, and to falsify reports of the services performed. The use of pay for performance therefore has both important advantages as well as potential pitfalls, so deserves further consideration in the African contexts using rigorous evaluation methodologies. These should pay particular attention to monitoring strategies to overcome the incentives issues just noted.

It should also be strongly emphasized that health worker motivation is not merely a function of financial incentives. Many other social and institutional factors play a role in eliciting effort (Franco, Bennett, and Kanfer 2002). Governments or health ministries may be successful in instilling a sense of mission in health care workers; productivity may be greater when practitioners are allowed greater autonomy and flexibility in their work; a shared code of professional responsibility may inspire better performance. For example, Tandler and Freedheim (1994) ascribe the dramatic success of a child health prevention program in Ceara, Brazil (which reduced infant mortality by 36% in just a few years) in large part to the first two of these factors, as well as to deliberate efforts by the government to elevate the prestige of the program's field workers in the communities they served. These factors are especially

²² However, an experimental evaluation of pay for performance incentives, also in Rwanda, is currently being conducted.

important when monitoring of worker performance is difficult (Filmer, Hammer, and Pritchett, 2000).²³ An oft-expressed concern is that the introduction of pay for performance incentives will erode these sources of professional cohesion and motivation. These issues, while important, are outside the scope of this review.

3.2.3 Contracting with private providers

Given limited public sector capacity to reach all those in need, a number of countries have instituted policies or pilot studies to contract out health services provision with the private sector. This approach, recommended most importantly by the World Bank in its 2004 World Development Report, has an additional potential benefit of reducing incompatible incentive problems that characterize the relationship between policymakers and health care providers, on the one hand, and between providers (i.e., owners of the health facility) and health workers, on the other. With regard to the first incentive problem, individual private facilities or networks can be held accountable by the government for results (which are specified in the contract), though again it should be noted that this is only possible for results that can be easily measured, such as the number of persons receiving a given service.²⁴ With regard to the second problem, within private organizations, it is usually much easier institutionally to reward or discipline employees based on performance than in the public sector.

Loevinsohn and Harding (2005) reviewed ten assessments of government contracting with private organizations to provide primary care or nutrition services. The methodologies used for these evaluations varied: there was one randomized control trial (in Cambodia), while the rest were either before-after evaluations with comparison groups (areas), before-after evaluations with no comparison areas, or comparisons with non-intervention areas without baseline data. The studies consistently report beneficial impacts of contracting in terms of coverage of services and (where measured) quality measures. In the Cambodia randomized trial, coverage for immunizations rose by 40% in areas under contracting compared with 19% in areas with standard public providers (Bhushan, Keller, and Schwartz 2002). Further, provisions in the contracts for targets to reach the poor were successfully implemented

In the two evaluations from Africa of community nutrition services—the Secaline program in Madagascar and the Community Nutrition Project in Senegal (both reported in Marek et al. 1999)—high coverage was achieved, and severe and moderate malnutrition rates among beneficiaries fell steadily in both cases over several years of implementation. It is not clear to what extent the latter trend reflects project benefits as opposed to changes in the composition of participants over time. However, in one community in Senegal where the project was active and pre- and post population surveys were available, severe malnutrition declined 6% (to zero) and moderate malnutrition declined 4% among all children over a 17-month period (though there was no comparison group to control for trends).

²³ Though in the Ceara program the authorities also strongly encouraged community monitoring of health worker performance.

²⁴ Note also that the Rwanda study described in the previous sub-section suggests that it is also possible to improve performance of public facilities with incentive schemes.

One thing these evaluations—all of which concerned large projects—clearly show is that contracting can be managed on a large scale in poor countries. Still, the evidence of successful outcomes should be treated cautiously given the varied nature of the study designs, which often lacked either comparison groups or had comparison groups but lacked baseline data. Also, it is not always clear what the gains are being compared to. For example, in the Madagascar and Senegal evaluations, the nutrition projects were new, publicly-funded initiatives to use contracting to increase children’s access to care in certain areas. The projects apparently achieved high coverage, but the relevant comparison is not to what coverage was before but to what a similarly ambitious direct provision by public providers could have achieved.²⁵

A striking commonality in all of these contracting schemes is the use of NGOs rather than private (for profit) providers. This is in line with the observation by many that NGO health workers tend to have high levels of intrinsic motivation to serve their clientele, thus overcoming a key agency problem.²⁶ Consistent with this, Reinikka and Svensson’s (2009) analysis of a Quantitative Service Delivery Survey in Uganda shows that religious not-for-profit (RNP) facilities hire medical staff below the market wage, are more likely to provide pro-poor services, and charge lower prices for services than for-profit facilities providing a similar measured quality of care. Both RNP and for-profit facilities provide better quality of care than government facilities, although government facilities have better equipment. Further, the authors exploit a ‘near’ natural experiment, a public financial aid program for not-for-profit (public and religious) centers whose implementation was staggered, to estimate the impacts of additional resources on provider behavior. The financial aid leads to more laboratory testing of suspected malaria and intestinal worm cases as well as lower outpatient fees, but only does so in religious-based facilities.

The evidence above suggests that contracting with (or subsidizing) NGOs, including religious-based providers, may avoid some of the incentive and agency problems that would normally characterize the relationship between the government and contracted private organizations. There is a lack of evidence on such arrangements with for-profit providers, but the above would lead us to expect more serious problems of monitoring and perverse incentives. With regard to expanding the role of NGOs, there are broader questions about the role of the state to which advocates of contracting have not given much attention. One concern is that having NGOs take over what is traditionally considered a basic function of the government, to provide essential health services (even if the activities are still publicly funded under the contracting scheme), will inhibit the development of the state’s capacity to eventually perform these functions. State capacity will be further diminished if NGOs attract scarce skilled managerial and front-line health personnel away from government.²⁷ Reliance

²⁵ It is important to note that in both countries these community-based projects were only implemented in communities that agreed to have them and were involved in their operations. Clearly the project communities may have had traits not present in other communities that made success more likely.

²⁶ Leonard and Leonard (2004) argue further that relative to the public sector, NGOs are in a better position to monitor their employees. Note that in some of these countries, there may also have been a lack of for-profit private providers operating on a large enough scale to qualify for the contracts.

²⁷ A counter-argument would be that by letting some functions be handled by the private or NGO sector, the government can effectively concentrate its limited skilled resources on building capacity in other areas. Of course, if NGOs siphon off many skilled individuals from the public sector, this will not happen.

on multiple NGOs to serve different needs or regions may lead to a fragmentation of the health system (Palmer et al. 2006).

If these factors are significant, there is a tradeoff between developing state capacity, which takes time, and being able to provide for the health needs of the population quickly by using NGOs that are likely already functioning in the country. In very poorly functioning states where capacity is extremely weak (e.g., ‘fragile states’) there is much less ambiguity over the need for non-governmental actors (see Liu et al. 2004).²⁸

3.2.4 Private sector and positive prices

While the evidence on price responsiveness presented in Section 3.1.1 lends support to the view that essential health services and products need to be very low cost or free to reach the poor, there is another side of the story: the supply side, namely, how providers respond to price incentives. For the private (for-profit) sector to play a role in service delivery, prices must be positive unless other means are found to compensate private sector actors. It is worth emphasizing that even for services or goods for which wide agreement exists on the need for public sector subsidization (for efficiency or equity reasons), there can be significant advantages to engaging the private sector for their provision. For example, malarial medicines are typically distributed at thousands of small general retail stores across a country. Having them available only in government clinics, which are less accessible, would place them out of easy reach of households who need these medicines quickly when a child or adult gets sick. Further, government provision in Africa is typically reliant to a very large extent on donor funding. Because such funding is highly variable, public supply is likely to be subject to interruptions. For many health products, vibrant commercial production and distribution networks may already exist or be easily developed.

The objective of insuring access at zero or highly subsidized prices need not be incompatible with a significant role for the private sector in providing health services or goods. First, the government can subsidize or contract with providers of health care or producers and distributors of health products. We saw above that contracting for basic health services is feasible in a wide variety of settings. In other cases, in particular health-related products, supply-side subsidization may be impractical or inefficient. For example, it does not allow for targeting specific groups (except perhaps by targeting specific areas), since the item or service will be equally available to all at the lower price. This concern can be addressed in principle by demand-side subsidies to households in the form of vouchers. Households would be given a voucher for a product or service, redeemable at any care provider or seller. This allows more easily for targeting: for example, vouchers can be given only to vulnerable households or individuals (e.g., pregnant women seen at antenatal clinics or visited at home by a health worker). Vouchers will not be suitable for all cases but may work well with well-defined, homogeneous products for which monitoring of quality is relatively simple.

²⁸Where middle income countries have achieved very high coverage for health services (notably in Asia, but also in Botswana), this has generally been through public delivery; the same pattern prevails in lower income areas or countries that have been able to achieve very wide coverage (e.g., Kerala state in India). This does not prove that the same pattern is possible or necessary in other contexts, especially where state capacity is weak, but it raises the burden of proof for national strategies relying strongly on private sector delivery.

An example of this is insecticide-treated bednets (ITNs), which currently is a prime locus of the debate over whether or not health goods and services should be provided for free. Recent experiences with ITNs in Africa also illustrate the potential and limitations of vouchers as a solution to the dilemma. I return to these issues below in Section 4.2. For now, it may be pointed out that many aspects of the supply response issue are difficult to investigate using experiments. This is the case with ITNs. The scale of any such experiment would potentially have to be very large, because producers and distributors of products such as bednets will only respond to real market demand. The time frame for the evaluation would have to be long, because it would take some time for production and distribution networks to get set up. Therefore it is usually necessary to investigate these issues using broad observational data, that is, by looking at changes in policies and outcomes at the country level or comparing policies and outcomes across countries (see, for example, Noor et al. 2009).

4. EVIDENCE ON CHILD HEALTH INTERVENTIONS

This section considers evidence for the effectiveness of two types of child health interventions, each addressing a major source of child mortality: water and sanitation improvements to reduce diarrheal disease, and insecticide-treated bednets to prevent malaria infection. There is a particularly large evaluation literature, including, more recently, a number of experiments dealing with these interventions. A detailed discussion of this evidence serves to illustrate many of the issues with regard both to policy and evaluation methodology brought up throughout this paper. Next, this section considers evaluations of ‘community-based’ interventions for child health, which have been designed to address different sources of under-five mortality (including, but not limited to, malaria and diarrheal illness). As noted in the Introduction, these programs are characterized by a common reliance on non-professional health workers to expand contacts of households with the health care system. As such, they represent efforts to address the problem of low service uptake and coverage, hence merit particular attention.

4.1 Water, Sanitation, and Child Diarrheal Illness²⁹

Reducing diarrheal illnesses in children is a longstanding global health objective. Diarrhea accounts for some 20 percent of deaths of children under five in developing countries (Kosek, Bern, and Guerant 2003). Persistent diarrhea may predispose children to malnutrition, may lead to impaired weight gain, and is associated with increased risk of death from childhood illnesses (Lancet 1991; Pelletier et al. 1995). Among other interventions, provision of piped water and sanitation infrastructure can drastically reduce child mortality from diarrheal illness and are thought to be responsible for most of the rapid decline in child mortality in the U.S. in the early twentieth century (Cutler and Miller 2005).

However, it is not usually feasible to supply piped water and indoor plumbing in rural areas of developing countries, where households are disbursed over wide areas. This means that other approaches are needed. For water, this could be community infrastructure investments such as well construction and well or spring purification, or at the household level, point of use purification. For sanitation, this could be community latrines, household latrines, or adoption of new behaviors such as regular hand washing with soap. A key aspect of these strategies is that behavior—of households or communities—is an essential ingredient in the success and sustainability of the interventions. Community wells or latrines must be maintained; households must recognize the value of routine hand washing and continue the practice, or they must be willing to purchase and use a water disinfectant product. In view of this concern, recent evaluations of sanitation and water have focused on questions of uptake and sustainability.

With regard to community water and sanitation infrastructure, a few early evaluations were randomized across communities or used comparison groups of communities. These studies indicated that investments such as borehole wells, water pumps, and pit latrines can reduce

²⁹ This discussion in this section draws on several reviews by others, most importantly the paper by Zwane and Kremer (2007).

child diarrheal incidence or improve weights for height of children. (Huttly et al. 1987 for Nigeria; Aziz et al. 1990 for Bangladesh). However, as Zwane and Kremer (2007) note in their review of the literature, the fact that these studies, which had only a handful of treatment and control villages, did not adjust for correlations of observations within each village means that the reported statistical significance of the findings may be substantially overstated.

More sophisticated recent community-level evaluations produce more reliable inferences (and use samples with substantially larger numbers of communities), but these are still rare. Kremer et al (2007) assess a spring protection intervention (a source intervention) in a sample of 175 communities in Western Kenya. The spring protection, which involves ensuring that the water flows directly out of a pipe rather than over potentially contaminated ground, led to large improvements in source water quality, as measured by the fecal indicator bacteria *E. coli*. Moderate gains were also seen in home water quality, and reported child diarrhea incidence fell by one quarter in the treatment communities. The survey was carefully designed to capture household characteristics and behaviors that may influence the effectiveness of the intervention. Although prior work (Esrey 1996) seemed to indicate that water infrastructure investments have little effects in the absence of complementary changes in sanitation infrastructure or behavior, Kremer et al. find that the amount of improvement in home water quality from the spring protection was not related to latrine coverage or baseline hygiene knowledge of the household—nor did the presence of spring protection result in any compensatory behavior by the household in terms of water collection, transport, storage practices, water quantity used, or other preventive health behaviors.³⁰

Community-level infrastructure for providing safe water and sanitation is a public good; so, too, is the maintenance of this infrastructure. Having mechanisms in place to ensure that, for instance, wells and latrines are maintained, is crucial for the sustainability of these investments. There is no shortage of evidence that local water infrastructure in rural areas of developing countries is poorly maintained (Ray 2004; Miguel and Gugerty 2005). It has been suggested that involving women more prominently in management of these community resources will be an effective strategy, both because women tend to be more willing than men to supply public goods (their time) and because they are the main users of water resources (Agarwal 2000). It has also been suggested that local control via user committees will result in better managed resources than centralized government control. However, while some studies show a correlation of community control and infrastructure quality, this question (and the role of women) has not been subjected to rigorous evaluation. For one thing, there is a strong possibility of heterogeneity and simultaneity biases in simple correlations. Communities where individuals tend to be well organized to manage local resources may have had better outcomes anyway, or it could be that when resources are well managed, more people are willing to become involved with their management (Zwane and Kremer 2007).

An alternative to community-managed infrastructure is to try to change household behavior—for example, encouraging households to purify their water using disinfectant chemicals, or to wash hands regularly with soap. Interventions to provide hand washing education and soap to

³⁰ This study is a good example of how an experimental evaluation can be designed not just to measure whether an intervention had an effect on the outcomes of interest (water quality in the home, child illness) but also why the intended outcome was or was not obtained.

households have been shown in numerous studies to reduce diarrhea incidence among children (see Curtis and Cairncross 2003). Among these studies are randomized trials conducted in urban settings in Pakistan (Luby et al. 2004), Bangladesh (Khan 1982), and Burma (Han and Hlaing 1989). The benefits from these interventions were sizable, including a 40% reduction in diarrhea incidence among under two's in the Burma study and 39% fewer days of diarrhea in the Pakistan study. For point-of-use water purification, numerous evaluations, including randomized trials, show that this behavior also can substantially reduce diarrhea incidence in children (Semenza et al. 1998; Quick et al 1999; Quick et al. 2002; Sobsey, Handzel, and Venczel 2003; Reller et al. 2003).

Having consistent findings over many contexts point to the promise of these interventions to improve child health and potentially, reduce mortality. However, as Zwane and Kremer (2007) point out, in many of these interventions, households received intensive encouragement and reinforcement of messages via frequent visits by fieldworkers. As a result, compliance tended to be high so that intention to treat estimates (the mean effects calculated over both those who followed the suggested practices and those who did not) showed strong results. However, such high levels of interaction are usually not feasible for scaled-up interventions. With the efficacy of these interventions established, emphasis needs to be directed at strategies to ensure uptake. There are without doubt many challenges to increasing—and sustaining—the practice of different health and hygiene behaviors. There is mixed evidence on whether changes in behavior induced by hygiene education and soap provision can be sustained long-term. Two non-randomized evaluations found persistence in follow-ups several years after the intervention (Wilson and Chandler 1993; Shordt and Cairncross 2004) while a randomized study (Hoque et al. 1996) found otherwise. Further, the results of some studies indicate that it is harder to change hygiene behavior among the poor, possibly reflecting complementarities of formal education and new health information. (Luby et al. 2004 is an example of this for hand washing promotion).

The Zambia study by Ashraf, Berry, and Shapiro (2007) discussed above provides evidence on how cost affects uptake for a point-of-use chlorine-based water disinfectant. As indicated, households were found to be very sensitive to cost in a door-to-door promotion intervention. Only about half the sample was willing to purchase the product at the (already subsidized) price at which the product was available in markets. While it was also found that a positive price tends to direct the product to customers who actually use the product, potentially eliminating waste, a higher price did not increase use among families with young children. These findings, together with findings for other products such as mosquito nets, suggest that very low or zero prices may be required to ensure high coverage of children, the main target population. At the same time, as noted earlier, there are potential difficulties ensuring adequate supply at very low or zero price.

Other strategies for sustained uptake of new hygiene and water purification practices (that do not require intensive and repeated door-to-door promotion) might include education interventions targeting women or mothers in children's health clinics or as part of nutrition programs. A different approach is to target communities rather than individuals. The Community Led Total Sanitation (CLTS), originally implemented in Bangladesh in 2000, emphasizes village-organized mobilization to educate about and discourage open defecation

and increase latrine use. In contrast to many earlier schemes, the ‘hardware’ component of CTLS—latrine construction—is intended to be self-financing by households or communities. This reflects concerns discussed above about how households value, hence use and maintain, goods that are given to them for free. After early implementation in India and Bangladesh, CTLS has been started in several countries in Africa, including Ethiopia and Kenya (Chambers 2009). There are very few rigorous assessments of CLTS in terms of latrine uptake, sustainability, and childhood illness, and none to date from Africa. However, one-year follow-up of a community-randomized study in Orissa, India, found that latrine uptake increased by 30% in the CLTS villages as a result of the campaign (Pattanayak et al. 2007). There were also indications that uptake reduced child diarrhea rates, but these results were not robust to model specification.

The impacts of latrine campaigns such as CLTS clearly involve health externalities (when a household substitutes latrine use for open defecation the disease environment improves for all) but also, potentially, externalities related to behavior. Social learning or peer effects may lead to a multiplied effect of an initial intervention whereby initially non-adopting (or non-targeted) households adopt the technology because their neighbors do. Dickinson and Pattanayak (2007) consider the role of these externalities in the Orissa CLTS intervention using two distinct strategies to deal with the endogeneity of other’s adoption to one’s own adoption. The first, based on the method of Bajari et al. (2006), uses the characteristics of neighboring households of household i to instrument adoption by those households: these characteristics should not directly affect adoption by i , hence can be excluded from the regression explaining i ’s adoption. Second, the intervention provided a subsidy for latrine building for households in each village below the poverty line. Dickinson and Pattanayak examine the effect of having a greater number of subsidy-receiving households in treatment villages on the adoption of latrines by households in the same villages that were above the line and thus not eligible for the subsidy. Increased uptake by the latter households when there are more eligible households (hence higher local intensity of uptake) would indicate the presence of positive externalities on demand for latrines.³¹ This effect is found, and consistent with this, the IV approach also yields large and significant impacts on own adoption of take-up by others in the same village. Hence externalities appear to have contributed to the substantial latrine uptake associated with the CLTS intervention.

4.2 *Bednets for Malaria*

About four-fifths of the one million malaria deaths recorded annually worldwide occur in Africa. Malaria is endemic in many countries of the region, particularly in West Africa. The disease imposes heavy economic costs in terms of illness in adults, but for young children the

³¹ This approach can be compared to the agricultural technology study by Duflo Kremer, and Robinson (2005) described in Section 3.1.4., which dealt with the endogeneity problem by randomizing training to selected individuals and observing the adoption behavior of the acquaintances of these individuals. In the Orissa study, eligibility for the latrine subsidy was not randomized but instead based on income, and instead of considering the impact of the behavior of neighbors or direct acquaintances, the study estimates the effect of the degree of uptake by others in the village as a whole. However, identification depends on the assumption that the variation in the density of subsidy-eligible households (essentially, variation in village poverty rates) is not associated with unmeasured village-level factors influencing take-up.

consequences are worse: the vast majority of deaths (90 percent) from malaria occur in children under five (WHO 2004). Pregnant women are another particularly high-risk group for malaria morbidity, and further, malaria in mothers is associated with low birthweight in children. For these reasons malaria prevention and treatment strategies tend to be directed at expectant mothers and young children.

Efficacy studies show that several interventions can reduce malaria mortality in children. A major recent development was the introduction of insecticide-treated mosquito nets (ITNs) which was shown in several clinical trials starting in the early 1990s to reduce child mortality (from any cause) by one-fifth and reduce malaria episodes by half (Lengeler 2004). Even more recently, Indoor Residual Spraying (IRS) with DDT, out of favor for several decades due to environmental concerns that have since been reassessed, has been reintegrated into the arsenal of prevention measures. On the side of treatment for those who contract malaria, new combination therapies that include Artemisinin are replacing earlier generations of anti-malarials which resistance has rendered ineffective.

While these prevention measures (or therapies) have been shown to be effective in a clinical sense, inadequate coverage remains a major barrier to achieving reductions in malaria morbidity and mortality. Use of ITNs has expanded dramatically since the results from early effectiveness trials became available and especially since governments made commitments to increase coverage; the Abuja Declaration of 2000 set a goal of 60% coverage of vulnerable groups, which was recently raised to 80%. However, despite gains since then, Noor et al. (2009) estimate that only a fifth of vulnerable young children in Africa are protected by ITN.

Therefore the issue of delivery and uptake—particularly of bednets—has increasingly drawn the attention of researchers and policymakers, and remains the subject of debate (Roberts 2007). As for most interventions, behavior on both the demand and supply side is important to the success of efforts to increase coverage, but with ITNs the issues are especially complex. Households first must be induced to buy nets (or accept them, if free). They must use them rather than leaving them idle, diverting them to other uses (a common practice), or reselling them. The nets eventually have to be replaced or re-treated (treated nets remain effective for up to five years). Further, the policy objective is to ensure that they are used to protect the most vulnerable individuals within the household, namely children and pregnant women, as well as to ensure that poor households have access to bednets. Finally, the potential for achieving high ITN coverage may be affected by externalities. Coverage may increase via peer effects but it may also be reduced via lower perceived need to protect oneself: use of bednets by others in the vicinity reduces the population of malaria carrying mosquitoes, hence one's own risk of becoming infected.³²

³² There is some evidence of a protective effect from the use of ITNs by neighboring households. For rural Kenya, Hawley et. al. (2003) found such an effect for non ITN-using housing compounds within 300 yards of a compound where the bednets were used. In contrast, no similar spillovers were observed in a study in southern Tanzania (Gosoni et al. 2008), although bednet ownership as expected was associated with reductions in all causes of child mortality in bednet-using households. The authors suggest that the lack of an effect on other households may be due to the small proportion of nets which were treated with insecticide. It is important to bear in mind that while spillovers in health protective benefits may dampen demand for ITNs, these positive health externalities are precisely what makes it possible to eliminate malaria as a threat without achieving complete coverage; it is estimated that 60% local ITN coverage would suffice.

How then to increase ITN coverage? The idea that bednets should not be distributed free of charge has been influential in policy discussions and actual policy in Africa. The stress is less on direct revenue benefits than on the idea (as discussed in Section 3.1.1) that a positive price (together with branding or social marketing) will make consumers feel that the product has value, hence be more likely to use it. Paying a positive price may induce use as well through a sunk cost effect, and a non-zero price may also weed out individuals who have relatively little need for the product or will not use it, reducing waste. Proponents of free distribution systems have argued, for their part, that even modest prices can dampen demand, especially among the poor.

The experimental studies by Cohen and Dupas (2008) and Hoffman (2008) cited earlier both attempt to address this question. Both find, first, that the demand for ITNs (made available to women during visits to antenatal clinics) is highly sensitive to cost, and second, that the price paid does not affect whether a bednet is actually used. In the Cohen and Dupas study in Western Kenya, while take-up was virtually 100% among women to whom the bednets were offered free of charge, it dropped sharply as price was varied upward. The estimates suggest that at the prevailing (already subsidized) price of ITNs available in the area through normal channels of about U.S. \$0.75, only 25% of women would purchase the nets. Nor did having a positive price select on women whose need was higher, as measured by anemia (hemoglobin levels), a common indicator of malaria in pregnant women.

As discussed in Section 3.1.1, these findings of high price elasticities are consistent with a number of other randomized studies of (and other evidence on) the demand for health products and services. They appear to offer strong support to proponents of free distribution, and the Cohen and Dupas paper in particular is regularly cited in the broader debate over free distribution vs. cost sharing—a debate that has made its way into the popular press.³³ Still, several issues remain. One is that in the Kenya experiment (as the authors acknowledge), the area in which the study was conducted was one in which an NGO, Population Services International, had long been active in promoting bednets via social marketing. This may have made the population generally aware of the value of bednets, hence willing to use free bednets even if this otherwise would be associated with low perceived value and use. If so, the high rates of utilization when nets were free actually may reflect in part the success of the previous social marketing (and cost-sharing) strategy.

The more general problem with (universal) free distribution is that the public sector may lack the resources and capacity to deliver the number of nets needed to achieve high coverage. As noted in Section 3.2.4, where public supply is substantially donor-funded there is the potential for disruption, which is an important concern because the nets must be replaced or retreated at regular intervals.³⁴ There is evidence that is possible to have well functioning private networks of manufacture and distribution of bednets in African countries (WHO 2002). For products such as ITNs, the harnessing of private retail networks is particularly advantageous: bednets can be, and in some countries are, sold at countless small retail stores. However, the

³³See, for example, http://www.boston.com/bostonglobe/ideas/articles/2007/11/11/a_handout_not_a_hand_up/

³⁴Since ensuring supply is a key issue in scale up and implementation, this is an external validity issue with the Cohen-Dupas study.

private sector requires incentives; if these come in the form of positive prices, many households, especially poor ones, will likely be priced out.

With appropriate methodological caveats, comparative country analysis can help in assessing the relative merits of different distribution strategies. A comprehensive recent assessment (Noor et al. 2009) of ITN coverage suggests that countries promoting free ITN distribution have achieved more coverage than those relying either on cost recovery (21% lower) or on “routine subsidized public-sector promotion” (11% lower median coverage). This evidence is suggestive that cost sharing is unable to achieve adequate coverage, possibly because poorer households do not participate.

Given the need to balance access to the poor with the benefits of commercial distribution (and to limit the financial burden on the public sector), many argue that there needs to be a mix of public and private provision, with public subsidies focused on the most vulnerable or poor. This can be described as the mainstream position on this issue, as it has been endorsed by the Roll Back Malaria Partnership consisting of WHO, UNICEF, UNDP and the World Bank (WHO 2002). Focused targeting of a subset of the population for free distribution directly through government channels should be compatible with a viable private sector supply network which would serve the bulk of the population. However, if a substantial portion of the population receives free bednets (or if leakage allows the non-eligible access to the benefit), there may be a ‘crowding out’ of the private sector as demand is depressed and prices fall for commercially available bednets.

In principle, this could be avoided with a voucher system. As in Tanzania’s National Voucher Scheme (TNVS), women visiting an antenatal clinic could be given a voucher covering the cost (or part of the cost) of a bednet, redeemable at any local retail outlet. Vouchers for health services and products are not without their own problems. They generally work best when the good or service in question is standardized so that assuring quality is relatively simple. They work less well when a service is complex and the need for the service and its quality are both hard to verify, as is the case for many health services. ITNs fairly clearly fall in the first category and so appear to be a good candidate for vouchers. Vouchers would allow customers to access this product for free or at low cost while maintaining incentives for commercial production and distribution.

At first glance, providing full cost vouchers for bednets would seem, from the point of view of consumers, to be essentially no different from free distribution. There is an important difference, however, in that an additional burden is placed on the consumer. In the case of bednets and antenatal clinics, instead of simply being given a net as part of her clinic visit, a woman would have to make an additional trip to a store or distribution center where the voucher can be redeemed. This may seem like a small additional (time) cost, but its potential importance should not be underestimated given evidence of high price elasticities as well as the opportunity costs of women’s time.

The most ambitious voucher scheme for ITNs in Africa is that of Tanzania mentioned above, which began in 2004 and was scaled up to all districts in the country by 2006. (Hanson et al. 2008). The program took advantage of the fact that Tanzania already had a relatively well

developed commercial distribution system for ITNs. Under the TNVS, all pregnant women, upon their first prenatal care visit, are entitled to receive a voucher that can be redeemed from a retailer for a bednet; the voucher is worth the equivalent of \$2.50 while most nets are about \$1.00 more than this, so there is still a degree of cost sharing. The voucher program succeeded in stimulating a viable private production and distribution system, with four domestic manufacturers and some 5,700 retailers selling the nets. Eventually 43% of all treated nets were obtained through vouchers. Coverage also rose, but remained well short of objectives of 80% among pregnant women and young children: a 2007 survey indicates that only 39% of pregnant women used any sort of a net (23% using an ITN). Fifty-six percent of infants were using any type of net and 34% were using an ITN (Hanson et al. 2008). Further, usage was much lower among the poor than the well-off (Hanson and Lengeler 2009). Clearly, the voucher system, despite its successes, did not 'solve' the take-up problem. The reasons may include the additional financial cost to households above the voucher amount as well as the time burden of redeeming the voucher. As a result of these shortcomings, temporary mass free distributions of ITNs were planned for 2009-2010. Of course, this competition from lower cost (free) distributions potentially jeopardizes the viability of the private supply network. The ultimate consequences of this policy change remain to be seen.

In sum, while the large empirical literature on malaria interventions in Africa indicates the clinical efficacy of several interventions, achieving adequate uptake and coverage remains a major challenge for research and policy. It appears that free distribution of treated bednets will increase coverage as well as utilization. Questions of how to ensure supply response under such a regime (or using a voucher system) remain, as do questions about the most cost effective combinations of preventative (bednets, spraying) and therapeutic interventions.

4.3 Community-based Programs

As discussed above, there is a strong clinical evidence base for the efficacy of many interventions to deal with the major illnesses affecting child health and mortality: respiratory illness, diarrhea, malaria, measles, and malnutrition (which likely underlies a large share of deaths from other causes). Given this evidence, emphasis has shifted significantly toward the meeting the challenge of effective delivery of these interventions so that they may have real public health impact (Bhutta et al. 2008). 'Community-based programs' refers to interventions that incorporate a central role not just for facilities and health professionals but for individuals drawn from the community itself. Typically these are 'community health workers' (CHWs) who after a modest amount of training are assigned to go door to door to provide health education and services. CHWs usually receive some pay, though programs operating with volunteers are not uncommon. In other cases community interventions involve workshops for targeted groups (e.g., mothers of young children) or the use of influential individuals in the community to help change health knowledge and behaviors. Since CHWs are not professionals, they can perform some activities (such as micronutrient supplementation, child growth monitoring, promotion of vaccinations and health behavior change) but not others (vaccinations, prenatal care). However, they can be trained to refer individuals to facilities when needed. They represent a relatively low cost means by which the interactions of households with the health system can be increased beyond what could be achieved with

traditional facility-based delivery. Community based programs thus potentially provide a way to substantially improve the uptake of health and nutrition services.

A number of the experimental studies reviewed above use elements of this approach, for example, the door-to-door promotion of water disinfectant in Zambia (Ashraf, Berry, and Shapiro 2008) and several interventions using field workers to promote hand washing. However, most of the evaluations of community-based programs discussed below differ from these studies in several respects. Many are not randomized, though they usually are able to compare trends in intervention and comparison areas. They tend to be relatively large (in part because the key outcome of interest is child mortality, a relatively rare event for which large sample sizes are needed to detect changes). Since the interventions being evaluated are to varying degrees already scaled up and using routine procedures for training and other aspects of implementation, the external validity concerns associated with smaller experiments are less relevant.

A number of community-based interventions, in Africa and elsewhere, have been able to reach most or even virtually all of the intended recipients in program areas. Coverage rates of 75% or better are not uncommon (Mason et al. 2006; Bang et al. 1999; Bang et al. 1990; Baqui et al. 2008; Fauveau et al. 1992; Mtango and Neuvians 1986) though substantially lower rates are reported in other cases.³⁵ Evaluations of health impacts usually measure intention to treat at the community level, capturing changes in the entire target population rather than just participants. No doubt reflecting the high coverage rates achieved using CHWs (or sometimes other means), many community based programs appear to have led to significant improvements in under-five mortality or other health indicators.

For example, a non-randomized community level evaluation in India (Bang et al. 1990) found that education about child pneumonia and case management provided by community health workers was associated with a 30% reduction in under-five mortality relative to comparison areas. Evaluation of two other community-based programs for prevention and treatment of child pneumonia, in Tanzania and Bangladesh, found reductions in mortality of 27% and 28-32%, respectively, relative to comparison areas (Mtango and Neuvians 1986; Fauveau et al. 1992). In the same district in India as the pneumonia study just mentioned, Bang et al. (1999) evaluated a project in which CHWs provided home-based neonatal care, including treatment of sepsis. After two years, infant mortality had fallen by almost 50% relative to comparison villages. In a community-randomized trial in Tigray, Ethiopia, home visits were made to teach mothers to recognize symptoms of malaria in their children and to provide antimalarial medication. Under-five mortality fell by 40% relative to control communities (Kidane and Morrow 2000). This contrasts with the lack of impact reported by Spencer et al. (1987) for a similar project in rural Kenya, but that study did not have baseline data to control for initial differences in intervention and comparison areas.

³⁵ It should be noted that 'coverage' is somewhat ambiguous here. It can refer simply to regular contact of a CHW with the targeted households or individuals (for purposes of growth monitoring, behavior change promotion or education, to make references as needed to facilities, etc.) or to shares of the target population actually receiving specific medical treatments (vitamin A supplementation, supplementary feeding for underweight children, referred visits to clinics for professional care). High values for both of these indicators are objectives of the programs.

Manandhar et al. (2004), using a community-randomized design, evaluate a community-based project to address neonatal mortality in Nepal consisting of frequent women's group meetings in which a trained facilitator discussed perinatal problems and how to deal with them. After two years, neonatal mortality had fallen 30% relative to control areas. Another cluster randomized evaluation (in Bangladesh) assessed a more standard home-visit based strategy to reduce neonatal mortality through education of mothers, monitoring of infants, and referral (Baqui et al. 2008). Neonatal mortality was reduced in the home-care arm by 34% during the last six months of the trial. Linnemayr and Alderman (2008) use a randomized design to assess a multi-faceted village-level nutrition intervention in Senegal. Positive village-level outcomes for nutrition were found only for very young children, but the authors suggest that the lack of any broader impact reflects imperfect randomization—some households in control communities had access to the intervention. Actual (individual) treatment status, instrumented by the exogenous assignment of treatment to villages, had more positive impacts on nutrition outcomes.

Reductions in mortality rates approaching 50% are impressive achievements and show that community-based programs, and specifically the use of community health workers, have the potential to significantly expand the reach and benefits of the health care system. How generalizable—and sustainable—are these successes? First, as mentioned, not all such programs were able to achieve high coverage, limiting their effectiveness. Second, even through they make use of low-paid non-professionals, community-based programs are not cheap, given the intensity of contact with households required as well as the need for adequate training and professional supervisory personnel. Focusing on interventions for improving the nutrition of young children and using rough estimates of effect and cost from a meta-analysis of such interventions, Mason et al. (1999, 2001) estimate for a sample of East and South Asian countries that complete coverage for these interventions would require approximately a 20% increase in public health care spending. This represents a large, though not infeasible, increase in spending to cover a major set of health needs of the population.³⁶

Besides cost, Mason et al. (2006) stress that for community-based interventions to be effective, a number of conditions must be in place. These include adequate health and administrative infrastructure, political commitment, relatively good status of women (because CHWs are typically women), and the presence of community cohesion and community organizations that can support the project. It may be that the seemingly high share of cases with good coverage and substantial reduction in mortality or other successful outcomes partially reflects the fact that community-based programs tend to be implemented in contexts where these conditions are in place.

Experience with the Integrated Management of Childhood Illness (IMCI) strategy is informative in this regard. IMCI is a major global health initiative, promoted since the mid-1990s by WHO, that aims to combine the management of the major causes of childhood deaths enumerated above—a 'horizontal' approach to replace the prevailing illness-specific

³⁶ To put this in African perspective, public spending per capita in Sub-Saharan Africa is about \$21, substantially higher than in South Asia (\$6.30) but lower than in East Asia and Pacific (\$42) (from World Development Indicators 2008).

‘vertical’ approach. In addition to health sector support and training of professionals, a major component of IMCI is the involvement of individuals and communities in efforts to change health behaviors. Therefore, while substantially more ambitious in scope than typical community-based programs, IMCI shares some key aspects with them. The approach has been endorsed by almost all developing countries and has been implemented, at least nominally, in many of them.

WHO carried out a major multi-country assessment of ICMI in Bangladesh, Brazil, Peru, Tanzania, and Uganda (Byrce et al. 2004, 2005). In Tanzania, a comparison of intervention and control districts (non-randomized) indicates that the program reduced under-five mortality by 13% after two years (Armstrong Schellenberg et al. 2004). In the Bangladesh evaluation, which was randomized by district, coverage indicators clearly rose in ICMI areas in terms of the share of ill children brought to health facilities (El Arifeen et al. 2004). These gains were made at relatively low cost. In the other country cases, ICMI was less successful in terms of quality improvements, coverage, or mortality outcomes. This was attributed mostly to imperfect or inadequate implementation. Indeed, the evaluation team noted that even selecting five countries for the evaluation from among the many more that had begun ICMI proved to be quite difficult, because (among other reasons) few had implemented the initiative to an extent that measurable results could be expected (Byrce et al. 2004).

Much of the discussion of IMCI has consequently centered on problems related to program implementation (Byrce et al. 2005), a discussion considerably enhanced by the fact that consideration of these issues was built into the evaluation designs. To varying degrees, poor health infrastructure, limited financial and human resources, organization weaknesses, and lack of political will acted as barriers to successful scale up of the program in many countries. These problems are far more general than for ICMI alone. We are back essentially to the external validity problem: will findings from pilot studies (experimental or not) hold when interventions are scaled up nation-wide? IMCI may be a more complex program than many other health interventions, hence more difficult to carry out at scale. Still, the careful comparative evaluation of large-scale implementations of this program (and a much wider literature on program assessment) suggest that we need to be very careful in assuming an affirmative answer. They also indicate that even as policy experiments come into wider use, program evaluation must always also rely on other approaches to assess scaled up programs that are able to capture realities in the field.

5. SUMMARY AND CONCLUSIONS

There have been some real gains in health care coverage and outcomes in Africa in recent decades. Further, for many diseases, especially those responsible for high rates of child mortality, the efficacy of many low cost interventions has been well established. However, for most types of services or illnesses, delivery of high quality services to target populations remains inadequate in Africa. The research discussed in this paper promises to aid this effort—or already has—by providing insights into the effectiveness of specific health interventions and by providing more general insights into consumer and health provider behavior. A good deal of this knowledge comes from recent randomized program evaluations. Among the key findings in this literature are:

- With respect to household or consumer behavior, a consistent finding is that poor households are sensitive to the cost of services and health-related products. While this argues against policies for cost sharing via user fees, other evidence suggests, as was initially envisioned in the Bamako Initiative, that fees coupled with quality improvements need not have detrimental impacts on uptake, even among the poor, and may increase it. Consistent with this are demand studies indicating that consumers recognize certain dimensions of service quality and respond to quality when making health care choices. However, in many cases in the wake of the Bamako Initiative, cost sharing was introduced without effective improvements in quality, and demand fell sharply.
- More recently, the idea of user fees has been promoted in more nuanced form, based on socially efficient differential pricing by type of service such that subsidies are highest for services with high positive externalities or high price elasticities. Some comparative country evidence suggests this can be feasible as well as equitable, but rigorous research in different contexts is needed.
- Externalities (with respect both to behavior and health itself) may significantly influence the impacts of many health programs. Many community randomized evaluation designs implicitly capture these spillovers, but less has been done to investigate their extent or how they operate. In part this reflects the difficulty of examining this issue with non-experimental data, and it is likely that more insight on the role of externalities will emerge as new experimental designs are carried out.
- On the supply side, a relatively recent focus on provider incentives has motivated a range of new approaches to improve service delivery. The number of rigorous evaluations of each remains low, but some conclusions can be drawn. Expectations have been high that beneficiaries themselves, via community monitoring, would be able to improve accountability hence performance of public health care (and education) providers. The handful of experimental evaluations so far provide at best mixed results, for reasons that may include collective action problems, weakness of local institutions, and lack of direct control of beneficiaries over resources and personnel.

- An alternative is to use direct financial incentives by basing payments to providers on their performance or output, measured for example as the number of patients seen by category of treatment. This strategy is open to a number of problems, including monitoring difficulties and perverse incentives on the part of providers. However, at least one evaluation (in Rwanda) suggests it may be effective in improving some (easily observed) aspects of performance.
- Other approaches to poor public health sector performance involve increasing the role of the private sector rather than directly reforming public care. Responding to demand, the private sector in Africa already provides a significant share of health services and products even without public subsidies. However, there is a tension between the need for incentives for private providers and the policy objective (arising from externalities and equity considerations) of providing households with low cost or free access. Different strategies attempt to deal with this by combining, in one way or another, private sector delivery with public financing. Contracting on a large scale with the private sector to provide certain services such as immunizations has been shown to be feasible in a number of countries in Africa and elsewhere. These experiences almost always involve NGOs rather than for-profit private providers; hence their success probably relies to some extent on intrinsic motivations of NGOs to provide high quality services rather than (only) on financial incentives. Against these successes, contracting out the government's traditional function as a provider of basic health services plausibly will prevent or delay the development of the public sector's capacity in the sector. This issue has largely been ignored and deserves consideration.
- Another means of combining public subsidization with private delivery is through the use of vouchers. This is most likely to work where monitoring of service quality is not essential and there is a well developed commercial network of provision or (in the case of health products) distribution. Insecticide-treated bednets provide a good illustration of this, and evidence clearly shows that private networks of production and distribution for this product can be developed in Africa. However, in the one country (Tanzania) where a system of vouchers for bednets was aggressively implemented on a national scale, delivery still fell short among the targeted population of pregnant women. As a result, the government was forced to resort, at least temporarily, to free public (universal) distribution. Hence the tension between private sector incentives and public subsidies for those in need is difficult to completely resolve.

A number of broader conclusions emerge from this review with respect to evaluation and methodology:

- Several interventions that are relatively simple in design and attempt to affect individual behavior, such as demand side price incentives for health services or products or door-to-door promotion of hand washing with soap, show a consistency in terms of outcomes across contexts. Other, more complex interventions that involve the behavior of both individuals and institutions (including institutions for collective behavior) do not. Examples are community monitoring and the Integrated

Management of Childhood Illness. Outcomes for the former will be affected by local institutions for collective action and the political power of public health sector personnel; the latter, by administrative capacity, existing health infrastructure, and political will as well as local institutions for outreach. Given how widely these factors vary across contexts, it is not surprising that outcomes also vary. Efforts to understand these constraints should be a routine part of program evaluation—and *ex ante*, should inform the design of programs to the extent possible.

- Regarding scale up, effective delivery of services at the national level depends on a range of factors, including financial resources, physical and human health service infrastructure, the organization of the health sector, and institutions for monitoring and incentives. The importance of these constraints comes up time and again in the broader evaluation literature (the IMCI evaluation is one example). Since they are not binding, or less binding, in small scale experiments or other pilot evaluations, the results from these studies are inevitably limited when it comes to predicting ultimate program outcomes. Therefore despite the advantages of randomized trials, other approaches are (usually) needed to capture the realities of scaled-up program implementation under routine conditions. For example, the discussion in this paper at times drew on relatively simple comparative analysis of trends in different countries or in regions within countries where there were sharp differences in policy evolution, with careful attention to potential confounding factors.
- At the same time, and somewhat paradoxically, experiments are well situated to investigate how constraints to effective delivery can be overcome. The randomized evaluations discussed in this paper are not clinical efficacy studies of specific medical treatments; instead, they are designed to shed light on how behavior—of households and health care providers—impact uptake and outcomes of programs and how this behavior might be changed. The range of factors explored includes price incentives (to households and to providers), health promotion and education, externalities and peer effects, community oversight, and contracting with private providers. Results from these and future experiments will surely contribute significantly to the design of more effective health interventions.

REFERENCES

- Agarwal, B. 2000. "Conceptualizing Environmental Collective Action: Why Gender Matters." *Cambridge Journal of Economics* 24(3): 283–310.
- Akin, J., S. C. Griffin, D. K. Guilkey, and B. M. Popkin. 1984. *The Demand for Primary Health Care in Developing Countries*. Totowa, NJ: Littlefield, Adams.
- Alderman, H., J. Hoddinott, and W. Kinsey. 2006. "Long Term Consequences of Early Childhood Malnutrition." *Oxford Economic Papers* 58(3): 450-74.
- Alderman, Harold, and Victory Lavy. 1996. "Household Responses to Public Health Services: Cost and Quality Tradeoffs." *World Bank Research Observer* 11(1): 3-22.
- Angeles, Gustavo, David K. Guilkey, and Thomas A. Mroz. 1998. "Purposive Program Placement and the Estimation of Family Planning Program Effects in Tanzania." *Journal of the American Statistical Association* 93(433): 884-99.
- Armstrong Schellenberg J. R., T. Adam, H. Mshinda, H. Masanja, G. Kabadi, O. Mukasa, T. John, S. Charles, R. Nathan, K. Wilczynska, L. Mgalula, C. Mbuya, R. Mswia, F. Manzi, D. de Savigny, D. Schellenberg, C. Victora. 2004. "Effectiveness and Cost of Facility-based Integrated Management of Childhood Illness (IMCI) in Tanzania." *Lancet* 364(9445): 1583-94.
- Ashraf, Nava, James Berry, and Jesse Shapiro. 2008. "Can Higher Prices Stimulate Product Use? Evidence from a Field Experiment in Zambia." NBER Working Paper No.13247.
- Ashraf, Nava, Dean Karlan, and Wesley Yin. 2006. "Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines." *Quarterly Journal of Economics* 121(2): 635-72.
- Attanasio, O., C. Meghir, and A. Santiago. 2005. "Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate Progreso." Working Paper EWP05/01, Institute for Fiscal Studies.
- Aziz, K., B. Hoque, K. Hasan, M. Patwary, S. Huttly, M. Rahaman, and R. Feachem. 1990. "Reduction in Diarrhoeal Diseases in Children in Rural Bangladesh by Environmental and Behavioral Modifications." *Transactions of the Royal Society of Tropical Medicine and Hygiene* 84(3): 433-38.
- Bajari, P., H. Hong, J. Krainer, and D. Nekipelov. 2006. "Estimating Static Models of Strategic Interactions." University of Minnesota Working Paper.
- Banerjee, Abhijit, Esther Duflo, Rachel Glennerster, and Dhruva Kothari. 2008a. "Improving Immunization Coverage in Rural India: A Clustered Randomized Controlled Evaluation of Immunization Campaigns with and without Incentives." Mimeo, Massachusetts Institute of Technology.

- Banerjee, Abhijit, Rukmini Banerji, Esther Duflo, Rachel Glennerster, and Stuti Khemani. 2008b. "Pitfalls of Participatory Programs: Evidence From a Randomized Evaluation in Education in India." NBER Working Paper No. 14311.
- Banerjee, Abhijit, Angus Deaton, and Esther Duflo. 2004. "Wealth, Health, and Health Services in Rural Rajasthan." *American Economic Review* 94(2): 326-30.
- Banerjee, Abhijit, and Esther Duflo. 2006. "Addressing Absence." *Journal of Economic Perspectives* 20(1): 117-32.
- Bang, Abhay T., Rani A. Bang, Sanjay B. Baitule, M. Hanimi Reddy, and Mahesh D. Deshmukh. 1999. "Effect of Home-based Neonatal Care and Management of Sepsis on Neonatal Mortality: Field Trial in Rural India." *Lancet* 354(9194): 1955-61.
- Bang, A. T., R. A. Bang, O. Tale, P. Sontakke, J. Solanki, R. Wargantiwar, and P. Kelzarkar. 1990. "Reduction in Pneumonia Mortality and Total Childhood Mortality by Means of Community-based Intervention Trial in Gadchiroli, India." *Lancet* 336(8709): 201-6.
- Baqui, Abdullah H., Shams El-Arifeen, Gary L. Darmstadt, Saifuddin Ahmed, Emma K. Williams, Habibur R. Seraji, Ishtiaq Mannan, Syed M. Rahman, Rasheduzzaman Shah, Samir K. Saha, Uzma Syed, Peter J Winch, Amnesty Lefevre, Mathuram Santosham, and Robert E. Black, for the Projahnmo Study Group. 2008. "Effect of Community-based Newborn-care Intervention Package Implemented through Two Service-delivery Strategies in Sylhet District, Bangladesh: A Cluster-randomised Controlled Trial." *Lancet* 371(9628): 1936-44.
- Behrman, Jere R., Hans-Peter Kohler, and Susan C. Watkins. 2002. "Social Networks and Changes in Contraceptive Use over Time: Evidence from a Longitudinal Study in Rural Kenya." *Demography* 39(4): 713-36.
- Bhushan, I., S. Keller, and B. Schwartz. 2002. *Achieving the Twin Objectives of Efficiency and Equity: Contracting Health Services in Cambodia*. ERD Policy Brief Series Number 6, Economic and Research Department, Asian Development Bank, Manila, Philippines.
- Bhutta, Z. A, T. Ahmed, R. E. Black, S. Cousens, K. Dewey, E. Giugliani, B. A. Haider, B. Kirkwood, S. S. Morris, H. P. S. Sachdev, and M. Shekar for the Maternal and Child Undernutrition Study Group. 2008. "What Works? Interventions for Maternal and Child Undernutrition and Survival." *Lancet* 371(9610): 417-40.
- Björkman, Martina, and Jakob Svensson. 2009. "Power to the People: Evidence from a Randomized Field Experiment of Community-Based Monitoring in Uganda." *Quarterly Journal of Economics* 124(2): 735-69.
- Blas, E., and Limbambala, M. 2001. "User Payment, Decentralisation and Health Service Utilisation in Zambia." *Health Policy and Planning* 16 (Suppl. 2): 19-28.

- Boerma, J. Ties, Jennifer Bryce, Yohannes Kinfu, Henrik Axelson, and Cesar G. Victora. 2008. "Mind the Gap: Equity and Trends in Coverage of Maternal, Newborn, and Child Health Services in 54 Countdown Countries." *Lancet* 371(9620): 1259-67.
- Bryce, Jennifer, Shams el Arifeen, George Pariyo, Claudio F. Lanata, Davidson Gwatkin, Jean-Pierre Habicht, and the Multi-Country Evaluation of IMCI Study . 2003. "Reducing Child Mortality: Can Public Health Deliver?" *Lancet* 362(9378): 159-64.
- Bryce, Jennifer, Nancy Terreri, Cesar G. Victora, Elizabeth Mason, Bernadette Daelmans, Zulfiqar A. Bhutta, Flavia Bustreo, Francisco Songane, Peter Salama, Tessa Wardlaw. 2006. "Countdown to 2015: Tracking Intervention Coverage for Child Survival." *Lancet* 368(9541): 1067-76.
- Bryce, Jennifer, Cesar G. Victora, Jean-Pierre Habicht, Robert E. Black, and Robert W. Scherpbier on behalf of the MCE-IMCI Technical Advisors. 2005. "Programmatic Pathways to Child Survival: Results of a Multi-country Evaluation of Integrated Management of Childhood Illness." *Health Policy and Planning* 20(Suppl. 1): i5-i17.
- Bryce, Jennifer, Cesar G. Victora, Jean-Pierre Habicht, J. Patrick Vaughn, and Robert E. Black. 2004. The Multi-country Evaluation of the Integrated Management of Childhood Illness Strategy: Lessons for the Evaluation of Public Health Interventions." *American Journal of Public Health* 94(3): 406-15.
- Buddlemeyer, H., and E. Skofias. 2003. "An Evaluation on the Performance of Regression Discontinuity Design on PROGRESA." IZA Discussion Papers 827, Institute for the Study of Labor (IZA).
- Bühler, Christoph, and Hans-Peter Kohler. 2003. "Talking about AIDS: The Influence of Communication Networks on Individual Risk Perceptions of HIV/AIDS Infection and Favored Protective Behaviors in South Nyanza District, Kenya." *Demographic Research* S1(13): 397-438.
- Burnham, G., G. Pariyo, E. Galiwango, and F. Wabwire-Mangen. 2004. "Discontinuation of Cost Sharing in Uganda." *Bulletin of the World Health Organization*, 82(3): 187-95.
- Caldwell, John C. 1986. "Routes to Low Mortality in Poor Countries." *Population and Development Review* 12(2): 171-220.
- Castro-Leal, Florencia, Julia Dayton, Lionel Demery, and Kaplana Mehra. 1999. "Public Social Spending in Africa: Do the Poor Benefit?" *World Bank Research Observer* 14(1): 49-72.
- Chambers, Robert. 2009. "Going to Scale with Community-led Total Sanitation: Reflections on Experience, Issues and Ways Forward." IDS Practice Paper 1, Institute of Development Studies at the University of Sussex, Brighton, UK.

- Chaudhury, N., J. Hammer, M. Kremer, and K. Muralidharan, and F. H. Rogers. 2006. "Missing in Action: Teacher and Health Worker Absence in Developing Countries." *Journal of Economic Perspectives* 20(1): 91-116.
- Cohen, Jessica, and Pascaline Dupas. 2008. "Free Distribution or Cost-Sharing? Evidence from a Randomized Malaria Prevention Experiment." NBER Working Paper No.14406.
- Conley, T., and C. Udry. 2001. "Social Learning through Networks: The Adoption of New Agricultural Technologies in Ghana." *American Journal of Agricultural Economics* 83(3): 668-73
- Cook, T. D., W. R. Shadish, and V. C. Wong. 2006. "Within Study Comparisons of Experiments and Non-Experiments: Can They Help Decide on Evaluation Policy." Draft, Northwestern University.
- Crook, Richard Charles. 2003. "Decentralisation and good goverance." In Blindenbacher, Raoul, and Arnold Koller, (eds.), *Federalism in a Changing World: Learning from Each Other*. Montreal: McGill-Queens University Press.
- Curtis, V., and S. Cairncross. 2003. "Effect of Washing Hands with Soap on Diarrhea Risk: A Systematic Review." *Lancet Infectious Disease* 3(5): 275–81.
- Cutler, D., and G. Miller. 2005. "The Role of Public Health Improvements in Health Advances: The 20th Century United States." *Demography* 42(1): 1-22.
- Deaton, Angus S. 2009. "Instruments of Development: Randomization in the Tropics, and the Search for the Elusive keys to Economic Development." NBER Working Paper No. 14690.
- Deininger, K. and Mpuga, P. 2004. "Economic and Welfare Effects of the Abolition of Health User Fees: Evidence from Uganda. World Bank Policy Research Working Paper 3276, World Bank, Washington, DC.
- Diaz, J. J., and S. Handa. 2006. "An Assessment of Propensity Score Matching as a Non Experimental Impact Estimator: Evidence from Mexico's Progresa Program." *Journal of Human Resources* 41(2): 319-45.
- Diop, F., A. Yazbeck, and R. Bitran. 1995. "The Impact of Alternative Cost Recovery Schemes on Access and Equity in Niger." *Health and Policy Planning* 10(3): 223-40.
- Dickinson, K., and S. K. Pattanayak. 2007. "Open Sky Latrines. Do Social Interactions Influence Decisions to Use Toilets?" RTI International Working Paper.
- Donabedian, A. 1980. "The definition of quality and approaches to its assessment." In *Explorations in Quality Assessment and Monitoring Vol. 1. The Definition of Quality and Approaches to Its Assessment*. Ann Arbor, MI: Health Administration Press, 1980.

- Duflo, Esther. 2001. "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment." *American Economic Review* 91(4): 795-813.
- Duflo, Esther. 2003. "Grandmothers and Granddaughters: Old Age Pension and Intrahousehold Allocation in South Africa." *World Bank Economic Review* 17(1): 1-26.
- Duflo, E., R. Glennerster, and M. Kremer. 2008. "Using randomization in development economics research: A toolkit." In Schultz, T. and J. Strauss, (eds.), *Handbook of Development Economic*, Vol. 4. Amsterdam: North-Holland, 3895-962.
- Duflo, Esther, and Michael Kremer. 2008. "Use of randomization in the evaluation of development effectiveness." In Pitman, George Keith, Osvaldo V. Feinstein, and Gregory K. Ingram (eds.), *Evaluating Development Effectiveness*, Vol. 7. Washington, DC: World Bank, pp. 205-231.
- Duflo, E. C., M. Kremer, and J. Robinson. 2006. "Understanding Technology Adoption: Fertilizer in Western Kenya." Draft, Harvard University.
- Easterly, William. 2009. "Can the West Save Africa?" *Journal of Economic Literature*, Forthcoming.
- El Arifeen, Shams, Lauren S. Blum, D. M. Emdadul Hoque, Enayet K. Chowdhury, Rasheda Khan, Robert E. Black, Cesar G. Victora, and Jennifer Bryce. 2004. "Integrated Management of Childhood Illness (IMCI) in Bangladesh: Early Findings from a Cluster-randomised Study." *Lancet* 364(9445): 1595-1602.
- Esrey, Steven A. 1996. "Waste, Water, and Well-Being: A Multicountry Study." *American Journal of Epidemiology* 143(6): 608-22.
- Fauveau, Vincent, M. K. Stewart, J. Chakraborty, and S. A. Khan. 1992. "Impact on Mortality of a Community-based Programme to Control Acute Lower Respiratory Tract Infections." *Bulletin of the World Health Organization* 70(1): 109-16.
- Filmer, D., J. Hammer, and L. Pritchett. 2000. "Weak Links in the Chain: A Diagnosis of Health Policy in Poor Countries." *World Bank Research Observer* 15(2): 199-224.
- Filmer, D., J. Hammer, and L. Pritchett. 2002. "Weak Links in the Chain II: A Prescription for Health Policy in Poor Countries." *World Bank Research Observer* 17(1): 47-66.
- Franco, L. M., S. Bennett, and R. Kanfer. 2002. "Health Sector Reform and Public Sector Health Work Motivation: A Conceptual Framework." *Social Science and Medicine* 54(8): 1255-66.
- Frankenberg, Elizabeth, Wayan Suriastini, and Duncan Thomas. 2005. "Can Expanding Access to Basic Healthcare Improve Children's Health Status? Lessons from Indonesia's 'Midwife in the Village' Program." *Population Studies* 59(1): 5-19.

- Frankish, J. 1986. "Day Hospital Fees and Accessibility of Essential Health Services." *South African Medical Journal* 70(5): 285-8.
- Gertler, P. 2004. "Do Conditional Cash Transfers Improve Child Health? Evidence from *Progresa's* Control Randomised Experiment." *American Economic Review* 94(2): 336-41.
- Gilson, L., S. Russell, and K. Buse. 1995. "The Political Economy of User Fees with Targeting: Developing Equitable Health Financing Policy." *Journal of International Development* 7(3): 369-401.
- Giné, Xavier, Dean Karlan, and Jonathan Zinman. 2009. "Put Your Money Where Your Butt Is: A Commitment Savings Account for Smoking Cessation." Draft, Yale University.
- Glick, P. 2009. "How Reliable Are Client Satisfaction Surveys? Evidence from Matched Health Care Facility and Household Surveys in Madagascar." *Social Science and Medicine* 68(2): 368-379.
- Glick, Peter. 2008. "Reproductive Health and Behavior, HIV/AIDS, and Poverty in Africa." Cornell Food and Nutrition Policy Program Working Paper No. 219, Cornell University, Ithaca, NY.
- Glazerman, S., D. Levy, and D. Myers. 2002. "Nonexperimental Replications of Social Experiments: A Systematic Review." Discussion Paper MPR Reference No.: 8813-300, Mathematica Policy Research, Inc., Princeton, NJ.
- Gosoni, Laura, Penelope Vounatsou, Adriana Tami, Rose Nathan, Hajo Grundmann, and Christian Lengeler. 2008. "Spatial Effects of Mosquito Bednets on Child Mortality." *BMC Public Health* 8: 356.
- Graff Zivin, Joshua S., Harsha Thirumurthy, and Markus Goldstein. 2006. "AIDS Treatment and Intra-household Resource Allocations: Children's Nutrition and Schooling in Kenya." NBER Working Paper No. 12689.
- Gwatkin, D. R., S. Rutstein, K. Johnson, E. Suliman, A. Wagstaff, and A. Amozou. 2007. "Socio-Economic Differences in Health, Nutrition, and Population within Developing Countries: An Overview." World Bank, Washington, DC.
- Han, A. M., and T. Hlaing. 1989. "Prevention of Diarrhoea and Dysentery by Hand Washing." *Transactions of the Royal Society of Tropical Medicine and Hygiene* 83(1): 2128-31.
- Hanson, K., T. Marchant, R. Nathan, H. Mponda, C. Jones, J. Bruce, H. Mshinda, and J. Armstrong-Schellenberg. 2008. "Vouchers for Scaling Up Insecticide-treated Nets in Tanzania: Methods for Monitoring and Evaluation of a National Health System Intervention." *BMC Public Health* 8: 205.

- Hanson, Kara, and Christian Lengeler. 2009. "Establishment of a Sustainable System of Net Distribution." *Bulletin von Medicus Mundi Schweiz* Nr. 112, April.
- Hawley, W.A., P. A. Phillips-Howard, F. O. ter Kuile, D. J. Terlouw, J. M. Vulule, M. Ombok, B. L. Nahlen, J. E. Gimnig, S. K. Kariuki, M. S. Kolczak, and A. W. Hightower. 2003. "Community-wide Effects of Permethrin-treated Bednets on Child Mortality and Malaria Morbidity in Western Kenya." *American Journal of Tropical Medicine and Hygiene* 68 (Suppl. 4): 121-7.
- Hayes R. J., and S. Bennett. 1999. "Simple Sample Size Calculation for Cluster-randomized Trials." *International Journal of Epidemiology* 28(2): 319-26.
- Heckman, J., H. Ichimura, and P. Todd. 1997. "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program." *Review of Economic Studies* 64(4):605-54.
- Heckman, J., H. Ichimura, J. Smith, and P. Todd. 1998. "Characterizing Selection Bias Using Experimental Data." *Econometrica* 66(5): 1017-98.
- Heckman, J., and J. Smith. 1995. "Assessing the Case for Social Experiments." *Journal of Economic Perspective* 9(2): 85-115.
- Heckman, J., and S. Urzua. 2009. "Comparing IV with Structural Models: What Simple IV Can and Cannot Identify." NBER Working Paper No. 14706.
- Heller, P. S. 1982. "A Model of the Demand for Medical and Health Services in Peninsular Malaysia." *Social Science and Medicine* 16(3): 267-84.
- Hill, Jenny, Jo Lines, and Mark Rowland. 2006. "Insecticide-treated Nets." *Advances in Parasitology* 61: 77-128.
- Hoffman, Vivian. 2008. "Psychology, Gender, and the Intrahousehold Allocation of Free and Purchased Mosquito Nets." Draft, Cornell University.
- Holla, Alaka, and Michael Kremer. 2009. "Pricing and Access: Lessons from Randomized Evaluations in Education and Health." Center for Global Development Working Paper Number 158, January.
- Hoque, B. A., T. Juncker, R. Sack, M. Ali, and K. Aziz. 1996. "Sustainability of a Water, Sanitation and Hygiene Education Project in Rural Bangladesh: A 5-Year Follow-Up." *Bulletin of the World Health Organization* 74(4): 431-7.
- Hussein, A., and Mujinja, P. 1997. "Impact of User Charges on Government Health Facilities in Tanzania." *East African Medical Journal* 74(12): 751-7.
- Huttly, S. R., A. D. Blum, B. R. Kirkwood, R. N. Emeh, and R. G. Feachem. 1987. "The Epidemiology of Acute Diarrhea in a Rural Community in Imo State, Nigeria." *Transactions of the Royal Society of Tropical Medicine and Hygiene* 81(5): 865-70.

- Imbens, Guido W. 2009. "Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)." NBER Working Paper No. 14896.
- Imbens, G., and J. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 61(2): 467-76.
- Jalan, J., and M. Ravallion. 2003. "Estimating the Benefit Incidence of an Antipoverty Program by Propensity Score Matching." *Journal of Business and Economic Statistics* 21(1): 19-30.
- Jütting, Johannes, Elena Corsi and Alois Stockmayer. 2005. "Decentralisation and Poverty Reduction." Development Centre Policy Insights No. 5, OECD, Paris.
- Khan, M. U. 1982. "Interruption of Shigellosis by Hand Washing." *Transactions of the Royal Society of Tropical Medicine and Hygiene* 76(2): 164-8.
- Kidane, Gebreyesus, and Richard H. Morrow. 2000. "Teaching Mothers to Provide Home Treatment of Malaria in Tigray, Ethiopia: A Randomised Trial." *Lancet* 356(9229): 550-5.
- Kipp, W., J. Kamugisha, P. Jacobs, G. Burnham, and T. Rubaale. 2001. "User Fees, Health Staff Incentives and Service Utilization in Kabarole District, Uganda." *Bulletin of the World Health Organization* 79(11): 1032-7.
- Kosek, M., C. Bern, and R. L. Guerant. 2003. "The Global Burden of Diarrheal Disease, As Estimated from Studies Published between 1992 and 2000." *Bulletin of the World Health Organization* 81(3): 197-204.
- Kremer, M., J. Leino, E. Miguel, and A. Zwane. 2007. "Spring Cleaning: Results from a Randomized Impact Evaluation of a Source Water Quality Improvement." Unpublished Working paper
- Lancet. 1991. "Diarrhoea and malnutrition." *Lancet* 338(8772): 921-2.
- Lengeler C. 2004. "Insecticide-treated Bed Nets and Curtains for Preventing Malaria." *Cochrane Database of Systematic Reviews*, Issue 2. Art. No.: CD000363.
- Leonard, K. L. 2008. "Is Patient Satisfaction Sensitive to Changes in the Quality of Care? An Exploitation of the Hawthorne Effect." *Journal of Health Economics* 27(2): 444-59.
- Leonard, David, and Kenneth Leonard. 2004. "The Political Economy of Improving Health Care for the Poor in Rural Africa: Institutional Solutions to the Principal-Agent Problem." *Journal of Development Studies* 40(4): 50-77.
- Leonard, K. L., and M. C. Masatu. 2005. "The Use of Direct Clinician Observation and Vignettes for Health Services Quality Evaluation in Developing Countries." *Social Science and Medicine* 61(9): 1944-51.

- Leonard, K. L., M. C. Masatu, and A. Vialou. 2007. "Getting Doctors to Do Their Best: The Roles of Ability and Motivation in Health Care." *Journal of Human Resources* 42(3): 682-700.
- Leonard, K. L., G. Mliga, and D. H. Mariam. 2002. "Bypassing Health Facilities in Tanzania: Revealed Preferences for Observable and Unobservable Quality." *Journal of African Economies* 11(4): 441-71.
- Levine, R. 2007. *Case Studies in Global Health: Millions Saved*. Sudbury, MA: Jones and Bartlett.
- Levy-Bruhl, D., A. Soucat, R. Osseni, J. M. Ndiaye, B. Dieng, X. De Bethune, A. T. Diallo, M. Conde, M. Cisse, Y. Moussa, K. Drame, and R. Knippenberg. 1997. "The Bamako Initiative in Benin and Guinea: Improving the Effectiveness of Primary Health Care." *International Journal of Health Planning and Management* 12 (Suppl 1): S49-79.
- Linnemayr, Sebastian, and Harold Alderman. 2008. "Almost Random: Evaluating a Large-Scale Randomized Nutrition Program in the Presence of Crossover." Policy Research Working Paper Series 4784, World Bank, Washington, DC.
- Litvack, Jennie I., and Claude Bodart. 1993. "User Fees Plus Quality Equals Improved Access to Health Care: Results of a Field Experiment in Cameroon." *Social Science and Medicine* 37(3): 369-83.
- Liu, X., D. Hotchkiss, S. Bose, R. Bitran, and U. Giedon. 2004. "Contracting for Primary Health Services: Evidence on Its Effects and a Framework for Evaluation." Bethesda, Maryland: Abt Associates, Partners for Health Reform Plus Project.
- Loevinsohn, B., and A. Harding A. 2005. "Buying Results? Contracting for Health Service Delivery in Developing Countries." *Lancet* 366(9486): 676-81.
- Lokshin, M., and R. Yemtsov. 2005. "Has Rural Infrastructure Rehabilitation In Georgia Helped The Poor?" *World Bank Economic Review* 19(2): 311-333.
- Luby, S., M. Agboatwalla, J. Painter, A. Altaf, W. Billhimer, and R. Hoekstra. 2004. "Effect of Intensive Hand Washing Promotion on Childhood Diarrhea in High-Risk Communities in Pakistan: A Randomized Control Trial." *Journal of the American Medical Association* 291(21): 2547-54.
- Maluccio, J. A., and R. Flores. 2005. "Impact Evaluation of a Conditional Cash Transfer Program: The Nicaraguan Red de Protección Social." Research Report No. 141, International Food Policy Research Institute, Washington, DC.
- Manandhar, D. S., D. Osrin B. P. Shrestha, N. Mesko, J. Morrison K. M. Tumbahangphe S. Tamang, S. Thapa, D. Shrestha, B. Thapa, J. R. Shrestha, A. Wade, J. Borghi, H. Standing, M. Manandhar, A. M. Costello, and Members of the MIRA Makwanpur trial team. 2004. "Effect of a Participatory Intervention with Women's Groups on

- Birth Outcomes in Nepal: Cluster-randomised Controlled Trial.” *Lancet* 364(9438): 970-9.
- Manski, C. 1990. “Nonparametric Bounds on Treatment Effects.” *American Economic Review Papers and Proceedings* 80(2): 319-23.
- Marek, Tonia, Issakha Diallo, Biram Ndiaye, and Jean Rakotosalama. 1999. “Successful Contracting of Prevention Services: Fighting Malnutrition in Senegal and Madagascar.” *Health Policy and Planning* 14 (4): 382-9.
- Mariko, M. 2003. “Quality of Care and the Demand for Health Services in Bamako, Mali: The Specific Roles of Structural, Process, and Outcome components.” *Social Science and Medicine* 56(6): 1183-96.
- Mason, John B., David Sanders, Philip Musgrove, Soekirman, and Rae Galloway. 2006. “Community health and nutrition programs.” In Jamison, Dean T., Joel G. Breman, Anthony R. Measham, George Alleyne, Mairam Claeson, David B. Evans, Prabhat Jha, Anne Mills, and Philip Musgrove (eds.), *Disease Control Priorities in Developing Countries*, 2nd Edition. Washington, DC: The World Bank.
- Mason, J. B., J. Hunt, D. Parker, and U. Jonsson. 1999. “Investing in Child Nutrition in Asia.” *Asian Development Review* 17 (1,2): 1-32.
- . 2001. “Improving Child Nutrition in Asia.” *Food and Nutrition Bulletin* 22 (3 Suppl.): 5-80.
- McPake, B., Hanson, K., and A. Mills. 1992. “Implementing the Bamako Initiative in Africa: A Review and Five Case Studies.” Public Health and Policy Publication No. 8, Department of Public Health and Policy, London School of Hygiene and Tropical Medicine, London.
- Meessen Bruno, Jean-Pierre I. Kashala, and Laurent Musango. 2007. “Output-based Payment to Boost Staff Productivity in Public Health Centres: Contracting in Kabutare District, Rwanda.” *Bulletin of the World Health Organization* 85(2): 108-15.
- Meuwissen, Liesbeth Emm. 2002. “Problems of Cost Recovery Implementation in District Health Care: A Case Study from Niger.” *Health Policy and Planning* 17(3): 304-313.
- Miguel, E., and M. K. Gugerty. 2005. “Ethnic Divisions, Social Sanctions, and Public Goods in Kenya.” *Journal of Public Economics* 89(11-12): 2325-68.
- Miguel, Edward, and Michael Kremer. 2004. “Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities.” *Econometrica* 72(1): 159-217.
- Montgomery, Mark R., and John B. Casterline. 1996. “Social Learning, Social Influence, and New Models of Fertility.” *Population and Development Review* 22 (Suppl.: Fertility in the United States: New Patterns, New Theories): 151-75.

- Mtango F. D., and D. Neuvians. 1986. "Acute Respiratory Infections in Children under Five Years. Control Project in Bagamoyo District, Tanzania." *Transactions of the Royal Society of Tropical Medicine and Hygiene* 80(6): 851-8.
- Mwabu, G., M. Ainsworth and A. Nyamete. 1993. "Quality of Medical Care and Choice of Medical Treatment in Kenya: An Empirical Analysis." *Journal of Human Resources* 28(4): 838-62.
- Mwabu, G., J. Mwanzia, and W. Liambila. 1995. "User Charges in Government Health Facilities in Kenya: Effect on Attendance and Revenue." *Health Policy and Planning*, 10(2): 164-70.
- Noor, Abdisalan M., Juliette J. Mutheu, Andrew J. Tatem, Simon I. Hay, and Robert W. Snow. 2009. "Insecticide-treated Net Coverage in Africa: Mapping Progress in 2000-07." *Lancet* 373(9657): 58-67.
- Palmer, N., L. Strong, A. Wali, and E. Sondorp. 2006. "Contracting Out Health Services in Fragile States." *BMJ* 332(7543): 718-21.
- Parker, Susan W., Petra E. Todd, and Kenneth I. Wolpin. 2006. "Within-Family Treatment Effect Estimators: The Impact of Oportunidades on Schooling in Mexico." Philadelphia: University of Pennsylvania.
<http://athena.sas.upenn.edu/~petra/papers/ksppaper.pdf>
- Pattanayak, S. K., K. L. Dickinson, J.-C. Yang, S. R. Patil, P. Praharaj, R. Mallick, and C. Poulos. 2007. "Nature's Call: Can a (Randomized) Social Mobilization Campaign Lead Households to Use Toilets and Reduce Diarrhea?" RTI International Working Paper 07-02, March.
- Pelltier, D., E. Frongillo, D. Schroeder, and J. Habicht. 1995. "The Effects of Malnutrition on Child Mortality in Developing Countries." *Bulletin of the World Health Organization* 73(4): 443-48.
- Pitt, Mark, and Nidhiya Menon. 2007. "Program Evaluation using Spatial Instruments: Evidence from Indonesia." Draft, Brandeis University.
- Pitt, Mark M., Mark R. Rosenzweig, and Donna M. Gibbons. 1993. "The Determinants and Consequences of the Placement of Government Programs in Indonesia." *World Bank Economic Review* 7(3): 319-48.
- Population Services International. 2003. "What is Social Marketing?" Available online at: http://www.psi.org/resources/pubs/what_is_smEN.pdf
- Pörtner, Claus C., Kathleen Beegle, and Luc Christiaensen. 2007. "Family Planning and Fertility: Estimating Program Effects Using Cross-sectional Data." World Bank, Washington, DC.

- Quick, R. E., A. Kimura, A. Thevos, M. Tembo, I. Shamputa, L. Hutwagner, and E. Mintz. 2002. "Diarrhea Prevention through Household-Level Water Disinfection and Safe Storage in Zambia." *American Journal of Tropical Medicine and Hygiene* 66(5): 584-9.
- Quick, R. E., L. V. Venczel, E. D. Mintz, L. Soletto, J. Aparicio, M. Gironaz, L. Hutwagner, K. Greene, C. Bopp, K. Maloney, D. Chavez, M. Sobsey, and R. V. Tauxe. 1999. "Diarrhea Prevention in Bolivia through Point-of-Use Disinfection and Safe Storage: A Promising New Strategy." *Epidemiology Infections* 122(1): 83-90.
- Rawlings, L., and G. M. Rubio. 2005. "Evaluating the Impact of Conditional Cash Transfer Programs." *World Bank Research Observer* 20(1): 129-55.
- Ray, I. 2004. "Water for All? Peri-Urban and Rural Water Delivery Options: The Case of India." Paper Presented to the University of California, Berkeley, Energy and Resources Group Fall Colloquium Series, 6 October, 2004.
- Reinikka, R. and J. Svensson. 2009. "Working for God? Evidence from a Change in Financing of Not-for-profit Health Care Providers in Uganda." *Journal of the European Economic Association*, forthcoming.
- Reller, E., C. Mendoza, M. Lopez, M. Alvarez, R. Hoekstra, C. Olson, K. Baier, B. Keswick, and S. Luby. 2003. "A Randomized Control Trial of Household-Based Flocculant-Disinfectant Drinking Water Treatment for Diarrhea Prevention in Rural Guatemala." *American Journal of Tropical Medicine and Hygiene* 69(4): 411-19.
- Ridde, V. 2003. "The Bamako Initiative in Burkina Faso: What Effect Did Fees-for-Services Have on Health Service Utilisation?" *Bulletin of the World Health Organization* 81(7): 532-8.
- Roberts, Leslie. 2007. "Combating Malaria: Battling Over Bed Nets." *Science* 318(5850): 556-9.
- Rodrik, Dani. 2008. "The New Development Economics: We Shall Experiment, but How Shall We Learn." Working Paper 2008-0142, Weatherhead Center for International Affairs, Harvard University, May.
- Schwartz, J. B., and I. Bhushan. 2004. "Improving Equity in Immunization through a Public-private Partnership in Cambodia." *Bulletin of the World Health Organization* 82(9): 661-7.
- Semenza, J., L. Roberts, A. Henderson, J. Bogan, and C. Rubin. 1998. "Water Distribution System and Diarrheal Transmission: A Case Study in Uzbekistan." *American Journal of Tropical Medicine and Hygiene* 59(6): 941-6.
- Shordt, K., and S. Cairncross. 2004. "Sustainability of Hygiene Behaviour and the Effectiveness of Change Interventions: Findings from a Multi-Country Research and

Implications for Water and Sanitation Programmes.” The Hague: International Water and Sanitation Centre.

- Sobsey, M. D., T. Handzel, and L. Venczel. 2003. “Chlorination and Safe Storage of Household Drinking Water in Developing Countries to Reduce Waterborne Disease.” *Water Science and Technology* 47(3): 221-8.
- Soucat, A., T. Gandaho, D. Levy-Bruhl, X. de Bethune, E. Alihonou, C. Ortiz, P. Gbedonou, P. Adovohekpe, O. Camara, J. M. Ndiaye, B. Dieng, and R. Knippenberg R. 1997. “Health Seeking Behavior and Household Expenditures in Benin and Guinea: The Equity implications of the Bamako Initiative.” *International Journal of Health Planning and Management* 12 (Suppl. 1): S137-63.
- Soucat, A., and A. Wagstaff. 2001. “HNP and the Poor: An Integrated Policy Framework for Improving the Outcomes for the Poor.” <http://siteresources.worldbank.org/INTPRS1/Resources/Presentations/session6.pdf>
- Spencer, H. C., D. C. O. Kaseje, W. H. Mosely, E. K. N. Sempebwa, A. Y. Huong, and J.M. Roberts. 1987. “Impact on Mortality and Fertility of a Community Based Malaria Control Programme in Saradidi, Kenya.” *Annals of Tropical Medicine and Parasitology* 81(Suppl. 1): S36–S45.
- Tendler, Judith, and Sara Freedheim. 1994. “Trust in a Rent Seeking World: Health and Government Transformed in Northeastern Brazil.” *World Development* 22(2): 1771-91.
- Thirumurthy, Harsha, Joshua Graff Zivin, and Markus Goldstein. 2008. “The Economic Impact of AIDS Treatment: Labor Supply in Western Kenya.” *Journal of Human Resources* 43(3): 511-52.
- Thomas, Duncan, Victor Lavy, and John Strauss. 1996. “Public Policy and Anthropometric Outcomes in Côte d’Ivoire.” *Journal of Public Economics* 61(2): 155-92.
- Thornton, Rebecca. 2008. “The Demand for and Impact of HIV Testing: Evidence from a Field Experiment.” *American Economic Review* 98(5): 1829-63.
- Todd, P. 2008. “Evaluating social programs with endogenous program placement and selection of the treated.” In Schultz, T. and John Strauss (eds.), *Handbook of Development Economics*, Volume 4. Amsterdam: North-Holland.
- United Nations. 2008. *The Millennium Goals Report 2008*. New York: United Nations.
- Victoria, Cesar G., Joanna Armstrong Schellenberg, Luis Huicho, João Amaral, Shams El Arifeen, George Pariyo, Fatuma Manzi, Robert W. Scherpbier, Jennifer Bryce, and Jean-Pierre Habicht. 2005. “Context Matters: Interpreting Impact Findings in Child Survival Evaluations.” *Health Policy and Planning* 20(Suppl. 1): i18-i31.

- Von Braun, Joachim, and Ulrike Grote. 2002. "Does decentralization serve the poor?" In Ahmad, Ehtisham and Vito Tanzi (eds.), *Managing Fiscal Decentralization*. New York: Routledge.
- Waddington, C., and K. Enyimayew. 1989. "A Price to Pay: The Impact of User Charges in Ashanti-Akim District, Ghana." *International Journal of Health Planning and Management* 4(1): 17-47.
- Waddington, C., and K. Enyimayew. 1990. "A Price to Pay, Part 2: The Impact of User Charges in the Volta Region of Ghana." *International Journal of Health Planning and Management* 5(4): 287-312.
- Willis, C., and C. Leighton. 1995. "Protecting the Poor under Cost Recovery: The Role of Means Testing." *Health Policy and Planning* 10(3): 241-56.
- Wilson, J., and G. Chandler. 1993. "Sustained Improvements in Hygiene Behaviour amongst Village Women in Lombok, Indonesia." *Transactions of the Royal Society of Tropical Medicine and Hygiene* 87(6): 615-6.
- World Bank. 2004. *World Development Report: Making Services Work for Poor People*. Washington, DC: World Bank/International Bank for Reconstruction and Development.
- World Health Organization. 2002. "Scaling-up Insecticide-Treated Netting Programmes in Africa. A Strategic Framework for Coordinated National Action." WHO/CDS/RBM/2002.43. Geneva: WHO, 2002
- World Health Organization (WHO). 2004. *The World Health Report 2004*. Geneva: WHO.
- Yates, R. 2004. "Voting with Their Feet: What Lessons Can Be Learnt from Increased Consumption of Public Services in Uganda?" *Health Policy and Development* 2(1): 48-51.
- Zwane, A. P., and M. Kremer. 2007. "What Works in Fighting Diarrheal Diseases in Developing Countries? A Critical Review." *World Bank Research Observer* 22(1): 1-24.