How Were the Reaching the Poor Studies Done?

Adam Wagstaff and Hugh Waters

All of the studies in this volume are concerned with one broad question: how well do specific health programs reach the poor? Yet they vary in a number of respects—in the countries and types of program studied, of course, but also in the methods used. In part, this reflects variation in the exact question asked, but there are also methodological differences among studies that ask the same question. This chapter provides a tour of the methodologies used in the volume and explores the meaning of some of the terms used.

Snapshots, Movies, and Experiments

Although all the studies look at the broad issue of how well programs reach the poor, they take different approaches, which can be divided into three general categories. The first group uses the snapshot approach and asks, for example: How do rates of utilization of services delivered by the program differ between the poor and the less poor? Are the users of the program services drawn disproportionately from the poor? Such questions are clearly of interest, as many programs are designed explicitly in the hope that the poor will be the primary beneficiaries, and evidence that the use of program services is higher among the less poor should cause policy makers to worry. The snapshot approach is the least demanding in terms of data; data are needed for only one point in time, to capture the current distribution of utilization by the poor and the less poor.
The second group of studies asks: has inequality between the poor and the less poor in the use of services increased or decreased over time? This approach lets the film roll for two or more periods and then halts and reviews the trend over time. It may find that in each “frame,” contrary to what the program’s designers had hoped, the poor are not the heaviest users of program services. But as the frames roll forward, it may turn out that the inequality, or gradient, narrows over time, and knowing this would be some consolation, even though there is scope for improvement. This movie approach is more data demanding than the first. Now data are needed at two points in time, capturing the distribution of utilization by income (or some other measure of living standards) as it was and as it is.

There remains a third question: is the distribution of service use more equal (or less unequal) under the program than it would have been without the program? This begs the question of the counterfactual—what the world would have been like in the absence of the program. For example, the research may be examining a large program financed and delivered by the government, and one extreme counterfactual would be what the supply of and demand for health care would have been under a pure private market. Or the program under study may be run by a nongovernmental organization (NGO), and a plausible counterfactual would be what the government would have done in the absence of the NGO. In this approach, the researcher takes a snapshot of the actual situation and then disappears into the art studio to paint a picture of what he or she thinks the distribution would have looked like without the program. The two are then compared.

This third approach—the snapshot coupled with an artist’s impression of an unknown world—is the most data demanding of the three. Data are now needed on the extant distribution and on the distribution in the counterfactual. Of course, this second distribution is never observed, so the researcher must not only be explicit about the scenario he or she has in mind but must also have some way of generating data that approximate the chosen scenario. There are various nonexperimental ways this can be done, but in this volume the tool used to generate the counterfactual data is the experiment. The data are collected before the experiment starts and again after it has been running for a while.

In the studies reported on here, the intervention groups and the control groups have been assigned on the basis of geography; some areas are under the program, and some are not. Ideally, this assignment would be random, and some chapters (for example, chapter 8, on the Cambodian health service reform) do report the results of a randomized experiment. More often than not, however, the assignment is not random. Evaluators are generally
measuring the effects of the programs on participants only and do not know what effect the program would have had on nonparticipants. Because nonparticipants differ from participants in ways that might influence the outcome, it cannot be assumed that they would have been affected in the same manner or to the same extent.

Biases may result. Suppose the intervention and control groups differ in certain key factors that influence the outcomes being measured. Insofar as these are observable to the researcher, methods such as regression analysis and propensity score matching can be used to reduce selection on observables. This leaves the possibility that there may be selection on unobservables. Individuals who self-select into a program may have unobservable characteristics related to preferences or health status that make them more likely than others to join the program and also influence their use of health services or other positive outcomes (Waters 1999). But by focusing on differences in changes over time between the treatment and control groups, researchers can reduce this bias to the extent that the unobservables in question stay constant over time.

Table 2.1 shows how the three categories apply to the chapters in this volume. Most chapters are "snapshots" of a single time period in which equity is measured. The second-largest group consists of "experiments" that attempt to measure the results of various interventions, while two of the chapters are "movies" that examine changes in equity over time.

Targeting, Leakage, and Benefit Incidence

Before turning to the practicalities of undertaking studies in this field, some additional terms and concepts need to be introduced.

Let us suppose we are assessing a program designed to reach the poor. The administrators of such a program would be happy if the poor were indeed benefiting from the program and if, as intended, the nonpoor were not benefiting from it. The term leakage is used to describe instances where the nonpoor benefit from the program (a type II error); a "leaky" program is a poorly targeted program. The administrator would also want to know about people who are supposed to be beneficiaries but are not being reached. The term undercoverage describes instances where, contrary to the program’s goal, the poor do not benefit from the program (a type I error).

Figure 2.1 illustrates the hypothetical case of a fee waiver program intended only for the poor. People in the top left-hand cell are poor and, correctly, receive the waiver. People in the top right-hand cell are also poor but do not receive the waiver (undercoverage). People in the bottom left-hand
These concepts can also be applied to an analysis of program beneficiaries. Leakage would then mean the fraction of program beneficiaries who are nonpoor, and undercoverage would mean the fraction of the poor who do not benefit from the program. Where all program beneficiaries benefit equally from a program, presenting the data along these lines makes sense.

Chapter 13, by Barros and colleagues, provides an example of this methodology. Among the four health and nutrition programs in Brazil that were evaluated is one targeted to very poor families (the Pastorate of the Child) and another that was initially implemented through a targeted approach but was designed to expand over time (the Family Health Program). The chapter begins by testing the hypothesis that considerable leakage and undercoverage exist in both programs. It then takes up the more interesting issue of what structural, programmatic, or social factors may be behind the inequitable distributions observed. Utilization and coverage data are presented across wealth quintiles. The implications of the analysis are particularly important for the wholly targeted program, since its coverage takes in a relatively wide wealth distribution even though its objective is to focus on children who are either undernourished or from extremely impoverished families.

The analysis by Valdivia in chapter 14 evaluates the targeting success of several nutritional programs in Peru. Using data from the Living Standards

---

**Figure 2.1. Leakage and Undercoverage in Targeting in a Fee Waiver Program**

<table>
<thead>
<tr>
<th>Income group</th>
<th>Waiver granted</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor</td>
<td>Yes</td>
<td>Correct</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>Type I error: undercoverage</td>
</tr>
<tr>
<td>Rich</td>
<td>Type II error: leakage</td>
<td>Correct</td>
</tr>
</tbody>
</table>

- cell, although not poor, nonetheless are granted the waiver (leakage). People in the bottom right-hand cell are not poor and do not receive the waiver—the correct outcome.

---
## Table 2.1. Questions Asked by the Studies Reported in This Volume

<table>
<thead>
<tr>
<th>Study type</th>
<th>Question</th>
<th>Chapter and topic</th>
</tr>
</thead>
</table>
| 1. Snapshot | How unequal is the distribution of service use between the poor and the less poor? | 5. Reproductive health training program for private medical providers in Kenya  
6. HIV counseling and testing services in public clinics in South Africa  
9. Utilization of health services provided by the Self-Employed Women’s Association (SEWA) in India  
13. Two universal and two targeted health programs in Brazil  
14. Targeted early childhood nutritional programs in Peru |
| 2. Movie    | Has inequality between the poor and the less poor in the use of services increased or decreased over time? | 7. Utilization of maternal health services across socioeconomic status groups within a rural community in Bangladesh between 1997 and 2001  
12. Changes in the targeting of health and nutrition programs in Argentina between 1997 and 2001 |
| 3. Experiment | Is the distribution of service use more equal (or less unequal) under the program than it would have been without the program? | 4. Integrated program of vaccination and bednet distribution in Ghana and Zambia  
8. Comparison of primary health services in Cambodia delivered through contractors with those delivered directly by the government  
10. Utilization and patient satisfaction related to a multifaceted health reform initiative in Uttar Pradesh State, India  
11. Participatory approach to reproductive health among rural and urban adolescents in Nepal |
Measurement Surveys (LSMSs), Valdivia initially shows the proportions of beneficiaries who are within and outside eligibility thresholds. He then moves on to a valuable examination of marginal incidence to explore whether expansion of the programs would produce equity results on the margin similar to those produced on average. Two of the initiatives evaluated—a school breakfast program and the Vaso de Leche (Glass of Milk) program—showed favorable pro-poor results at the margin despite suboptimal targeting on average, suggesting that even where initiatives have significant leakage (with between 40 and 50 percent of beneficiaries outside the target group), their expansion may lead to disproportionately pro-poor results.

But what if “the program” consists of the government’s entire spending on the health sector? Can we really analyze the extent to which the program reaches the poor by using the concept of a program beneficiary? Is someone who uses a government clinic just once a year as much a beneficiary as someone who makes extensive use of primary care, outpatient, and inpatient facilities? We could get a highly distorted picture if the poor, or the nonpoor, make much more extensive use of the program’s services than does the other group.

The idea in benefit-incidence analysis (BIA) is to calculate the benefits associated with the program and see how they are distributed across the population, paying particular attention to the distribution between the poor and the better off. Benefits, usually expressed in monetary terms, are based on records of service utilization. For example, if “the program” in question is the government’s entire spending program for the health sector, the study would ask how many primary care visits, how many hospital outpatient visits, and how many hospital inpatient days each individual or household had in the period covered by the study. Each of these would then be converted to a monetary amount by multiplying the number of visits or days by the amount of government spending or subsidy involved. This may vary from one individual to the next—for example, poorer individuals may be exempt from fees at government facilities, so their subsidy per visit is larger than the subsidy to better-off people who have to pay toward the cost of the visit.

Armed with a measure of benefit accruing to each person, we can compute leakage in terms of program benefits, and this would indicate the share of program benefits accruing to nonpoor people. BIA studies often present the complement of this number—the share of program benefits going to the poor, sometimes known as the benefit-incidence ratio. This is a direct measure of targeting success. (Leakage is in a sense an indirect measure because a higher number indicates worse targeting performance.)
In chapter 12, Gasparini and Panadeiros employ BIA as part of their broader examination of individuals who use publicly financed health and nutrition programs, but rather than convert benefits into monetary terms, they integrate their findings into a gradient that expresses the degree of inequality. It is to this technique that we now turn.

**Gradients and Inequalities**

Sometimes the focus is not so much on juxtaposing the experiences of the poor and the nonpoor as on examining a gradient. It could be a gradient in health outcomes, or in service utilization, or in the benefits from government subsidies, and it could span the distribution of income or wealth or some other measure of living standards. But the common concern is with a gradient that captures inequality.

Let us suppose we have a living standards measure. (We take up the question of how to obtain such a measure in the next section.) We rank households according to the measure, starting with the poorest, and divide the sample into equal groups—say, five equal groups, or quintiles. Our interest is then in the gradient in the given health indicator across the five quintiles at a particular time, or in changes in the gradient over time, or in differences in the gradient between the actual situation and the situation under the hypothetical counterfactual.

Looking at gradients at a moment in time is fairly straightforward. Figure 2.2 shows rates of underweight across four income groups, using data from Brazil.

**Figure 2.2. Changes in the Distribution of Underweight Children, Ceará, Brazil**

![Graph showing changes in underweight rates across income groups from 1987 to 1994.]

Source: Based on Victora and others (2000).
from Ceará, Brazil, for 1987 and 1994. In both years there is an appreciable income gradient in the probability of child underweight, with poorer children suffering substantially higher rates of undernutrition. This much is easy to see. We could produce similar charts for health outcomes such as malnutrition, or for measures of utilization such as visits to hospital or full immunization of a child. When the variable is a measure of service utilization and the gradient is downward-sloping, the program is said to be pro-poor.

The analysis of South African voluntary counseling and testing (VCT) programs described in chapter 6 by Thiede, Palmer, and Mbatsha is an example of this approach. The authors present the distribution of VCT clinic utilization across wealth quintiles. Other studies in this volume that use these types of graphic representation include chapter 10, which analyzes health improvements in India, and chapter 4, which evaluates an integrated bednet and vaccination campaign.

Looking at figure 2.2, it is easy to say that in both years a gradient in malnutrition favored the better off. It is harder to say whether the gradient became steeper or shallower between 1987 and 1994. A similar problem might arise when comparing the distribution of utilization under a particular program with utilization in a counterfactual case. One device that makes it easier to answer such questions is the concentration curve. In figure 2.3 the x-axis shows the cumulative percentage of the sample (children, in this case) ranked by per capita income, wealth, or whatever measure of living standards is being used, starting with the poorest. On the y-axis is plotted the cumulative percentage of whatever variable is being investigated (in this case, underweight), corresponding to the cumulative percentage of the sample. So, if the poorest 20 percent of children accounts for 30 percent of all underweight children, the ordinate for the y-axis is 30 percent. If the variable being investigated has higher values among the poor (as it does in this example), the resultant curve—the concentration curve—will lie above the 45 degree line. The latter is known as the line of equality because this is the shape the concentration curve would take if everyone had the same value of the indicator whose distribution we are investigating.

If the health indicator being considered is an undesirable outcome such as being malnourished, a concentration curve above the line of equality is a bad thing from the point of view of equity. If the health indicator is a good outcome, then, arguably, from the point of view of equity we would want the concentration curve to lie above the line of equality. The farther from the line of equality the concentration curve is, the worse things are from an equity standpoint. So, figure 2.3 provides an answer to the question of...
whether inequality in undernutrition in Ceará had decreased or increased; the answer is that it had increased.

Chapter 9 by Ranson and colleagues provides an example of this approach. The authors examine a health program of the Self-Employed Women’s Association (SEWA) in India to assess whether services are in fact utilized by the poor. Frequency distributions of utilization by socioeconomic status are presented, and concentration curves are used to assess equity of utilization of the services offered by SEWA. Rather than display changes across time, the concentration curves in this study allow for comparisons of equity between SEWA’s mobile reproductive health units, tuberculosis detection and treatment services, and women’s education services in urban and rural communities.

Comparing two concentration curves is straightforward, but comparing multiple concentration curves is a little hard on the eyes. The curves may also cross, making comparison difficult. The concentration index is a useful

Figure 2.3. Concentration Curves Showing Changes in the Distribution of Underweight Children, Ceará, Brazil

Source: Authors’ calculations based on Victora and others (2000).
way of reducing the strain on the eye, and it acts as a tiebreaker in the event of intersecting concentration curves. The index, quite simply, is twice the area between the curve and the line of equality. The convention is to use a minus sign in front of the index when the curve is above the line of equality and a plus sign when it is below the line of equality. Of course, whether a negative concentration index is good or bad from an equity standpoint depends on whether the variable in question is something good, such as receiving health care when needed, or something bad, such as being malnourished.

The analysis in chapter 9, for example, presents concentration indexes for each of the SEWA program’s health services, using the minus sign convention. All three services had concentration curves of utilization above the line of equality, yielding indexes in the range of –0.33 to –0.37. In chapter 8, Schwartz and Bhushan use concentration indexes to analyze the level of inequality for selected health service indicators in Cambodia. Here, the authors are primarily interested in whether contracting out health services affects inequality.

**Measuring Living Standards: Pitfalls for the Unwary**

To obtain numbers for leakage and undercoverage and to graph a gradient and a concentration curve, we first need to know how to measure the variables underlying these concepts. For most readers, the measurement of living standards will be the big unknown (by contrast, measuring health service utilization, health outcomes, and so on will be fairly familiar), and so it is on this that we focus. Four approaches to measurement are possible, using income, expenditure, consumption, or a wealth proxy.

The *income approach* is, on the face of it, straightforward and appealing. Many surveys ask respondents a simple question along the lines of how much was your household income in the past year? Respondents might be asked to place their income in a bracket rather than report the exact amount. Both practical and conceptual issues arise. Will people respond truthfully, especially if they think their answers may leak back to the authorities? Will respondents be able to recall accurately all household members’ incomes, from all sources, for the past year? Should the question be asked in stages, focusing first on labor income and then on unearned income, including transfers? How should in-kind income such as gifts of food from neighbors be treated? Should income be measured before or after taxes? Should expenses associated with running the family business be deducted from income to derive net income, and if so, how are data on such expenses to be
collected? How should the expenses associated with a piece of equipment such as a tractor be handled? The information collected to date on household income in developing countries is typically of poor quality, not least because of these problems.

Chapter 12, by Gasparini and Panadeiros, provides an example of a useful income approach to measurement when neither expenditure nor consumption data are available. The authors utilize two large living standards surveys to obtain data on utilization of targeted health and nutrition programs in Argentina. Although these surveys do not provide insight into expenditure, the authors estimated household welfare using household income, adjusting for household size through an “equivalence” scale. Utilization rates for the programs were then analyzed across the distribution of this income measure.

It is not clear, conceptually, whether income is the best measure of living standards. Income simply tells us how much money is coming into the household; it may not give an accurate picture of the household’s living standard. Take the case of pensioners. In the developing world, any pension is probably small, and so even if people have pension income, they may be using up their savings in their old age.

An alternative that gets around this problem is to look at expenditure. Many surveys ask about household spending patterns and aggregate up to a total spending figure for the household. But here too there are technical and conceptual issues. Over what period should expenditure be measured? A week? A month? A year? Should the period vary according to type of purchase? How should “lumpy” purchases, such as a car or a television set, be handled?

A big problem with the expenditure approach is that many households in the developing world grow a great deal of the food they consume, and food can make up a large fraction of total consumption. A household may look poor from an expenditure standpoint because it gets its food from its own plot. Yet it may enjoy a reasonable standard of living. Another problem is that subsidies may give a distorted picture of living standards. For example, one household may be able to live rent free because the head of household is a doctor at the village clinic or a state-enterprise worker. Judged by expenditure, the household would appear poor because it pays no rent, but that would be the wrong inference.

For these and other reasons, the World Bank’s Living Standards Measurement Study (LSMS) decided early on to measure living standards in terms of consumption instead of income or expenditure (Deaton and Grosh 2000; Deaton and Zaidi 2002). A household’s food consumption is defined
as the sum of its own produce and any produce it buys from others or is
given by others. A similar principle applies to other items of consumption;
if, for example, the household has a small business producing handicrafts
and keeps some for its own use, these products are counted as consump-
tion, and the proceeds from sales of the remaining products show up in the
household’s consumption of other items (for example, school uniforms
bought from the proceeds). The consumption approach does not measure
expenditures on consumer durables; rather, it attributes to each durable a
use value that depends on the purchase price and the expected life of the
durable. It does not look at the amount of rent a family pays for its accom-
modations but at the imputed rent—the amount it would have had to pay
had it rented its accommodations at the market rate. Everyone, whether liv-
ing in a rent-free government house, a home the person built and owns, or a
rented house, is assigned a positive imputed rent.

The consumption approach pioneered by the LSMS overcomes many of
the objections that can be leveled at the income and expenditure approaches
to measuring living standards. But it has a major drawback: it is complex and
time consuming. LSMS consumption questions typically stretch over many
pages of a household survey questionnaire. This is fine if assessing living
standards is the main goal of the exercise, but it is a little cumbersome if that is
only one of the aims. It is perhaps partly for this reason that other large-scale
household survey exercises such as the Demographic and Health Survey
(DHS) have eschewed the consumption approach. In fact, the DHS, some-
what surprisingly, has eschewed all approaches to measuring living stan-
dards, leaving researchers in the frustrating position of having excellent
household data on maternal and child health variables but no income, expen-
diture, or consumption data that would, for example, allow them to compare
immunization rates among the poor with those among the better off.

This situation led researchers to look at the nonhealth information col-
clected in the DHS and ask whether it could be used to construct an ad hoc
proxy measure of living standards. They concluded that it could (Filmer
and Pritchett 1999, 2001). A long list of variables was assembled from the
DHS, covering attributes of the household’s dwelling (type of floor; materi-
als used for the roof and floor), water and sanitation facilities enjoyed by the
household (piped water in the house, piped water in the yard, or water from
a pump), and ownership of various household durables (radio, television
set, bicycle, or car). The information allows researchers to construct a
weighted sum of these indicators. Clearly, an infinite number of weighting
schemes could be applied. For example, should a car be given twice or three
times the weight of a bicycle?
To select the weights, researchers typically employ principal component analysis (PCA). In extracting the first principal component, the weights are selected in such a way that no other weighted sum of the living standards indicators has a larger variance. The same is done for the second principal component, except that this second linear combination must be completely uncorrelated with the first component just extracted. The same procedure is followed for the third and fourth principal components, and so on.

When PCA is used to generate a proxy living standards measure, researchers retain just the first principal component, and this becomes the proxy measure of wealth or living standards. So, for example, when the method is applied to DHS data for Indonesia, possession of a bicycle gets a weight of 0.0285, and possession of a car receives a weight of 0.07262, or 2.5 times that of the bicycle (see Gwatkin and others 2000). The weights that emerge from the PCA exercise vary from one country to the next and one year to the next, reflecting circumstances and customs, as well as the number and type of indicators available in the DHS or in whatever survey is being used.

The PCA approach to the measurement of living standards is the one most often used in the chapters in this volume. Some of the studies utilized actual DHS data to which the PCA approach was applied, while others computed living standards measured by applying the PCA approach to data sets other than those produced by the DHS. The studies of voluntary HIV/AIDS counseling and testing in South Africa (chapter 6) and of private sector reproductive health services in Kenya (chapter 5) are examples of the use of DHS information.

Given the specificity of the programs evaluated in this volume, the second approach—use of non-DHS data sets—is far more common. In chapter 13 the authors measure inequality by applying PCA to both the Brazil DHS and data from regional surveys to construct an asset measure. In chapter 8 Schwartz and Bhushan construct a household wealth index, using PCA applied to baseline and follow-up surveys in Cambodia. The study of reproductive health among Nepalese adolescents in chapter 11 applies the PCA approach to baseline and endline surveys of urban and rural communities where a targeted, community-based, participatory health program had been initiated for disadvantaged youths. The resulting asset measures were used to assess inequalities in the effectiveness of the youth programs. Chapter 10 utilizes the same PCA asset approach to construct a socioeconomic status indicator from a family health survey in Uttar Pradesh State, India, and chapter 7 applies PCA to 1996 census data collected from a rural area of the Ganges-Meghna Delta in Bangladesh. Given the diversity of data and
research questions addressed in these chapters, the PCA approach to living standards measurement can be a very robust tool when using non-health-related data.

This ad hoc approach to living standards measurement—in contrast to the income, expenditure, and consumption approaches—does not produce a cardinal measure of living standards, let alone a monetary one. That is, it simply yields a normalized score with zero mean and a variance of one, not a number with a dollar (or other monetary) sign in front of it. This approach cannot be used to determine whether people are poor in the sense that they live on less than a dollar a day. Nor can we make comparisons across countries; we cannot say that a household with a score of –0.75 living in Indonesia is poorer than one with a score of –0.25 living in India. What we can do is to rank households within a country. We can say that within a specific sample a household with a score of –0.75 is poorer, according to this measure, than one with a score of –0.25. From the point of view of capturing inequalities in health outcomes between poor and less-poor people, the approach seems to work reasonably well.7

Where Do the Data Come From?

One issue remains: where do we get the data to operationalize these ideas? It would be a shame, after all, to have come this far and not be able to put the ideas into effect. Two approaches are possible: we could rely on nationally representative household survey data, or we could collect new data and, if necessary, use nationally representative survey data to put this new information in context.

Many studies in this field make use of routine national surveys. Examples include chapters 10 and 13, which draw on India’s National Family Health Survey and the Brazil DHS, respectively. Such surveys contain the necessary information on use of services and living standards, allowing the requisite tables and graphs to be made. Table 2.2 displays the data sources for the studies in this volume.

Researchers who are able to make use of routine surveys are in luck—no data-collection costs are incurred, and the focus can be entirely on data analysis. The study of reproductive health in chapter 7, for example, relies solely on census data, birth records, and routine maternal health care utilization data. When, however, the objective is to assess a specific program, especially one that operates only at the local level, a routine national survey may be of little use. It may not include enough—or even any—households in the program area, and it may not ask about use of the program’s services. Then, a different strategy is called for.
Table 2.2. Data Sources Used by the Studies, by Chapter

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Data sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>4. Ghana and Zambia: Achieving Equity in the Distribution of Insecticide-Treated Bednets through Links with Measles Vaccination Campaigns</td>
<td>Exit interviews at vaccination/bednet sites; follow-up community-based household surveys in both countries</td>
</tr>
<tr>
<td>5. Kenya: Reaching the Poor through the Private Sector—A Network Model for Expanding Access to Reproductive Health Services</td>
<td>Survey of health providers; exit interviews of individuals at both case and control sites; household interviews of women of reproductive age; Demographic and Health Survey (DHS) data for socioeconomic status measurement</td>
</tr>
<tr>
<td>6. South Africa: Who Goes to the Public Sector for Voluntary HIV/AIDS Counseling and Testing?</td>
<td>Survey of individuals at voluntary counseling and testing (VCT) sites; in-depth interviews and focus groups; South Africa DHS</td>
</tr>
<tr>
<td>7. Bangladesh: Inequalities in Utilization of Maternal Health Care Services—Evidence from Matlab</td>
<td>Local census data; birth records; maternal health service utilization database</td>
</tr>
<tr>
<td>8. Cambodia: Using Contracting to Reduce Inequity in Primary Health Care Delivery</td>
<td>Baseline (1997) and follow-up (2001) household surveys within case and control health districts</td>
</tr>
<tr>
<td>9. India: Assessing the Reach of Three SEWA Health Services among the Poor</td>
<td>Exit surveys of individuals at Self-Employed Women’s Association (SEWA) clinics; focus group of nonusers; in-depth interviews of health workers; DHS data for socioeconomic status measurement</td>
</tr>
<tr>
<td>10. India: Equity Effects of Quality Improvements on Health Service Utilization and Patient Satisfaction in Uttar Pradesh State</td>
<td>Baseline (1999) and follow-up (2003) exit interviews of outpatients (for assessment of patient satisfaction); Uttar Pradesh data from the National Family Health Survey for socioeconomic status measurement</td>
</tr>
<tr>
<td>11. Nepal: The Distributional Impact of Participatory Approaches on Reproductive Health for Disadvantaged Youths</td>
<td>Small-scale baseline (1999) and endline (2003) surveys of households and adolescents within two rural and two urban areas</td>
</tr>
<tr>
<td>13. Brazil: Are Health and Nutrition Programs Reaching the Neediest?</td>
<td>Brazil DHS; three surveys of specific populations residing in the states of Santa Catarina, Sergipe, and Rio Grande do Sul</td>
</tr>
<tr>
<td>14. Peru: Is Identifying the Poor the Main Problem in Reaching Them with Nutritional Programs?</td>
<td>Living Standards Measurement Survey (LSMS)</td>
</tr>
</tbody>
</table>
In such a case, the program administrators could organize an exit survey of users and ask them about their dwelling type, their ownership of household durables, the water supply they use, and so on. As table 2.2 shows, this proved a useful tactic in several studies, since the specific experiences and characteristics of program users were important. Armed with these data, the administrator could apply the PCA weights from a national survey (for example, the country’s DHS) and generate for each user of the program’s facilities a score on a proxy wealth index. From the national survey, the administrator will know the cutoff points on the wealth index that separate the poorest quintile from the second poorest, the second poorest from the middle quintile, and so on. She can then place each user in the national wealth distribution and see what fraction of the program’s users comes from the country’s poorest quintile, what fraction from the second-poorest quintile, and so on. If, say, 65 percent comes from the country’s poorest quintile and the rest from the second poorest quintile, the administrator should probably feel fairly satisfied that her program is indeed reaching the poor.

Or should she? Perhaps the administrator would want to check how the population her program serves compares with the national population. If 75 percent of the local population is in the country’s poorest 20 percent, the fact that just 65 percent of her users comes from the nation’s poorest quintile would be something of a disappointment. To find out whether this is so, the administrator would have to obtain data on the local population—not just users of her program’s facilities but nonusers as well. The more local is the focus of the program, the harder this is likely to be. Suppose the program operates across an entire state of a large country. The administrator might be lucky and find a national survey that was conducted in her state and, moreover, is representative at the state level. She could then sort the state’s population into state wealth quintiles and locate her users in the state’s wealth distribution rather than the nationwide wealth distribution. But if the program is more tightly focused or there is no representative survey at the state level, she has little option but to conduct a household survey of her own, in addition to or instead of the user exit survey. The household survey would allow her to collect data from all sampled households on wealth indicators, use the national PCA-based weights to generate a wealth score for each sampled household, and form wealth quintiles for the local population. Using the cutoff points for this local distribution, she will be able to answer more satisfactorily the question of whether she is indeed reaching the poor in her locality.

Chapters 4, 8, and 11 are interesting examples of studies that grapple with these methodological issues. The study of Cambodia in chapter 8 uses
baseline and follow-up surveys that were specific to the areas served by the contracting program being evaluated. Chapter 11, which has a tightly focused analysis, uses community-based surveys. Although the limited size of these surveys poses some analytical limitations, their use allows for more comprehensive qualitative analysis. In chapter 4 the household surveys carried out to follow up on the integrated program of vaccination and bednet distribution allow the authors to search for discrepancies between ownership and use of the bednets.

Conclusions

In designing and conducting program evaluations, researchers need to weigh the benefits and disadvantages of these different approaches. As this chapter makes clear, measurement of success in reaching the poor involves a series of trade-offs.

One relates to program design. Estimates of program distribution made at a single point in time (the snapshot approach) require just one set of data but are subject to potential bias related to changes over time and to the selection of participants. Data collection at more than one point in time (the movie approach) and estimation of the potential effect of the program on nonparticipants (the experiment approach) require more complex data collection and analytical techniques.

Another trade-off is in the measurement of living standards or economic status. Consumption is a more accurate indicator than income or expenditure, but it is more time consuming to construct and entails more complex data collection. Principal component analysis requires only data on household assets. It provides a proxy for living standards, but one that is relative only, with no information on absolute levels of well-being. As more information from household surveys becomes available, researchers can take advantage of these data to conduct increasingly accurate and insightful research that can inform policy aimed at reaching the poor with important health benefits.

In evaluating the relative equity of programs making special efforts to reach the poor, it is important to not lose sight of the overall value of the benefit—and the extra cost of the special efforts themselves. Indicators such as the benefit-incidence ratio and the concentration index measure relative distribution only. They do not take into account the absolute value of the benefit being transferred, the administrative cost of the transfer, or the opportunity cost of the resources used to finance the benefit and its transfer.
The financing source of the benefit is also important. Public funds have an opportunity cost, and the success or failure of programs should be viewed in this light. An appropriate question would be whether the results are more or less pro-poor than the results from alternative programs that might have been funded.

In brief, there is no such thing as a single ideal approach. This is no cause for despair; it simply reflects the world’s complexity, as attested by the studies in this volume. Many methods now exist that provide highly illuminating assessments of how well health programs serve the disadvantaged groups we most want to reach. They need to be used more often.

We promised a tour of the numerous methodological issues that chapter authors have had to grapple with in preparing their Reaching the Poor studies. The tour has probably seemed a little dry—like being given a lecture on the key decisions in wine making without getting close to any wine. It is hard in such a situation to gain a sense of how any of it matters. Perhaps the best thing to do is to bookmark this chapter and refer back to it while reading the rest of the volume. We hope that, as with drinking wine, knowing what you are looking for will lead to greater appreciation.

Notes


3. Just how far above the line of equality the concentration curve should be depends on the distribution of need. The more heavily concentrated is need among the poor, the farther above the line of equality we would want the concentration curve to be. Much of the literature in the field—although not in this volume—juxtaposes distributions of use and need in an attempt to assess health equity (Wagstaff, Paci, and van Doorslaer 1991; Wagstaff, van Doorslaer, and Paci 1991; van Doorslaer and others 1992, 2000; Wagstaff and van Doorslaer 2000).


5. For a guide to living standards measurement in the context of health equity, including the use of principal components, see http://siteresources.worldbank.org/INTPAH/Resources/Publications/Quantitative-Techniques/health_eq_tn04.pdf.
6. Further details on the LSMS may be found at http://www.worldbank.org/lsms/.

7. For example, the concentration indexes for child malnutrition are quite similar across 20 countries, irrespective of whether the children are ranked by consumption or by a PCA-based wealth index.

References


