

(Prepared for the Handbook of Field Experiments)

Impacts and Determinants of Health Levels in Low-Income Countries

Pascaline Dupas and Edward Miguel

Final Draft, April 2016

Abstract: Improved health in low-income countries could considerably improve wellbeing and possibly promote economic growth. The last decade has seen a surge in field experiments designed to understand the barriers that households and governments face in investing in health and how these barriers can be overcome, and to assess the impacts of subsequent health gains. This chapter first discusses the methodological pitfalls that field experiments in the health sector are particularly susceptible to, then reviews the evidence that rigorous field experiments have generated so far. While the link from in utero and child health to later outcomes has increasingly been established, few experiments have estimated the impacts of health on contemporaneous productivity among adults, and few experiments have explored the potential for infrastructural programs to impact health outcomes. Many more studies have examined the determinants of individual health behavior, on the side of consumers as well as among providers of health products and services.

Contents

1. Introduction.....	3
2. Methodological Section.....	5
2.1. Experimenting to estimate impacts of health improvements: beware of externalities.....	7
2.2. Experimenting to understand the determinants of health behavior: beware of measurement	13
2.3. Research transparency, registration, and pre-analysis plans.....	15
3. Experimental estimates of the impact of health on individual productivity.....	17
3.1. Impacts of Adult Health and Nutrition on Productivity.....	18
3.2. Impacts of Child Health and Nutrition on Education	23
3.3. Impacts of Child Health and Nutrition on Later Outcomes	28
4. Environmental/Infrastructural Determinants of Health.....	31
5. Demand for health products and healthcare	35
5.1. Pricing Experiments.....	36
5.2. Liquidity experiments: credit and cash transfer experiments	48
5.3. Information Experiments.....	50
5.4. Schooling Experiments.....	59
5.5. Non-monetary Cost Experiments.....	60
5.6. Incentive Experiments	63
5.7. Psychology Experiments.....	64
5.8. Taking stock: How important is the role of present bias in explaining observed preventive health behaviors?	66
5.9. Commitment Experiments	70
6. Supply of Health Care	72
6.1. Experimental audit studies	73
6.2. Monitoring Experiments.....	76
6.3. Improving the quality of Informal Providers	83
7. Conclusion	86
References Cited	89

1. Introduction

The links between health and economic development are many and varied. One of the most robust stylized facts of economic development is that higher income levels correlate strongly with longer life spans, lower infant mortality, and reduced illness throughout the life course (Deaton 2013). Infectious diseases that kill millions in poor countries are largely unknown in the world's wealthy societies, while sophisticated new curative technologies, procedures, and pharmaceutical advances often appear in wealthy economies years before they are available in low-income regions. The recent ravages of HIV/AIDS in Sub-Saharan Africa – which has killed tens of millions, and counting – have only deepened the divide between the global health haves and have-nots. Even the briefest introspection makes it obvious that health levels are critical determinants of human wellbeing even beyond their impact on economic productivity, and that the health gaps across countries are a major contributor to global inequities.

This is a powerful and well-known pattern, but its underlying causes are not obvious or entirely clear. There are many channels that could plausibly contribute to the link between wealth and health. On the one hand, higher incomes may allow individuals, households and whole societies to invest more resources in better health prevention and treatment (as well as better nutrition), leading to improved health outcomes. On the other, individuals with poor health may not be able to work as long, as hard, or as effectively as their healthier peers, leading to lower incomes and living standards for themselves, and potentially for their households and offspring. Adding further complexity is the possibility that both economic outcomes and health levels might be jointly determined by a third factor, such as the effectiveness of government public good provision – which itself might be a function of the quality of government institutions – or the design of foreign aid programs. All of these mechanisms also potentially interact in complex ways.

The research community has worked over the past few decades to better understand each of these different potential relationships and mechanisms, and this survey is an attempt to draw together the evidence from experimental studies on these important topics. Making sense of the bidirectional relationship between health and income is critical for both our scholarly understanding of the world in which we live, and for the effective design of public policies. As we discuss in this chapter, the evidence base is growing on the impacts and determinants of health levels in developing countries, but there remain many knowledge gaps, and we highlight areas where further research would be useful. In doing so, we build on the many recent surveys and chapters that have tackled related issues.

One fundamental relationship that has been the focus of many studies is the causal link between improved health and economic productivity. This is actually a challenging relationship to estimate rigorously, given the bidirectional relationships described above. Several recent experimental studies have begun to make progress, and have shown large impacts of gains in both in utero and child health on later educational and economic outcomes (Glewwe and Miguel 2008). With a handful of notable exceptions, there is limited experimental evidence tracing links from child health gains to adult economic outcomes due to the paucity of long-term panel datasets in low-income settings (although this is starting to change with more sophisticated longitudinal data collection approaches). In the meantime, some of the best evidence in this area will have to come from natural experiments or other non-experimental designs (Bleakley 2010a). There is even less evidence on links between health levels and contemporaneous adult productivity. This is harder to estimate than it might appear in part due to difficulties both in intervening to improve adult health, as well as in measuring individual output in many economic occupations (e.g., household subsistence farming, self-employment and informal work) in poor countries, although a new body of evidence is attempting to overcome these challenges.

There has been much more high-quality experimental research on the demand for health, with many studies documenting that even poor households in developing countries spend large sums on acute health care. However, a growing body of evidence indicates that demand for many preventive technologies, as well as a range of new and seemingly useful health products, is far lower than might seem optimal (Dupas 2011a).

There has recently been active debate over the right “model” of health behavior among the poor, with an excellent recent survey (Kremer and Glennerster 2011) arguing that the low demand for prevention is consistent with the importance of present biased preferences for many individuals. We assess the patterns in the exploding literature in this area, describe evidence that we believe is consistent with this interpretation as well as other patterns that we believe may be more consistent with alternative theoretical models and interpretations. Overall, our view on the role of present bias is more nuanced and uncertain than Kremer and Glennerster (2011).

There is a growing consensus that supply-side constraints, both in terms of access and the quality of health provision, could be a major determinant of poor health outcomes in developing countries (Das and Hammer 2014). The lack of high quality health care in many settings is important in its own right, and also affects interpretation of other findings in this literature. The lack of high quality service options could be a key driver of the low demand for healthcare observed in many settings (other than in emergency curative situations, where the consequences of non-treatment are especially severe and immediate). The generally low quality of health provision in poor countries also makes it more challenging to study the link between health and economic

productivity, to the extent that even large-scale health policy reforms do not “deliver” better population health. Experimental studies that aim to improve individual health by working through existing health institutions may be stymied by the limited capacity of those institutions.

Throughout this chapter, we survey the high-quality evidence from developing countries on these topics, highlight holes in the existing literature, and point the way forward for future research in the area. Before we survey the evidence, we review some key methodological issues in Section 2 that pertain to field experiments in health and that are critical to reading the evidence. Section 3 surveys experimental estimates of the impact of health on individual productivity, including studies in child as well as adult interventions. Section 4 surveys recent experimental studies on the environmental determinants of health, including clean water interventions. Section 5 discusses the exploding literature on the demand for health and healthcare, while Section 6 surveys work on the supply of health care. The final section concludes.

2. Methodological Section

Experiments in the health sector have been prominent among the field experiments carried out in development economics over the past two decades, and they have highlighted a number of important methodological issues related to the estimation of externalities, variable measurement, and pre-registration and research transparency. We discuss each of these issues in turn in the three subsections that follow.

Before launching into the detailed discussion, a few observations about the differences between field experiments conducted by development economists and those carried out among clinical trialists and epidemiologists are in order. One key difference relates to the goals of their studies, in relation to the widely used distinction between *efficacy trials* and *efficiency trials*. Efficacy trials are designed to capture the impact of an intervention under the most controlled and ideal circumstances possible, while efficiency trials capture effects under more authentic real-world conditions (Singhal et al 2014).

In reality, many studies lie somewhere in between these two extremes, with both partial study control and some degree of realism. While medical researchers typically carry out both types of studies, and often make a sharp distinction between the two, most recent field experiments in economics have tended to be closer to efficiency trials. Many of these studies have been carried out in close collaboration with government or non-governmental organization (NGO) partners,

and have been implemented as “real” programs, rather than experiments that are carried out directly by the researchers themselves, as in many efficacy trials.

Even more importantly, there is a major difference in the role of theory in the design of experiments by economists versus health researchers. Economists (and other social scientists) often design experiments to shed light on underlying theoretical mechanisms, to inform ongoing theoretical debates, and measure and estimate endogenous behavioral responses. These behavioral responses may shed light on broader issues beyond the experimental intervention at hand, and thus could contribute to greater external validity of the results.

This distinction between the types of studies carried out by medical researchers versus development economists working on health topics has a number of important implications that will become apparent in the course of this chapter. One has to do with the quality “standards” and perceptions of the “risk of bias” in a particular design. For medical trialists accustomed to the CONSORT standards or other medical efficacy trial reporting guidelines, studies that do not feature double-blinding, and thus run the “risk” of endogenous behavioral responses to the medical intervention, are considered less reliable than those studies that employ double-blinding (for a detailed discussion, see Eble, Bourne and Elbourne 2015). While a few of the studies conducted by economists surveyed below do feature double-blinding (most notably Thomas et al 2003, 2006), in nearly all settings blinding participants to their status is either logistically difficult (for instance, if government partners are unwilling to distribute placebo treatments to some of their population) or even impossible.

To illustrate, how would you provide a placebo treatment in a study investigating the impact of the distribution of anti-malarial bednets? Even in settings that might seem promising for placebo treatments, such as the community-level deworming treatments discussed below at several points, blinding participants to their status is basically impossible. Deworming generates side effects (mainly gastrointestinal distress) in roughly 10% of those who take the pills, so community members in a placebo community would quickly deduce that they were in fact not receiving real deworming drugs if there are few or no cases of side effects.

As noted above, endogenous behavioral responses are often exactly what we economists (and other social scientists) set out to measure and estimate in our field experiments, and thus are to be embraced rather than rejected as symptomatic of a “low quality” research design that is at “high risk of bias”. Finally, economists’ interest in many cases in the cost-effectiveness, economic returns, or fiscal implications of particular real-world health interventions once again make efficacy trials of inherently less interest in most cases than more realistic effectiveness trial approaches. In fact, the differences in outcomes between efficacy trials and effectiveness trials are of great interest to

social sciences, since understanding why an intervention that “works” in a highly controlled settings might “fail” in a more realistic setting can shed light on the functioning – and limitations – of existing organizations and institutions.

Taken together, it is clear to us that the experimental literature on health interventions in economics (and increasingly in other social sciences such as political science) often has very different objectives than medical, public health and epidemiological research, and thus different methodologies are often called for. Researchers working on health topics in development economists have not simply been able to import existing medical trial methods wholesale, but have instead been quite innovative in developing new approaches to estimating externalities, in measurement, and regarding issues of pre-registration and transparency, as discussed in the three subsections below.

2.1. Experimenting to estimate impacts of health improvements: beware of externalities

Field experiments in development economics focusing on the health sector have been innovative in adapting and creating new approaches to estimating treatment externalities. Treatment externalities go by many different names in different subfields and disciplines – including spillovers, contamination, herd immunity, and indirect effects, among others – and they have been of interest to statisticians and epidemiologists for a long time (for classic treatments, refer to Cox 1958, Rubin 1990, Fine 1993, and Rosenbaum 2007). However, despite their theoretical importance for the health field, especially in low-income societies where infectious diseases account for a large share of the disease burden, these issues have received far less empirical attention in public health and epidemiological research.

As surveyed in a recent synthetic review (Benjamin-Chung et al. 2015), the rapid growth in empirical studies of treatment externalities in epidemiology began after 2000, at roughly the same time that such empirical treatments became more common in economics, although there were a handful of earlier empirical treatments of the issue (for instance, Paul et al 1962 and Cooper and Fitch 1983). This literature has tended to focus on low income settings in Asia, Latin America and Africa in both the public health literature and in economics. As Benjamin-Chung et al (2015) show in their detailed review of the existing literature, both these early public health studies, as well as most recent treatments (for instance, Forleo-Neto et al 1999, Ali et al 2005, 2008, 2013), tend to focus on “herd immunity” effects in vaccine treatment programs, and they estimate treatment spillovers using the “treatment coverage mean”, i.e., the proportion of individuals in the area who received treatment, as the key measure of exposure. They then examine whether there

is a lower risk of later infection in sites where more people received treatment relative to areas where fewer people were treated. These studies provide consistent support for the existence of positive vaccine spillover benefits among those who did not receive the vaccine themselves.

This is an empirically sensible approach but it has a number of immediate limitations. First, many studies using this approach typically leave the question of how and why particular sites had lower vaccine coverage than other areas unanswered. This is potentially problematic to the extent that coverage rates are affected by omitted variables (“confounders”) such as the local disease environment, capacity of local health institutions, and perhaps local attitudes towards particular diseases, all of which could both affect coverage as well as population health outcomes. For instance, it is plausible that areas where populations have less awareness of, or support for, treatment of a disease might both have lower coverage rates and higher subsequent infection (although there are other possible confounders that would lead to bias in the opposite direction). Second, this approach is typically quite imprecise about the geographic area that is relevant for spillovers, the choice of geographic area often appears quite ad hoc, and different studies use different definitions of a community or site, thus leading to a lack of comparability across studies. Taken together, this implies that the evidence base within public health regarding the extent of treatment spillovers is not extremely solid, and moreover, the evidence generated so far has tended to focus on a narrow set of treatments, namely, vaccinations.

While it may be surprising that there has not been more empirical research on the empirical estimation of spillover effects within public health and epidemiology (despite the theoretical centrality of these issues in these fields, i.e., they are even in the name “epidemiology”), we speculate that it may be the result of a tendency in most empirical health studies to focus on “standard” RCT empirical approaches that compare treatment to control groups directly, and that tend to regard any sort of externality effect as a source of “contamination” that is to be avoided or minimized, rather than as a key element of our understanding of the overall treatment effect. Economists who have worked on these topics in health have instead been more open to embracing the estimation of externalities, perhaps in part because norms regarding the “right” way to carry out field experiments are less established in economics (given how recently these tools have been adopted in the field), and also given the importance of spillover effects within public economics theory to potentially rationalize public subsidies for health treatments and interventions (Dybvig and Spatt 1983).

The Miguel and Kremer (2004) paper on school-based deworming impacts in Kenya is among the first health studies in development economics to experimentally estimate externality effects. In their main analysis, they exploit the variation in deworming treatment status generated by the experimental assignment of schools to early versus late treatment (in a phase-in, or stepped wedge,

research design), and show that this generates extensive variation in local “exposure” to treated schools within 3 km and up to 6 km away from sample schools. Their econometric approach conditions on the total density of local school pupil population within a particular geographic distance, and conditioning on this quantity, the experimental design implies that the number of treated schools is experimentally assigned and should thus be orthogonal to other local observables and unobservables. They cannot reject that the observed characteristics of schools with lots of “exposure” to other local treatment schools are the same as for schools with little such exposure. Thus this analytical approach – which Benjamin-Chung et al 2015 terms “estimation of spillovers conditional on treatment density” – addresses both of the limitations of most existing empirical research on spillovers in public health described above. In particular, the assignment of exposure to treatment spillovers is assigned experimentally (and thus should be largely free of the possible omitted variable bias, or confounding, that could affect most existing vaccine studies in epidemiology), and this approach also makes precise the extent of externalities within precisely defined geographic distances away from a particular site.

Formally, Miguel and Kremer (2004) estimate the following econometric model:

$$(eqn. 1) \quad Y_{ijt} = a + \beta T_{it} + X_{ijt}'\delta + \sum_d (\gamma_d \cdot N_{dit}^T) + \sum_d (\phi_d \cdot N_{dit}) + e_{ijt}.$$

Y_{ijt} is the individual health or education outcome, where i refers to the school, j to the student, and t the time period; T_{it} is an indicator variable for school assignment to deworming treatment; and X_{ijt} are school and pupil characteristics, and time controls. N_{dit} is the total number of pupils in primary schools at distance d from school i in year t , and N_{dit}^T is the number of these pupils in schools randomly assigned to deworming treatment. In their example, $d=03$ denotes schools that are located within three kilometers of school i , and $d=36$ denotes schools that are located between three to six kilometers away. Individual disturbance terms were assumed to be independent across schools, but are allowed to be correlated for observations within the same school. In this framework $\beta + \sum_d (\gamma_d \bar{N}_{dit}^T)$ is the average effect of deworming treatment on overall infection prevalence in treatment schools, where \bar{N}_{dit}^T is the average number of treatment school pupils located at distance d from the school. Under spatial externality models in which a reduction in worm prevalence at one school affects neighboring schools and this in turn affects their neighbors, some externalities would spill over to even greater distances, in which case equation 1 yields a lower bound on treatment effects, a point that Baird et al (2014) show formally. β captures both direct effects of deworming treatment on the treated, as well as any externalities on untreated pupils within the treatment schools.

Miguel and Kremer (2004) also use another source of variation to estimate spillover effects within treated school communities. Within communities, a subset of the population was not assigned to treatment, namely, older girls for whom the deworming drugs were considered potentially dangerous at the time of the original study (due to potential embryotoxicity), and other students simply did not receive treatment, usually because they did not attend school on the announced day of treatment or did not receive parental consent for treatment. The comparison of subsequent infection rates among those in treatment schools who did not themselves take deworming drugs, compared to those who did take the drugs, is potentially problematic due to non-random selection into treatment, and any such differences lack a reliable counterfactual (since time trends or other secular changes might affect both groups).

However, Miguel and Kremer (2004) are able to exploit the fact that the same treatment inclusion rules were used in subsequent years of the program as later treatment groups were phased into deworming, and they compare infection rates among those who did not receive deworming treatment when it was available in their school, to those in other schools who were not yet offered deworming but who we know did not receive treatment when later offered the opportunity. This approach potentially addresses much of the “selection” into deworming treatment, as long as patterns of selection remain roughly constant across years 1 and 2 of the study. This is a “within-cluster spillover effect”, and when focusing on the excluded older girls, Benjamin-Chung et al (2015) term it a “within-cluster spillover effect among ineligible”.

There is evidence for large and statistically significant externality effects on both worm infection rates, and on subsequent school participation rates, using both sources of variation in Miguel and Kremer (2004), namely, the spillover estimates conditional on local treatment density, and the within-cluster spillover effect. These effects are concentrated within school communities, and extend out to at least 3 km away from treatment schools.

In a follow-up study in the same area of Kenya, Ozier (2014) also generates within-cluster spillover effect estimates among ineligible, by focusing on children who were 0-2 years old when the program was launched, and thus were too young to have directly received deworming treatment. However, they were potentially affected by epidemiological spillovers generated by deworming treatment, since treating infected individuals means they are less likely to pass on worm larvae into the environment through fecal matter (the usual route of transmission for intestinal helminth infections). Ozier finds evidence that the youngest children (those under 2) gain substantially a full 10 years after deworming treatment in terms of their cognitive performance and academic test scores, with average gains of roughly half a school year of learning. This finding reinforces the results in Miguel and Kremer (2004) about the potentially large magnitude of positive deworming

treatment externalities in an area with high rates of worm infections; infection rates at baseline in this region of western Kenya were over 90%.

An implication of these externality effects is that research on infectious diseases – or other types of health or economic interventions – that does not account for externalities is likely to underestimate total program effects, both by potentially understating differences across the treatment and control groups (if the control group is gaining relative to the counterfactual of no exposure to spillovers), and by missing out on the spillovers entirely, thus doubly undercounting effects. The existence of treatment externalities thus makes cluster randomized designs – that treat most or all individuals in a given area, and consider the entire unit “treatment” in the analysis – more attractive than individually randomized designs in such settings, since treatment spillover benefits are at least partially “internalized” within the cluster (although as Miguel and Kremer 2004 show, some spillovers may extend beyond the cluster and these could be important to consider as well). We discuss this issue in greater detail below, but the presence of sizeable treatment externalities is a possible explanation for why several of the early studies of deworming treatment impacts on growth and cognition – all of which randomized across individuals within the same community or school – tended to find quite small (although typically positive) effects (see Dickson et al 2000), namely, that the control group gained considerably from the intervention, dampening effects. In contrast, both of the large cluster randomized experiments on deworming discussed below (the Miguel and Kremer 2004 study, as well as the Alderman et al 2007 project in Uganda) find both large short-run and long-run impacts of deworming on participant outcomes. The fact that many of the early deworming RCT’s were conducted by nutritionists (rather than epidemiologists) might help explain the minimal attention paid to these issues in those studies.

A large number of studies within economics – including both health and non-health studies – have subsequently utilized the same basic empirical approach as Miguel and Kremer (2004) in order to estimate the magnitude of treatment externalities. Some of these studies modify the estimator to focus on the share of individuals within one’s social network that are affected by a treatment, rather than relying on geographic distance per se, as in the original analysis. In the health sector, this includes studies of mental health (Baird et al 2013), water treatment (Ziegelhofer 2012), learning about HIV results (Godlonton and Thornton 2012), community monitoring of health clinic performance (Bjorkman and Svensson 2009), risky sexual behavior (Dupas 2011b), child nutrition (Zivin et al 2011, Fitzsimons et al 2012), family planning (Joshi and Schultz 2013), and malaria prevention (Tarozzi et al 2014, Dupas 2014), as well as a study of take-up of the deworming treatments themselves within a social network (Kremer and Miguel 2007), among many other related research studies.

As might be expected given the diversity of health conditions and behaviors that have been examined using these methods, the magnitude and range of externalities vary considerably across cases. However, it is worth mentioning the estimated effects in some of these cases. The case of malaria is particularly important, given how widespread the condition is in many low-income regions (especially in Africa) and its contribution to the total global burden of disease. Both of the malaria studies in economics mentioned above find suggestive evidence that positive externalities are generated when households use insecticide treated bednets, although neither has adequate statistical power to reach definitive conclusions (Tarozzi et al 2014, Dupas 2014). In contrast to deworming, the malaria spillover benefits tends to be localized within a community, and it appears to within 20 to 30 meters from the household using the net (Tarozzi et al 2014). This information on the magnitude and geographic extent of spillovers can be important for both public health planners, as well as for those considering the desirability of large public subsidies for these, or other, health interventions.

As alluded to above, In other recent work economists have moved beyond studying epidemiological spillovers directly (as in the deworming and malaria cases), and have begun to explore spillovers through social networks in terms of technology adoption and behavioral change (as in Kremer and Miguel 2007 and Dupas 2011b, for instance), and also the possibility that spillover could occur through channels other than epidemiology or social influence. For instance, one direct way that externalities might occur is through the sharing of medical treatment between those assigned to treatment and those assigned to control; in the case of a treatment such as iron supplementation which has limited side effects, this is something that might be considered quite low risk among participants (see the discussion of Thomas et al 2003, 2006 and Bobonis et al 2006 for studies on iron supplementation in this literature).

Recent research has made methodological progress in understanding how to most efficiently estimate externality effects, and how to address the possibility of nonlinearities in the relationship and complementarities with local treatment decisions. Bhattacharya et al (2013) exploit experimental variation combined with detailed geospatial information to estimate how the local subsidy rates faced by others affect insecticide-treated mosquito nets (ITN) use in Kenya, and show that there are important non-linearities in the subsidy incidence. The issue of possible nonlinearities in social effects are raised as a possibility in both Miguel and Kremer (2004) and Kremer and Miguel (2007) but in neither study was there sufficient statistical power to reject linear specifications. Baird et al (2014) discuss the optimal design of experiments to estimate spillover effects in settings where it is possible to randomize the intensity of treatment within clusters, and then randomly assign individual treatment conditional on this intensity. They include calculations

of statistical power to detect externality effects given program parameters, which is useful for those prospectively designing experiments with this aim.

In areas beyond health, spillover effects and related general equilibrium effects are increasingly being studied in a wide range of sectors including in the study of cash transfer programs, microfinance programs, and beyond, demonstrating the analytical usefulness of these approaches to economics research as a whole; Muralidharan and Sundararaman (2015) present an application to education, and Angelucci and Di Maro (2015) provide further discussion of such studies across subfields within development economics.

2.2. Experimenting to understand the determinants of health behavior: beware of measurement

Like other field experiments in development economics, experiments focusing on health topics have often relied on original data collection – including individual and household survey data, biomarker data, as well as data from clinics and schools – in the analysis. As they were with research design issues, these studies have also been highly innovative in their development of new data collection methodologies, as well as in clarifying some of the potential biases that could arise from these different types of original data collection. We discuss these different concerns – namely the direct effect of being surveyed on responses, social desirability bias, and the reliability of health self-reports – in turn in this subsection.

The simplest and potentially most pervasive form of bias from original data collection would occur if any act of being surveyed itself affected subsequent responses and, even more importantly, behaviors. Zwane et al (2011) quantifies the possible extent of this bias using data from multiple data collection activities in development economics, all of which featured some randomized variation in the frequency with which different groups of households or individuals were surveyed followed by administrative data collection on the outcome of interest, and show that the experience of being surveyed can often affect subsequent behavior in health studies, as well as in microfinance projects.

In the context of the health data that was featured in their analysis, the authors show the randomly chosen individuals in rural Kenya communities who were surveyed more frequently regarding their children’s health status (here, the diarrhea outcomes and other health dimensions for infants) were significantly more likely to change their behavior in the direction of making more investments in their children’s health, specifically, in the use of a point-of-use chlorine disinfectant for household water. These behavioral responses were large in magnitude and statistically significant among the households surveyed at high frequency (biweekly) relative to those surveyed

infrequently (every six months), with a near doubling in use of chlorine disinfectant. This response also appears to have led to large reductions in reported diarrhea, and they are large enough to change the estimated effect of an ongoing water investment campaign (namely, spring protection) in the same region. Taken together, the authors suggest that frequent surveys may serve as a reminder to households to engage in particular health practices, similar to the effect that has been documented for explicit reminders through mobile phone and other means (for instance, see Pop-Eleches et al 2011).

This has extremely important implications for health studies. While many economics studies collecting original data utilize relatively infrequent data collection (presumably for reasons of cost), some like those discussed in Zwane et al (2011) do make use of high frequency data collection, and such approaches are actually the “standard” in many public health studies, such as those studying diarrhea outcomes in children (the health data in Zwane et al was modeled on these approaches). Data reliability might be improved to the extent that data can be collected less frequently from a larger sample of individuals, or to the extent that more “passive” forms of data collection, such as from administrative records or “big data” sources (such as mobile phone usage), rather than high frequency enumerator visits. An alternative that is increasingly employed in field experiments in development economics is the creation of a “pure control” group of households or communities who are not contacted by the research team or surveyed until the very end of the study, when outcome measures are collected; for an example of a study that uses this approach, see Muralidharan and Sundararaman (2011). These individuals are thus unlikely to have been affected by the process of data collection, and any such bias on the “regular control” group can also be quantified in this way.

A related but distinct concern with original data collection relying on surveys is the possibility that respondents will provide answers that they think the enumerators want to hear, what is known as social desirability bias, or experimenter demand effects. These are widely discussed in laboratory data collection in experimental economics, and are increasingly recognized as a major concern in field experimental data collection settings.

In many settings where sensitive health information is collected, researchers are increasingly creating “private” situations within the data collection encounter for them to enter in such data in a way that cannot be immediately verified by the enumerator (for instance, see Baird et al 2008). These concerns may be particularly pronounced when it comes to reproductive health and sexual health topics. To address these concerns, scholars have recently been quite creative in employing enumerators who are more likely to elicit truthful responses from respondents, e.g., Robinson and Yeh (2011, 2012) hire former sex workers to privately survey other sex workers on their sexual practices and decision-making.

An alternative approach that creates privacy for respondents is a survey technique called list randomization. List randomization, also known as the item count or unmatched count technique, allows respondents to report on potentially sensitive behavior without allowing the researcher or surveyors to identify individual responses. In practice, some proportion of survey respondents are randomly selected to receive a short list of statements (e.g., general health choices and outcomes, say) and asked to report how many, but not which, statements are true. The remainder of the survey respondents are presented with the same list of statements and one key additional statement designed to capture sensitive behavior (e.g., regarding sensitive sexual behavior). By subtracting the mean number of true statements in the first group from the mean number of true statements in the second group, researchers can estimate the proportion of the sample that engages in the sensitive behavior. This approach has been widely used to study health behaviors in many contexts (see Droitcour et al 1991, LaBrie and Earleywine 2000, Chong et al. 2013), as well as sensitive behavioral choices in other spheres (Karlan and Zinman 2011), though an obvious limitation of this method is that it only generates group-level outcomes, not individual-level ones.

Even when techniques such as creating a private space for survey respondents, employing more appropriate enumerators, and list randomizations are used, there remain important concerns about the validity of self-reported health behaviors, especially in sensitive areas such as sexual and reproductive health. A growing number of studies have documented a sharp divergence between self-reported sexual behavior and objectively measured infection status. In data collected from over 10,000 adolescents in Western Kenya, Duflo et al. (2015) find that 4.6% of girls and 4.8% of boys who report that they never had sex test positive for Herpes Simplex Virus type 2 (HSV2), a sexually transmitted infection, suggesting pervasive misreporting. Gong (2015) uses field experimental data from Kenya and Tanzania in the context of an HIV/AIDS related testing and information campaign, and shows that self-reported sexual behavior becomes less risky for individuals who were informed that they had tested HIV-positive, even while their incidence of STI infections – a more reliable measure of risky sex – increases significantly.

2.3. Research transparency, registration, and pre-analysis plans

There is growing awareness in economics fields that current research methods and practices can sometimes produce misleading bodies of evidence (Miguel et al. 2014), and in many ways this growing awareness in the social sciences parallels earlier trends in medical research, as we discuss below. For instance, there is growing evidence documenting the prevalence of publication bias in economics, as well specification searching, and widespread inability to replicate empirical findings. While some of these issues have been widely discussed within economics for some time (see Leamer (1983), Dewald, Thursby, and Anderson (1986) and DeLong and Lang (1992), among others),

there has been a recent flurry of activity documenting these problems, and also generating ideas for how to make research more transparent and reproducible.

A leading proposed solution to the problem of publication bias is the registration of empirical studies in public registry. This would ideally be a centralized database of all attempts to conduct research on a certain question, irrespective of the nature of the results, and such that even null (not significant) findings are not lost to the research community. The most high profile attempt at a registry within Economics, and indeed, across the social sciences, is the new AEA Randomized Trial Registry (Katz et al 2013), which was launched in May 2013.

In another example of intellectual exchange across economics and health, the AEA registry was explicitly inspired by existing registries for medical trials. Clinical trials began being registered in large numbers in the 1990s, but the proportion registered has increased dramatically since roughly 2005, when more stringent requirements were placed on medical researchers seeking to publish in leading medical journals, as well as by government medical regulatory authorities. While recent research in medicine finds that the registry has not eliminated all under-reporting of null results or other forms of publication bias and specification searching (Laine et al 2007, Mathieu et al 2009), they do over time help constrain inappropriate practices, and at a minimum the existence of a registry allows the research community to quantify the extent of these problems. It also helps scholars locate studies that are delayed in publication, or are never published, helping to fill in gaps in the literature and thus resolving some of the problems of “disappearing” null results identified in Franco et al (2014).

Though it is too soon after the adoption of the AEA’s trial registry to measure the impact, the registry is being used by many empirical researchers: in its first two years, over 500 studies conducted in over 80 countries had been registered, and the number continues rising each month. In addition to the AEA registry, several other registries have recently been created across the social sciences, although they have received fewer studies and less attention so far. These include registries created by the International Initiative for Impact Evaluation (3ie) for international development studies (the Registry for International Development Impact Evaluations, RIDIE, launched in September 2013), and the Experiments in Governance and Politics (EGAP) registry, also created in 2013.

Parallel to the trend in the pre-registration of studies, support has grown for including pre-analysis plans (PAP’s) that can be posted and time stamped even before data are collected or are otherwise available for analysis in prospective studies (Miguel et al 2014). While there were scattered earlier cases of pre-analysis plans being utilized in the social sciences (most notably Neumark 2001), the

numbers of published papers using pre-specified analysis has grown rapidly in the past few years, mirroring the rise of studies on the AEA registry. Some of these early uses of pre-analysis plans in Economics are in health economics, most notably the influential Oregon Health Insurance experiment studied in Finkelstein et al (2012). This is not unexpected given how widespread pre-registration of studies and analysis plans has become within medical research. However, most economics studies using pre-analysis plans have been within development economics (see Casey et al 2012 and Olken et al 2014 among others). Casey et al (2012) show how the lack of a pre-analysis plan might have provided sufficient latitude for an unscrupulous researcher to report a wide range of different – and erroneous – conclusions using the same data, heightening concern about the possible extent of specification searching and biased reporting even in studies using randomized experimental designs.

There remain many open questions about whether, when, and how pre-analysis plans could and should be used in Economics research, with open debates about how useful they are in different subfields of the discipline (Olken 2015, Coffman and Niederle 2015)). Yet even among these authors, who are critical about widespread adoption of pre-analysis plans in all cases, there appears to be a growing consensus that, in certain situations – such as large-scale randomized trials that are expensive or difficult to repeat, and/or cases where a government, policymaker, or corporation has a vested interest in the outcome – pre-analysis plans can increase the credibility of reporting and analysis. To follow-up on Leamer’s (1983) famous pun, pre-analysis plans can help keep the “con” out of randomized controlled trials.

3. Experimental estimates of the impact of health on individual productivity

There is a growing literature within Economics that uses experimental variation to estimate the impact of health status on various measures of individual productivity. In this section, we focus on the experimental studies on this topic, and largely ignore the vast observational literature on these issues, a literature that spans economics, public health, epidemiology, and medical trials.

It is useful to divide the emerging experimental health economics literature on this topic into three groups: first, those studies that directly examine the impact of improved health status on current adult labor productivity and other economically relevant outcomes; second, those studies that examine the impact of improved child health and nutrition on current educational and other outcomes; and finally, those studies that estimate longer-term persistent effects of earlier health

investments on later productivity measures and other life outcomes. We consider these in turn below.

3.1. Impacts of Adult Health and Nutrition on Productivity

An important early experiment that estimated the effect of adult health status on contemporaneous measures of productivity and individual well-being is Thomas et al. (2003, 2006), the Work and Iron Status Evaluation (WISE). The intervention aimed to address iron deficiency anemia (IDA), one of the world’s most widespread health and nutritional problems. IDA is well-known to lead contribute to physical weakness and lower aerobic capacity, and thus could plausibly affect individual labor productivity. The WISE study features a randomized evaluation of iron supplementation (weekly supplements of 120 mg of iron) plus deworming to a large sample of over 17,000 adults in Indonesia, with ages ranging from 30 to 70 years old. Since roughly 30 percent of the sample were infected with intestinal helminths at baseline, the impact of the intervention should be interpreted as the combined effect of iron and deworming.

It is worth noting that WISE features an unusual study methodologically within economics – although not medical research – in that it was carried out as a double-blinded experiment, i.e., the control group received placebo pills of identical appearance. This might limit any behavioral responses to the treatment that are due to the fact that beneficiaries know they are receiving treatment. While this sort of design is considered the ideal for medical trials, it is debatable whether it constitutes a similar “gold standard” for social science research studies, where endogenous behavioral responses are often central to the theoretical framework motivating a given study. Double-blinding is possible for a relatively simple intervention, such as the iron supplementation and deworming in Thomas et al (2003, 2006), but it is also often logistically infeasible in more complicated interventions, or those in which participants themselves are called upon to make decision (for instance, in the studies of technology take-up discussed below).

Thomas et al (2003, 2006) follow-up participants in the WISE study for 6 months, and focus on the intention-to-treatment estimates of program impact. They first document that iron status does improve significantly in the treatment group, with particularly large gains among those whose baseline hemoglobin (Hb) level was particularly low (below 12 g/dL, a common cut-off for anemia). The heterogeneity in Hb gains as a function of baseline deficiency motivates an estimation strategy that is similar to a difference-in-difference-in-differences (triple difference) approach: outcomes are compared between the treatment and control groups, over time (post-treatment versus baseline), across groups that had relatively low Hb at baseline (below 12.5 g/dL) versus relatively high Hb. This approach provides more statistical power than the simple treatment versus control difference, since a large share of individuals in the treatment group, namely those with relatively high Hb at baseline, do not experience any gains in Hb as a result of

the intervention, and thus would not necessarily be expected to experience any gains in productivity.

In their 2006 working paper (Thomas et al. 2006), the authors report evidence of sizeable and statistically significant gains in a range of economic and wellbeing outcomes, with effects particularly large for males (although they are generally of the same sign for females, although smaller in magnitude). The probability that individuals are not working falls significantly by between 3 to 5 percentage points for both males and females, there is suggestive evidence that total earnings increase for males, and statistically significant gains in self-employed total earnings and hourly earnings (which is similar to a wage measure) for males, as well. There are also substantial gains in psycho-social outcomes for both genders, with males finding having less difficulty sleeping and having more energy and leisure time, and females feeling less anxious. Given the relatively low cost of iron supplementation and deworming, and authors argue that this investment could have a high economic return. Given these returns, a question remains why individuals are not already doing making these sorts of investments in iron, deworming or improved nutrition more broadly in the absence of the intervention. It is also worth noting that even where there is an intervention, take-up of the nutritional supplements can remain limited (see the discussion on the demand for double-fortified salt in India in section 5.3). Thus while the rigorous research design, large sample size, and rich set of outcome measures make the preliminary evidence of the WISE study some of the most provocative estimates of the causal impact of improved adult health on contemporaneous economic productivity to date. As with all experiments the external validity of the results is unknown, and we hope that additional studies of this kind will be conducted in different settings to help further our knowledge on this important topic.

Iron deficiency anemia is a pervasive but rarely fatal health condition. Thirumurthy et al (2008) usefully consider a much more severe disease, HIV/AIDS, and estimate impacts of the introduction of anti-retroviral (ARV) treatment on individual labor productivity measures in a Kenyan site. This study is not an RCT, but given the paucity of evidence on the productivity impacts of contemporaneous health in developing countries, it is worth briefly describing. The introduction of ARV's at the individual level is determined by a "cut-off" value of the individual CD4 count (which captures how compromised the individual immune system is), and thus the study's design exploits experimental variation, although for ethical reasons treatment was not provided in a randomized fashion across individuals. Incorporation into ARV treatment during this period (when ARV treatment was still rare in Kenya) was typically life-saving and thus any labor earnings can plausibly be considered a treatment effect relative to the counterfactual. The authors more conservatively compare earnings after treatment to those immediately before treatment,

which is arguably a lower bound on true effects. Due to the relatively high cost of treatment and limited number of individuals incorporated into the sample over the study period, they compare 266 households with at least one HIV positive individual to 503 other households representative of the local population.

Thirumurthy et al (2008) estimate large and statistically significant gains of over 50% in individual labor supply up to six months after the start of ARV treatment, with large impacts on total earnings. There are also important within-household effects, as the labor supply of other individuals in the households, including children, fall after an adult begins treatment, suggesting that adult health status has important externalities for others in the household and may affect the human capital accumulation of the next generation. While perhaps not surprising given that most of the treated individuals would have passed away in the absence of ARV treatment, this study provides further evidence on the important economic consequences of large health shocks.¹

An important share of the morbidity burden among adults in poor countries, especially Sub-Saharan Africa, come from malaria. Yet there is little rigorous evidence to date on the productivity impact of malaria. To the best of our knowledge, two field experiments were set-up in the last decade to estimate those. In what follows we describe their findings but also the challenges they have met. Indeed, the dearth of evidence on this issue stems primarily from the inherent difficulty in measuring productivity precisely without very large samples, and from the difficulty in changing health but nothing else at the same time, i.e., of isolating the health channel.

One of the two field experiments we are aware of is Fink and Masiye (2015), who randomized access to bednets among cotton-growing households in a rural area of Zambia with highly endemic malaria. They compare a control arm to an arm where households could obtain additional bednets for free, and to an arm where they could get bednets at subsidized prices (through an agricultural loan program). On average, 2.4 nets were distributed in the free distribution group and 0.9 in the net loan group. As found previously in other contexts (Dupas 2009), 90% of farmers used their nets at follow-up in both intervention arms, and as a result the interventions led to a large drop in self-reported all-cause morbidity (by 40–42%) and the odds of self-reported confirmed malaria by 53–60% (Fink and Masiye 2012).

A central methodological issue is that measuring the productivity of farming households in low income settings is challenging. Fink and Masiye (2015) rely on self-reported yield and find an increase in total farm output of 14.7%, suggesting large positive impacts of reducing malaria cases.

¹ Fox et al (2004) provide related non-experimental evidence on the labor productivity effects of HIV/AIDS, in their case on a Kenyan tea plantation.

However, computing yields is non-trivial for farm households, since assessing the value of unpaid labor inputs across individuals within the household is difficult, and there may also be a concern that health gains might have influenced the accuracy of reporting, even beyond any actual gains in yields.

These estimated gains are obtained conditional on the baseline yield, which unfortunately appears quite imbalanced across arms, possibly owing to the relatively small sample size. The randomization was clustered at the farming inputs distributor level, so a cluster corresponded to 11 randomly selected farmers from among a given distributor's clients, with only 49 clusters in total, 15 assigned to control, 15 to the free net arm, and the remainder (19) to the loan program. At baseline, farms in the free net and net loan arms were, on average, larger and more productive than farms in the control group, despite relative balance with respect to household head characteristics. Given this, it is not clear that controlling for baseline values in yield – a de facto difference-in-differences design – is sufficient to recover the causal impact of the bednet on productivity, since it remains possible that these baseline differences also translate into differential trends over time, reflect other unobservables, or capture differences in how susceptible the households are to rainfall shocks. The existence of reported yield data over only a single season is also potentially problematic, if climatic conditions in that year might have disproportionately favored particular types of farmers, and thus particular treatment arms (given the baseline imbalance), making it difficult to even argue that effects are likely to be lower or upper bounds. Thus while the analysis suggests large positive impacts of treating malaria on productivity, there are reasons to be cautious in our interpretation of the findings.

The second field experiment attempting to estimate the productivity impacts of malaria was conducted by Dillon, Friedman and Serneels (2014) with sugar cane cutters in Nigeria. Here the difficulty in observing productivity is less of an issue, since sugar cane cutters are paid a piece rate for every measured “rod” of cane cut, where a “rod” (approximately two meters in length) is a physical standard that a worker's supervisor carries to the final dropping point once completed by the worker. At the end of each day, the worker's output for that day is entered on his personal ‘blue card’ and is signed off on by both the supervisor and worker. The plantation thus keeps records of the daily output (quantity cut), the days worked, and the total earnings for each worker, which means that getting access to this data provides the researchers with a high-quality and objective measure of productivity.

The key constraint facing the research team was instead in the type of health intervention that could feasibly be carried out. It seems that the ideal research design, randomizing access to malaria prevention over the entire period across workers, was not possible. Instead, in the experiment, 800 workers who had been hired for a 6-week harvest season by a large sugar cane plantation were

called in for a “medical visit” at some point over the 6-week period, with the exact date on which the visit happened randomized across workers. During the medical visit, workers were tested for malaria and, if positive, treated with highly effective antimalarial treatment. On average around 30% of workers were found positive and subsequently treated. Ideally, there would be a control group of individuals also positive but not treated, and comparing those to the treated would provide an estimate of the impact of malaria detection and treatment on labor productivity. Unfortunately, the experimental design did not generate that, as those not sampled for the medical visit were simply not tested at all, making it impossible to assess their malaria status.

The best the researchers can do here is to compare workers who were tested early on in the study period to those tested later on. Doing so, they find a large intention-to-treat effect: those sampled for an early test are 15% more productive on the plantation during the three weeks following the medical visit compared to those tested later on. This is a large effect, and remarkably similar to the nearly 15% increase in farm productivity after bednet distribution reported in the Zambia study discussed above (Fink and Masiye 2015). (There are some possible discrepancies, though, since the larger measured malaria reductions in the Zambia data might also imply larger productivity impacts, to the extent that the farm labor tasks in both cases are largely comparable.) Moreover, the intervention evaluated in the Nigerian case is somewhat peculiar: it provides each worker a one-time testing and treatment opportunity on a pre-assigned day regardless of how they feel on that day. This is quite different from most real-world malaria policies one might imagine. In particular, this approach will likely have a smaller productivity impact than making malaria testing and treatment free and easily available for workers on any day that they feel ill, suggesting that the 15% effect in Dillon, Friedman and Serneels (2014) is a lower bound of providing treatment on a permanent basis. That said, the intervention studied may have the advantage of catching and resolving some cases that affect productivity but are not sufficiently severe to lead workers to seek treatment on their own.

Other recent studies have exploited empirical settings outside of the family farm where individual worker productivity can be accurately measured. One of the most noteworthy recent contributions is Adhvaryu et al (2014a). This study uses data from garment-factory workers in India, whose productivity is accurately measured by managers in the natural course of running the assembly line. The health issue they focus on is exposure to fine particulate matter (PM2.5) in the factory, which is monitored at high frequency by multiple sensors on the factory floor. Worker rotation to different tasks within the factory leads to differential exposure, as does natural variation in pollution levels. Exposure to fine particulate matter leads to a range of respiratory and cardiovascular problems in the short-run, including acute constriction of blood vessels, and long-run exposure is linked with severe health and mortality risks.

The central finding is that higher levels of PM2.5 exposure significantly reduce factory worker productivity: an increase of one standard deviation increase in PM2.5 air pollution (roughly 45 $\mu\text{g}/\text{m}^3$) reduces productivity by 6%. This provides further evidence that higher morbidity – in this case, along a health dimension different from other existing studies – is associated with lower earnings. The authors also report that workers working under experienced factory managers also appear to suffer less from high pollution levels, although the precise mechanisms underlying this pattern are unclear. Air pollution is an unfortunate fact of life for hundreds of millions in the rapidly growing cities of Asia and Latin America, and increasingly in Africa, and one that appears to only become more severe over time, suggesting that these findings have broad scientific import and policy relevance.²

3.2. Impacts of Child Health and Nutrition on Education

A distinct sub-literature estimates effects of child health and nutritional investments on contemporaneous educational outcomes; Glewwe and Miguel (2008) provide a thorough review of both the experimental and non-experimental research in this area. Here we focus on a selection of the experimental studies in this area. This is actually a vast literature that crosses many disciplines, and it is beyond the scope of this survey to cover all relevant studies. We focus mainly recent studies within economics, but also discuss some related contributions from other field.

Many of the earliest randomized studies by nutritionists and other public health researchers focused on the impacts of specific nutrients that were lacking in children’s diets. Studies in India and Indonesia by Soemantri, Pollitt, and Kim (1989), Soewwondo, Husaini, and Pollitt (1989), and Seshadri and Gopaldas (1989) found large and statistically significant impacts on cognitive development and school performance of iron supplementation among anemic children, but a study by Pollitt and others (1989) found no such impact in Thailand. See Nokes et al (1998) for a more complete survey of the related iron supplementation literature.

Other early studies focused on parasitic infections, especially intestinal parasites. Kvalsig, Cooppan, and Connolly (1991) examined whipworms and other parasites in South Africa and found that drug treatments had some effect on cognitive and education outcomes, but some impacts were not statistically significant. Nokes and others (1992) evaluated treatment for whipworms in Jamaica and concluded that some cognitive functions improved from the drug

² In a companion paper, Adhvaryu et al (2014b) estimate large negative effects of higher temperatures on garment-factory worker productivity, a finding with potential importance for the possible economic productivity impacts of future climate change and global warming.

treatment, but others, particularly those related to academic performance in schools, appeared not to have changed substantially. Overall, the early experimental literature on the impact of treatment for intestinal parasites on child growth and cognition did not reach strong conclusions, as argued in the Dickson et al (2000) survey and in the more recent Cochrane review on the topic (Taylor-Robinson et al 2012). One possible reason why many of the early experimental deworming studies show limited impacts is that they carried out randomized treatment within school communities, creating the possibility that positive treatment externalities experienced by children in the control group lead to attenuated treatment effects, as discussed earlier and in Miguel and Kremer (2004). Many of these studies also have relatively small sample sizes, such as 210 children in the South African study and 103 in the Jamaican study. Other experimental studies (not reviewed here) include education interventions combined with health interventions, so the impact of the health intervention by itself cannot be credibly assessed.

Other studies have focused on general food supplementation to supply calories and protein. The most well-known of these is the INCAP study (Pollitt et al 1993, Martorell et al 1993) initiated in four Guatemalan villages in 1969, two of which were randomly selected to receive a porridge (atole) high in calories and protein while the other two villages received a drink (fresco) with less calories and no protein. Follow-up studies over the next three decades appear to show sizeable effects on later cognitive outcomes from providing the atole to mothers and young children, and we discuss these in greater detail below.

A number of recent randomized experiments have also been carried out by economists on the impact of health interventions on educational outcomes. These studies also typically evaluate actual interventions carried out by real-world non-governmental organizations (NGOs) or governments, and as such the findings of these studies may be of particular interest to policymakers in less developed countries. These are in contrast to several of the studies discussed above, which were often small-scale researcher implemented interventions. Many of these evaluate school-based health or nutrition interventions which some have argued may be among the most cost-effective approaches for delivering health and nutrition services to children in less developed countries (Bundy and Guyatt 1996).

Miguel and Kremer (2004), discussed at length in Section 2.1 evaluates the impact of school-based mass treatment for intestinal worms using inexpensive deworming drugs. The study is based on a sample of 75 primary schools with a total enrollment of nearly 30,000 children, a much larger sample size than most other studies in this literature. The sampled schools were drawn from areas where there is a high prevalence of intestinal parasites among children. Worm infections – including hookworm, roundworm, whipworm and schistosomiasis – are among the most widespread diseases in less developed countries: recent studies estimate that 1.3 billion people

worldwide are infected with roundworm, 1.3 billion with hookworm, 900 million with whipworm, and 200 million with schistosomiasis. Infection rates are particularly high in Sub-Saharan Africa (Bundy, et al. 1998; WHO 1993). Geohelminths – hookworm, roundworm, and whipworm – are transmitted through poor sanitation and hygiene, while schistosomiasis is acquired by bathing in infected freshwater. School-aged children typically exhibit the greatest prevalence of infection and the highest infection intensity, as well as the highest disease burden, since morbidity is related to infection intensity (Bundy 1988).

The educational impacts of deworming are considered a key issue in assessing whether the poorest countries should accord priority to deworming, but until recently research on these impacts has been inconclusive (see Dickson et al. 2000 for a survey). Indeed, earlier randomized evaluations on worms and education suffer from several important methodological shortcomings that may partially explain their weak results. Earlier studies randomized the provision of deworming treatment within schools to treatment and placebo groups, and then examine the impact of deworming on cognitive outcomes. However, the difference in educational outcomes between the treatment and placebo groups understates the actual impact of deworming if placebo group pupils also experience health gains due to local treatment externalities (due to breaking the disease transmission cycle). The earlier studies also failed to adequately address sample attrition, an important issue to the extent that deworming increases school enrollment.

The study by Miguel and Kremer finds that absenteeism in treatment schools was 25% (7 percentage points) lower than in comparison schools and that deworming increased schooling by 0.14 years per pupil treated (on average). This is a large effect given the low cost of deworming medicine; the study estimates an average cost of only US\$3.50 per additional year of school participation. The finding on absenteeism does not reflect increased school attendance on the part of children who attend school only to receive deworming drugs, since drugs were provided at only two pre-announced days per year, and attendance on those two days is not counted in the attendance analysis. There is no statistically significant difference in treatment effects between female and male students.

Somewhat surprisingly, despite the reduction in absence no significant impacts were found on student performance on academic tests. It is unclear what exactly is causing this discrepancy, although one possibility is that the program led to more crowded classrooms and that this may have partially offset positive effects of deworming on learning in the treatment schools.

The schooling data in Miguel and Kremer (2004) are noteworthy from a measurement perspective. School attendance was collected at sample schools by survey enumerators on unannounced days four to five times per year, rather than relying on school registers (which are often thought to be

unreliable) or on parent reports in household surveys, as done in most of the previous literature. Efforts were also made to follow children who transferred to other schools in the same Kenyan district. This yields a more detailed and reliable measure of school participation than the data available from most other studies. Bobonis et al (2006) and Vermeersch and Kremer (2004) use similar measures of school attendance.

The authors found that child health and school participation – i.e., attendance, where dropouts are considered to have an attendance rate of zero – improved not only for treated students but also for untreated students at treatment schools (roughly a quarter of pupils in treatment schools chose not to receive the deworming medicine) and for students at nearby primary schools located within 3 kilometers. The impacts on neighboring schools appear to be due to reduced disease transmission brought about by the intervention, an epidemiological externality. Econometric identification of the cross-school treatment spillovers on the worm infection rate relies on the randomized design of the project: conditional on the total local density of primary school pupils, there is random exogenous variation in the number of local pupils assigned to deworming treatment through the program. A key finding of the paper is that failure to take these externalities (or spillovers) into account would lead to substantial underestimation of the benefits of the intervention and the cost effectiveness of deworming treatment.

Miguel and Kremer (2014) document a coding error in the construction of the variables used to measure treatment externalities at distances between 3-6 km from each school; correcting this issue weakens the statistical significance of externality effects on worm infections at this distance but does not affect other results in the original paper.³

Bobonis, Miguel, and Sharma (2006) conducted a randomized evaluation in India of a health program that provided iron supplementation and deworming medicine to pre-school children age 2–6 years in 200 preschools in poor urban areas of Delhi. Even though only 30% of the sampled children were found to have worm infections, 69% of children had moderate to severe anemia according to international standards. After 5 months of treatment, the authors found large weight gains and a reduction of one-fifth in absenteeism, a treatment effect similar to the estimated school participation effect in the Miguel and Kremer (2004) study in Kenyan primary schools. The authors attempted to obtain estimates after 2 years, but high sample attrition and apparently non-random enrollment of new children into the preschools complicated attempts to obtain unbiased longer term impact estimates. It also does not present data on any type of child learning, and thus is limited to examining anthropometric outcomes and school enrollment and attendance.

³ For further discussion of the implications of the updated Miguel and Kremer (2004) findings, refer to Aiken et al (2015), Davey et al (2015), and Hicks et al (2015).

Finally, because all children received a combined treatment of iron supplements and deworming medicine, the India study cannot distinguish between the separate impacts of these two treatments, similar to the Thomas et al (2003, 2006) studies discussed above.

An arguably cleaner test is provided by Chong et al (2016), who randomized access to materials promoting iron supplementation among a small sample of 219 Peruvian secondary schoolchildren, in an area where local stores had also been stocked with iron pills. Baseline rates of anemia were fairly high in this population, at over 40%. The authors find that the informational encouragement design was effective, such that the treatment group consumed significantly more iron over the three month intervention period; this in itself is an interesting finding, given the many challenges in changing health behaviors that we detail in this chapter. More importantly, the schooling performance of the sample is then tracked using administrative records, and the data indicate that among anemic children in the iron supplementation treatment group, academic test scores rose by a sizeable 0.4 standard deviation units, and that grade progression and aspirations both also improved.

Other recent studies paint a more mixed picture of the impacts of child health and nutritional interventions on educational outcomes. Clarke et al (2008) examine the impact of intermittent preventive treatment (IPT) of malaria among schoolchildren in a region of western Kenya with high perennial transmission of the disease. The treatment occurred roughly every four months with a combination of sulfadoxine-pyrimethamine and amodiaquine, and outcomes were compared to a placebo group over the course of a year. This approach is reminiscent of the malaria screening and treatment approach in Dillon, Friedman and Serneels (2014), described above, in that it does not depend on individuals seeking out treatment during periods when they are already ill. Randomization occurred among roughly 5,000 children aged 5-18 within 30 sample primary schools.

The study produced evidence of large positive impacts of IPT on anemia, as well as on some cognitive outcomes after one year. In particular, there are significant gains in two class-based tests of sustained attention. There is no evidence of impacts on hyperactive-compulsive behavior or on educational achievement. There are two issues worth keeping in mind when interpreting the results. The first is the relatively high rate of attrition over the course of one year, at 27%, although rates appear balanced between treatment arms. A second is the fact that there was simultaneously mass treatment for intestinal helminth infections in the sample schools, which the studies mentioned above have shown lead to school participation gains in a nearby area of western Kenya. Thus the treatment effects in the Clarke et al (2008) study are conditional on this deworming regime, and it remains unclear if anemia and educational gains would have been larger or smaller in magnitude in the absence of deworming. The possibility that treatments for infectious

diseases might serve as complements (or substitutes) is important from a public policy standpoint, but little is known about these interactions in practice; this remains a promising area for future research.

The expansion of large-scale national social welfare programs, often linked to conditional cash transfers, in many Latin American and other developing countries has provided opportunities to explore alternative ways to improve early child development outcomes. Amarante et al (2016) exploit an eligibility discontinuity in the design of a cash transfer program in Uruguay to show that receipt of a large transfer leads to substantial reductions in the incidence of low birth weight among the children of beneficiaries. Beyond the cash itself, the contact with recipients induced by the program allows for opportunities to deliver additional interventions. Attanasio et al (2014) study one such effort in Colombia as part of its *Familias en Acción* program, which carried out large-scale micronutrient supplementation as well as psycho-social stimulation of children aged 12-24 months, and assessed nutritional, health, and cognitive impacts. The stimulation consisted of weekly home visits, which are relatively expensive, as well as biweekly supplementation with iron, zinc, vitamin A, folic acid, and vitamin C, all carried out over the course of the 18 month study. The research consisted of a 2x2 factorial design.

Attanasio et al (2014) find that the nutritional supplementation alone did not have any impacts on measured child height, weight, hemoglobin, cognitive scores or receptive language, nor did it have any additional benefit in combination with stimulation. Psycho-social stimulation, on the other hand, did lead to significant improvements in the cognitive and language outcome measures, echoing similar findings from earlier work in Jamaica (Gertler et al 2014). It remains possible that micronutrient supplementation would have larger impacts in poorer populations; recall that Colombia is a middle-income country and the study sample consisted of households receiving a generous cash transfer. Further research is needed to shed light on the generality of this finding.

3.3. Impacts of Child Health and Nutrition on Later Outcomes

A third group of studies estimates long-run impacts of child health interventions on life outcomes, where again we focus on experimental studies in development economics.

We first examine a growing number of studies estimating long-run impacts of deworming. New evidence is rapidly accumulating on the positive long-run educational and socio-economic impacts of child deworming. A key lesson of Miguel and Kremer (2004) is that traditional individual-level randomized designs will miss any spillover benefits of deworming treatment, and this could

contaminate estimated treatment effects. Thus cluster randomized designs provide better evidence. Three new working papers with such cluster randomized designs estimate long-run impacts of child deworming up to 10 years after treatment; these effects on long-run life outcomes are arguably of greatest interest to public policymakers.

A main puzzle with the Miguel and Kremer (2004) Kenya deworming study is that increased school participation (primarily attendance, but also reduced dropping out) is not reflected in students' academic test scores or cognitive test scores. The authors present some cost-benefit analyses at the end of the paper that suggest that the intervention is cost-effective, but it is unclear exactly how to interpret these if the intervention does not increase learning of basic skills.

This issue is addressed in the follow-up study, Baird et al (2015), which collects information on a wide range of adult life outcomes. Baird et al. (2015) followed up the Kenya deworming beneficiaries from the Miguel and Kremer (2004) study during 2007-2009 and find large improvements in their labor market outcomes. The paper employs a conceptual framework building on the seminal health human capital model of Grossman (1972), which interprets health care as an investment that increases future endowments of healthy time. Bleakley (2010b) further develops this theory, arguing that how the additional time is allocated will depend on how health improvements affect relative productivity in education and in labor. Pitt, Rosenzweig, and Hassan (2012) further note that time allocation will also depend on how the labor market values increased human capital and improved raw labor capacity, and that this in turn may vary with gender. They present evidence consistent with a model in which exogenous health gains in low-income economies tend to reinforce men's comparative advantage in occupations requiring raw labor, while leading women to obtain more education and move into more skill-intensive occupations.

Consistent with Pitt, Rosenzweig, and Hassan (2012), the Kenya deworming program increased education among women and labor supply among men, with accompanying shifts in labor market specialization. Ten years after deworming treatment, women who were eligible as girls are 25% more likely to have attended secondary school, halving the gender gap. They reallocate time from traditional agriculture into cash crops and entrepreneurship. Men who were eligible as boys stay enrolled for more years of primary school, work 17% more hours each week, spend more time in entrepreneurship, are more likely to hold manufacturing jobs, and miss one fewer meal per week. Since deworming treatment is cheap (at less than US\$1 per person per year), the authors estimate a large annualized financial internal rate of return of at least 32.2%. Many studies argue that early childhood health gains in utero or before age three have the largest impacts (for instance, Almond and Currie 2010) and some have argued that health interventions outside a narrow biological window of child development will not have major effects. This evidence suggests that health interventions among school-aged children, which are too late in life to affect cognition or

height, can have long-run impacts on labor market outcomes by affecting the amount of time people spend in school or work.

There are several noteworthy methodological features of the Baird et al (2015) article. First, it remains unusual for studies to combine experimental designs and long-run 10 year follow-up longitudinal data, and in this case most individuals in the sample were between 19 to 26 years old at the follow-up. Second, the rate of attrition was quite low in the follow-up Kenya Life Panel Survey (KLPS). KLPS tracked a representative sample of approximately 7,500 respondents who were enrolled in grades 2-7 in the Kenya deworming schools at baseline. Survey enumerators traveled throughout Kenya and Uganda to interview those who had moved out of local areas. The effective survey tracking rate in KLPS overall is 82.7%, and 84% among those still alive. These are high tracking rates for any age group over a decade, and especially for a mobile group of adolescents and young adults. Tracking rates are nearly identical and not significantly different in the treatment and control groups.

While the primary school children in the Miguel and Kremer (2004) sample were probably too old for deworming to have major impacts on brain development, and there was no evidence of such impacts, Ozier (2014) estimates cognitive gains 10 years later among children who were 0 to 2 years old when the deworming program was launched and who lived in the catchment area of a treatment school. These children were not directly treated themselves but could have benefited from the positive within-community externalities generated by mass school-based deworming. Ozier (2014) estimates average test score gains of 0.3 standard deviation units, which is equivalent to roughly half a year of schooling. This provides further strong evidence for the existence of large, positive, and statistically significant deworming externality benefits within the communities that received mass treatment.

Croke (2014) finds positive long-run educational effects of a program that dewormed a large sample of 1 to 7 year olds in Uganda, with statistically significant average test score gains of 0.2 to 0.4 standard deviation units on literacy and numeracy 7 to 8 years later. These are similar to the effect magnitudes estimated by Ozier (2014) in Kenya. The Ugandan program is one of the few studies to employ a cluster randomized design, and earlier evaluations of the program had found large short-run impacts on child weight (Alderman et al., 2006; Alderman, 2007).

The long-run impacts of the well-known INCAP experiment in Guatemala are described in Hodinott et al (2008), Maluccio et al (2009), and Behrman et al (2009). As mentioned above, INCAP provided substantial nutritional supplementation to two villages while two others served as a control, and the authors find evidence of very large and statistically significant gains in male wages of one third, improved cognitive skills among both men and women, and even positive

intergenerational effects on the nutrition of beneficiaries' children up to 35 years after the original project. This is a highly unusual and exceptional data collection effort, and it provides further evidence that childhood health and nutrition gains can have large returns in terms of adult labor productivity.

The pioneering INCAP study also has some limitations. In one sense, it has a sample size of only four villages since the intervention did not vary within villages, and it is unclear if all the existing studies fully account for the intracluster correlation of respondent outcomes in their statistical analyses, thus perhaps leading them to overstate the statistical significance of their findings. Second, strictly speaking, the control group also received an intervention, the fresco drink, albeit one with a relatively small benefit compared with what was received in the treatment group. Third, within each village receipt of the atole or fresco was voluntary, which implies that those who were treated were not a random sample of the population within each village. This means that the most convincing estimation strategy may be an intention to treat analysis, rather than direct estimation of the effect of child health on education. Finally, sample attrition is a major concern in both the 1988-89 follow-up and the most recent surveys, as more than one quarter of the original sample were apparently lost by 1988-89 and roughly 40% were lost by the time of the 35 year follow-up survey.

4. Environmental/Infrastructural Determinants of Health

The Adhvaryu et al (2014a) study on the impact of indoor air pollution on Indian factory worker productivity discussed above highlights the links between environmental conditions, health, and incomes, and the importance of environmental issues potentially extends far beyond their setting. In the United States and other countries, public health measures such as improved sanitation, provision of clean drinking water, and hookworm and malaria eradication campaigns have been shown to have played a key role in the massive improvements in child health and decreases in mortality observed during the 20th century (Cutler and Miller 2005; Currie 2000; Bleakley 2007, 2010; Galiani, Gertler, and Schargrodsky 2005; and see Deaton 2013 for a review). The implementation of many such “infrastructural” public health investments is not only costly (at times prohibitively so, especially in low population density areas such as much of rural Africa), it also likely requires a minimal level of state capacity that may still elude many low-income countries. In this section, we discuss the relatively limited number of field experiments in

developing countries that have recently studied the impact of environmental investments on health outcomes.

Kremer et al. (2011a) run an experiment to estimate the health consequences of improving water quality at the source in rural Kenya. Naturally occurring springs are an important source of drinking water in the area of their study (Western Kenya), yet few of them are “protected” – i.e., the water seeps from the ground, which implies that the water that pools is often contaminated with runoff from the surrounding area (often containing human and animal fecal matter and other sources of pollution). In contrast, a spring is considered protected if its source is sealed off and encased in concrete so that water flows out from a pipe, where it can be collected by consumers before the water touches the ground.

Kremer et al. (2011a) exploit the randomized protection of 100 springs out of 200 in order to estimate the impacts of improved water quality on child health outcomes. They find that spring protection dramatically improves water quality: through follow-up visits conducted among a representative sample of households that had been identified as regular spring users at baseline, the authors find that fecal contamination of home-stored water was one quarter lower for those in the catchment area of a protected spring compared to those in the catchment area of an unprotected spring. As a result child health improves, with a 25% decrease in child diarrhea attributable to the spring protection intervention. Note that the study was not statistically powered to detect mortality impacts. The authors also tested for, but did not find any evidence of, epidemiological spillovers (in terms of water contamination) within 3 km of the protected springs.

The study goes on to study the demand for cleaner water, using a travel cost methodology developed in environmental economics to value non-market amenities. Specifically, Kremer et al (2011a) estimate how much the usage of protected springs increases (relative to other water sources) after the intervention; large increases in usage, and especially a willingness to walk longer distances to reach such sources, would be consistent with increased demand for these sources. About three quarters of water trips were collected at the reference source (which was randomized into spring protection or control) at baseline, so while limited, there is still scope for some switching, and it is from this switching (or lack thereof) that the authors can back out a “revealed preference” estimate of willingness to pay for clean water. Using a discrete choice multinomial logit framework, the authors document that use of protected springs does increase significantly, but that magnitudes are not particularly large. Making standard assumptions on the value of water collector time, they estimate that the revealed preference willingness to pay for improved water quality is relatively low, at only US\$2.96 per year.

The authors then go on and translate this willingness to pay for water quality into a valuation for health. This exercise is challenging because the value placed on water quality may differ based on what the water is used for. While households may be willing to walk extra distance for clean drinking water, they may not do so for laundry water. Kremer et al. (2011a) do not have data on which water trip is for drinking water, thus they assume that the benefit from an additional trip to the clean water source is constant. In other words, they assume households trade off the cost of every water trip against child health benefits, potentially not only those made for drinking water. This is a relatively strong assumption since, as mentioned above, households in their sample conduct most of their water trips to the reference source to start with, and it is unclear how much of the water they fetch at other (unclean) sources is actually for drinking.⁴ Under this assumption, they estimate that willingness to pay is only US\$0.89 to avert one child diarrhea episode. Even stronger assumptions allow them to translate the diarrhea estimate into an implied mean value of averting one child statistical death of only US\$769, far lower than is commonly assumed by health policymakers, although this is plausibly an underestimate (given the discussion above). In Section 5, we discuss the broader literature on the demand for health in developing countries, in which we discuss alternative and more direct approaches to estimate the demand for health products.

Besides access to clean water, access to sanitation (in particular connection to sewerage) has been shown to be an important determinant of health (Cutler and Miller 2006). Yet the type of sanitation possible in low-density settings does typically not involve sewage, and its potential for health impacts is unclear. In rural Orissa, in India, Clasen et al. (2014) estimate the impact of latrine construction and find only modest impact on latrine usage and no impacts on health. Similar results were found in a comparable experiment in rural Madhya Pradesh, also in India (Patil et al. 2014).

Duflo et al. (2016) argue that these null results, rather than reflecting a low demand for health, may be due to the fact that sanitation by itself is not enough, and that water access may be a necessary complement. Indeed, maintaining a clean latrine without access to sufficient water quantity is difficult, likely reducing latrine usage. Moreover, given epidemiological externalities (such as those discussed earlier in the chapter), the private returns to sanitation may be limited if the majority of neighbors do not also have access to sanitation. This would mean that a sanitation program can only have a meaningful impact if combined with access to water *and* implemented at large scale (i.e., taken up nearly universally). Duflo et al. (2016) estimate the

⁴ The authors mention that their estimated valuation does not vary with the intensity of usage of the source at baseline (see footnote 12 in Kremer et al. 2011a), which could be indicative that the benefit is not decreasing with the quantity of clean water.

effect of such a scheme, which provided household-level water connections, latrines, and bathing facilities to all households at once in approximately 100 Indian villages. The program was not randomized, but a before-after comparison suggests large impacts on hygiene and reductions in diarrhea rates.

In a cluster-randomized trial in rural Bangladesh, Guiteras et al. (2015) find that providing subsidies to the majority of the landless poor households increases ownership and usage of latrines among both subsidized households and their neighbors, which is also consistent with the presence of important epidemiological externalities connecting individual investment and usage decisions with those of neighbors. They did not measure health outcomes, therefore it is not possible to know whether the resulting decrease in open defecation had impacts on health despite the absence of a complementary water provision intervention.

Further research could usefully document the nature of any interactions between water provision, sanitation access, and improved hygiene both for beneficiaries themselves and in terms of local spillovers. A large-scale intervention and study will likely be necessary to precisely estimate impacts on the full range of relevant health outcomes, including infant mortality.

Another feature of the environment that is becoming an increasingly important factor in life expectancy even in disease-ridden countries is road safety. Road deaths are the leading cause of death for people ages 15 to 29 worldwide. 90% of the world's fatalities on the roads occur in low- and middle-income countries, even though these countries have approximately half of the world's vehicles (World Health Organization 2015). Habyarimana and Jack (2011, 2015) estimate the impact of a pilot and subsequent large-scale interventions to encourage passengers of 14-seater minibuses (the ubiquitous form of public transport in Kenya) to “heckle and chide” the bus driver when his behavior compromises their safety, in particular, when overspeeding or overtaking without visibility. The encouragement came in the form of evocative messages on stickers placed inside the minibuses. They find a large impact on insurance claims for minibuses randomized to receive the stickers, of between one-quarter and one-third depending on the stickers used, as well as a decrease on the average maximum speeds and average moving speeds of the vehicles. In contrast, radio programs with the same encouragement seem to have no impact the weeks they are aired. Together, these results suggest that overcoming collective action problems is particularly important for reducing dangerous road behavior, but nudges may need to be very salient to be successful. They are an extremely cheap way to save lives however: Habyarimana and Jack (2015) estimate the cost-effectiveness of the most impactful stickers to be between \$10 and \$45 per disability adjusted life-year saved.

5. Demand for health products and healthcare

As discussed in section 3, health is an input: it matters for how productive one can be. It is also a direct component of well-being (a consumption good, in the terminology used by economists). Both of these are reasons for individuals to invest in their health and that of their children. Health investments include preventive behavior, from getting vaccinated to wearing a seatbelt to avoiding risky sexual contacts, as well as prompt treatment of illness episodes, and diagnosis and management of chronic conditions. In a standard model with no market failures, the demand for any such health input or behavior is a function of its benefits, its costs (both monetary and non-monetary), as well as the horizon over which both benefits and costs are accrued.

But many assumptions made in the standard model may not hold in low-income settings. Households in developing countries are often liquidity constrained, and they often lack information, or the education to process information, on the potential returns to various health investments. Even when information can be processed, judging whether it is correct may be difficult. As first noted by Arrow (1963), households' uncertainty about the true source of their health ailment makes learning about the quality of health care services, or of specific treatment or prevention tool, very challenging. This is particularly true for primary care where most illnesses are self-limiting, i.e., will get better over time (think of the common cold), and therefore signals are particularly noisy.

A second important issue for decision-making is widespread skepticism about outside advice, especially in societies where trust in the health care sector has been eroded for historical reasons, as in India where it is often argued that the forced sterilization effort carried out by Indira Gandhi's government during the "state of emergency" in the 1970's has created lingering distrust regarding government family health initiatives.

More generally, psychological factors are an important part of the health decision-making process for households in developing countries as well as in wealthy societies. Thinking about one's own mortality or the mortality of loved ones is unpleasant. People may thus tend to push such thoughts out of their mind unless there is no way to avoid them, such as when there is an acute illness episode or when an outside event makes health concerns highly salient. This has important implications for the design of health programs. A given intervention could have different impacts depending on whether it is implemented when health is at the top of people's mind or not.

There has been a large increase in the number of randomized experiments aimed at understanding the role of the more standard "economic" constraints and their implications for public health

policy. Less research has been carried out on the trust and psychological factors directly, although considering those factors is useful for interpreting the results of existing experiments that act on the economic environment, as we discuss below. For example, the timing of an intervention relieving household liquidity constraints (e.g., whether or not it takes place while the household is suffering from an acute illness) could matter greatly for the take-up and impact of such an intervention.

In this section we review field experiments on the demand for health, sorting them not by the outcome they are examining but by the behavioral factor that they focus on, e.g. price factors, non-monetary costs, cash on hand, etc. Because field experiments are often designed to study a number of such factors at once, often with a multiplicity of randomized treatment arms, our organizational scheme implies that we sometimes discuss the results of a given experimental study in multiple subsections. For other excellent recent reviews that are organized by health outcome (or sector) rather than behavioral factors, we refer the reader to Ahuja, Kremer and Zwane (2010) for a review of randomized evaluations on safe water access, and Jayachandran (2015) for a recent review on gender inequality.

5.1. Pricing Experiments

A number of field experiments have examined the role of prices in the adoption of health products and services. They typically do so by randomizing the price at which a household can access a product, and comparing take-up across price points, thereby tracing out the demand curve and estimating the price elasticity at different price points. The price elasticity is an important parameter because private health investments often have social externalities. Identifying that private demand is low may therefore justify government subsidies. For subsidies to not be wasteful, however, they have to strike a delicate balance: they have to maximize the likelihood that a needy person can access the health products or services that could benefit her, while also minimizing the likelihood that the subsidy accrues to those for whom the returns to the subsidy are infra-marginal, either because they would have invested in the product privately anyway, or because they are unlikely to make effective use of products they receive at a highly subsidized price. This is a serious concern theoretically since households that are unwilling to pay a high monetary price for a product may also be unwilling to pay the non-monetary costs associated with daily use of the product, or may not actually need the product at all. Indiscriminate subsidies would then undermine the screening or allocative benefits of prices. What's more, subsidies could also reduce the potential for psychological effects associated with paying for a product, such as a "sunk cost" effect in which people, having paid for a product, feel compelled to use it (an issue we return to below).

Below we review about half a dozen field experiments conducted over the last 15 years that have shed light on these issues. Before we go into their details, we discuss the relative merits of the various methods used to estimate willingness to pay at different price points.

Methods to estimate the demand curve

The most commonly used willingness to pay (WTP) elicitation method outside of field experiments is stated WTP: people are simply asked how much they would be willing to pay for the product. The main problem with this measure is that it is not incentivized, therefore respondents may not think hard enough before providing their answer. Different individuals may also interpret the question differently if not asked precisely enough: some may report what they would pay if they had access to credit, some may report a low WTP if they think their answer may affect future subsidy policies, and others may exaggerate their willingness to pay to please the survey enumerator, etc. For this reason, researchers have moved towards field experiments in order to observe the “true” demand at each price point. For this, two main methods have been used, TIOLI and BDM, which we now discuss in turn and compare to each other.

Take-it-or-leave-it, or TIOLI, experiments consist in randomizing the price that an individual faces, and observing whether that individual actually purchases the product at that price or not. This is a straightforward revealed preference mechanism.

The BDM mechanism, named after theorists Becker, DeGroot, and Marschak (1964), is an incentive-compatible elicitation mechanism with real stakes that can be used to elicit individual willingness to pay as follows. People are asked to state the maximum they would be willing to pay for a product, i.e., make a bid, and to put forward their bid amount. Then a price is randomly drawn from a known distribution, and those who had bid at or above the randomly drawn price have to use the money they had put forward to purchase the product at that price (they keep the balance if they were willing to pay more than the price); while those who bid below the price cannot purchase the product. This mechanism is incentive-compatible, that is, it is a dominant strategy for expected utility maximizers, since those who bid less than their true value risk failing to buy the product when the price drawn is low enough that they would in fact prefer to do so. Conversely, bidding above one's true value entails the risk of buying when the price is higher than one would actually be willing to pay.

Berry et al. (2012) discuss the merits of each method in detail and compares them in field trials in Ghana. TIOLI is straightforward to implement through door-to-door experiments, voucher distribution or retail-level subsidies. Importantly, TIOLI can be done in a way that allows people time to think through and save for the product, for example by having a fixed TIOLI price in place for a certain time period, or distributing vouchers redeemable for a given number of months,

as in Dupas (2009). In contrast, BDM can only elicit immediate willingness to pay, unless it is done over credit contracts. But BDM has the advantage of telling us, for each individual in the sample, what their exact willingness to pay is, whereas TIOLI only informs us of the share of the sample willing to pay at least a certain price. Thus TIOLI studies generally require much larger sample sizes. They also cannot easily be used to test for heterogeneity in outcomes based on individual willingness-to-pay without additional experimental features, such as a second, surprise randomization, as in Karlan and Zinman (2009), a method subsequently applied in the health sector by Ashraf et al. (2010) and Cohen and Dupas (2010), two studies we discuss further below. In contrast, BDM generates randomized variation in access to the product *within each observed willingness to pay* group: conditional on being willing to pay a given price, whether the study participant gets to acquire the product is random, and this allows researchers to estimate whether the returns to owning the product are heterogeneous by underlying willingness to pay (Berry et al. 2012, Chassang et al. 2012).

While BDM has the potential to generate richer data, the quality of this data is unclear, especially in resource-constrained settings where the population whose willingness to pay is elicited has low numeracy skills. Berry et al. (2012) assess the validity of the BDM mechanism in Ghana, comparing the demand curve for water filters obtained through BDM with that observed through TIOLI (disabling the time feature of TIOLI, i.e., forcing people to decide on their TIOLI offer immediately). They find that even after shutting off the time dimension, BDM systematically under-predicts willingness to pay relative to TIOLI. The magnitude of this under-prediction is not negligible, and appears to increase with price. Namely, the demand under BDM is 20 percent lower than under TIOLI at the lowest price considered (USD 1.40, a tenth of the retail price), 34% lower when the price is USD 2.80, and 45% lower when the price is USD 4.20. Berry et al. (2012) remain agnostic as to the reason why BDM appears somewhat inaccurate (if we take TIOLI as reflective of the “true” demand), though through additional experimental treatments, they can rule out that the difference between the two mechanisms is driven by either strategic bidding under BDM (i.e., people stating a low willingness to pay in the hope of influencing future prices, in particular the possibility of a greater NGO subsidy) or anchoring under TIOLI (if the TIOLI price influenced people’s willingness to pay). More work is needed to understand when and how the BDM approach can provide more accurate estimates of willingness to pay. In the meantime, most pricing experiments have been conducted using TIOLI.

As more pricing experiments take place, it should also become possible to better understand the circumstances under which stated willingness to pay, which is obviously far cheaper to elicit than either TIOLI or BDM experiments, gives a “sufficiently good” approximation of the demand curve. In Figure 1 we use the dataset underlying the experiment in Dupas (2009), as well as the dataset

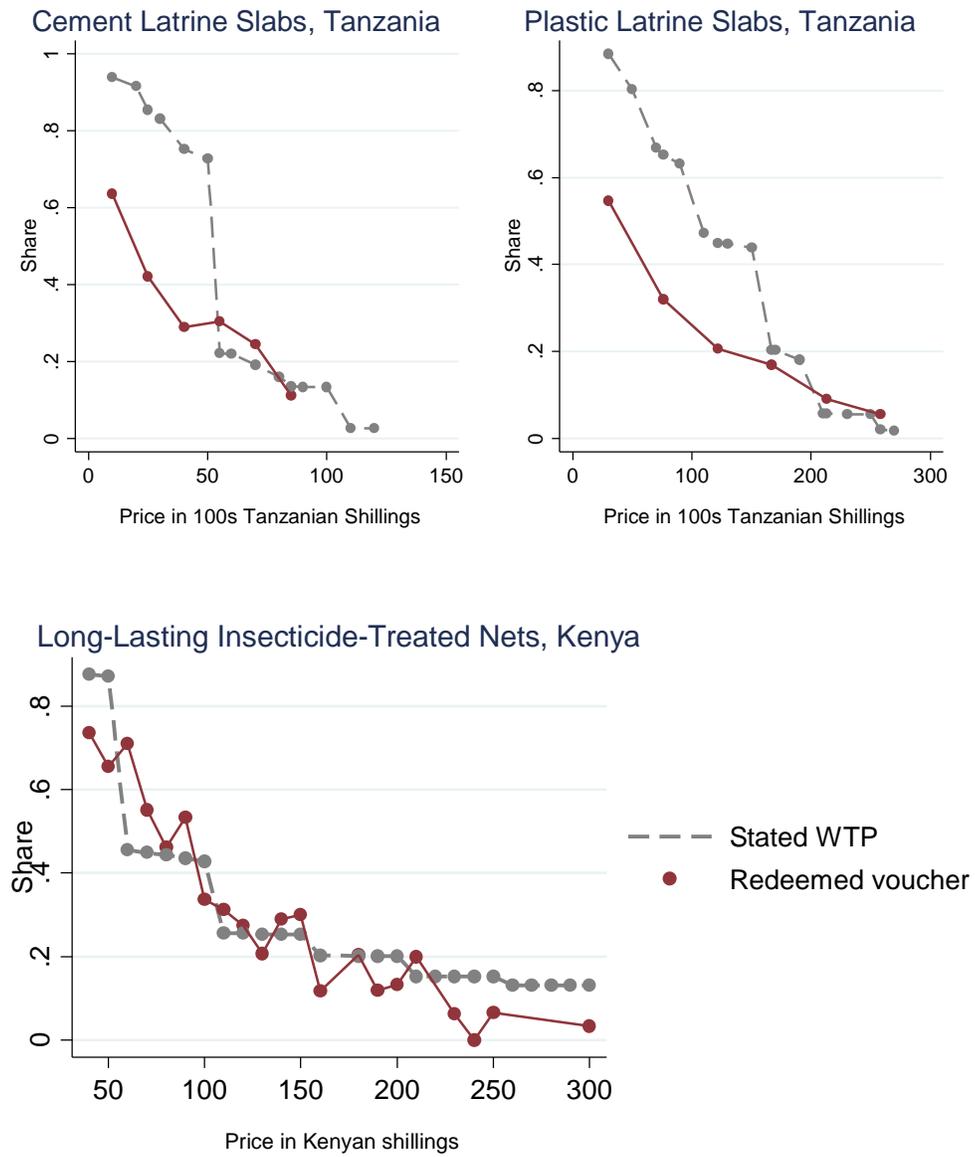
in Peletz et al. (2016), to compare stated willingness to pay to actual take-up in the TIOLI experiments. Dupas (2009) looks at the demand for bed nets, which are fairly well known products at the time of the study. Peletz et al. (2016) estimate demand for less common latrine slabs – concrete or plastic slabs with footholds and a hole for urine and feces, which can be used to cover latrine pits and are considered as health-enhancing as they reduce the risk of contamination of the home compound with fecal matter.

In both studies households were first asked at baseline how much they would be willing to pay for the product. In Figure 1, for each of the three products studied, the share reporting a WTP above each price point is plotted in grey dashed lines. In the subsequent TIOLI experiments, households in both studies had three months to redeem a voucher for the product at a specific price that was randomly varied across households. Only six price points were offered in the latrine slabs studies, while a wider range of prices was covered in the bednet study. The observed redemption rates at each price point are shown in solid maroon lines.

There are three interesting findings to note in Figure 1. First, there is a lot of “rounding” in stated willingness to pay: people tend to report amounts in multiples of 50 in Kenya and 5,000 in Tanzania, which creates the illusion of very large drops in demand at certain price thresholds, when in reality demand is likely to be much smoother. Second, the accuracy of elicited WTP can be high, as in the bednets case in Kenya, but in some other contexts it seems quite inaccurate: in Tanzania where familiarity with latrine slabs was low, elicited willingness to pay vastly overstates demand. Third, there is little evidence of strategic under-reporting of willingness to pay. This is particularly striking in the case of the slabs study, where the survey eliciting willingness to pay was carried out by NGO workers introducing themselves as interested in understanding how to improve health in the community, and yet stated willingness to pay was higher than observed take-up, likely due to wishful thinking regarding own take-up.

Kremer et al (2011a) also estimate large discrepancies between stated preference valuations of access to a protected spring (elicited two different ways) and the revealed preference travel cost approach described in section 4, on the order of two- to five-fold differences. As discussed above, however, the travel cost approach relies on strong assumptions and therefore it is not clear what share of the discrepancy comes from stated preferences overestimating willingness to pay and what share of the discrepancy comes from the travel cost approach underestimating willingness to pay.

Figure 1. Observed Take-Up in TIOLI experiments vs. Stated WTP at Baseline



Data sources: Peletz et al. (2016) for slabs and Dupas (2009) for nets

Results of Pricing Experiments

We now turn to reviewing over half a dozen randomized pricing experiments conducted to date. The great majority of those concern the pricing of *preventive* health products.

One of the earliest randomized TIOLI experiment for a health product in a poor country was however for a treatment product – a deworming drug -- and took place in 2001 in Western Kenya. Among 50 primary schools enrolled in a free deworming program in 2000, Kremer and Miguel (2007) randomly selected 25 that moved to a *cost-sharing* program: parents now had to contribute a fee in order for their children to receive the deworming pill(s) on deworming day. Parents had to pay the fee at the school in advance of the deworming day, and were informed of this fee one or two months prior to treatment day. The researchers found that the share of children receiving deworming medication on the day the NGO visited the school for mass deworming was only 18 percent in the cost-sharing schools, compared to 75% in the schools who kept the free program, despite the fact that deworming remained heavily subsidized, i.e., the fee charged per child was just 20% of the actual program cost on average. Interestingly, parents of sicker pupils were no more likely to pay for deworming drugs, suggestive no screening effect of the cost-sharing program. While these results suggest that demand is highly sensitive to price, understanding why it is the case in this specific context is somewhat difficult. It could be that parents had gotten “used to” the free program and resented the introduction of the cost-sharing fee, and therefore their demand was lower than what it would have been had a free program never been implemented in the first place. It could also be, as the authors hypothesize is the case, that the perceived private value of deworming is lower than the fee charged, perhaps due to the treatment externalities that they document. Subsequent pricing experiments have adopted more nuanced designs in order to disentangle these mechanisms from each other.

Cohen and Dupas (2010) use a two-level randomized TIOLI design to estimate (1) the demand curve for a new health product in rural Kenya: long-lasting antimalarial bed nets (LLINs); and (2) the distinct roles of the screening and psychological sunk cost effects that price may have on their usage. LLINs cost \$7, and they prevent bites from malaria-carrying mosquitos while sleeping. The experiment, conducted in 2007, randomized the price at which prenatal clinics offered nets to pregnant women. Clinics charged either 0 (free distribution), 15, 30 or 60 US cents (note that the highest price point considered, 60 US cents, still represented a 92% subsidy).

This first level of randomization, at the clinic level, involved only 20 observations (20 clinics), something which has implications for inference, as discussed in Cohen and Dupas (2010). The second level of randomization was at the individual level. Namely, a random subset of women who had agreed to purchase the net for 30 or 60 cents were subsequently given a surprise rebate right after they had given their payment to the clinic’s cashier. Cohen and Dupas (2010) find that

demand is very sensitive to price: the likelihood that pregnant women acquired a net fell from 99% to 39% when price increased from 0 to 60 US cents (with the demand at the intermediate price points of 15 and 30 US cents at 92% and 72%, respectively). This suggests that while there is no discontinuity at zero (it was not the shift from free provision to any positive price that makes demand drop, but rather larger price increases), demand is on the whole quite price sensitive, with very low demand rates at prices that are still heavily subsidized.

They do, however, find that the rate at which pregnant women used the net (measured through home observation visits two months after distribution) was relatively high (60%); and it was completely independent of the price they had paid for the net, whether initially or after the surprise rebate. This suggests that there is neither a screening nor a sunk cost effect of prices in their context. Thus coverage (the share of pregnant women sleeping under a bed net), and hence its potential for public health outcomes, increases very rapidly as the price goes down.

In another TIOLI experiment conducted with a sample of households with school-aged children, also in Kenya, Dupas (2014a) found that demand for LLINs becomes slightly less price sensitive if subsidies are provided in the form of vouchers that households have three months to redeem at local retail shops: the demand at \$0.60 becomes 73%. But overall price remains the primary driver of the demand, with the purchase rate dropping to just around 33% when the price reaches \$1.50 (still an 80% subsidy) and to 6% when the price reaches \$3.50 (corresponding to a 50% subsidy). Various marketing strategies (e.g., making the morbidity burden or treatment costs salient, targeting mothers, eliciting verbal commitments to invest in the product) failed to change the slope of the demand curve (Dupas, 2009). But here again, the price paid did not matter for subsequent usage. In fact, home observation visits showed that usage of bed nets acquired through a subsidized voucher was extremely high, rising from 60% at a three-month follow-up to over 90% after one year, and that was the case across all price groups, including recipients of fully subsidized nets. A similar level of bed net usage (90%) irrespective of initial price paid was observed in rural Zambia (Fink and Masiye, 2015), suggesting that this result is at least somewhat general.

The finding that demand is price sensitive has been established by TIOLI experiments for products other than deworming drugs and bed nets. In 2010, Meredith et al. (2013) randomized the subsidy level households faced for rubber shoes to prevent worm infections in Kenya, and in 2008 they randomized the price of soap and vitamins in Uganda, Guatemala, and India. In all contexts, they found that demand was very sensitive to price. In a TIOLI experiment in urban Zambia, Ashraf, Berry and Shapiro (2010), also find that demand for a bottle of water purifying solution (diluted chlorine) is sensitive to price, dropping from around 80 percent at the price point of 9 US cents (a 62.5% subsidy) to only about 50 percent at the full market price of 25 US cents. That experiment also used a two-stage randomization design but in this case both randomizations took

place at the household level. Specifically, not just the surprise rebate but also the initial offer price was randomized across households. Using this design, they test for both the screening and sunk cost fallacy effects of prices. As in Cohen and Dupas (2010), they found no evidence of a use-inducing sunk-cost effect, but found some evidence of a screening effect of prices. Specifically, those households who had selected themselves into paying a higher price were more likely to have used the purification solution within 6 weeks of acquiring it, while those who had received a higher subsidy were more likely to still have it on their shelf, possibly because they were keeping it for later or, as per the authors' interpretation, because they were less likely to ever plan to use it for a health purpose.

The studies discussed above suggest that price is often not a good mechanism to target subsidies for health prevention tools to those who most need them. If anything, higher prices seem to create too many errors of exclusion, and to prevent the positive spillovers on disease transmission that may justify subsidies in the first place.

The evidence regarding health *treatment* products is somewhat different, however. Cohen, Dupas and Schaner (2015), in a TIOLI experiment conducted in 2009 in the same area of Kenya as the bed net studies mentioned above, find that price can be (to some extent) used as a targeting mechanism to allocate malaria treatment. Targeting of malaria treatment is very important because of the negative spillovers that the overuse of antimalarials can generate: it can delay or preclude proper treatment for the true cause of illness, waste scarce resources for malaria control, and may contribute to drug resistance among malaria parasites, making treatment of malaria harder in the long-run. The reason why, within essentially the same population, price can be effective at targeting treatment when it is not effective at targeting prevention products (like bed nets) is that demand for treatment appears much less price-sensitive (especially among the poor) than demand for prevention. What's more, conditional on experiencing malaria-type symptoms, adults are much less likely to be malaria positive than children, but as with most treatments, the price per antimalarial dose for adults (who need to take more pills) is higher than the price for children. Consequently, at a given price per pill, children (the key target for the subsidy in this case) are on a flatter portion of the demand curve.

In addition to furthering our understanding of how price can be used to target health products in the developing world, that study makes two other contributions: (1) it highlights the trade-off inherent to subsidies for medications in environments with weak health system governance (which prevents conditioning the subsidy on a formal diagnostic test result), and (2) it points out that bundling subsidies for medications with subsidies for diagnostic tests has the potential to improve welfare impacts, a point we come back to in section 5.3 below when we discuss information experiments.

While the studies above have mostly focused on the effect of price on contemporaneous demand, some field experiments have been specifically designed to look at the dynamic effects of prices. The questions here are the following: Can a one-off subsidy be enough to trigger learning and generate sustained adoption? Or is there a risk that people are unwilling to pay for a product they once received for free, as a cursory look at the Kremer and Miguel (2007) deworming cost-sharing results could suggest? This could happen if people, when they see a product being introduced for free, start to feel “entitled” to receive the product for free (i.e., they would “anchor” around the subsidized price).

To gauge the relative importance of these effects, Dupas (2014a) examines the long-run effects of the one-time bed net subsidy vouchers households received in the study mentioned above. Specifically, the research team came back a year after the first pricing experiment had been done, and implemented an additional stage in the study in which all households received a second subsidy voucher, but this time they all faced the same price of \$2.30 (a 70% subsidy). By observing how the take-up rate of the second, uniformly-priced bed net varies as a function of the price a household faced in the first year, Dupas (2014a) can test whether being exposed to a large or full subsidy in the first year (which, as discussed above, considerably increased adoption at that time) reduces or enhances willingness to pay for the bed net a year later.

Dupas finds that a larger initial subsidy *enhances* willingness to pay for a bed net a year later, suggesting the presence of a positive learning effect which dominates any potential anchoring effect. Interestingly, the learning effect trickles down to others in the community: households facing a positive price in the first year are more likely to purchase a bed net when the density of households around them who received a free or highly subsidized bed net is greater. Though once bed net ownership is widespread, the transmission risk starts to decrease and the returns to private investments decrease, and accordingly those with more subsidized neighbors in year 1 were less likely to invest in year 2. Dupas (2014a) also tests for cross-product effects of price subsidies, namely, whether getting a subsidy for the bed net led households to expect subsidies for a different health product, namely, a water purification product. She finds no such effect: willingness to pay for the water product a few months after being exposed to a subsidy for the bed net was not lower among households who had received a full or very high subsidy.

Karlan et al. (2014) adopted a similar research design to study the relative importance of learning vs. anchoring effects in free distribution programs for a wider set of products, as well as the importance of who implements the free distribution. They conducted their experiment in northern Uganda. In a first round of door-to-door visits, they offer households one of three products either for free or for sale at the prevailing market price. The three products were chosen to differ in their

scope for learning, and included a pain reliever widely known at baseline (thus with little to no scope for learning); a deworming drug that was moderately well-known and which has some side effects, so that the learning effect, as in Kremer and Miguel (2007), might be expected to be negative; and a new, largely unknown treatment for childhood diarrhea for which the authors expected positive learning. The second door-to-door visits took place two to three months later and were conducted by a different set of sales agents. Households were offered the same product in round 2 as they had been offered in round 1, except for some randomly selected households who were offered a fourth product (a water purification solution) to test for the presence of any cross-product effects.

As in Dupas (2014a), Karlan et al (2014) find no evidence of cross-product effects, but they argue that the patterns of demand they observe in round 2 are consistent with anchoring playing an important role: when there is no scope for learning (i.e., the pain killer case), they find that demand is lower after free distribution than after a sale. This is the case irrespective of whether the free distribution was done by an NGO or by a for-profit firm advertising the free distribution using a standard marketing tool (a “free trial” to help people learn a product is worth their money). The results are somewhat weak, however, with no significant differences in the treatment effects across products. One potential concern with the design of this study was that the authors did not collect information on the demand for the three products outside their experimental door-to-door visit, even though those products were available for a similar, if not lower, price at local shops.

This highlights a challenge facing many recent health demand studies: measuring impacts often requires measuring demand both within and outside the experiment. A good example of that is Cohen et al. (2015) discussed above, which measured access to Artemisinin Combination Therapy (ACT, the latest and most effective antimalarial) through not only the drug shops involved in the study but also local health facilities, something critical in their context given the scope for crowding out. This can be difficult if recall bias is a concern, especially when the time lag between baseline and endline is long. An alternative is the design used in Dupas (2014a), who offered a product (a long-lasting insecticide treated bed net) unavailable on the market at the time of the experiment. Having perfect control over the supply means that observing the demand in the experiment provides Dupas (2014a) with a complete picture of the demand in both years. The drawback of that design is that when the product is not available on the market, the option value of experimenting with the product in round 1 may be lower, as households were not aware they would have a second chance to obtain the product from the experimenters, thus the take-up in round 1 may have been an underestimate of the take-up that would have prevailed in a real-world market environment.

Figure 2 graphically presents the TIOLI-estimated willingness to pay at various price points for 15 country-product combinations. For all products shown, the price points shown are at or below the market price. We indicate the country and year during which the TIOLI experiment was done for each product in the legend, as well as the market price. There are 6 broad types of preventative health products: soap, vitamins, bed nets, water filters, latrine slabs, and chlorine-based products to purify water.

Putting all of these results together in one figure yields a number of interesting patterns. First, and most importantly given that this is often misunderstood (see, for example the inaccurate treatment of this literature in the “Health Chapter” of the 2015 World Bank Development Report, World Bank 2015), demand at small but non-zero prices is often substantial: at the USD 0.50 price point, demand is over 65% for 9 out of 13 country-products for which this price point was included. At the USD 1.5 price point, it is over 50% for 6 out of 13 country-products (the price of chlorine is below USD 1.50 so we exclude the two chlorine studies from the denominator). This is important to note because many earlier review papers have been interpreting the evidence from pricing experiments as suggesting a sharp drop as soon as the price is not zero, but this is definitely not the case. In fact, there are only two products for which there is a sharp drop in demand as soon as the price is not zero: this was documented for deworming in Kenya in 2001 (Kremer and Miguel 2007) and for chlorine in Kenya in 2004 (Kremer et al. 2011). In the deworming case, the authors show that the return to private investment in deworming is possibly negative given the large externalities, and possibly in part due to the side effects associated with treatment. More speculatively, parents needed to make the payment for their children to be dewormed at the school in advance of the treatment day, so one potential contributor to the low take-up could be a lack of trust that the payment would indeed be followed by treatment.

In the chlorine case, the striking fact is that demand is only 60% even at zero price. This suggests that chlorine is a product for which a substantial portion of the population has little to no demand, possibly owing to the residual taste it sometimes leaves in water. The fact that small prices lead to a large drop in demand could come from the fact that many people who do not value the product because of the taste still take it if it is free in order to use it for other purposes (e.g., cleaning), hence the drop from zero to a positive price may not directly reflect a drop in health usage. Or it could be that people take the first free sample and once they have learned they do not value the product, they stop taking it even if it is free (indeed in Kremer et al. 2011, demand for chlorine at non-zero prices was observed two months after everyone had received a free sample bottle for free).

While there may not be anything “special” about the price point of zero in many cases, it is evident from Figure 2 that investment in preventive health is highly sensitive to price even when the price is below the prevailing market level. Moreover, the evidence from these studies concerning the characteristics of those who select into paying higher prices suggests that prices are usually not a very effective allocation mechanism in the sense that they fail to target those who appear to need the products the most (see Dupas, 2014, for a review).

Kremer and Glennerster (2011) present a framework highlighting that the price sensitivity documented above may be due to liquidity constraints, lack of information, non-monetary costs, or behavioral biases such as present bias and limited attention. A number of experiments have generated randomly varying access to liquidity, convenience or information, sometimes interacted with random variation in prices, to estimate the relative role of these potential factors. The evidence to date suggest that information about a product is necessary but not sufficient, and in particular information does not appear to substitute for higher subsidies, while reducing non-monetary costs and increasing liquidity often matter a lot for take-up. We review this evidence below, before discussing their implications for the scope of behavioral factors at play in section 5.8.

5.2. Liquidity experiments: credit and cash transfer experiments

In the pricing experiments discussed above, households commonly report lacking access to credit as a reason for not being able to take up the subsidized products offered. To test whether such reported liquidity constraints are indeed a serious barrier, one would need to allow a random subset of households to purchase health products on credit. While researchers who have exploited the random introduction of microcredit have typically not found impacts on health expenditures (see Banerjee, Karlan and Zinman 2015 and references therein), this may due to the coarseness of their data on health investments, and/or to the fact that most microfinance institutions focus on business loans rather than consumption loans, and that “flypaper effects” (the fact that money sticks where it first “hits”, like a fly on a flypaper) are common (Fafchamps et al. 2014).

The first studies (to our knowledge) to directly study demand for health products at full price when credit constraints are relaxed are Devoto et al. (2012) and Tarozzi et al. (2014). Devoto et al. (2012) identified low-income households not connected to the water grid in the city of Tangiers in northern Morocco, and randomized which households were told about a credit program to purchase a water connection. They see an impressive take-up rate of 69 percent, despite the fact that the cost of the connection (which varies with distance to the water mains) averages around \$1,000, an amount that they would have to repay over five years. Tarozzi et al. (2014) randomized access to ITNs on credit across villages in the state of Orissa, India. They find that 52% of

households offered full-price ITNs on credit purchased at least one ITN (and all of them fully repaid the loan). In contrast, in a follow-up cash sales study, they find that only around 11% of households purchase at least one ITN absent any credit. Fink and Masiye (2012, 2015) also examine the demand for full-price bed nets when people are offered a zero-interest loan in the context of rural Zambia. They find that households offered bed nets on credit acquire 0.8 nets on average, a demand comparable in magnitude to that observed in Tarozzi et al. (2014).

Another set of studies looks at the impacts of cash transfers on health choices. In their pricing experiment described in section 5.1, Meredith et al. (2013) gave households a randomly determined amount of cash (cash drops in the form of payouts for incentivized risk preferences elicitation) at the same time they distributed discount vouchers for rubber-sole shoes aimed at protecting children from worm infection. The market price for the shoes was about 85 Ksh (\$1.13) at the time, and discount vouchers varied from a low (20 Ksh) to a high (80 Ksh) discount. The cash drop varied from 0 to 200 Ksh, with a mean of 35. The researchers use the variation in cash drop amount to estimate the effect of liquidity on purchase.

They find a large and significant effect of the cash drop on demand: on average, every additional 100 Ksh in randomized cash payout increases the probability of voucher redemption by 22 percentage points. Since the cash drop was very small relative to lifetime income, at about 4% of weekly income on average, its effect on demand reflects a cash-on-hand effect rather than an income effect. Importantly, since households had to travel to a local store to redeem a voucher in order to obtain the product, the cash drop effect is arguably unlikely to be driven by an experimenter demand effect (whereby study households would feel compelled to use the money obtained from the experimenter to purchase the product sold by the experimenter). That said, some form of experimenter demand effect cannot be ruled out, as it may have been in the implicit priming embedded in the subsidy. The subsidy may have made health particularly salient in participants' mind, and that is why they chose to spend the extra cash on a health product, rather than something else.

Indeed, in contexts where individuals are not primed to think about health, the impact of cash drops on health investments has been mixed. Banerjee et al. (2015) find that a "graduation program" that combines the transfer of a productive asset with consumption support, training, savings encouragement and basic health services to very poor households can lead to sustained income improvements for these households, but they do not find particularly meaningful impact on physical health despite the income change. It is important to note that their data comes from 6 countries with quite different underlying health burdens, however, and the researchers do not

present data on health investments tailored to each context (e.g., bed nets for Ghana or water filters for India).

Haushofer and Shapiro (2013) find that provision of fairly large unconditional cash transfers to households in Kenya does not affect the health of under-five children as measured by their height, weight, and upper-arm circumference, but here again they do not report any information on investments in preventative health behavior such as bed nets for adults (since at the time of the experiment under-five children were supposedly covered by a government distribution scheme), latrine upgrades, etc. In contrast, in an experiment that primed people to save for health, Dupas and Robinson (2013) find that investment in preventive health increased significantly for those who gained access to a simple saving technology.

One study focuses more specifically on the reproductive health impacts of providing cash transfers to adolescent girls in Malawi. Baird et al. (2012) randomly assigned 175 enumeration areas (EAs) to three groups: girls in 46 EAs received conditional cash transfers (CCTs) if they achieved 80 percent school attendance; those in 27 EAs received unconditional cash transfers (UCTs); 88 EAs served as a comparison group and did not receive transfers. The cash transfers significantly lowered the prevalence rates of HIV and the herpes simplex 2 virus (HSV2). For example, 1.2 percent of girls enrolled in school at baseline who received transfers (CCT or UCT) tested positive for HIV at 18 months relative to 3.0 percent of girls in the comparison group. Self-reported sexual behavior was also lower among girls who received transfers; 3 percent of girls who received transfers reported having sex at least once per week, compared to 7 percent in the comparison group. These results suggest that financially empowering school-aged girls can have substantial effects on their sexual and reproductive health choices. Interestingly, the authors find that the amount of the money transferred did not itself matter, nor the share of the transfer directly transferred to girls vs. their parents.

5.3. Information Experiments

Even when liquidity constraints are alleviated, adoption of high-return health products or behaviors is often not 100%. A potential explanation for this could be that individuals lack of information on the health costs or benefits of different products or behaviors. In this section, we review information experiments showing that (exhaustive) information is necessary but often not sufficient to generate take-up.

Impact of information on Willingness to Pay

In their anti-worm rubber-sole shoes pricing experiment in Kenya, Meredith et al. (2013) find that health workshops did not affect total demand nor the price gradient in demand. In their complementary evidence from India, Guatemala and Uganda, they find an effect of the health information in only one of 6 country-product combinations they experimented with – namely, a health script delivered at the time households could obtain subsidized soap in India flattened the effect of price on demand. The authors discuss that this one result may be driven by an experimenter demand effect, however, since in this specific case the purchase decision was contemporaneous with the cash drop.

In China, Rozelle et al. (2014) find that information and training on the importance of wearing eye glasses for children needing correction had no impact on take-up absent subsidies, but a significant impact on usage of free eye glasses. Their context is the following: While eye examinations with close to 20,000 primary school students suggested that about 16% needed glasses, only 15% of those had glasses at the time of the baseline. The study distributed free eyeglasses to a random subset of those who needed them (the randomization was done at the school level), and find that usage (observed through unannounced spot checks in the classroom) doubles among those diagnosed as needing glasses – but it remains stubbornly low at just around a third. A fairly intense training program (showing a short documentary-type film, handing out a set of cartoon-based pamphlets for students, and a lecture and handout for parents and teachers) had no impact on eyeglasses wearing in the pure control group (without free distribution) but was found to increase usage in the free distribution group by 17 percentage points or 30% compared to free distribution alone.

Ashraf, Jack and Kamenica (2013) also interact the subsidy level and information provision, but the information provided concerns the relative merits of one product over another, rather than absolute information about the returns to preventing or treating a condition compared to doing nothing. Specifically, in a door-to-door marketing campaign in urban Zambia, sampled households were offered the option to buy one of two water purification products, a product well known in the area and available at retail stores (called “Clorin”) and a similar product from another brand, which people had never seen before and which was not available at any local stores. The price of the familiar product was fixed at 800 Kwacha, the standard retail price. The price of the unfamiliar product was randomized across households, and varied from 0 to 1200 Kwacha. In addition to the price randomization, the information given about the unfamiliar product was randomly varied: half of all households were provided no information, while the other half were told that the unfamiliar product is similar to and as effective as the familiar product.

Ashraf, Jack and Kamenica (2013) find no overall impact of the information treatment on the demand for the new product, or on total demand. However, the demand curve for the unfamiliar product becomes steeper when information is provided, and it pivots exactly around the price at which the familiar product is available. Demand for the familiar product (as a function of the price of the unfamiliar product) pivots the opposite way, and total demand for water purification products does not increase significantly in the presence of information. This apparent complementarity between information and subsidies can be interpreted as follows: in the absence of any information, people tend to take the price of the unfamiliar product as a signal of its quality, so they are not completely turned off by high prices, while they are somewhat turned off by low prices. When information is provided, the signaling content of the price diminishes. As a result, people are less likely to be turned off by low prices, and more likely to be turned off by high prices (in particular, there is now no reason why they would pay more for the unfamiliar product than the price of the familiar product, since the information reveals that the two products are the same). The effect of information provision is thus to encourage more people to switch from the familiar product to the unfamiliar product at low prices, and to deter more people to do the switch at high prices.

On the whole, the few existing studies examining the impact of information on willingness to pay have found limited impacts on total level of investments in health products, but suggest that the impact of subsidies on health can be heightened when the subsidy is accompanied with information.

Impact of Information on Health Behavior Change

A number of experiments have studied the impact of information on health behavior change. In Kenya, as part of the provision of deworming medication discussed above (in Miguel and Kremer 2004), Kremer and Miguel (2007) randomly varied whether schoolchildren received information on how to avoid intestinal worm infections. The information was provided in the classroom by a mixture of trained teachers and NGO staff, and focused on preventative behaviors such as washing hands, wearing shoes, and avoiding infected fresh water. One year later, data on pupil cleanliness and shoe wearing (as observed by the research team) as well as self-reported data on exposure to fresh water showed no effect of the education campaign.

Also in Kenya, Duflo, Dupas and Kremer (2015) and Dupas (2011b) examined the impact of providing different types of HIV/AIDS information to primary school students. In a randomly selected subset of 328 schools, teachers were trained on how to implement the national HIV/AIDS curriculum, which focuses on abstinence as the only prevention method available for adolescents. Duflo, Dupas and Kremer (2015) find that the training greatly increased the likelihood that

teachers teach about HIV in the classroom, and two years after the training students whose teachers had been trained had greater knowledge about the disease. The intervention did not reduce childbearing rates among girls, however, suggesting that it did not decrease the likelihood that girls engaged in unprotected sex. It also did not reduce the risk of STI as measured after 6-7 years.

Within the 328 schools, Dupas (2011b) randomly selected a separate 71 schools to receive an information session that discussed the role of cross-generational sex in the spread of HIV, and the relative risk of HIV infection by gender and partner's age. In many African countries, HIV prevalence increases with age among men. Therefore sex with older partners, which in many cases occurs in relationships with so-called "sugar daddies", substantially increases the risk of HIV infection for adolescent girls. This information was provided by a trained facilitator, introducing herself as working for a local NGO, to upper grade students in the selected 71 schools. This "relative risk" information intervention, which provided adolescents with information on how to reduce their exposure to HIV conditional on being sexually active rather than only exhorting them to abstain, led to a 28 percent decrease in teen pregnancy among school-going adolescent girls, and was driven by a reduction in cross-generational sex (with male partners five or more years older). Together, the results of these two experiments suggest that providing specific information is more effective than general exhortation at changing sexual behavior.

Rather than experimenting with the *content* of the information provided, two recent studies have experimented with the *delivery method* for HIV information. Indeed, recent advances in communication technology mean that information does not need to be delivered in-person by either a teacher or an outside facilitator. In Colombia, Chong et al. (2013) looked at the impact of an online sexual health education course provided through schools. Researchers partnered with Profamilia, a large Colombian NGO, to randomly provide a course to one third of 138 ninth-grade classrooms from 69 public schools in 21 cities. One third of the classrooms were randomly assigned to the comparison group, which did not receive the program, while the remaining one third of classrooms did not participate in the course but were located in the same schools as the classrooms that did receive it. The course increased overall sexual health knowledge by 0.38 standard deviations, and increased positive attitudes towards condom use. There was no impact on self-reported sexual behaviors, but there was a reduction of 5.2 percentage points (83 percent) in self-reported sexually transmitted infections among females who were already sexually active before the program, suggesting that some students adopted safer sex practices. The reliance of the study on self-reported sexual behavior is somewhat problematic, however, for reasons discussed in section 2 above.

In Uganda, Jamison et al. (2013) tested the impact of increasing access to information about sexual and reproductive health for the general population (not just schoolchildren) via a text messaging service about risky sexual behavior. Among 60 villages, marketing teams encouraged individuals in a random subset to use a new mobile phone-based information system through which users could send questions and receive responses on sexual and reproductive health. Usage among these villages was fairly high at 40 percent, but the service had no discernible impact on villagers' sexual or reproductive health knowledge. The intervention did, however, lead to an overall increase in the incidence of risky sexual behavior and self-reported promiscuity, particularly among men, while women reported increased abstinence. Qualitative information sheds some light on the potential causes of this mixed impact: men and women both reported that married women who learned about the risks associated with having an unfaithful partner insisted their husbands be faithful and get STI tested. According to these reports, some husbands did not comply, leading women to deny them sex and men to seek sex from alternative partners. Overall, individuals in treatment villages perceived their sexual behavior to be riskier, which could indicate an actual increase in risky behaviors, or could indicate that the information service increased accurate assessment of health risks. Unfortunately the researchers here again do not have objective measures of the risk level, such as biomarkers of sexually transmitted infections or pregnancy, to tease out these two potential explanations.

In Malawi, Godlonton et al. (2014) estimate the impact on sexual behavior of an information campaign about the relationship between circumcision and HIV status. They study the impact on men who are not circumcised at baseline, as well as those who are. For the former group, they find that the information increased correct knowledge about relative risk, reduced risky sexual activity, and increased condom use. Specifically, uncircumcised men in the treatment group were 25 percent less likely to have sex each month, and 58 percent more likely to use a condom. For the latter group, which learned they were better protected, there was no evidence of riskier sexual behavior. While uncircumcised men reported an increased willingness to have their male descendants circumcised, overall take-up of adult male circumcision was low. Researchers also found that the circumcision information campaign, though it increased correct understanding about how male circumcision can partially protect males against HIV transmission, also increased the incorrect belief among participants that male circumcision protects females against infection as well (which it does not). These results suggests that information alone is not enough to increase the demand for male circumcision, and that one has to be careful in the way information is delivered to mitigate the risk that incorrect beliefs are propagated.

Also in Malawi, Chinkhumba et al. (2014) conducted a randomized experiment to study the impact of information and prices on the demand for medical male circumcision. 1,634 men were

given vouchers for a subsidized circumcision at a nearby clinic; the researchers randomly assigned different values to the vouchers, with subsidies ranging from 8% to 100% (i.e. free) of the full price. The study randomly selected half of the men to receive comprehensive information about the biological relationship between male circumcision and HIV risk. Results were collected through both self-reports and clinic records. Information increased the number of circumcisions by 66 percent (1.4 percentage points), but overall the rate of circumcision was extremely low: no one offered the full price was circumcised, and only 3.1% of those offered a free circumcision elected to take up the procedure.

Banerjee, Barnhardt and Duflo (2015) compare the impacts of two information interventions on the take-up of salt fortified with both iron and iodine (Double Fortified Salt, or DFS) in rural Bihar, India. The extremely high rates of anemia in the region suggest that widespread adoption of DFS could have important health impacts for adults as well as children. Absent any information intervention, however, just under 10% of households use DFS about two to three years after the introduction of the product at a subsidized price in their villages (and only 20% have ever tried it). This suggests a potential role for information interventions.

The authors compare a light touch information intervention – namely, a flyer hand-delivered at respondents’ homes, informing them of where DFS can be bought locally – to a fully-fledged “infotainment” intervention. The infotainment intervention consisted of a specifically designed movie which lasted 26 minutes and was shown during the intermissions of two screenings of a classic Bollywood movie, first at night in the center of the village and again the next day at a school or health center. The infotainment movie aimed to showcase the health benefits of adequate iron consumption and the availability of iron in DFS in an entertaining way. The movie was modeled on sitcoms and starred real actors from the local movie industry, and the main character in the movie is a short and scrawny man who dreams of having a tall and strong son. During a prenatal care visit, his pregnant wife learns of the importance of taking iron supplementation for the health and future physical development of the unborn child, and convinces him to purchase DFS.

In light of the low baseline take-up absent any intervention, the movie intervention had a large impact: at follow-up, conducted between 7 and 16 months after the movie was shown in the village, DFS use was 5.5 percentage points (57%) higher in treatment villages. Having “ever used” DFS was 11.5 percentage points (22%) higher. Based on observed viewership for a random subset of the screenings, the authors estimate that someone had seen the movie in roughly 20% of households on average, so the impact of the infotainment intervention in “per viewer” terms is very large. In contrast, the “light touch” encouragement (the home-delivered flyer) had no

apparent effect on usage. This suggests that a heavy touch is needed for information on the importance of DFS to sink in.

Interestingly, in a parallel experiment the authors tested the importance of incentives for retailers in the diffusion of DFS. In the treatment arm for that experiment, retailers selling DFS were given financial incentives in the form of higher markups on the DFS. This intervention had an impact on take-up of the same magnitude as that of the infotainment, but not because retailers were effective at changing households' perception of the importance of consuming DFS: retailers did not appear to try to convince households of the importance of DFS for health, or if they tried, they did not succeed. Instead, they pushed the DFS on households by claiming to have no other salt in stock. The fact that shopkeepers cannot be successful information messengers for new health products is perhaps not surprising, since their credibility as a neutral information agent is limited unless they can prove that they are not making more profit from selling DFS compared to other types of salt. I.e., by construction greater financial incentives to boost sales of a new health product will likely undermine the potential for shopkeepers to be effective knowledge agents about the product.

Impact of Tailored Information on Behavior Change

All the studies above concern generic information. In contrast, Prina and Royer (2014) study the impact of providing tailored (individual-specific) information – namely, the impact of providing parents with body weight report cards for their own school-aged children. The report cards included information on a child's height and weight as well as their weight classification (i.e., underweight, healthy weight, overweight, or obese). This intervention increased parental knowledge and shifted parental attitudes about children's weight, but did not lead to meaningful changes in parental behaviors or children's body mass index, even when the body weight information was accompanied by information on the health risk of obesity. The authors provide evidence that social norms matter: if the report card included information on the distribution of weights in the classroom, then the larger the fraction of overweight children in the child's class, the less likely a parent was to report that her overweight child weighed too much. As the authors note, this implies that as obesity rates increase, programs aimed at reducing obesity may become less and less successful, as local reference points for appropriate body weights may rise.

Also looking at the impact of tailored information, Madajewicz et al. (2007) test the impact of informing households in Bangladesh about the safe/unsafe status of the arsenic concentration of their local well water. They find that the information treatment increased water source switching: 60% of households informed that they were using unsafe wells changed wells, compared with only 8% of households in control areas changing wells within the same time period. In contrast with

Dupas (2011b) in the context of HIV risk information, Benneer et al. (2013) show that adding to such bright line message (“this well is safe/unsafe”) some information about the dose response of arsenic contamination (namely, the fact that switching to a well with a lower level of arsenic concentration is always better than not switching, even if the switch-into well is not below the national “safe” standard) does not help – it does not increase switching, in fact if anything it decreases it, although the authors cannot reject the null hypothesis of no additional effect of the dose response information.

Diagnostic test results are another form of tailored information. Access to such tests is fairly limited in many developing countries where testing is expensive and hard to access. The experiment in Cohen, Dupas and Schaner (2015) included randomized access rapid malaria diagnostic tests (RDTs) among households in Western Kenya, an area where overtreatment with malaria appears very common due to poor access to reliable diagnostic tests. They find that conditional on seeking care for presumed malaria illness, those who learned their malaria status through an RDT were 40 percentage points less likely to buy malaria medicine than those who did not know their status at the time of purchase, reflecting the fact that only 36 percent of adults seeking malaria treatment in response to a presumed malaria episodes do not actually have malaria. Still, they find that compliance with a negative test result is not perfect: just around half of adults testing negative for malaria still went on to purchase an antimalarial drug. This cautiousness in learning from one’s own test results exemplifies the complexities associated with learning in the very noisy health environment that many people find themselves in, where information on the reliability of the test itself is limited. In an environment with at least three unknowns – the true underlying cause of an illness episode, the relative efficacy of drugs given a cause, and the accuracy of diagnostic tests – establishing over time the reliability of information provided by local health “experts” is extremely challenging. Since many diseases are self-limiting, a true non-malaria episode nevertheless treated with an anti-malarial treatment may appear to benefit from the treatment (and may thus be misperceived as a false negative on the test) even though it would have resolved equally rapidly without treatment.

Delavande and Koehler (2012) and Gong (2015) exploit randomized access to HIV testing to study the impact of learning one’s HIV status on sexual behavior in Malawi, Kenya and Tanzania, prior to the introduction of anti-retroviral therapy. HIV risk and sexual behavior is another domain where there are a number of often unknown parameters – the status of partners, the transmission rate per unprotected act conditional on having sex with an infected partner, and one’s own status. As a result, and as previously argued by Boozer and Philipson (2000), the impact of HIV testing on subjective beliefs and risky behavior is theoretically ambiguous. Learning about one’s status at a point in time provides a joint signal about the transmission risk and previous partners’

statuses. Depending on priors regarding those, the HIV test result can thus lead individuals to update their beliefs about the transmission parameter in opposite directions. Those who know they had an infected partner in the past may revise downward their beliefs about the transmission risk. At the same time, a test result can change one's time horizon -- in particular, assuming they trust the test enough, a negative test result would increase expected life expectancy for those who were pessimistic about their status before. In contrast, those who get a positive test results would suddenly have "nothing to lose" (especially absent treatment).

Delavande and Koehler (2012) exploits the randomized access to HIV test results from the 2004 experiment by Thornton (2008), discussed in section 5.5, to look at some of these pieces of the puzzle in the context of Malawi. They find surprising results: those who learned they were HIV negative in 2004 are more pessimistic about their status in 2006 than those who did not learn their status, and also have less precise beliefs. Those who learned they were HIV positive did not change their beliefs about their own status, and if anything revised downwards their beliefs about the transmission rate (the authors suggest this could be because a number of them jointly learned their spouse was HIV negative). Ultimately they find no impact on risk taking among those who learn they are negative and a decrease in risky behavior among those who learn they are positive. They do not perform the analysis separately by baseline prior regarding one's own risk however, as they only have data on subjective expectations at follow-up. They also do not have objective measures of risky behavior.

Gong (2015) is able to overcome both of these shortcomings in his re-analysis of a randomized study of voluntary counseling and testing (VCT) for HIV, conducted in the mid-90s in Kenya and Tanzania (The Voluntary HIV-1 Counseling and Testing Efficacy Study Group, 2000). In this earlier study, baseline data on expectations were collected, as well as data on STI incidence over the 6 months that followed the randomized VCT intervention. Gong (2015) tests whether STI incidence at the 6-month follow-up varies as a function of treatment status among 4 subgroups: (1) those who are HIV- and anticipated they were; (2) those who are HIV- but thought they were not; (3) those who are HIV+ and anticipated they were; and (4) those who are HIV+ but thought they were not. The author finds significant impact of the VCT treatment (learning your true HIV status) among those who are "surprised" by the test result (people in groups 2 and 4). In particular, those surprised by a negative test result are less likely to get an STI. While those surprised by a positive result are more likely to get an STI. The author argues that these differences in STI incidence are due to differences in sexual behavior -- those surprised by a negative test result adopt safer behaviors while those surprised by a positive test result adopt riskier behaviors. In contrast, unverifiable self-reported sexual behavior data finds decrease in risky behavior for all groups.

While the set-up in Gong (2015) is very appealing, a weakness is that the sample is limited in size. What's more it suffered from important attrition between testing and follow-up, and the subgroup analysis was an afterthought. In fact, the team of researchers who initially implemented the randomized study had estimated that “the study was not powered to detect significant differences between groups in the rate of incident sexually transmitted disease” (The Voluntary HIV-1 Counseling and Testing Efficacy Study Group, 2000). Thus one concern with Gong (2015)'s analysis could thus be one of ex-post data mining, since doing the analysis by subgroups based on priors was not pre-specified.

Impact of Targeted Information

A final question regarding the role of information on health behavior concerns whom the information should target. Ashraf, Field and Lee (2014) explore this question in the context of the demand for family planning in Zambia. Women in the study received vouchers that granted appointments with a family planning nurse at the local government clinic. Information explaining all methods of family planning, including “concealable” methods such as injectables, was provided along with the vouchers. Women were randomized into two treatment groups. In the “individual” arm of the study, women were given these vouchers alone. In the “couples” arm, women were given these vouchers in the presence of their husbands. In all other respects, the experimental protocol in the individual and couples arms was identical. The authors find that take-up of the voucher was significantly lower when information about family planning services was provided in the presence of husbands: women who received the voucher in the presence of their husbands were 9 percentage points (18 percent) less likely to use the voucher to obtain an appointment at a family planning clinic. This gap was larger (12 percentage points) among couples with divergent fertility preferences, in particular, where the husband reported wanting more children than the wife, strongly suggesting that individual health choices cannot always be easily separated from within-household bargaining issues or other forms of interaction among household members.

5.4. Schooling Experiments

Educational attainment is an important predictor of health, even conditional on income, in most countries, irrespective of their level of development (Strauss and Thomas, 1995; Lleras-Muney and Cutler, 2011). While there are a number of potential channels through which years of schooling could have a causal impact on health – for example, if educational attainment increases how much information people have or are able to process about the health impacts of various behaviors, or if it changes individual's discount factor or intrinsic valuation of health – disentangling those from the reverse causal channel, in which better health increases the incentive to invest in education, is methodologically difficult.

Jensen and Lleras-Muney (2012) follow-up a randomized intervention that increased schooling among men in the Dominican Republic and estimate a reduction in the incidence of their heavy drinking and smoking in the short run. They argue that the effect came about mainly by changing subjects' resources and peers (since at the time of follow-up, those with more education were still in school and thus less likely to work).

While there is no randomized experiment conducted over a sufficiently long timeframe to provide causal estimates of the effect of education on health in the long run, evidence on medium run effects can be found in Duflo, Dupas, and Kremer (2016), who randomized access to secondary education in Ghana by offering scholarships to a random subset of students who had received admission into senior high school but had difficulty enrolling due to financial hardship. The scholarship, offered in January 2009, produced a large difference in enrollment and completion of secondary school: among boys (girls), 75% (65%) of the scholarship winners completed senior high school compared to 47% (36%) of the non-winners. Since 2009, Duflo, Dupas and Kremer (2016) have been keeping track of the 2,064 youths in their study sample (a third of whom were randomized into the scholarship treatment) in order to study how this gap in educational attainment translated into gaps in later life outcomes, including health. Longer term effects on objective health levels are yet to be measured, but as of the 7-year follow-up they find significantly higher rates of preventive health behaviors, such as bed net use, condom use, handwashing with soap and use of mosquito repellent, among those who were offered the scholarship. A likely channel for this finding is the information channel: they find little evidence for price or income effects, as their sample is not yet well integrated in the labor market at the time of the follow-up, and also find no effect of education on individual discount factors. Scholarship winners scored an average of 0.16 standard deviation higher on a reading and math test administered in 2013, and were also more likely to engage with the media (e.g. read the newspaper), suggesting that schooling differences also led to cognitive skills gains, another potential channel.

5.5. Non-monetary Cost Experiments

For some products or services, demand remains low even at very low or even zero prices. For example, in the poor district of Udaipur in India, despite the fact that immunization services were offered for free in public health facilities, Banerjee et al. (2010) estimate that only 2% of children aged between 1 and 2 had received the recommended basic package of immunizations. As discussed above, in Malawi, only 3% of uncircumcised adult males who received a voucher for a free circumcision at the local clinic underwent the surgery (Chinkhumba et al. 2014). This could be because of non-monetary costs associated with take-up, such as time costs (e.g., walking 60

minutes to reach the facility where the free service is available), hassle costs (having to fill out complicated paperwork to receive a subsidy), and cultural barriers (for males thinking of getting circumcised as a means of reducing vulnerability to HIV). Some of these non-monetary costs can be experimentally varied, in particular, distance and convenience. Other, most notably cultural psychic “costs”, cannot be directly experimented on, but their relative importance can sometimes be backed out by experimenting with financial incentives, the idea being that if a relatively small financial incentive increases adoption, then cultural barriers cannot be too important.

Back to the example from Udaipur, India, Banerjee et al. (2010) run an experiment to test the hypothesis that the reliability of the supply of free services may actually be at fault. The premise for this hypothesis is the observation made by Banerjee, Deaton and Duflo (2004) (and discussed earlier in the chapter) that public facilities in charge of providing free immunization are characterized by very high absenteeism: spot checks conducted over a year suggested that 45 percent of the health staff were absent from their health posts, typically leading the health post to close, on any given workday. Because there was no predictable pattern to this absenteeism, obtaining all five shots included in the basic immunization package could require twice as many attempted visits to the public health facility.

In the experiment, some villages served as controls and other were randomly selected to receive a reliable, well-advertised “immunization camp”. The researchers found that the adding a reliable camp boosted full immunization rates from 6% to almost 18%. Impressive as the tripling of take-up in the immunization camps experiment may be, at only 18 percent fully immunized, the take-up rate remained among the lowest in the world. Could that be due to cultural barriers? In the experiment, a third group of villages was randomly selected to receive incentives for parents, in addition to the immunization camps. Specifically, parents were given a kilogram of lentils per immunization, and a set of plates for a child fully immunized. The incentive treatment increased immunization rates from 18 percent to 39 percent, which suggest that cultural barriers may not be decisive since they can be overcome with a fairly small handout. So what is the main barrier? We discuss the potential interpretations of this and other incentives experiment in the next section.

Thornton (2008) conducted a field experiment in rural Malawi that randomized the distance that individuals had to travel in order to obtain results of an HIV test, as well as whether they received a financial incentive to seek their results. This field experiment took place in 2005, at a time that preceded the introduction of rapid HIV tests, thus people had to do two visits to the testing center in order to learn their status – a first visit to get their blood drawn and a second visit a few weeks later to fetch their result. At the time Thornton conducted her experiment, the prevailing conventional wisdom in HIV prevention circles was that demand for knowing one’s status was

very low due to the high psychic costs of learning one was HIV positive: since there was no access to antiretroviral therapy (ARV) in Malawi at the time, learning one's positive status was akin to a death sentence. If psychic costs were indeed high, then providing a financial incentive and reducing the time costs of fetching one's results to a minimal level would likely have only a small effect on the demand for test results.

Thornton (2008) found the exact opposite: providing a financial incentive increased the likelihood that individuals sought their HIV test results from 35% to 78%. Reducing the distance that one had to travel to get results also increased the share of individuals seeking their results. Absent any incentives, those living within 1.5 kilometers from the center where results could be picked up were 6.4 percentage points more likely to seek their HIV results than those living more than 1.5 kilometers away. These large impacts thus teach us two things: that distance matters, i.e., time costs are not to be neglected; and that psychic costs were, on the other hand, much less important than believed at the time.

In their study of the demand for contraceptives in Zambia, the vouchers that Ashraf, Field and Lee (2014) distributed granted appointments for ordinary family planning services (provided routinely at government clinics), but with a guarantee that the wait would be less than one hour, and that the modern contraceptive method of their choice would be available. Take up of the voucher was high (47 percent), indicating, as in Banerjee et al. (2011), that unreliable supply and its associated substantial time costs may be important barriers to the take-up of health services.

The fact that non-monetary costs matter can sometimes be used to improve targeting. Recall the results from the pricing experiments discussed in section 5.1, which on the whole suggested that in environments where people face serious liquidity constraints, as in most of the developing world, price is not a particularly good screening mechanism. It fails at allocating scarce products to those who have higher returns to these products, as many people with a high valuation for a product may not be able to afford it. On the other hand, under free distribution, the product may be wasted on people with a low valuation for the product, and for products where there is a high share of low-valuation individuals in the population, this may be very costly.

Imagine that a provider delivers a year's supply of water-treatment product to a household, and the household members learn within a few days that they hate the taste of chlorinated water and stop using the product. In such cases, where households need to learn their own valuation, imposing some non-monetary cost that households have to pay to access the free product may be efficient. This is what the literature refers to as an "ordeal mechanism". The provider may, for example, require that those who want a year's supply go to a store to redeem coupons every month for 12 months. A field experiment conducted in Western Kenya in 2007-2008 suggests that

such a micro-ordeal can help target free products only to those who will use them. Dupas et al. (2016) provided households with the opportunity to obtain enough free samples of chlorine solution for the treatment of drinking water for a whole year, but varied the effort required to obtain the samples. Households randomly allocated to treatment arm 1 received a free supply of chlorine delivered directly at their home, while households randomly allocated to treatment arm 2 were given 12 coupons which could be redeemed for chlorine at a local shop over the course of a year. The researchers compare chlorine usage across arms and find no difference in rates of usage, during the year that followed the distribution, between those who were required to redeem coupons, compared to those who were given chlorine directly. They estimate that under reasonable assumptions regarding distribution costs, the results imply a significant increase in cost-effectiveness, with no negative impact on usage, of imposing a non-monetary price on the acquisition of a health good such as chlorine, which is not valued equally by all households.

Rozelle et al (2014) perform a similar micro-ordeal experiment when estimating the demand for and usage of eyeglasses for schoolchildren in China. Recall from the discussion in section 5.2 that this study, performed in 2012-2013, had found that usage of free eyeglasses (delivered at school) among primary school children in China was lower than 50%. This suggests that eyeglasses are more like chlorine than bed nets – the cost of using them outweighs the perceived benefits for a large share of the population. This is therefore an area where screening the unlikely non-users prior to free distribution could be highly cost-saving. Rozelle et al. (2014) adopt the voucher approach of Dupas et al. (2015) as a “micro-ordeal” – parents had to redeem a voucher for free eyeglasses for their child at an optical store in the county’s center, about 25 kilometers away on average -- and find that it successfully screens out households whose child would not use the eye glasses were they delivered for free at school. When interacted with the information training program discussed earlier, however, the ordeal mechanism appears counterproductive: it reduced the impact of the information. This result is interesting and suggests additional research on micro-ordeal is necessary to better understand when they are desirable. Indeed, it could be that the micro-ordeal mechanism undermined the information treatment in Rozelle et al. (2014) – if recipients interpreted the voucher scheme as a lower level of endorsement of the product on the part of the program implementers, compared to the endorsement that comes with free distribution at school.

5.6. Incentive Experiments

Earlier we briefly mentioned two experiments that included incentives as one of their treatment arms: the Udaipur immunization study (Banerjee et al., 2010) and the HIV testing study (Thornton 2008). In both cases, small financial incentives were provided to encourage take-up of

a specific health behavior: immunizing children, and learning one's HIV status, respectively. In both cases, the small financial incentives had a large impact on take-up. Banerjee et al. (2011) interpret this as possible evidence of present bias: the natural tendency to delay an action that is slightly costly today even if it has high payoffs in the future, a tendency which can be overcome if the incentive is sufficient to transform the cost into a positive *immediate* benefit.

While these two studies consider incentives rewarding a specific behavior (a specific input in the health production function), more recent studies have looked at the impact of output-based incentives, namely, incentives paid based on achieving a certain health outcome. de Walque et al. (2012) estimated the impact of offering varying amounts of a cash incentive to remain STI-free among adults aged 18-30 years old in Tanzania. They randomly assigned 2,409 individuals to one of three groups: 1) a "high-value" conditional cash transfer (CCT) group that received \$20 for testing negative for curable STIs; 2) a "low-value" CCT group that received \$10 for testing negative for curable STIs; and 3) a comparison group that received no transfer. STI tests were conducted for all groups every four months over one year. Over the course of the first year, the number of people who tested positive for STI infection significantly decreased among people who received the high-value CCT, but no reduction was found for the group that received the low-value CCT. One downside of this study is that it was not powered to detect effects on HIV infection, given lower prevalence (relative to the treatable STIs).

In a similar study in Malawi, Kohler and Thornton (2012) randomly provided cash transfers of random amounts to 1,307 participants, which ranged from no cash to approximately US\$16, conditional on maintaining one's HIV negative status for one year. Researchers conducted interviews with participants throughout the year to collect data on sexual behavior. The promise of financial incentives of any amount had no effect on subsequent self-reported sexual behavior or HIV status. However, receiving cash after the final round of HIV testing had significant effects on respondents' self-reported behavior: specifically, men were 9 percentage points more likely to engage in riskier sex, and women were 6.7 percentage points less likely to do so. As in the case of the incentivized immunization experiment discussed above, these results suggest that money given in the present may have stronger effects on behavior than rewards in the future – but sometimes for worse, as in this case.

5.7. Psychology Experiments

A host of studies in developed country contexts have explored how insights from psychology might be harnessed to increase adoption of preventive health behaviors. Indeed, in contexts where information, access and affordability are (largely) non-issues, the role of human behavior becomes the primary driver of health outcomes, and numerous surveys document that individuals have

difficulty keeping up with their intentions when it comes to many health choices, including maintaining a regular exercise regimen, dieting, preventive screenings, etc. This has been coined the “want/should” conflict, pervasive in many domains besides health (Bitterly et al., forthcoming).

One insight from psychology concerns the role of *planning* in solving the “want/should” conflict. It has been argued that forming precise plans can reduce forgetfulness and procrastination by linking intended behaviors with a concrete future moment and course of action. Consistent with this, randomized experiments in the United States have found that prompting people to form plans about where and when (not only the precise day, but even the precise time) they will complete an intended health behavior significantly increases the likelihood that they follow through with it, be it vaccination (Milkman et al., 2011) or preventive screening (Milkman et al., 2013).

Related prompts have been attempted with a cross-cutting design in some of the pricing experiments described above. In particular, Dupas (2009) evaluated the impact of having individuals verbally commit to purchase the bed net product: a randomly selected half of all the households were asked about their intention to redeem the voucher for the bed net, and over 92% said “yes”. They were then asked to state who would sleep under it once they had bought it, and also asked to estimate how many days they would need before they were in a position to purchase the net. At the end of this discussion, the enumerator asked: “Ok, so this means that if I stop by the shop on *[date after the date provided by respondent]* I will find your deemed voucher there? Can you promise? Do we have a deal”? This verbal commitment intervention had no impact on actual redemption behavior. In particular, it made no difference to the slope of the demand curve.

This could be because people were victims of the so-called “planning fallacy” (Buehler, Griffin, and Ross 2002): it seems they underestimated the time it would take them to save enough to afford the bed net. Kremer and Miguel (2007) similarly had earlier found no impact of an explicit verbal commitment intervention embedded as a treatment arm within their deworming take-up experiment in Kenya.

In some cases, the key issue may not be a want/should conflict but rather individual undecidedness. An extensive literature in psychology and marketing suggests that decision-making can be affected by frames or cues that do not add information about a product, but can be effective at persuading undecided individuals to invest in it. The aforementioned experiment in Dupas (2009) evaluated the effects of framing, besides those of price and planning. At the time they received their first voucher, households in her experiment were exposed to a randomly assigned marketing message. The “health framing” group emphasized the morbidity and mortality due to

malaria which could be avoided by using the bed net. The “financial framing” group emphasized the financial gains households would realize (from averting medical costs and loss of daily income) if they could prevent malaria. A third group received no marketing message. Neither of the two framing options (health or financial) had any impact on bed net take up. Combined with the finding that usage is quasi-universal among free bed net recipients, this finding suggests that the price sensitivity of bed net demand is largely due to liquidity constraints rather than undecidedness.

A third important factor in health behavior is how much attention people have to devote to it. Adherence to daily health treatment regimens is often imperfect, and “I forgot” is a common answer to the question of why a patient did not take his or her treatment as planned. Against this back-drop, Pop-Eleches et al. (2011) estimate the effect of reminders on adherence to antiretroviral treatment among AIDS patients in Kenya. 430 adult patients who had initiated anti-retroviral therapy (ART) within three months were assigned to either a comparison group that received no reminders, or to one of four groups that received either: 1) a short message sent daily, 2) a long message (that included words of encouragement) sent daily, 3) the short message sent weekly, or 4) the long message sent weekly. Treatment adherence was monitored through bottles with a medication-event-monitoring cap.

The results indicate that 53 percent of those in the weekly reminder groups achieved treatment adherence greater than 90 percent, compared to just 40 percent in the comparison group. Yet daily reminders had no impact on individual ART adherence, suggesting that patients respond less to a frequently repeated stimulus. The type of message also had no impact on drug adherence. There is a need for future research to understand why daily reminders in this context were ignored (in particular, were they so frequent to get treated as “spam” and not even opened?) while weekly reminders had a substantial effect. It would also be useful to explore the role of different types of framing beyond those explored here.

5.8. Taking stock: How important is the role of present bias in explaining observed preventive health behaviors?

In a chapter entitled “Improving Health in Developing Countries: Evidence from Randomized Evaluations” written for the *Handbook of Health Economics*, Kremer and Glennerster (2011, henceforth KG) review a subset of the experiments described in sections 5.1 to 5.8 (namely, those available at the time of writing) and conclude that they “provide considerable support for a present

bias model.” Specifically, they argue that present bias is useful in explaining why small prices or convenience barriers can dramatically reduce take-up of cost-effective approaches to prevention and non-acute care. In this section we revisit their conclusions in light of the accumulating evidence in this area, including the latest work.

The main argument underlying KG’s conjecture that present bias is an important factor is the following. In the standard human capital investment model they propose, in which individuals invest in a given health behavior or product if the expected discounted private benefit exceeds the cost, there is no reason to expect that a large proportion of the population would switch into using a given preventive product by the change from a low positive price to a zero price, unless there are large disutility costs that almost exactly offset the benefits. As they write, “the odds that this would occur for multiple products in multiple settings are particularly low” unless the discount rate is very high for a large group of people, but they rightfully argue that this would not be consistent with the fact that many people in the studies considered also invest in their children’s education, which likely generates only modest and very delayed returns.

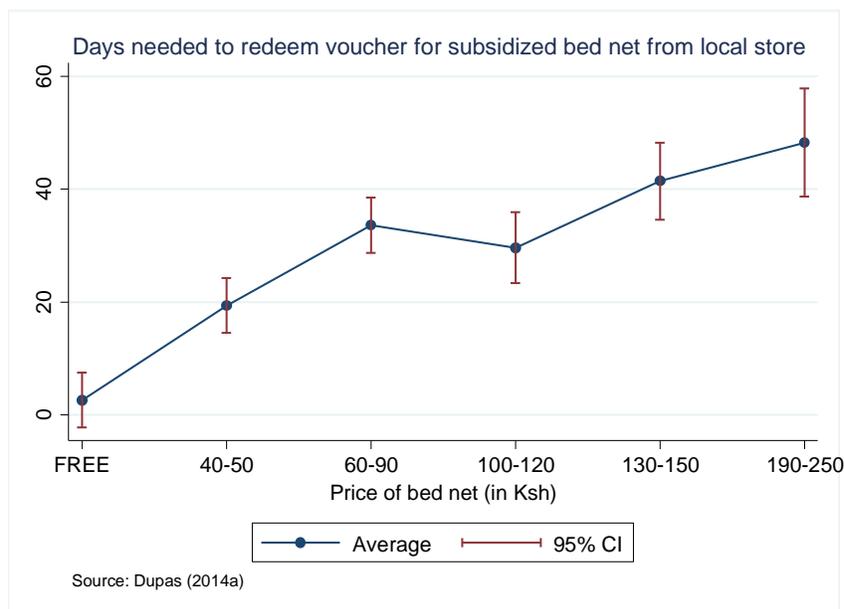
In contrast, in a model with present bias, a small change of price away from zero can have a large impact since the small cost is given a lot of weight against the discounted future benefit. Present bias could also explain why small incentives, such as the lentils provided in the Banerjee et al. (2010) immunization experiment in rural India, could lead to a sharp increase in take-up: the authors of that study write that “Providing the lentils helps to overcome procrastination because the lentils make the occasion [the visit to the immunization center] a small “plus” rather than a small “minus”. Thus, in the case of preventive care, small barriers might turn out to have large implications.”

While we agree with KG’s theoretical discussion, our interpretation of the evidence from the growing body of pricing experiments discussed in section 5.1 differs somewhat from theirs. In particular, as shown in Figure 2, for most products there does not appear to be anything special about zero, although there are exceptions (namely, for deworming and chlorine in Kenya, as noted above).⁵

⁵ Note that many experiments did not include free distribution (price of zero), which does limit interpretation somewhat. However, for most of the studies the smallest price point considered is fairly low and demand at that price very substantial. E.g. the fact that 93% of Kenyan households purchased slippers at the price point of 5 Ksh (USD 0.08) in Meredith et al. (2013), suggests that there is not a discontinuity in demand at 0.

Although the pricing experiments may not provide clear evidence that present bias (in terms of consumption) is a primary barrier to adoption of health products, one could also look for empirical evidence on the importance of present bias (and related self-control issues) in situations where monetary costs are unimportant but effort or hassle costs might cause procrastination. TIOLI experiments carried out with coupons (vouchers) redeemable at a pre-specified location (in contrast to the door-to-door visit experiments) provide a way to examine that issue. In particular, we would expect present-bias to lead many people to procrastinate in redeeming their voucher until it is about to expire. Yet the data on the timing of voucher redemption shown in Figure A1 of Dupas (2014a), reproduced here as Figure 3, suggests that most recipients of free bed net vouchers redeemed them immediately, suggesting any present bias in effort (if present) was not sufficiently large to overwhelm the positive expected health benefits. Similarly, none of the other studies that used vouchers (Meredith et al., 2013, Dupas et al. 2015, Dupas et al. 2016, Peletz et al., 2016) report a bunching in redemption around the expiry date. This is a strike against the present-bias interpretation.

Figure 3. Delays in redemption of vouchers for subsidized nets, by price



It is also potentially useful to revisit the present bias interpretation of the immunization incentives discussed in Banerjee et al. (2010). In that experiment, 78% of parents took their child to obtain the first injection without the incentive, and 75% took their child to obtain a second injection. The impact of the lentils incentive was thus mainly to “reduce the number of children dropping out after three injections.” This suggests that the hassle cost (immunization was free, but it took

some time and effort to go to the immunization camp which was up to 5 km away and where there could be some waiting time; moreover, the child might develop a fever afterwards) are not large enough to deter take-up of the first couple of injections. What is an issue is the fact that completion of the whole course (5 rounds of injections) is low (only 25% without incentives compared to 40% with incentives). The drop-off after three injections could thus be due to a lack of understanding that the later rounds of injections (the booster shots) are actually important. Observing the impact of immunization on health may be difficult in noisy environments where illness episodes are very common and due to many different causes – the difference in the incidence of health shocks among immunized children vs. non-immunized children over the first year of life may not be perceptible with the naked eye, and this may contribute to low perceived benefits of immunization. In such a context, the incentives could possibly have been perceived as a signal that completing the full course is in fact important, the same way transfers to parents labeled as “for education” raised the perceived returns to education in Morocco (Benhassine et al., 2015). Or it could be simply be that if the perceived benefits are very low, and the cost of going to the immunization camp is low enough that people are nearly indifferent, then even small incentives are sufficient to shift their decision.

Possibly the most clear-cut evidence in favor of the present bias model would come from empirical evidence of an individual taste for commitment. The definition provided in Bryan, Karlan and Nelson (2010) is that a commitment device is “any arrangement, entered into by an individual, with the aim of making it easier to fulfill his or her own future plans”. In the next section (5.8), we discuss three studies of commitment products for health behavior change from developing country contexts, all three finding that a subset of the population does take-up commitment schemes against future selves – though whether these work at affecting health outcomes is unclear. This emerging evidence from developing country contexts is reminiscent of a number of studies from developed country contexts, finding strong evidence of a demand for commitment to healthier lifestyles. For example, DellaVigna and Malmendier (2007) show that individuals purchase a flat-fee membership to the gym as an (often unsuccessful) commitment device to increase future attendance: in their data, 89% of new gym users sign up for a monthly contract, even though given their actual attendance frequency, 80% would have saved money choosing the pay-per-visit contract, holding constant the number of visits. The role of present bias in health decision-making may thus grow in developing countries as the burden of disease tilts towards non-communicable diseases (NCDs) like diabetes and hypertension, which have seen a substantial rise in prevalence (or at least, detection) in countries as varied as India, China, and Mexico in recent years, as well as in parts of Sub-Saharan Africa. Daily disease management, in particular lifestyle changes in terms of diet and exercise, are key to decreasing the rate of complications and avert early mortality

from NCDs. Thus to the extent that self-control and procrastination issues are particularly important for diet and exercising, there will be a growing need for health experiments in low-income countries testing behavioral change interventions that tackle time inconsistency in preferences.

5.9. Commitment Experiments

A growing body of empirical evidence from many domains suggests that people seek commitment devices to help themselves follow through on their plans. Commitment devices can be “hard” or “soft”. An experiment involving a hard commitment device for health is the study of the Committed Action to Reduce and End Smoking (“CARES”) program, a voluntary commitment savings program to stop smoking, designed and tested by Giné et al. (2010) in collaboration with a bank in the Philippines. The basic design of the program was as follows: a smoker could open a bank account and deposit a self-selected amount of his own money that would be forfeited unless he passes a urine test indicating smoking cessation after 6 months. Regular smokers were recruited for the study off the street. All subjects received an informational pamphlet on the dangers of smoking, and a tip sheet on how to quit.

The commitment contract was taken up by 11% of smokers offered the account, and on average participants made deposits every two weeks and had a balance of 585 (US\$11) pesos after 6 months, some 535 pesos more than the minimum balance and approximately six months’ worth of cigarette spending. Individuals who were *offered* a CARES contract were 3.3 to 5.8 percentage points more likely to pass a urine test (negative for nicotine) after six months than those in the comparison group, and were 3.4 to 5.7 percentage points more likely to pass it after 12 months, a substantial effect considering the well-known difficulties of quitting and the fact that only 8.9 to 14.7 percent of comparison individuals passed the test. This represents an over 35 percent increase in the likelihood of smoking cessation compared to baseline. Treatment on the treated effect estimates imply a 30% to 65% quit rate for this population relative to the control group. However, the overall welfare impact of offering the commitment contract program is unclear. A large proportion (66 percent) of smokers who voluntarily committed to CARES ended up failing to quit, and thus lost the money they had deposited.

Schilbach (2015) estimates the demand for commitment for *sobriety* among cycle-rickshaw drivers in the city of Chennai, India. The sample for the study was restricted to rickshaw drivers with a significant average daily consumption of hard liquor at baseline. At some point in the study, the

drivers in the sample were given the choice between either (a) a guaranteed payment of 150 Rupees or (b) a guaranteed payment of 90 Rupees and an additional 30 Rupees conditional on them passing a breathalyzer test. The payment scheme under option (b) is *de facto* an incentive to be sober. Since the maximum possible value of the payment under the incentive scheme is lower than under option (a), everyone should choose option (a), unless people *value* the incentive embedded in option (b). Schilbach (2015) finds that about 30% of those given the choice chose the incentive payment – thus losing at minimum 30 Rupees (\$0.50) for sure (and since many of them failed to pass the test, the actual loss was greater than that). When the choice was between option (a') with 120 guarantee vs. the same option (b) as above, the share taking up the incentive scheme was around 50%. Here again, a large share of those who took up the commitment device failed to pass the breathalyzer test and thus ended up earning less money than if they had chosen option (a'). The effect of the sobriety incentives on actual drinking behavior was modest – the incentives reduced day-time drinking significantly but not overall drinking, due to substitution towards night time drinking.

“Softer” commitment devices, which do not require that individuals put their own money on the line (be it savings or potential earnings), were studied by Dupas and Robinson (2013), who examine the determinants of health savings. They find a large demand for a commitment savings product that earmarks savings for health emergencies, as well as for a simple product (namely, a lockable box) that people can use to store their health savings. Both products have a large and significant impact on health investments, with the former decreasing the risk that an illness goes untreated and the second increasing spending on preventive health, compared to a control group that was also primed to save for health but not provided any specific tools to facilitate saving. In contrast, very little was saved in a product that locked up savings for preventative health with no possibility to access the savings in times of emergency, suggesting that some degree of flexibility is essential in product design.

6. Supply of Health Care

Most of the experiments described above in Section 5 examine determinants of the demand for *preventive health*. This is because demand for acute care is quite high, as discussed previously in Dupas (2011a) and Kremer and Glennerster (2011).

A growing body of evidence also documents important gaps in both access and quality in the delivery of health services in developing countries, especially for the poor (World Bank 2004; see Das and Hammer 2014 for a review). Major issues identified to date concern: absenteeism among public health providers (Chaudhury et al. 2006, Banerjee, Deaton and Duflo 2004, Banerjee et al. 2008); limited knowledge and training, as well as an important “know-do” gap among health professionals (see Das, Hammer and Leonard 2008 for a review); limited availability of diagnostic testing, leading to high rates of inappropriate treatment (Banerjee, Deaton and Duflo 2004; Cohen, Dupas and Schaner 2015); and drug quality problems (Bennett and Yin, 2014; Nayyar et al., 2012). Another common concern is that of corruption among health providers, although quantitative evidence on this is limited, and the evidence to date is perhaps less pessimistic than anticipated (Dizon-Ross, Dupas and Robinson 2016). The majority of experimental economics research on these issues has focused on testing the effectiveness of interventions aimed at improving quality by changing how providers are monitored and/or incentivized. Only a handful of experimental studies have seriously attempted to tackle the problems of limited diagnostic availability and drug quality.

The healthcare market in most countries in the developing world is comprised of government facilities with trained professionals, adjacent to a myriad of loosely regulated informal providers, from quack doctors to drug shops staffed by individuals with no formal pharmaceutical training but from whom medical advice is regularly sought. Incentives for healthcare providers at public facilities to come to work, or to perform well while at work, are generally seen as very weak (World Bank 2004; Das and Hammer 2014). In contrast, private providers face at least some market discipline, but their ability to perform may be limited by inadequate medical training, and given the traditional information asymmetry in the health sector, patients’ ability to avoid low quality informal providers may also be limited. In this set-up, there are two ways to improve the quality of the health services that the majority of the poor have access to. The first is to better improve incentives for trained providers in the public sector. The second is to improve the quality of informal providers through better training.

Below we first review the set of studies that document the types of problems observed in the health care market, before turning to experiments that aimed to provide solutions to some of these problems.

6.1. Experimental audit studies

The last fifteen years have seen a great deal of innovation in how empirical audit studies can be used to measure the quality of health care provision. An important innovation is the use of a type of audit called the standardized patient (SP) method, which has long been used for training purposes but has now been adapted to allow quality measurements in “business as usual” conditions, and is considered the gold standard in assessing the quality of medical care delivered in outpatient settings. Standardized patients are people from a local community who are trained as actors to present a given “case” (symptoms) to a health provider. They are trained to find ways to refuse examinations that would reveal that they are in fact lying about some symptoms (e.g., temperature checks if they report having a fever). After the visit with the provider, they then record the details of their interaction, and in particular, any questions asked by the provider, any examinations done, and the diagnosis that was pronounced (if any). This method and other measurement tools are well described and discussed in the World Bank manual *Are You Being Served? New Tools for Measuring Service Delivery* (World Bank 2008), to which we refer the reader for more details. In what follows, we describe some of the key insights generated by this new type of measurement tool.

Das et al. (2012) use the SP method with a representative sample of providers to document provider quality in both rural and urban areas of Madhya Pradesh in 2008–2009. The study revealed a number of issues, including that: the majority of health care providers are not medically trained; overall quality is poor in both urban and rural settings and across all types of providers, with the average provider spending less than 4 minutes with a patient, and low rates of diagnosis (let alone correct diagnosis), low rates of correct treatment, and massive levels of overtreatment. Similar results were found in a replication of the study among 48 providers in a Chinese province (Sylvia et al. 2015).

Following on this work, Das et al. (2016) use the SP method to compare the quality of care provided by a given provider depending on whether the patient visited him in his private practice or at the public facility. Indeed, even though it is illegal, around 61% of public providers with a medical degree moonlight (i.e., run a separate private practice on the side). In private clinics, providers are paid on a fee-for-service basis by their customers. They can also earn a profit from selling medication. In contrast, as public servants doctors are paid a fixed salary, and the drugs they prescribe are supposed to be provided for free at the public clinic. Patient demand for

treatment appears very high in the context studied, however, so there are possibly demand-driven incentives to over-prescribe among both public and private providers.

There are 71 providers in this Das et al. (2016) study that were visited by standardized patients in both their public and private clinics, and analysis generates three main findings: (1) providers exert greater effort when the SP visits their private rather than public practice; (2) the likelihood that the “correct treatment” is prescribed is higher in the private clinic, and the likelihood that a palliative treatment (which relieves pain but does not solve the problem) is prescribed is lower; and (3) the likelihood of incorrect or even harmful treatment is identical across the two settings. The total number of drugs prescribed is also identical across the two settings. The authors conclude that the incentives generated through customer accountability in the fee-for-service private market lead to greater provider effort and the higher rate of correct treatment being prescribed, in stark contrast to the solely administrative (and ipso facto non-existent) accountability in the public sector. Interestingly, this gain in provider performance does not come at the expense of increased overtreatment: the likelihood of incorrect treatment does not increase in the private clinics relative to the public setting, but rather as private providers exert greater effort in examining the patient (and thus acquire tighter priors on the likely illness), they are more likely to prescribe the correct drug.

A country where medical over-treatment is considered a particularly important issue is China, where antibiotic abuse in particular has been well-documented. Using simulated (standardized) patients, Currie, Lin and Zhang (2011) provide evidence that a large share of this over-prescription behavior is initiated by the physician: while none of their simulated patients required antibiotics given the symptoms they experienced, 62% were prescribed antibiotics. The question then becomes what incentives physicians face that leads them to over-prescribe. There are two leading hypotheses: first, they could be responding to what they think is the patients’ wish, or second, they could be responding to financial incentives. The setup in China is such that doctors do indeed have strong financial incentives to over-prescribe drugs: while wage employees at public hospitals, they get paid a bonus that represents typically over a third of their salary, and this bonus is in part based on sales that they generate at the attached hospital pharmacy.

Several pieces of evidence suggest that financial incentives are indeed a key factor behind drug over-prescription. First, Currie et al. (2011) find that physicians tend to prescribe more expensive rather than less expensive antibiotics. Second, Currie et al. (2011) carried out a randomization across the simulated patients regarding whether the patients would express knowledge of appropriate antibiotic use to the physician (during the visit). They find that such demonstrated knowledge by the patient reduces the incidence of unwarranted antibiotic prescription by 25

percentage points (40%). While this could be consistent with physicians having vastly mistaken beliefs about the extent to which consumers demand antibiotics, the authors argue it more likely reflects the fact that once physicians realize the patient will not buy the prescribed antibiotic in any case (due to their medical knowledge), the financial incentive to overprescribe is gone. To further test this point, in a follow-up audit study experiment Currie et al. (2014) randomized (within physician) whether the simulated patients: (A) said nothing special (the control group); (B) directly asked the doctor for an antibiotic prescription; (C) asked for a prescription (not specifically antibiotics) but indicated that he/she would buy any drugs prescribed in another pharmacy, thereby eliminating the financial incentive for the physician; (D) asked specifically for antibiotics and indicated that he/she would buy any drugs prescribed elsewhere. As in the previous study, over-prescription was very high, with 55% of physicians prescribing antibiotics when the patient said nothing (the control). This rate increased to a staggering 85% when patients specifically requested antibiotics, but only if the physician expected the prescription to be filled in the hospital pharmacy. If the patient indicated that he/she would purchase the drugs elsewhere, antibiotics were prescribed in only 14% of the cases, even when antibiotics had been specifically requested by the patient. Among patients who did not request antibiotics but indicated that they would buy any type of drug elsewhere, antibiotics were prescribed at a rate of 10%, statistically indistinguishable from the 14% if antibiotics had been requested. These results strongly indicate that high rates of antibiotic prescription are not mainly driven by patient demand or provider ignorance, but rather by providers' misaligned financial incentives.

These audit studies above have focused on physician behavior. Dizon-Ross, Dupas and Robinson (2016) audit the performance of other health workers, primarily nurses and midwives, as they are asked to implement a bed net distribution program targeted to pregnant women, an increasingly common scheme in Sub-Saharan Africa. They conduct these audits in Kenya and Uganda (where the distribution scheme is a government program), and in Ghana, where the distribution scheme is sponsored by a non-governmental organization. In all three cases, they find relatively satisfactory performance levels among providers, with the vast majority of eligible beneficiaries receiving the subsidized bed net as intended. They also measure whether health workers respond to bribe attempts from ineligibles by sending "mystery clients" -- undercover enumerators posing as ineligible individuals trying to obtain a bed net. They find that a very small minority of mystery clients were successful at obtaining a bed net from prenatal centers, suggesting that in the context of an easily observable targeting rule (only pregnant woman are eligible) health workers comply with it, the same way Chinese physicians stop overprescribing when they know their patient are knowledgeable about appropriate treatment (Currie et al. 2011).

Another important dimension of health care quality that can be measured through experimental audit studies concerns drugs, in particular, the prevalence of counterfeits. Bennett and Yin (2014) sent mystery shoppers to small drug stores in Hyderabad, India. The mystery shoppers bought two common off-patent antibiotics, which were then analyzed in a lab. They found that 6% of the samples fell below pharmacopeia standards. Among so-called local (cheaper) brands, the share counterfeited was as high as 22%. In a comparable study, Björkman, Svensson and Yanagizawa-Drott (2014) tested the quality of antimalarials (ACTs) in Uganda, and found that around 30% were fake.

6.2. Monitoring Experiments

To better align the incentives of trained providers with the overall objective of improving health outcomes, the interventions studied to date can be grouped into four primary types: district-level contracting, input-based incentives (e.g., nurses are paid a bonus, or avert a fine, if their absenteeism is sufficiently low), output-based incentives (e.g., providers are paid a bonus if health outcomes in their community are sufficiently high), and decentralization (giving greater monitoring power to local communities). We discuss each in turn below.

First, however, we note that many of the same types of interventions have also been assessed in the education sector in low income countries, and a number of findings in that literature are likely to be relevant to health. For brevity, we do not survey the education experiments here, however, and instead refer the interested reader to the chapter by Muralidharan in this volume. We also note that there have been many related provider performance experiments in the field of public health, although they are also not our focus here.

District-level contracting

In what is arguably among the most innovative experiments ever carried out in the health sector, in 1999 Cambodia launched a brand new approach to solving the low quality problem: it contracted out management of government health services to private (international) nonprofit organizations (NGOs) in some districts. In 8 districts randomly selected out of 12 districts involved in this at-scale “pilot”, NGOs could bid for the government contract, which was paid for through increased public-health expenditures. Contracts were ultimately signed in only 5 of the 8 districts, as in two districts no bids were received, and in a third district the bid was too expensive and was not chosen by the government.

In the 5 contracted districts, contractors were responsible for health services at all levels, from district hospitals down to remote health posts. Performance on eight key (targeted) service-

delivery indicators (most of them related to maternal and child health) was measured, and the contract was renewed yearly based on these outcomes. The idea behind this district-level contracting scheme is that it can strengthen incentives for government workers while reducing potentially harmful incentives associated with private fee-for-service provision, such as the incentive to over-prescribe antibiotics discussed above.

Bhushan et al. (2007) study whether the contracting in Cambodia indeed improved health performance by using the randomized assignment to contracting as an instrument for a district actually having a private contractor. They use baseline survey data from 1997 and follow-up survey data from 2003 in the analysis, and find meaningful improvements in health-care service delivery, especially on the targeted indicators, such as receipt of vitamin A by children under five, which increased by 21 percentage points, and receipt of antenatal care by pregnant women, which increased by 33 percentage points. These improvements also did not come at the expense of non-targeted health services, which experienced no decline. Instead, the gains on the targeted indicators appear to have come about through improvements in management quality: absenteeism of providers and stock outs of drugs and other equipment fell in districts where the NGOs were in charge. This increase in public facility service reliability in turn increased demand for these services, with residents increasing the number of visits made to public facilities and reducing their visits to (often expensive) informal providers, such as traditional healers.

Top-down, Input-based incentives

Absenteeism among public health providers has been shown to be a very significant concern in many parts of the developing world (Chaudhury et al. 2006). Since provider presence is obviously a necessary if not sufficient condition for health care services to function, a number of experiments in the health sector (alongside with a number of experiments in the education sector, plagued by the same absenteeism problem) consider the effectiveness of programs incentivizing provider presence. We discuss them in this section.

Banerjee et al. (2008) evaluated an incentives program for Assistant Nurse Midwives (ANM) at Primary Health Subcenters in Udaipur District in the Indian state of Rajasthan. The program was implemented collaboratively by a non-profit organization and the state and local health administrations, with the goal of improving ANM's attendance at rural subcenters. Indeed earlier research had established that due to pervasive absenteeism among ANMs, health centers were closed 56 percent of the time during regular business hours (Banerjee, Deaton and Duflo, 2004). The program tested consisted of monitoring ANM attendance and "punishing" absenteeism: ANMs absent for more than 50 percent of the time on monitored days would have their pay reduced

proportional to the number of absences recorded that month, and ANMs absent for more than 50 percent of the time on monitored days for a second month would be suspended from government service. The program was implemented in 49 randomly selected subcenters. In those centers, the ANM was required to stamp a register secured to the wall of the subcenter three times a day: once at 9 am, once between 11 am and 1 pm, and once at 3 pm, using a tamper-proof time/date-stamping machine. Researchers then measured the impact of the program on ANM performance through random unannounced visits to the 49 “treatment” subcenters and 51 control subcenters.

The results of this intervention are mixed. In the short-run, the incentive scheme was highly successful, doubling attendance, from around 30 to 60%. The program was not popular with nurses, however, who complained heavily to the local health administration about the pay deductions. The share of “missed stamps” due to either an (intentionally) broken time clock or excused absence increased considerably over time, and at 16 months after program inception, the absence rates were comparable between treatment and control centers. What the researchers take away from these mixed results is that, on the one hand, nurses are responsive to properly administered incentives, but on the other hand, incentive systems can be very difficult to properly administer, due to the perennial question of “who monitors the monitor?” Ultimately, the decision to monitor and incentivize public sector employees is a political one, and there may be a variety of political economy explanations for why these programs are opposed by either public employee unions, or the public at large.

A similar experiment took place a few years later with primary health care center staff in Karnataka (Dhaliwal and Hanna, 2013). Instead of an NGO, the program was designed and implemented by the National Rural Health Mission (NRHM) of Karnataka, the lead state department for the delivery of health services. Instead of stamps, the monitoring system relied on fingerprints taken at the beginning and end of each day, and instead of proportional pay deductions, the penalty was a loss of paid vacation days, although in practice the penalty was rarely imposed. These researchers found that the health staff monitoring system increased attendance among medical staff by 18%, but not among doctors. They also find a large, 26% decrease in the incidence of low birth weight, confirming that provider attendance is potentially a critical input in the health production function, although effects on a range of other health outcomes are mixed. The mechanism through which birth weight was affected appears not to have been through an increase in prenatal care attendance, but rather an increase the likelihood that prenatal clients received iron folic acid tablets. Once again, as in the earlier Indian health monitoring and incentive experiment, the program did not appear to be politically sustainable.

Top-down, Output-based incentives

While provider attendance may be the first step, it may not always lead to improved population health outcomes. For this reason, more recent monitoring experiments have based the rewards on outcomes rather than on inputs such as provider attendance. Basing incentives on actual health outcomes is difficult, however, for reasons discussed in Miller and Babiarz (2014), in particular, the fact that provider behavior is only one of many factors that determine health outcomes, and that health outcomes can be difficult and expensive to measure. Given this, outcomes over which performance-pay contracts are written tend to relate to the utilization or coverage of specific easily observable health services, e.g., the share of children who are immunized, the share of pregnant women seeking prenatal care, the share of deliveries that take place at the facility, etc. The potential downside of contracting over such specific and narrow indicators is that providers may devote too much effort to those, at the expense of activities related to non-contracted indicators which may be just as important for the production of health but simply harder to measure (in a version of the multi-tasking problem).

At the time of writing, we are aware of two economics field experiments that have directly tested the impact of performance-based incentives in the health sector, one in Indonesia and one in Rwanda. In both cases, the incentives were at the group level (not at the individual provider level), and the performance mattered for the total budget available to the providers, rather than for their own personal gain. (As far as we know, performance-based incentives for individual health workers have not yet been studied using field experimental methods.)

The Indonesia experiment estimated the effect of incentivized community-based block grants that aimed to improve both health and education (Olken, Onishi and Wong 2014). The program, known locally as *Generasi*, provided villages with annual block grants of \$8,000 to \$14,000, and villages were encouraged to use the funds to make progress on 12 pre-specified maternal and child health indicators, including prenatal visits, delivery by trained midwives, childhood immunizations, and child growth monitoring. For the experiment, which was conducted jointly with the Government of Indonesia, 264 subdistricts were randomized into either control or to one of the two versions of the *Generasi* program: the “incentivized” version with a pay-for-performance component, or the otherwise identical, “non-incentivized” version without pay-for-performance incentives. In the first year of the program, villages in all groups received program funds based on their size and demography. In the second year, the allocation rule stayed the same for the non-incentivized villages, but for the incentivized villages 20 percent of the funds were distributed based on the village’s performance on the 12 indicators during year one. Impacts were measured over the two years.

The pay-for-performance incentives led to an increase in the labor hours of midwives, the major providers of maternal and child health services in the area. Likely as a result, the targeted maternal and child indicators were somewhat higher in incentivized villages than in non-incentivized villages, but the overall effect was quite modest, with a gap of just 0.04 standard deviations on average. The main impacts were on the number of prenatal visits (an 8.2 percent increase) and regular monthly weight checks for children under five (+4.5 percent). The effect of the incentives varied with the baseline levels of service delivery, however, and effects were stronger in the poorer provinces not on Java. Interestingly, no detrimental effects of the incentive scheme were detected on non-targeted health indicators, at least to the extent they could be measured.

The Rwanda experiment was conducted in partnership with the government as it launched a national pay-for-performance scheme to supplement primary health center budgets (Basinga et al., 2011). As a pilot, the program was supposed to be launched first in 80 facilities from 8 randomly chosen districts, with 8 districts (86 facilities) assigned to a comparison group. Under the program, facilities received payments as a function of their performance on 14 maternal and child health-care output indicators, including many of those used in the Indonesia study. Performance was assessed as follows: facilities in the program had to submit monthly activity reports which were then audited against the facility's records. The specific payment amounts differed for each service, between US\$0.09 for an initial prenatal visit and US\$4.59 for an institutional childbirth delivery. Facilities in the control group received funding as a function of their size and the demographic characteristic of their catchment area.

Unfortunately, the experimental design was compromised somewhat before the start of the study however, due to a change in district boundaries that required that the research team switched treatment and control status for 8 districts. The final design is thus more of a quasi-experimental design, and the authors use a difference-in-difference estimation strategy to study impacts. They estimate large positive impacts on some of the targeted indicators. In particular, the incentives led to a 23% increase in the number of institutional deliveries, a 56% increase in the number of preventive care visits by children aged 0 to 2 years, and a massive 132% increase in the number of preventive care visits by older children. They also found a 0.16 standard deviations increase in prenatal quality as measured by compliance with Rwandan prenatal care clinical practice guidelines, but no change in the quantity of prenatal care sought or in rates of full immunization among children.

One of the mechanisms underlying the estimated effects appear to be an increase in health provider productivity. The researchers measured productivity as “the gap between provider knowledge and actual practice of appropriate prenatal care clinical procedures” (Gertler and Vermeersch, 2013).

This gap appears substantial in the control group: while providers know 63 percent of the appropriate clinical protocols for prenatal care on average (based on correct answers when asked), they appear to only deliver about 45 percent of the appropriate protocols. This 18 percentage point gap was reduced by 4 percentage points in the incentivized facilities. The gap is much larger at baseline among providers who have better knowledge and skills, and the impact of the pay-for-performance incentives is larger for this subgroup.

Bottom-Up: Beneficiary oversight

While monitoring coupled with incentives – whether carrots or sticks, and whether input or output based – can be successful at improving provider performance, as demonstrated in the studied surveyed above, these programs can often be costly to implement. The monitoring costs may become prohibitive in remote areas, and as discussed earlier, they often generate the problem of “who monitors the monitor?”, as well as a political backlash among the staff who are not subjected to the monitoring. For this reason, an obvious alternative would be to make the monitor the person who is the direct beneficiary of the gains to be had, namely, the patient herself. In other words, citizens, as *clients* of healthcare providers, have a direct interest in seeing their performance improve and this could translate into a willingness to expend some effort (or cost) carrying out monitoring. The difficulty here is that of monitoring ability: how do patients know if their doctor is actually making the right diagnosis or correct prescription? Health care is one of many domains – in the popular consciousness, along with auto repair, plumbing, etc. – where clients often find it difficult to evaluate the performance of the informed “expert” providing the service. For this reason, the impact of increasing monitoring by beneficiaries by itself may be limited.

The existing experimental evidence to date on this issue comes from two experiments conducted in Uganda. In the first experiment conducted in nine Ugandan districts, Bjorkman and Svensson (2009) partnered with an NGO that focused on increasing the local accountability of health providers. The experiment was conducted with 50 communities (with one facility each), with 25 treatment and 25 control communities. In the treatment communities, the NGO first created “report cards” on the quality of services at the health facility, based on information generated through facility audits as well as household interviews conducted by the researchers. A unique report card was established for each facility, and it contained (1) information on key areas subject to improvement, including utilization, access, absenteeism, and patient-clinician interaction; and (2) comparisons vis-à-vis other local health facilities, and with the national standard for primary health care provision. The report cards were written in the local vernacular languages and included graphics to help communicate the key points to non-literate residents.

The NGO then facilitated three sets of meetings: a provider staff meeting, a community meeting, and an interface meeting. The staff meeting was fairly short and consisted of sharing and

discussing the content of the report cards. The community meeting gathered around 150 community members (the NGO made sure all stakeholders were represented, including the young, elderly, and women). Participants were asked to critically review the quality of the health services available to them locally in an open discussion, and through this process the NGO disseminated the information on the report cards. Participants were then encouraged to identify concrete steps the local providers could take to improve quality, as well as actions community members could take to monitor the providers taking those steps. The discussion and proposed solutions were summarized at the end of the meeting in an action plan. The content of the action plans differed across communities, but the researchers note that high rates of absenteeism, long waiting times, weak attention by the health staff, and differential treatment across residents were common to many of the 25 communities in the treatment group. The interface meeting encouraged community members and health workers to discuss patient rights and provider responsibilities. At the end of the interface meeting, the community and the facility staff reached an agreement on the way forward. This shared action plan was called a “community contract” and it spelled out concrete steps for the provider to take and specific ways that the community members would monitor them. The NGO came back after six months to conduct two additional meetings, a community and an interface meeting, to discuss the progress to date and fine-tune the action plan.

To estimate impacts, the researchers conducted surveys one year after the first set of three meetings had taken place. They surveyed households (the same households surveyed prior to the intervention and whose information was used for the report cards), as well as the health staff. They also collected administrative records from the facilities and performed visual checks. They find large positive impacts on both the quality of care and on health outcomes. A year after the first round of meetings, health facilities in treatment communities had taken significant steps to reduce waiting times, in particular through the introduction of numbered waiting cards (20% of the treatment facilities had them, compared to only 4% of control) and a 13% reduction in absenteeism, leading to a reduction in waiting times of 12 minutes on average. This is despite the fact that utilization of general outpatient services in the local public facility was 20 percent higher in the treatment communities, with households shifting away from traditional healers and self-treatment. In particular, immunizations increased for all age groups, especially newborns, and prenatal care attendance also increased, contributing to a 0.14 z-score increase in infant weight and a remarkable 33 percent reduction in under-5 mortality. Encouragingly, these large effects persisted over time: four years after the initial intervention, researchers went back to collect new data, and found that the treatment communities still had significantly higher rates of health care utilization, better adherence to clinical guidelines by providers, and better health outcomes, including reduced child mortality and increased weight-for-age and height-for-age for children (Bjorkman-Nyqvist, de Walque and Svensson 2014).

The results of this first experiment suggest that beneficiary oversight can work, but it is notable that in this study beneficiaries were given information that they could act on, as well as, even if implicitly, information on how to stay informed (the report cards gave them concrete items to look for when monitoring). Does beneficiary control work absent this fairly costly provision of information? The second experiment, by the same researchers in the same Ugandan context, suggests that the answer is probably no (Bjorkman-Nyqvist, de Walque and Svensson 2014). In new communities, the researchers designed an intervention that mimicked everything in the first experiment except for the report cards. The new intervention is called “participation only”, in contrast to the “participation and information” treatment of the first experiment, and in this setting the staff and community meetings started without any quantitative (or even qualitative) information being provided by the NGO.

The researchers found no significant impact of this “participation only” intervention on health provision or outcomes. The authors conclude that information provision is key, and theorize that it enables clients to better distinguish between health workers’ actual effort versus factors that also matter for outcomes but are outside the health workers’ control. This information thus makes effective monitoring possible, since the client knows what to focus his monitoring efforts on, and this beneficiary monitoring is what then leads to improved health worker performance.

6.3. Improving the quality of Informal Providers

Although the experiments described above suggest that increasing the accountability of public health providers can improve their productivity, the extent to which this will affect population health outcomes depends on the “market share” of these public providers. While services such as prenatal care are rarely provided outside regulated facilities, curative primary care is commonly provided by informal private sector providers with at best minimal medical training. The large role played by poorly trained “quack” doctors has been well documented in India, in particular (see Das and Hammer, 2014, and references therein). In Sub-Saharan Africa, it is common for households to procure medication through retail sector drug shops without consulting public providers first (Cohen, Dupas and Schaner 2015). In such contexts, while increasing widespread availability of quality public care may be the goal in the long run, improving the quality of the care and/or medical advice provided by informal providers may be essential in the short run. In this section, we discuss two recent economics experiments, one in India and one in Uganda, of programs designed to improve the quality of services and products available outside the formal sector.

In West Bengal, India, Banerjee et al. (2015) estimate the impact of offering training to existing informal providers (IPs). The training program included 72 sessions of 2 hours, spread over 9 months, and was taught by certified medical doctors, but no training certificate was issued upon completion. The training course covered multiple illnesses, and emphasis was placed on basic medical conditions, triage, and avoidance of harmful practices, accompanied by frequent patient simulations. Informal providers could continue operating their clinics throughout the training since the training demanded only 4 hours of time per week. Nevertheless, take-up of the training was not universal: out of 360 providers initially asked whether they were interested in the training program, 304 (84.5%) initially signed up. Half of those 304 were then randomly assigned to start the training immediately (the treatment group). Of these 152 treatment IPs, 20 (or 13%) quit the program within three sessions, bringing take-up to just above 70%. The attendance rate over the training period was then 64% among those IPs that took up the program.

To measure impacts, the researchers used unannounced standardized patients. As described above, this method is considered the “gold standard” in care quality measurement because it does not suffer from observation and recall bias, and because it generates estimates of both the quality of the diagnosis (since illnesses are pre-specified in the study design) and of the treatment prescribed conditional on the diagnosis. In this West Bengal experiment, SP’s were trained to depict symptoms of either angina, asthma, or dysentery in a child asleep at home. These three conditions require different dimensions of care (angina requires referral, asthma requires identification of a chronic lung disease, and dysentery requires the provision of ORS) which were all supposed to be impacted by the training. The data collection through SPs started three months after the completion of training, and SPs were completed for 267 of the 304 providers in the study sample. Additional data was collected through clinical observations.

The researchers found a significant, positive impact of the training on the quality of care. Being assigned to the training group improved case-specific checklist adherence by 4.2 percentage points (on a base of 27.3% in the control group) and the likelihood of correct treatment by 7.8 percentage points (from 52% to 59.8%). Prescription of antibiotics (unnecessary in all cases) remained unchanged at very high levels (close to 50%), though interestingly, such unnecessary or even harmful practice is even more common in the public sector. Patients may thus mistakenly expect to be prescribed medicines in all cases, leading trained informal providers to continue prescribing them in order to keep their clients satisfied. Overall though, the results of this West Bengal experiment suggest that training existing IPs can improve the quality of care for rural populations with little access to care from fully qualified providers in either the public or private sector.

When existing providers are absent, or unwilling to go for training, an alternative is to encourage entry of new higher quality providers. Björkman-Nyqvist et al. (2014) evaluate the impact of market-based community health care program led by two NGOs, BRAC and Living Goods, in Uganda. In treatment villages (107 villages randomly chosen out of 204), the NGOs recruited and trained one woman per village to be a “Community Health Promoter” as well as an incentivized sales agent, conducting home visits to not only educate households on essential health behaviors but also sell preventive and curative health products at 20-30% below prevailing retail prices, with the woman earning a margin on product sales. Data collected from households three years after the rollout of the intervention in treatment villages suggests that the introduction of these trained informal providers considerably increased care seeking and resulted in a large 25% reduction in under-five mortality.

One potential channel through which the community health promoter program improved outcomes could have been by influencing other actors to improve the quality of services and products that they provided or sold. In a companion paper, the researchers document that this may have been the case with respect to drug quality (Bjorkman-Nyqvist, Svensson and Yanagizawa-Drott 2014). The quality of the antimalarials that community health promoters were allowed to sell was strictly monitored by the NGOs, and the authors argue that this reduced the likelihood that antimalarials sold at drug shops in the treatment villages were counterfeits, through a pro-competitive effect. Exploiting the randomized assignment of the program across villages, and using data on drug quality from a subset of villages, they estimate that the introduction of the community health promoter in the village led to an increase in the share of authentic artemisinin-based antimalarials sold by incumbent drug shops of 11-13 percentage points, corresponding to a roughly 50% decrease in the share of fake drugs.

The take-away from this experiment is not entirely clear-cut, however. Its results suggest that subsidizing high-quality products in the private sector improves health outcomes, but the cost of such subsidization could be prohibitive, especially if one includes the cost that the implementing NGOs had to incur for quality control purposes. Moreover, whether this subsidization requires that the promoters are paid through their sales rather than a fixed salary is unclear. The NGOs running the programs consider the implicit piece-rate pay system to be a critical feature of their model (which they call “entrepreneurial”), but the experiment was not designed to estimate the role of the incentive-pay in observed impacts. Introducing these sorts of incentive pay schemes may be more difficult in other settings, especially in the public sector health context.

If the market is sufficiently large, subsidization may not be needed, however: pharmacy chains are common in developed countries and increasingly common in developing ones, and thanks to economies of scale, it seems that they can often guarantee better quality at low prices. Bennett

and Yin (2014) discuss how chains can improve quality by purchasing in bulk from trusted manufacturers, establishing independent distribution networks, employing licensed pharmacists, and advertising to raise consumer awareness. Studying the impact of the entry of chain pharmacies in many markets of Hyderabad, India, they find that the entry of chain pharmacies selling only high quality products led to market-wide impacts on quality just like those observed in the Ugandan experiment.

7. Conclusion

We have surveyed the large and growing literature using field experiments to study issues of health and health care access, usage and impacts in low income countries. There has been a veritable explosion of research in this area: 20 years ago, there was almost no experimental research in development economics, and we are in a position today where a (very long) chapter such as this one is only able to cover a small share of the published research literature in any depth. As we have argued, research progress has been particularly pronounced in the study of the demand for health products and services, the quality of health care, and in certain aspects of the question of health impacts. However, there remain a number of glaring gaps in the literature, and promising areas for future investigation, and we briefly survey these topics in this concluding section.

One of the most important areas of inquiry is how current adult health status affects contemporaneous labor productivity. However, this is also an area where there remain relatively few well-designed field experiments (with Thomas et al 2003, 2006 being notable exceptions). One possible explanation is that relatively few field experiments on health in low income settings have taken place with private sector commercial partners, those who would be most likely to have access to a large pool of workers who might serve as participants in such a health intervention study. (As discussed above, most experiments on health in development economics have been conducted with government or NGO partners.) This remains a topic of great intellectual and public policy importance, and one where further research could have high returns.

A related limitation of existing research is the relative lack of work studying the long-run impacts of earlier health investments. While research evidence is beginning to accumulate in this area – including the long-term follow-ups to the well-known INCAP (Hoddinott et al 2008) and Kenya deworming (Baird et al 2015) cases – few studies are successfully able to combine field experimental research designs with long-term longitudinal data collection with high tracking rates. Such studies require long time horizons and prospective planning, not to mention large research budgets, and few have been successfully carried out. Putting in place the data collection plans to

follow-up the participants in the large number of recent health experiments in development economics going forward could have a large research payoff. In the absence of such studies, most of our understanding about the long-run impacts of child health status comes from studies that rely on natural experiments (such as weather variation, as in the Maccini and Yang 2009 study from Indonesia), or variation in exposure to conflict or other large-scale political shocks (as in the work of Alderman et al 2006, Bundervoet et al 2009, Leon 2012, among many other recent papers), but in these cases the interpretation of resulting impacts is arguably less transparent than in a well-designed field experiment.

Another broad area where further research would be useful is in the area of mental health, as well as the interaction between psychology and economics. While a growing number of studies, including many surveyed above, include measures of mental health, depression, anxiety and wellbeing as outcomes of particular health interventions (including in the WISE project discussed in Thomas et al 2003, 2006 study, the KLPS data used in Baird et al 2015, among others), relatively few studies by economists explore the effects of mental health interventions on economic or other life outcomes, or study the heterogeneity in the impacts of other interventions (e.g. information, cash drop) by underlying psychological status, such as stress level. This is despite the fact that there is mounting evidence that many mental health problems are widespread in both high and low income societies, and that there is a strong correlation in the cross-section between mental health and socio-economic outcomes (see the evidence in Das et al 2008 and the references therein). The growing emphasis on psychological issues in economics overall and in the study of poverty, and in particular how particular psychological or neurophysiological processes affect the decision-making of the poor (for instance, see recent work by Mani et al 2013 and Haushofer and Fehr 2014), suggests that this is an area that is also ripe for further intellectual exploration.

Beyond deepening our understanding of links between health outcomes (along various dimensions) and economic outcomes, it would also be valuable to direct more research energy to understand the large-scale health systems reforms that have occurred in many low income countries during the past decade. For instance, many less developed countries have expanded government supported health insurance programs – including in Ghana, India, Indonesia, Mexico, and Rwanda, among many other countries – but relatively little academic health economics research has examined the performance, structure, and incentives produced by these reforms, or the resulting behavioral responses by households and individuals. This is in stark contrast to the health economics literature focusing on wealthy countries like the U.S., where much research energy has focused precisely on broader institutional and organizational issues in the health sector. While there is already some important work in this area (e.g., Gertler and Vermeersch 2013, Miller

and Babiarz 2014), development economists working on health could learn much from the public finance and industrial organization economists working on related health systems issues in the U.S., Europe, and other high income regions.

Over the past two decades, as field experiments have become an increasingly common tool used by applied researchers in development economics and health (and other fields of economics), we have often looked to the methods and approaches used in clinical trials as a model on which to base our own work. That was certainly the case with much of the early experimentation in economics (as discussed in Duflo et al 2007), and has continued with the increasing use of study pre-registration and pre-analysis plans over the past few years (Miguel et al 2014). There has been extensive and productive learning across disciplines. However, the increasing sophistication of the methods used to study health issues in low income countries, and the innovations in research design and measurements described above, now make it more likely, in our view, that considerable learning will flow in the opposite direction, from economics and other social science disciplines back into medical research. This appears especially likely given the growing awareness in global health research that the study of health policies, systems and individual behavior cannot be approached in the same way as traditional medical treatment efficacy trials (for a discussion of the rise of “implementation science” in health research, see Madon et al 2007, and Mwisongo et al 2011). It is thus our hope that the discussion in this chapter will not only be of interest to development economists already working on health in low income settings, but may also prove useful to scholars in other fields or disciplines who are engaged with these issues.

References Cited

Adhvaryu, Achyuta, Namrata Kala, and Anant Nyshadham. (2014a). “Management and Shocks to Worker Productivity: Evidence from Air Pollution Exposure in an Indian Garment Factory”, unpublished working paper, University of Michigan.

Adhvaryu, Achyuta, Namrata Kala, and Anant Nyshadham. (2014b). “The Light and the Heat: Productivity Co-benefits of Energy-saving Technology”, unpublished working paper, University of Michigan.

Ahuja, Amrita, Michael Kremer, and Alix Peterson Zwane (2010). “Providing Safe Water: Evidence from Randomized Evaluations.” *Annual Review of Economics* 2:237–56.

Aiken A, Davey C, Hargreaves J, Hayes R. Re-analysis of health and educational impacts of a school-based deworming program in western Kenya: a pure replication. *Int J Epidemiol* 2015; doi:10.1093/ije/dyv127.

Alderman, H., J. Hoddinott, and B. Kinsey. (2006). “Long term consequences of early child malnutrition”, *Oxford Economic Papers*, 58(3): 450-474.

Alderman, H., J. Konde-Lule, I. Sebuliba, D. Bundy, A. Hall. (2006). “Increased weight gain in preschool children due to mass albendazole treatment given during ‘Child Health Days’ in Uganda: A cluster randomized controlled trial”, *British Medical Journal*, 333, 122-6.

Alderman, Harold. (2007). “Improving nutrition through community growth promotion: Longitudinal study of nutrition and early child development program in Uganda”, *World Development*, 35(8), 1376-1389.

Ali M, Emch M, von Seidlein L, Yunus M, Sack DA, et al. (2005) Herd immunity conferred by killed oral cholera vaccines in Bangladesh: a reanalysis. *Lancet* 366: 44–49. doi:10.1016/S0140-6736(05)66550-6.

Ali M, Emch M, Yunus M, Sack D, Lopez AL, et al. (2008) Vaccine Protection of Bangladeshi Infants and Young Children Against Cholera: Implications for Vaccine Deployment and Person-to-Person Transmission. *Pediatr Infect Dis J* 27: 33–37. doi:10.1097/INF.0b013e318149dffd.

Ali M, Sur D, You YA, Kanungo S, Sah B, et al. (2013) Herd protection by a bivalent-killed-whole-cell oral cholera vaccine in the slums of Kolkata, India. *Clin Infect Dis*: cit009. doi:10.1093/cid/cit009.

- Almond, D., and J. Currie, Human Capital Development before Age Five. NBER Working Paper #15827, (2010).
- Amarante, Veronica, Marco Manacorda, Edward Miguel, and Andrea Vigorito. (2016). “Do cash transfers improve birth outcomes? Evidence from matched vital statistics, program, and social security data”, *American Economic Journal: Economic Policy*.
- Anderson, Michael L. 2008. “Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects.” *Journal of the American Statistical Association* 103 (484): 1481–95.
- Angelucci, Manuela, and Vincenzo Di Maro. (2015). “Program evaluation and spillover effects”, unpublished work paper, University of Michigan.
- Angrist, Joshua D, and Jörn-Steffen Pischke. 2010. “The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics.” *Journal of Economic Perspectives* 24 (2): 3–30. doi:10.1257/jep.24.2.3.
- Arrow, Kenneth J. 1963. Uncertainty and the Welfare Economics of Medical Care. *American Economic Review* 53(5): 941—973.
- Ashraf, Nava, Jim Berry and Jesse Shapiro (2010). “Can Higher Prices Stimulate Product Use? Evidence from a Field Experiment in Zambia.” *American Economic Review* 100(5)
- Ashraf, N., B.K. Jack and E. Kamenica. (2013) “Information and subsidies: Complements or substitutes?” *Journal of Economic Behavior and Organization* 88: 133-139.
- Ashraf, Nava, Erica Field, and Jean Lee. 2014. "Household Bargaining and Excess Fertility: An Experimental Study in Zambia." *American Economic Review*, 104(7): 2210-37.
- Attanasio, Orazio, Camila Fernandez, Emla O A Fitzsimons, Sally M Grantham-McGregor, Costas Meghir, Marta Rubio-Codina. (2014). “Using the infrastructure of a conditional cash transfer program to deliver a scalable integrated early child development program in Colombia: cluster randomized controlled trial”, *British Medical Journal*, 349:g5785.
- Baird, S, J Hamory Hicks, and E Miguel. (2008). “Tracking, Attrition and Data Quality in the Kenyan Life Panel Survey Round 1 (KLPS-1)”, University of California CIDER Working Paper.
- Baird S, De Hoop J, Özler B (2013) Income shocks and adolescent mental health. *J Hum Resour* 48: 370–403.

Baird, Sarah, Aislinn Bohren, Craig McIntosh, and Berk Ozler. (2014). “Designing Experiments to Measure Spillover Effects”, unpublished working paper, George Washington University.

Baird, Sarah, Joan Hamory Hicks, Michael Kremer, and Edward Miguel. (2015). “Worms at Work: Long-run impacts of a child health investment”, *NBER Working Paper #21428*.

Banerjee, Abhijit, Sharon Barnhardt and Esther Duflo. 2015. Movies, Margins and Marketing: Encouraging the Adoption of Iron Fortified Salt. *Insights in the Economics of Aging*, edited by David A. Wise. NBER Book Series, Forthcoming from University of Chicago Press.

Banerjee, Abhijit, Abhijit Chowdhury, Jishnu Das, Reshmaan Hussam (2015). “The Impact of Training Informal Providers on Clinical Practice in West Bengal, India: A Randomized Controlled Trial”. Working paper.

Banerjee, A, A Deaton, and E Duflo (2004). “Health Care Delivery in Rural Rajasthan,” *Economic and Political Weekly* 39, no. 9: 944-949.

Banerjee, Abhijit V. ., Esther Duflo and Rachel Glennerster (2008). “Putting A Band-Aid On A Corpse: Incentives For Nurses In The Indian Public Health Care System”. *J Eur Econ Assoc*; 6(2-3): 487–500.

Banerjee, Abhijit , Esther Duflo, Rachel Glennerster and Dhruva Kothari (2010). “Improving Immunization Coverage in Rural India: A Clustered Randomized Controlled Evaluation of Immunization Campaigns with and without Incentives.” *BMJ* 340:c2220.

Banerjee, Abhijit, Esther Duflo, Nathanael Goldberg, Dean Karlan, Robert Osei, William Parienté, Jeremy Shapiro, Bram Thuysbaert, and Christopher Udry (2015). “A multifaceted program causes lasting progress for the very poor: Evidence from six countries”. *Science* 15 May 2015 Vol. 348 no. 6236 . DOI: 10.1126/science.1260799

Banerjee, Abhijit, Dean Karlan, and Jonathan Zinman. 2015. "Six Randomized Evaluations of Microcredit: Introduction and Further Steps." *American Economic Journal: Applied Economics*, 7(1): 1-21.

Barber, Sarah L., and Paul J. Gertler (2010). “Empowering Women: How Mexico's conditional cash transfer programme raised prenatal care quality and birth weight.” *Journal of development effectiveness* 2, no. 1 (2010): 51-73.

Basinga, Paulin, Paul J Gertler, Agnes Binagwaho, Agnes L B Soucat, Jennifer Sturdy, and Christel M J Vermeersch (2011). “Effect on Maternal and Child Health Services in Rwanda of

Payment to Primary Health-Care Providers for Performance: An Impact Evaluation.” *Lancet* 377: 1421-28.

Barham, Tania. (2011) “A healthier start: the effect of conditional cash transfers on neonatal and infant mortality in rural Mexico.” *Journal of Development Economics* 94, no. 1: 74-85.

Benhear, Lori, Alessandro Tarozzi, H B Soumya, Alex Pfaff, Kazi Matin Ahmed and Lex van Geen (2013). “Impact of a Randomized Controlled Trial in Arsenic Risk Communication on Household Water-source Choices in Bangladesh”. *Journal of Environmental Economics and Management*, 65(2), 225-240.

Behrman, Jere R., Maria Cecilia Calderon, Samuel Preston, John Hoddinott, Reynaldo Martorell and Aryeh D. Stein. (2009). “Nutritional Supplementation of Girls Influences the Growth of their Children: Prospective Study in Guatemala”, *American Journal of Clinical Nutrition*, 90 (November), 1372-1379.

Benhassine, Najy, Florencia Devoto, Esther Duflo, Pascaline Dupas and Victor Pouliquen (2015). “Turning a Shove into a Nudge: A ‘Labeled Cash Transfer’ for Education”. *American Economic Journal: Economic Policy* 7(3): 86-125.

Benjamin-Chung, Jade, Jack Colford, David Berger, Benjamin Arnold, Veronica Jimenez, Diana Tran, Ashley Clark, Eugene Konagaya, Lauren Falcao, Jaynal Abedin, Alan Hubbard, Stephen Luby, Edward Miguel. (2015). “The Identification and Measurement of Health-Related Spillovers in Impact Evaluations: A Systematic Review”, unpublished working paper.

Bennet, Daniel, and Wesley Yin (2014). “The Market for High-Quality Medicine: Retail Chain Entry and Drug Quality in India.” Mimeo, University of Chicago.

Berry, Jim, Greg Fischer and Raymond Guiteras (2012). “Eliciting and Utilizing Willingness to Pay: Evidence from Field Trials in Northern Ghana”. Mimeo, London School of Economics.

Bhattacharya, Debopam, Pascaline Dupas, and Shin Kanaya. (2013). “Estimating the Impact of Means-tested Subsidies under Treatment Externalities with Application to Anti-Malarial Bednets”, unpublished working paper, University of Oxford.

Bhushan, Indu, Erik Bloom, David Clingingsmith. Rathavuth Hong, Elizabeth King, Michael Kremer, Benjamin Loevinsohn and J. Brad Schwartz (2007). "Contracting for Health: Evidence from Cambodia". Mimeo.

Bitterly, T.B., R. Mislavsky, H. Dai, and K.L. Milkman (forthcoming). "Dueling with Desire: A Synthesis of Past Research on Want/Should Conflict." In W. Hoffman and L. Nordgren (eds.) *The Psychology of Desire*.

Bjorkman-Nykvist, M., A. Guariso, J. Svensson, D. Yanagizawa-Drott. (2014). "Evaluating the impact of the Living Goods entrepreneurial model of community health delivery in Uganda: A cluster-randomized controlled trial." Working paper.

Bjorkman, Martina, and Jakob Svensson. (2009). "Power to the People: Evidence From a Randomized Field Experiment on Community-Based Monitoring in Uganda." *The Quarterly Journal of Economics* 124(2): 735-69.

Bjorkman-Nyqvist, Martina, Jakob Svensson and David Yanagizawa-Drott (2014). "Can Good Products Drive Out Bad? Evidence from Local Markets for Antimalarial Medicine in Uganda". Working paper.

Bjorkman-Nyqvist, Martina, Damien de Walque and Jakob Svensson (2014). "Information is Power: Experimental Evidence of the Long-Run Impact of Community Based Monitoring," World Bank Policy Research Paper Series No.7015.

Bleakley, Hoyt (2007). "Disease and Development: Evidence from Hookworm Eradication in the American South." *Quarterly Journal of Economics*, 122:1, 73-117.

Bleakley, Hoyt (2010a). "Malaria Eradication in the Americas: A retrospective Analysis of Childhood Exposure", *American Economic Journal: Applied Economics*, 2(2):1-45.

Bleakley, Hoyt (2010b). Health, Human Capital, and Development. *Annual Review of Economics* 2, 283-310.

Bobonis G, Miguel E, Sharma CP. (2006). "Iron deficiency, anemia and school participation," *Journal of Human Resources*, 41(4), 692-721.

Boozer, M. and Philipson, T.J. (2000). 'The impact of public testing for immunodeficiency virus', *Journal of Human Resources*, vol. 35(3), pp. 419-46.

Buehler, R., Griffin, D., and Ross, M. (2002), "Inside the Planning Fallacy: The Causes and Consequences of Optimistic Time Predictions," In T. Gilovich, D. Griffin, and D. Kahneman (Eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment*, 250-270. Cambridge, UK: Cambridge University Press.

- Bundervoet, T. P. Verwimp, and R. Akresh. (2009). "Health and civil war in rural Burundi", *Journal of Human Resources*, 44(2): 536-563.
- Bundy, D.A.P. (1988). "Population Ecology of Intestinal Helminth Infections in Human Communities." *Philosophical Transactions of the Royal Society of London. Series B.* 321 (1207), 405-420.
- Bundy, D, Guyatt J. (1996) Schools for Health: Focus on Health, Education, and the School-Age Child. *Parasitology Today*; 12:1-16.
- Bundy, D.A.P., Chan, M-S., Medley, G.F., Jamison, D., and Savioli, L. (1998). "Intestinal Nematode Infections." In *Health Priorities and Burden of Disease Analysis: Methods and Applications from Global, National and Sub-national Studies*. Harvard University Press for the World Health Organization and the World Bank.
- Casey, Katherine, Rachel Glennerster, and Edward Miguel. 2012. "Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan." *The Quarterly journal of economics* 127(4): 1755-1812.
- Chassang, Sylvain, Gerard Padró i Miquel and Erik Snowberg. (2012). "Selective Trials: A Principal-Agent Approach to Randomized Controlled Experiments." *American Economic Review* 102(4):1279—1309.
- Chinkhumba, Jobiba, Susan Godlonton and Rebecca Thornton (2014). The Demand for Medical Male Circumcision. *American Economic Journal: Applied Economics* 2014, 6(2): 152–177.
- Chong, Alberto, Dean Karlan, Marco Gonzalez-Navarro and Martin Valdivia (2013). "Effectiveness and Spillovers of Online Sex Education: Evidence from a Randomized Evaluation in Colombian Public Schools." Working paper.
- Chong, Alberto, Isabelle Cohen, Erica Field, Eduardo Nakasone, and Maximo Torero. (2016). "Are there nutrient-base poverty traps? Evidence on iron deficiency and schooling attainment in Peru", *American Economic Journal: Applied Economics*.
- Clark, SE, et al. (2008) "Effect of intermittent preventive treatment of malaria on health and education in schoolchildren: a cluster-randomised, double-blind, placebo-controlled trial", *Lancet*, 372(9633): 127-138.
- Clasen, Thomas, Sophie Boisson, Parimita Routray, Belen Torondel, Melissa Bell, Oliver Cumming, Jeroen Ensink, et al. 2014. Effectiveness of a rural sanitation programme on

diarrhoea, soil-transmitted helminth infection, and child malnutrition in Odisha, India: a cluster-randomised trial. *The Lancet Global Health* no. 2:e645-e653.

Cochrane, Susan H. 1979. "Fertility and Education: What Do We Really Know?" Washington, D.C.: World Bank.

Cohen, J. and P. Dupas (2010). "Free Distribution or Cost-Sharing? Evidence from a randomized malaria experiment." *Quarterly Journal of Economics* 125: 1-45.

Cohen, J., P. Dupas and S. Schaner (2015). "Price Subsidies, Diagnostic Tests, and Targeting of Malaria Treatment." *American Economic Review* 105(2): 609-645.

Cohen, Jessica, Günther Fink, Kathleen Maloney, Katrina Berg, Matthew Jordan, Theodore Svoronos, Flavia Aber & William Dickens (2015). "Introducing rapid diagnostic tests for malaria to drug shops in Uganda: a cluster-randomized controlled trial". *Bulletin of the World Health Organization*, Volume 93, Number 3, 133-208.

Cooper E, Fitch L (1983) Pertussis: Herd Immunity and Vaccination Coverage in St. Lucia. *The Lancet* 322: 1129–1132. doi:10.1016/S0140-6736(83)90637-2.

Cox DR (1958) *Planning of experiments*. Oxford, England: Wiley. 308 p.

Croke, Kevin. (2014). "The long run effects of early childhood deworming on literacy and numeracy: Evidence from Uganda", unpublished working paper, Harvard University.

Currie, Janet. (2000). "Child health in developed countries,," in *Handbook of Health Economics*, Vol. 1B, ed. AJ Culyer, JP Newhouse, pp. 1053-90. Amsterdam: North-Holland.

Currie, Janet, Wanchuan Lin, Wei Zhang (2011). Patient Knowledge and Antibiotic Abuse: Evidence from an Audit Study in China. *Journal of Health Economics* 30, 933-949

Currie, Janet, Wanchuan Lin and JuanJuan Meng (2014). "Addressing Antibiotic Abuse in China: An Experimental Audit Study," *Journal of Development Economics* 110: 39-51.

Cutler, David, and Adirana Lleras-Muney (2014). Education and Health: Insights from International Comparisons. in *Encyclopedia of Health Economics*, Anthony J. Culyer (ed.), Vol 1. San Diego: Elsevier; 2014. pp. 232-245.

Cutler, David M., and Grant Miller (2005), "The Role of Public Health Improvements in Health Advances: The Twentieth-Century United States," *Demography* 42, no. 1: 1-22.

Das, J., J. Hammer and K. Leonard (2008). “The Quality of Medical Advice in Low-Income Countries” *Journal of Economic Perspectives*, 22(2): 93-114. 2008.

Das, J., Q.T. Do, J. Friedman, and D. McKenzie. (2008). “Mental health patterns and consequences: Results from survey data in five developing countries”, World Bank Policy Research Working Paper #4495.

Das, J., and J. Hammer (2014). “Quality of Primary Care in Low-Income Countries: Facts and Economics”. *Annual Review of Economics* Vol. 6: 525–553.

Das, Jishnu, Alaka Holla, Veena Das, Manoj Mohanan, Diana Tabak, and Brian Chan. 2012. “The Quality of Medical Care in Clinics: Evidence from a Standardized Patients Study in a Low-Income Setting.” *Health Affairs*, 31(12): 2274–2784.

Das, Jishnu, Alaka Holla, Aakash Mohpal and Karthik Muralidharan (2016). Quality and Accountability in Health: Audit Evidence from Primary Care Clinics in India. *American Economic Review*.

Davey C, Aiken A, Hayes R, Hargreaves J. Re-analysis of health and educational impacts of a school-based deworming programme in western Kenya: a statistical replication of a cluster quasi-randomized stepped-wedge trial., *Int J Epidemiol* 2015; doi:10.1093/ije/dyv128.

Deaton, Angus (2013). *The Great Escape: Health, Wealth and the Origins of Inequality*. Princeton University Press.

Delavande, Adeline, and Hans-Peter Kohler (2012). “The impact of HIV Testing on Subjective Expectations and Risky Behavior in Malawi,” *Demography*, 49, 3:1011-36.

DellaVigna, Stefano, and Ulrike Malmendier (2006). “Paying Not To Go To The Gym”, *American Economic Review*, Vol. 96, pp. 694-719.

Dewald, William G., Jerry G. Thursby, and Richard G. Anderson. 1986. “Replication in Empirical Economics: The Journal of Money, Credit and Banking Project.” *The American Economic Review* 76 (4): 587–603.

DeLong, J. Bradford, and Kevin Lang. 1992. “Are All Economic Hypotheses False?” *Journal of Political Economy* 100 (6): 1257–72.

Dhaliwal, Iqbal and Rema Hanna (2013). “Deal with the Devil: The Successes and Limitations of Bureaucratic Reform in India”. Working paper.

Dickson R, Awasthi S, Williamson P, Demellweek C, Garner P. Effect of Treatment for Intestinal Helminth Infection on Growth and Cognitive Performance in Children: Systematic Review of Randomized Trials”. *British Medical Journal*; 320, 1697-1701, 2000.

Dillon, Andree, Jed Friedman and Peter Serneels (2014). “Health information, treatment, and worker productivity”. World Bank Policy Research working paper No. WPS 7120.

Dizon-Ross, R., P. Dupas and J. Robinson (2016). “Governance and Effectiveness of Health Subsidies”. NBER working paper 21324.

Droitcour, Judith, Rachel A. Caspar, Michael L. Hubbard, and Trena M. Ezzati. (1991) “The Item Count Technique as a Method of Indirect Questioning: a Review of its Development and a Case Study Application,” in: Beimer, P.B., Groves, R.M., Lyberg L.E., Mathiowetz N.A., Sudman S. (Eds.), *Measurement Errors in Surveys*. John Wiley & Sons, Inc., Hoboken, New Jersey, pp. 185-211.

Duflo, Esther, Pascaline Dupas and Michael Kremer (2015). “Education, HIV and Early Fertility: Experimental Evidence from Kenya”. *American Economic Review*, Vol. 105(9), pp. 2257-97.

Duflo, Esther, Pascaline Dupas and Michael Kremer (2016). “Does Secondary Education Transform Young Adults' Lives? Experimental Evidence from Ghana”. Mimeo, Stanford University.

Duflo, Esther, Rachel Glennerster, and Michael Kremer. 2007. “Using randomization in development economics research: A toolkit.” *Handbook of development economics* 4 (2007): 3895-3962.

Duflo, Esther, Michael Greenstone, Raymond Guiteras, and Thomas Clasen (2015). “Toilets *Can* Work: Short And Medium Run Health Impacts Of Addressing Complementarities And Externalities In Water And Sanitation.” Mimeo.

Dupas, P. (2009). “What matters (and what does not) in household’s decision to invest in malaria prevention?”, *American Economic Review* 99(2): 224-230.

Dupas, P. (2011a). “Health Behavior in Developing Countries”, *Annual Review of Economics*, 3: 425-449.

Dupas, P. (2011b). “Do Teenagers Respond to HIV Risk Information? Evidence from a Field Experiment in Kenya”, *American Economic Journal: Applied Economics*, 3 (1), 1-36.

- Dupas, P. (2014a). “Short-Run Subsidies and Long-Run Adoption of New Health Products: Evidence from a Field Experiment”. *Econometrica* 82(1), pp. 197-28
- Dupas, Pascaline (2014b). “Getting Essential Health Products to Their End Users: Subsidize, but How Much?” *Science* Vol. 345, Issue 6202, pp. 1279-1281, 12 September 2014
- Dupas, P. V. Hoffmann, M. Kremer and A. Zwane. (2016). “Targeting health subsidies through a non-price mechanism: A Randomized Controlled Trial in Kenya”. Forthcoming, *Science* 26 August 2016.
- Dupas, Pascaline, and Jonathan Robinson (2013). “Why Don’t the Poor Save More? Evidence from Health Savings Experiments”. *American Economic Review* 103(4): 1138-71.
- Dybvig PH, Spatt CS (1983) Adoption externalities as public goods. *J Public Econ* 20: 231–247.
- Eble, Alex, Peter Boone, and Diana Elbourne. (2015). “On minimizing the risk of bias in randomized controlled trials in Economics”, unpublished working paper, Brown University.
- Epidemiology*. 2010. “The Registration of Observational Studies—When Metaphors Go Bad:” July, 1. doi:10.1097/EDE.0b013e3181eafbcf.
- Fafchamps, Marcel, David McKenzie, Simon Quinn and Chris Woodruff (2014). “Microenterprise growth and the flypaper effect: Evidence from a randomized experiment in Ghana”. *Journal of development Economics* 106, 211-226.
- Fine PE (1993) Herd immunity: history, theory, practice. *Epidemiol Rev* 15: 265–302.
- Fink, Günther and Felix Masiye (2012). “Assessing the impact of scaling-up bednet coverage through agricultural loan programmes: evidence from a cluster randomised controlled trial in Katete, Zambia”. *Trans R Soc Trop Med Hyg* 106 (11): 660-667. doi: 10.1016/j.trstmh.2012.07.01
- Fink, G. & Masiye, F. (2015). Health and Agricultural Productivity: Evidence from Zambia. *Journal of Health Economics*. Volume 42, July 2015, Pages 151–164
- Finkelstein, Amy, et al. "The Oregon Health Insurance Experiment: Evidence from the First Year." 2012. *The Quarterly journal of economics* 127(3): 1057-1106.
- Fitzsimons E, Malde B, Mesnard A, Vera-Hernández M (2012) Household responses to information on child nutrition: experimental evidence from Malawi. Available: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2034133. Accessed 22 April 2014.

Forleo-Neto E, Oliveira CF de, Maluf EMCP, Bataglin C, Araujo JMR, et al. (1999) Decreased Point Prevalence of Haemophilus influenzae Type b (Hib) Oropharyngeal Colonization by Mass Immunization of Brazilian Children Less Than 5 Years Old with Hib Polyribosylribitol Phosphate Polysaccharide—Tetanus Toxoid Conjugate Vaccine in Combination with Diphtheria-Tetanus Toxoids—Pertussis Vaccine. *J Infect Dis* 180: 1153–1158. doi:10.1086/315018.

Fox, Matthew, et al. (2004). “The Impact of HIV/AIDS on Labour Productivity in Kenya”, *Tropical Medicine and International Health*, 9(3), 318-324.

Franco, Annie, Neil Malhotra, and Gabor Simonovits. 2014. “Publication Bias in the Social Sciences: Unlocking the File Drawer.” *Science* 345 (6203): 1502–5.

Galiani, Sebastian, Paul Gertler, and Ernesto Schargrodsy. 2005. Water for Life: The Impact of the Privatization of Water Services on Child Mortality. *Journal of Political Economy* no. 113 (1):83-120.

Gertler, Paul, and Christel Vermeersch (2013). “Using Performance Incentives to Improve Health Outcomes”. NBER WP 19046.

Gertler, Paul, et al. (2014). “Labor market returns to an early childhood stimulation intervention in Jamaica”, *Science*, 344: 998-1001.

Giné, Xavier, Dean Karlan, and Jonathan Zinman. 2010. "Put Your Money Where Your Butt Is: A Commitment Contract for Smoking Cessation." *American Economic Journal: Applied Economics*, 2(4): 213-35.

Glewwe, Paul, and Edward Miguel. (2008). “The Impact of Child Health and Nutrition on Education in Less Developed Countries”, *Handbook of Development Economics Volume 4*, (eds.) T. Paul Schultz and John Strauss, Elsevier.

Godlonton, Susan, Alister Munthali and Rebecca Thornton (2014) “Responding to Risk: Circumcision, Information, and HIV Prevention.” Forthcoming in the *Review of Economics and Statistics*.

Godlonton S, Thornton R (2012) Peer effects in learning HIV results. *J Dev Econ* 97: 118–129.

Gong, Erick. (2015). “HIV Testing and Risky Sexual Behavior”, *Economic Journal*, 125: 32-60.

Grossman, Michael. On the Concept of Health Capital and the Demand for Health. *Journal of Political Economy* 80, 223-255 (1972).

Guiteras, Raymond, James Levinsohn and Ahmed Mushfiq Mobarak (2015). “Encouraging sanitation investment in the developing world: A cluster-randomized trial”. *Science* 348, 903-906.

Habyarimana, James and William Jack (2011) Heckle and chide: Results from a randomized road safety intervention in Kenya. *J Public Econ* 95(2011):1438–1446.

Habyarimana, James and William Jack (2015). Results of a large-scale randomized behavior change intervention on road safety in Kenya. *Proceedings of the National Academy of Sciences* 4661–4670

Haushofer, Johannes, and Ernst Fehr. (2014). “On the psychology of poverty”, *Science*, 344, 862-867.

Haushofer, Johannes, and Jeremy Shapiro (2013). “Household Response to Income Changes: Evidence from an Unconditional Cash Transfer Program in Kenya”. Mimeo, Princeton University.

Hicks, Joan Hamory, Michael Kremer, and Edward Miguel. “Commentary: Deworming externalities and schooling impacts in Kenya: A comment on Aiken et al (2015) and Davey et al (2015)”, *Int J Epidemiol* 2015; doi:10.1093/ije/dyv129.

Hoddinott, John, John A. Maluccio, Jere R Behrman, Rafael Flores and Reynaldo Martorell. (2008). “The Impact of Nutrition During Early Childhood on Income, Hours Worked, and Wages of Guatemalan Adults”, *Lancet*, 371 (February), 411-416.

Hoffmann, Vivian (2009). “Intrahousehold Allocation of Free and Purchased Mosquito Nets”, *American Economic Review* 99(2): 236-41.

Jayachandran, Seema (2015). “The Roots of Gender Inequality in Developing Countries”. *Annual Review of Economics* Vol. 7 (2015): 63-88.

Jamison, Julian, Dean Karlan and Pia Raffler (2013). Mixed Method Evaluation of a Passive mHealth Sexual Information Texting Service in Uganda (May 2013) *Information Technologies & International Development*, 9(3).

Jensen, Robert, and Adriana Lleras-Muney (2012). “Does Staying in School (and not Working) prevent Teen Drinking and Smoking?” *Journal of Health Economics*, 31(4): 644-657.

Joshi S, Schultz TP (2013) Family Planning and Women’s and Children’s Health: Long-Term Consequences of an Outreach Program in Matlab, Bangladesh. *Demography* 50: 149–180.

- Karlan, Dean, and Jonathan Zinman. (2011). "List randomization for sensitive behavior: An application for measuring use of loan proceeds", *NBER Working Paper #17475*.
- Karlan, Dean, Greg Fischer, Maggie McConnell and Pia Raffler (2014) "To Charge or Not to Charge: Evidence from a Health Products Experiment in Uganda". Mimeo, Yale University.
- Kremer, M, and R Glennerster. (2011). "Improving health in developing countries: Evidence from randomized evaluations", *Handbook of Health Economics*, vol. 2, (eds.) Mark V Pauly, Thomas G. McGuire, Pedro P. Barros, pp. 201-315.
- Kremer, M. and E Miguel (2007). "The Illusion of Sustainability." *Quarterly Journal of Economics* 122 (3): 1007-1065.
- Kremer, Michael, Jessica Leino, Edward Miguel and Alix Peterson Zwane, 2011a. "Spring Cleaning: Rural Water Impacts, Valuation, and Property Rights Institutions," *Quarterly Journal of Economics*, 126(1), 145-205.
- Kremer, Michael, Edward Miguel, Sendhil Mullainathan, Clair Null and Alix Zwane (2011b). Social Engineering: Evidence from a Suite of Take-up Experiments in Kenya. Mimeo, Emory University.
- Kvalsig JD, Cooppan RM, Connolly KJ. The effects of parasite infections on cognitive processes in children. *Ann Trop Med Parasitol* 1991;73:501-6.
- LaBrie, Joseph W., and Mitchell Earleywine. 2000. "Sexual Risk Behaviors and Alcohol: Higher Base Rates Revealed Using the Unmatched-Count Technique. *Journal of Sex Research* 37:321-26.
- Laine, Christine, Richard Horton, Catherine D. DeAngelis, Jeffrey M. Drazen, Frank A. Frizelle, Fiona Godlee, Charlotte Haug, et al. 2007. "Clinical Trial Registration — Looking Back and Moving Ahead." *New England Journal of Medicine* 356 (26): 2734-36.
- LaLonde, Robert J. 1986. "Evaluating the Econometric Evaluations of Training Programs with Experimental Data." *The American Economic Review* 76 (4): 604-20.
- Leamer, Edward E. 1983. "Let's Take the Con Out of Econometrics." *American Economic Review* 73 (1): 31-43.
- Leon, Gianmarco. (2012). "Civil conflict and human capital accumulation: The long-term effects of political violence in Peru", *Journal of Human Resources*, 47(4): 991-1022.

- Ma, Xiaochen, Sean Sylvia, Matthew Boswell, and Scott Rozelle (2014). “Ordeal Mechanisms and Training in the Provision of Subsidized Products in Developing Countries.” Mimeo, Stanford University.
- Maccini, Sharon, and Dean Yang. 2009. “Under the Weather: Health, Schooling, and Economic Consequences of Early-Life Rainfall.” *American Economic Review*, 99(3): 1006-26.
- Madajewicz, M., A. Pfaß, A. van Geen, J. Graziano, I. Hussein, H. Momotaj, R. Sylvi, and H. Ahsan (2007). “Can information alone change behavior? Response to arsenic contamination of groundwater in Bangladesh”. *Journal of Development Economics* 84 (2): 731-754.
- Madon, T., K.J. Hofman, L. Kupfer, and R.I. Glass. (2007). “Implementation science”, *Science*, 318 (5857): 1728-1729.
- Maluccio, John A., John Hoddinott, Jere R. Behrman, Reynaldo Martorell, Agnes Quisumbing, Aryeh D Stein. (2009). “The Impact of Improving Nutrition During Early Childhood on Education among Guatemalan Adults.” *Economic Journal*, 199 (537) 734-763.
- Mani, Anandi, Sendhil Mullainathan, Eldar Shafir, and Jiaying Zhao. (2013). “Poverty Impedes Cognitive Function,” *Science*, 341, 976-980.
- Martorell R, Habicht JP, Rivera JA. History and Design of the INCAP Longitudinal Study (1969-1977) and its Follow-up (1988-89). *Journal of Nutrition*; 125: 1027S-1041S, 1995.
- Mathieu, Sylvain, Isabelle Boutron, David Moher, Douglas G. Altman, and Philippe Ravaut. (2009). “Comparison of registered and published primary outcomes in randomized controlled trials”, *JAMA*, Sept. 2, 2009, 302(9), 977-984.
- McEwan, P. J. (2014). “Improving learning in primary schools of developing countries: A meta-analysis of randomized experiments.” Forthcoming, *Review of Educational Research*.
- Meredith et al. (2013). “Keeping the doctor away: Experimental Evidence on Investment in Preventive Health Products.” *Journal of Development Economics* 105: 196-210.
- Miguel, Edward, and Michael Kremer (2004). “Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities”, *Econometrica*.
- Miguel, Edward, and Michael Kremer (2014). “Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities, Guide to Replication of Miguel and Kremer (2004)”, *CEGA Working Paper #39*.

- Miguel, E., C. Camerer, K. Casey, J. Cohen, K. M. Esterling, A. Gerber, R. Glennerster, et al. 2014. "Promoting Transparency in Social Science Research." *Science* 343 (6166): 30–31. doi:10.1126/science.1245317.
- Milkman, K.L., J. Beshears, J.J. Choi, D. Laibson, and B.C. Madrian (2011). "Using Implementation Intentions Prompts to Enhance Influenza Vaccination Rates." *Proceedings of the National Academy of Sciences*, Vol. 108, 10415-10420.
- Milkman, K.L., J. Beshears, J.J. Choi, D. Laibson, and B.C. Madrian (2013). "Planning Prompts as a Means of Increasing Preventive Screening Rates." *Preventive Medicine*, Vol. 56, 92-93.
- Miller, G., K. S. Babiarz, Pay-for-performance incentives in low- and middle-income country health programs. In *Encyclopedia of Health Economics*, Vol. 2, A. J. Culyer, Ed. (Elsevier, San Diego, CA, 2014), pp. 457–483.
- Muralidharan, Karthik, and Venkatesh Sundararaman. (2011). "Teacher Performance Pay: Experimental Evidence from India", *Journal of Political Economy*, 119(1): 39-77.
- Muralidharan, Karthik, and Venkatesh Sundararaman. (2015). "The Aggregate Effect of School Choice: Evidence from a Two-Stage Experiment in India", *Quarterly Journal of Economics*, 130(3): 1011-1066.
- Mwisongo, A, L Wang, T Madon, S Owusu-Agyei and M^A González Block. (2011) Current and foreseeable themes in implementation research for disease control. Chapter 9, Implementation research for the control of infectious diseases of poverty. Geneva: World Health Organization.
- Nayyar, G., J. G. Breman, P. N. Newton, and J. Herrington, 2012, "Poor-quality Antimalarial Drugs in Southeast Asia and Sub-Saharan Africa", *The Lancet Infectious Diseases*, 12(6):488-496.
- Neumark, David. 2001. "The Employment Effects of Minimum Wages: Evidence from a Prespecified Research Design The Employment Effects of Minimum Wages." *Industrial Relations: A Journal of Economy and Society* 40 (1): 121–44.
- Nokes C, Grantham-McGregor S, Sawyer A, Cooper E, Bundy D. Parasitic helminth infection and cognitive function in school children. *Proceedings: Biological Sciences*, 247(1319):77–81. 1992.

Nokes C, van den Bosch C, Bundy D. The Effects of Iron Deficiency and Anemia on Mental and Motor Performance, Educational Achievement, and Behavior in Children: A Report of the International Nutritional Anemia Consultative Group. USAID: Washington DC. 1998.

Olken, Benjamin, Junko Onishi and Susan Wong (2014). Should Aid Reward Performance? Evidence From a Field Experiment on Health and Education in Indonesia. *American Economic Journal: Applied Economics* 6(4): 1-34.

Ozier, Owen. (2014). “Exploiting Externalities to Estimate the Long-Term Effects of Early Childhood Deworming”, World Bank Policy Research Working Paper #7052.

Patil, Sumeet R., Benjamin F. Arnold, Alicia L. Salvatore, Bertha Briceno, Sandipan Ganguly, John M. Colford, Jr, and Paul J. Gertler. 2014. The Effect of India's Total Sanitation Campaign on Defecation Behaviors and Child Health in Rural Madhya Pradesh: A Cluster Randomized Controlled Trial. *PLoS Med* no. 11:e1001709.

Paul JR, Horstmann DM, Riordan JT, Opton EM, Niederman JC, et al. (1962) An oral poliovirus vaccine trial in Costa Rica. *Bull World Health Organ* 26: 311–329.

Peletz, R., Cock-Esteb, A., Ysenburg, D., Haji, S., Khush, R., Dupas, P. (2016). The supply and demand of improved sanitation: results from randomized willingness-to-pay experiments in rural Tanzania. In preparation.

Pitt, M.M., M. R. Rosenzweig, N. Hassan, Human Capital Investment and the Gender Division of Labor in a Brawn-Based Economy. *American Economic Review* 102, 3531–3560 (2012).

Pollitt E, Hathirat P, Kotchabhakadi N, Missel L, Valyasevi A. Iron deficiency and education achievement in Thailand. *Am J Clin Nutr* 1989(3);50:687–97.

Pollitt E, Gorman K, Engle P, Martorell R, Rivera J. Early supplemental feeding and cognition. *Monographs of the Society for Research in Child Development*, Serial No. 235, Chicago: University of Chicago Press. 1993.

Pop-Eleches C, Thirumurthy H, Habyarmina J, Graff Zivin J, Goldstein M, DeWalque D, MacKeen L, Haberer J, Sidle J, Ngare D, Bangsberg D. Mobile phone technologies improve adherence to antiretroviral treatment in resource-limited settings: a randomized controlled trial of text message reminders. *AIDS* 2011; 25(6):825-834.

Prina, Silvia, and Heather Royer (2014). “The Importance of Parental Knowledge and Social Norms: Evidence from Weight Report Cards in Mexico”. Forthcoming, *Journal of Health Economics*.

- Robinson, Jon, and Ethan Yeh. (2011). "Transactional Sex as a Response to Risk in Western Kenya," *American Economic Journal: Applied Economics* 3 (1): 35-64.
- Robinson, Jon, and Ethan Yeh. (2012). "Risk-Coping through Sexual Networks: Evidence from Client Transfers in Kenya," *Journal of Human Resources* 47 (1): 107-145.
- Rosenbaum PR (2007) Interference Between Units in Randomized Experiments. *J Am Stat Assoc* 102: 191-200. doi:10.1198/016214506000001112.
- Rosenthal, Robert. 1979. "The File Drawer Problem and Tolerance for Null Results." *Psychological Bulletin* 86 (3): 638-41. doi:10.1037/0033-2909.86.3.638.
- Rubin DB (1990) Formal mode of statistical inference for causal effects. *J Stat Plan Inference* 25: 279-292. doi:10.1016/0378-3758(90)90077-8.
- Seshadri S, Gopaldas T. Impact of iron supplementation on cognitive functions in preschool and school-aged children: the Indian experience. *Am J Clin Nutr* 1989;50(3):675-86.
- Schulz, Kenneth F., and David A. Grimes. 2005. "Multiplicity in randomised trials II: subgroup and interim analyses." *The Lancet* 365.9471: 1657-1661.
- Simonsohn, Uri, Leif D. Nelson, and Joseph P. Simmons. 2014. "P-Curve: A Key to the File-Drawer." *Journal of Experimental Psychology: General* 143 (2): 534-47.
- Singal, Amit G, Peter D R Higgins, and Akbar K Waljee. (2014). "A Primer on Effectiveness and Efficacy Trials", *Clinical and Translational Gastroenterology*, 5, e45; doi:10.1038/ctg.2013.13
- Soemantri AG, Pollitt E, Kim I. Iron deficiency anemia and education achievement. *Am J Clin Nutr* 1989; 50(3):698-702.
- Soewondo S, Husaini M, Pollitt E. Effects of iron deficiency on attention and learning processes of preschool children: Bandung, Indonesia. *Am J Clin Nutr* 1989;50(3):667-74.
- Strauss, John and Duncan Thomas. 1995. "Human resources: empirical modeling of household and family decisions," in J. R. Behrman and T. N. Srinivasan (eds.), *Handbook of development economics*, volume 3A, Amsterdam: North Holland Press.
- Sylvia Sean, et al. (2015). "Survey using incognito standardized patients shows poor quality care in China's rural clinics." *Health Policy and Planning* 30(3):322-33.

Tarozzi, Alessandro, Aprajit Mahajan, Brian Blackburn, Dan Kopf, Lakshmi Krishnan, Joanne Yoong (2014). “Micro-loans, Insecticide-Treated Bednets and Malaria: Evidence from a Randomized Controlled Trial in Orissa (India)”. *American Economic Review*. 104(7):1909-41 .

Taylor-Robinson DC, Maayan, N, Soares-Weiser, K. Garner, P. (2012) “Deworming drugs for treating soil-transmitted intestinal worms in children: effects on nutrition and school performance.” Cochrane Database of Systematic Reviews.

The Lancet. 2010. “Should Protocols for Observational Research Be Registered?” 375 (9712): 348. doi:10.1016/S0140-6736(10)60148-1.

The Voluntary HIV-1 Counseling and Testing Efficacy Study Group (2000). Efficacy of voluntary HIV-1 counselling and testing in individuals and couples in Kenya, Tanzania, and Trinidad: a randomised trial. *The Lancet*, Volume 356, No. 9224, p103–112, DOI: [http://dx.doi.org/10.1016/S0140-6736\(00\)02446-6](http://dx.doi.org/10.1016/S0140-6736(00)02446-6).

Thirumurthy H, Goldstein M, Graff Zivin J. “The economic impact of AIDS treatment: labor supply in western Kenya.” *Journal of Human Resources*, 2008; 43:511-552.

Thomas, Duncan, et al. (2003). “Iron Deficiency and the Well-Being of Older Adults: Early Results from a Randomized Nutrition Intervention”, unpublished manuscript, UCLA.

Thomas, Duncan, et al. (2006). “Causal Effect of Health on Labor Market Outcomes: Experimental Evidence”, unpublished manuscript, UCLA.

Thornton, Rebecca (2008). The Demand for, and Impact of, Learning HIV Status. *American Economic Review*, 98(5): 1829–1863, 2008.

Vermeersch C, Kremer M. School Meals, Educational Achievement and School Competition: Evidence from a Randomized Evaluation. Unpublished working paper, World Bank and Harvard University.

World Bank (2008). *Are you being served?: New tools for measuring service delivery*, edited by Samia Amin, Jishnu Das, Markus Goldstein. The World Bank, Washington, D.C. DOI: 10.1596/978-0-8213-7185-5.

World Bank (2015). World Development Report 2015: Mind, Society and Behavior. The World Bank, Washington, D.C.

World Health Organization. (1993). The Control of Schistosomiasis. Second Report of the WHO Expert Committee. (Technical Report Series 830). WHO, Geneva.

World Health Organization (2015). Road traffic injuries. Fact sheet N°358. October 2015. Last accessed 21 December 2015 at <http://www.who.int/mediacentre/factsheets/fs358/en/>.

Ziegelhöfer Z (2012) Down with diarrhea: Using fuzzy regression discontinuity design to link communal water supply with health. Graduate Institute of International and Development Studies Working Paper. Available: <http://www.econstor.eu/handle/10419/77433>. Accessed 30 July 2014.

Zivin JG, Thirumurthy H, Goldstein M (2009) AIDS treatment and intrahousehold resource allocation: Children's nutrition and schooling in Kenya. *J Public Econ* 93: 1008–1015.

Zwane, Alix Peterson, Jonathan Zinman, Eric Van Dusen, William Pariente, Clair Null, Edward Miguel, Michael Kremer, Dean Karlan, Richard Hornbeck, Xavier Giné, Esther Duflo, Florencia Devoto, Bruno Crepon, Abhijit Banerjee. (2011). "Being surveyed can change later behavior and related parameter estimates", *Proceedings of the National Academy of Sciences*.