Simple Poverty Scorecard® Poverty-Assessment Tool Yemen

Shiyuan Chen and Mark Schreiner

22 May 2009

This document is at `SimplePovertyScorecard.com`

Abstract

The Simple Poverty Scorecard-brand poverty-assessment tool uses 10 low-cost indicators from Yemen's 2005/6 Household Budget Survey to estimate the likelihood that a household has consumption below a given poverty line. Field workers can collect responses in about ten minutes. The scorecard's accuracy is reported for a range of poverty lines. The scorecard is a practical way for pro-poor programs in Yemen to measure poverty rates, to track changes in poverty rates over time, and to segment clients for differentiated treatment.

Acknowledgements

This paper was funded by the Ford Foundation via a grant to the Grameen Foundation. Data are from Yemen's Central Statistics Organization. Thanks go to Frank DeGiovanni, Syed Hashemi, Brendan Janet, Tony Sheldon, Srinivasan Thirumalai, and Jeff Toohig. "Simple Poverty Scorecard" is a Registered Trademark of Microfinance Risk Management, L.L.C. for its brand of poverty-assessment tools.

Simple Poverty Scorecard® Poverty-Assessment Tool

			Name	Identifier
Interview ID:				
Interview date:		Participant:		
Country:	YEM	Field agent:		
Scorecard:	001	Service point:		
Sampling wgt.:		Number of household members:		

Indicator	Value	Points	Score
1. How many members does the household have?	A. Ten or more	0	
	B. Eight or nine	8	
	C. Seven	11	
	D. Six	18	
	E. Four or five	21	
	F. Three	26	
	G. One or two	38	
2. How many household members aged 12 to 18 currently attend school?	A. Not all	0	
	B. All	2	
	C. No children aged 12 to 18	8	
3. What is the main material used for the floor?	A. Concrete, mud/soil, stone, or other	0	
	B. Floor tiles or marble	15	
4. What is the main material used for the ceiling?	A. Wood and mud, wood, hay, mud and hay, metal sheets and mud, or other	0	
	B. Reinforced concrete, wood and concrete, or metal sheets	4	
5. What type of toilet does the household have?	A. Non-flush toilet, other, or no toilet	0	
	B. Flush toilet	5	
6. How many rooms are there in the house, excluding bathrooms and kitchens?	A. One	0	
	B. Two	2	
	C. Three	3	
	D. Four	5	
	E. Five	7	
	F. Six or more	13	
7. Does the household or any of its members own a TV?	A. No	0	
	B. Yes	6	
8. Does the household or any of its members own a gas cylinder?	A. No	0	
	B. Yes	4	
9. Does the household or any of its members own a radio/cassette recorder?	A. No	0	
	B. Yes	2	
10. Does the household or any of its members own a washing machine?	A. No	0	
	B. Yes	5	

Score:

Simple Poverty Scorecard® Poverty-Assessment Tool Yemen

1. Introduction

Pro-poor programs in Yemen can use the Simple Poverty Scorecard poverty-assessment tool to estimate the likelihood that a household has consumption below a given poverty line, to estimate a population's poverty rate at a point in time, to track changes in a population's poverty rate over time, and to segment participants for differentiated treatment.

The direct approach to poverty measurement via surveys is difficult and costly, asking households about a lengthy list of consumption items (such as "How much wheat did you buy from the market in the past week? How much corn? . . .")

In contrast, the indirect approach via the scorecard is simple, quick, and inexpensive. It uses ten verifiable indicators (such as "What is the main material used for the ceiling?" or "Does the household or any of its members own a TV?") to get a score that is highly correlated with poverty status as measured by the exhaustive survey.

The scorecard differs from "proxy means tests" (Coady, Grosh, and Hoddinott, 2002) in that it is tailored to the capabilities and purposes not of national governments but rather of local, pro-poor organizations. The feasible poverty-measurement options for these organizations are typically subjective and relative (such as participatory

wealth ranking by skilled field workers) or blunt (such as rules based on land-ownership or housing quality). Measurements from these approaches are not comparable across organizations nor across countries, they may be costly, and their accuracy and precision are unknown.

Suppose, for example, that an organization wants to know what share of its participants are below a poverty line (say, USD1.25/day at 2005 purchase-power parity for the Millennium Development Goals, or the poorest half of people below the national poverty line as required of USAID microenterprise partners). Or suppose an organization wants to measure movement across a poverty line through time (for example, to report to the Microcredit Summit Campaign). In these cases, what is needed is a consumption-based, objective tool with known accuracy. While consumption surveys are costly even for governments, many small, local organizations can implement an inexpensive poverty-assessment tool that can serve for monitoring, management, and targeting.

The statistical approach here aims to be understood by non-specialists. After all, if managers are to adopt the scorecard on their own and apply it to inform their decisions, they must first trust that it works. Transparency and simplicity build trust. Getting "buy-in" matters; proxy means tests and regressions on the "determinants of poverty" have been around for three decades, but they are rarely used to inform decisions, not because they do not work, but because they are presented (when they are presented at all) as tables of regression coefficients incomprehensible to lay people (with

cryptic indicator names such as "LGHHSZ_2", negative values, and many decimal places). Thanks to the predictive-modeling phenomenon known as the "flat max", simple scorecards are usually about accurate as complex ones.

The technical approach here is also innovative in how it associates scores with poverty likelihoods, in the extent of its accuracy tests, and in how it derives formulas for standard errors. Although these techniques are simple, they have rarely or never been applied to poverty-assessment tools.

The scorecard is based on the 2005/6 Household Budget Survey (HBS) conducted by Yemen's Central Statistics Organization (CSO). Indicators are selected to be:

- Inexpensive to collect, easy to answer quickly, and simple to verify
- Strongly correlated with poverty
- Liable to change over time as poverty status changes

All points in the scorecard are non-negative integers, and total scores range from 0 (most likely below a poverty line) to 100 (least likely below a poverty line). Non-specialists can collect data and tally scores on paper in the field in five to ten minutes.

The scorecard can be used to estimate three basic quantities. First, it can estimate a particular household's "poverty likelihood", that is, the probability that the household has per-capita consumption below a given poverty line.

Second, the scorecard can be used to estimate the poverty rate of a group of households at a point in time. This is simply the average poverty likelihood among the households in the group.

Third, the scorecard can be used to estimate changes in the poverty rate for a group of households (or for two independent representative samples of households from the same population) between two points in time. This estimate is the change in the average poverty likelihood of the group(s) of households over time.

The scorecard can also be used for targeting. To help managers choose the most appropriate targeting cut-off for their purposes, this paper reports several measures of targeting accuracy for a range of possible cut-offs.

This paper presents a single scorecard whose indicators and points are derived from household consumption data and Yemen's national poverty line. Scores from this one scorecard are calibrated to poverty likelihoods for seven poverty lines.

The scorecard is constructed and calibrated using two sub-samples of the data from the 2005/6 HBS, and its accuracy is validated on another sub-sample.

While all three scoring estimators are unbiased when applied to the population from which they were derived (that is, they match the true value on average in repeated samples from the same population from which the scorecard was built), they are—like all predictive models—biased to some extent when applied to a different population.[1]

Thus, while the indirect scoring approach is less costly than the direct survey approach, it is also biased. (The survey approach is unbiased by assumption.) There is bias because scoring must assume that the future relationship between indicators and

[1] In the context of the scorecard, examples of "different populations" include nationally representative samples at a different point in time or non-nationally representative sub-groups (Tarozzi and Deaton, 2009).

poverty will be the same as in the data used to build the scorecard.[2] Of course, this assumption—ubiquitous and inevitable in predictive modeling—holds only partly.

When applied to the validation sample with bootstrap samples of $n = 16{,}384$, the difference between scorecard estimates of groups' poverty rates and the true rates at a point in time is +1.5 percentage points for the national line, and the average absolute difference is 1.0 percentage points across all seven lines. These differences are due to sampling variation and not bias; the average of each difference would be zero if the whole 2005/6 HBS were to be repeatedly redrawn and divided into sub-samples before repeating the entire process of building and calibrating scorecards.

The 90-percent confidence intervals for these estimates are ±0.7 percentage points or less for estimates of a poverty rate at a point in time. For $n = 1{,}024$, the 90-percent intervals are ±2.7 percentage points or less.

Section 2 below describes data and poverty lines. Section 3 places the new scorecard here in the context of existing exercises for Yemen. Sections 4 and 5 describe scorecard construction and offer guidelines for use in practice. Sections 6 and 7 detail the estimation of households' poverty likelihoods and of groups' poverty rates at a point in time. Section 8 discusses estimating changes in poverty rates through time. Section 9 covers targeting. The final section is a summary.

[2] Bias may also result from changes in the quality of data collection, from changes over time to poverty lines, from imperfect adjustment of poverty lines to account for differences in cost-of-living across time or geographic regions, or from sampling variation across consumption surveys.

2. Data and poverty lines

This section discusses the data used to construct and test the scorecard. It also presents the poverty lines to which scores are calibrated.

2.1 Data

The scorecard is based on data from the 13,136 households in the 2005/6 HBS conducted April 2005 to March 2006. This is Yemen's most recent available national consumption survey.

For the purposes of the scorecard, the households in the 2005/6 HBS are randomly divided into three sub-samples (Figure 2):

- *Construction* for selecting indicators and points
- *Calibration* for associating scores with poverty likelihoods
- *Validation* for measuring accuracy with data not used in construction or calibration

2.2 Poverty rates and poverty lines

2.2.1 Rates

As a general definition, the *poverty rate* is the share of people in a given group who live in households whose total household consumption (divided by the number of household members) is below a given poverty line.

Beyond this general definition, there two special cases, *household-level poverty rates* and *person-level poverty rates*. With household-level rates, each household is

counted as if it had only one person, regardless of true household size, so all households are counted equally. With person-level rates (the "head-count index"), each household is weighted by the number of people in it, so larger households count more.

For example, consider a group of two households, the first with one member and the second with two members. Suppose further that the first household has per-capita consuption above a poverty line (it is "non-poor") and that the second household has per-capita consumption below a poverty line (it is "poor"). The household-level rate counts both households as if they had only one person and so gives a poverty rate of 1 $\div$ (1 + 1) = 50 percent. In contrast, the person-level rate weighs each household by the number of people in it and so gives a poverty rate of 2 $\div$ (1 + 2) = 67 percent.

Whether the household-level rate or the person-level rate is relevant depends on the situation. If an organization's "participants" include all the people in a household, then the person-level rate is relevant. Governments, for example, are concerned with the well-being of people, regardless of how those people are arranged in households, so governments typically report person-level poverty rates.

If an organization has only one "participant" per household, however, then the household-level rate is relevant. For example, if a microlender has only one borrower in a household, then it might prefer to report household-level poverty rates.

Figures A1 and A2 at the end of this document report poverty rates and poverty lines for Yemen at both the household-level and the person-level, by urban/rural and by governorate. The scorecard is constructed using the 2005/6 HBS and household-level

lines, scores are calibrated to household-level poverty likelihoods, and accuracy is measured for household-level rates. This use of household-level rates reflects the belief that they are relevant for most pro-poor organizations.

Organizations can estimate person-level poverty rates by taking a household-size-weighted average of the household-level poverty likelihoods. It is also possible to construct a scorecard based on person-level lines, calibrate scores to person-level likelihoods, and measure accuracy for person-level rates, but it is not done here.

2.2.2 Poverty lines

Using the 2005/6 HBS, the World Bank (2007) derives household-level poverty lines using a cost-of-basic-needs approach. The first step is to estimate each household's caloric needs, accounting for the age and sex of each member. The second step is to derive the average cost per calorie reported by HBS households in the lowest two quintiles of nominal per capita consumption. The third step is to multiply a given household's caloric requirement by the cost per calorie. This produces a household-specific national food line that, when weighted to be nationally representative, is YR124/person/day.

To derive a national line, World Bank (2007) uses the 2005/6 HBS to estimate how food consumption as a share of total consumption changes as total consumption increases. It then uses this estimate to define the national poverty line as the sum of a given household's food line and the estimated non-food consumption that corresponds to

that food line.[3] When weighted to be nationally representative, this gives an average national line of YR179/person/day.

Because local pro-poor organizations may want to use different or various poverty lines, this paper calibrates scores from its single scorecard to poverty likelihoods for seven lines:

- National
- Food
- USAID "extreme"
- USD1.25/day 2005 PPP
- USD2.50/day 2005 PPP
- USD3.00/day 2005 PPP
- USD4.00/day 2005 PPP

The USAID "extreme" line is defined as the median consumption of people (not households) below the national line (U.S. Congress, 2002).

The USD1.25/day 2005 PPP line is derived from:

- 2005 PPP exchange rate for "individual consumption expenditure by households" (International Comparison Project, 2008): YR91.06 per USD1.00
- National Monthly Consumer Price Index from CSO[4]. The average CPI in 2005 is 100.3746, and the average CPI in the survey period, April 2005 to March 2006, is 103.5691.[5]

[3] What this paper calls "national poverty lines" are called "lower poverty lines" or "poverty lines" by World Bank (2007). That report also refers to the construction of "upper poverty lines", but these upper lines are not provided with the 2005/6 HBS data, and World Bank (2007) does not use them in that report.

[4] `cso-yemen.org/publication/price/Report2008/ priceBybase2005.xls`, accessed 20 May 2009.

[5] Because it is not directly documented, this paper assumes that consumption in the 2005/6 HBS has not been adjusted for Yemen's 20-percent price inflation during the 12 months of fieldwork. For lack of better alternatives, this paper adjusts consumption using the average CPI for the months while the survey was in the field.

Given this, the USD1.25/day 2005 PPP line for Yemen as a whole from April 2005 to March 2006, is (Sillers, 2006):

$$(\text{2005 PPP exchange rate}) \cdot \text{USD1.25} \cdot \left(\frac{\text{CPI}_{\text{Apr 05-Mar 06}}}{\text{CPI}_{\text{2005 average}}} \right) =$$

$$\left(\frac{\text{YR91.06}}{\text{USD1.00}} \right) \cdot \text{USD1.25} \cdot \left(\frac{103.5691}{100.3746} \right) = \text{YR117.45}.$$

The USD2.50/day, USD3.00/day, and USD4.00/day 2005 PPP lines are multiples of the USD1.25/day line.

The 2005 PPP lines just discussed apply to Yemen as a whole. They are adjusted for household differences using:

- L, a given all-Yemen 2005 PPP poverty line
- π_i, national poverty line for household i
- w_i, person-level weight for household i
- N, number of households in the 2005/6 HBS

The cost-of-living-adjusted 2005 PPP poverty line L_i for household i is then:

$$L_i = \frac{L \cdot \pi_i}{(\sum_{i=1}^{N} \pi_i \cdot w_i) / \sum_{i=1}^{N} w_i}.$$

For each of the seven poverty lines, Figure A1 and A2 shows the all-Yemen lines as well as the person-weighted average lines L_i by urban/rural and by governorate. The differences in local average lines reflect differences in local prices as well as differences in household composition. This paper uses the national household poverty lines to construct the scorecard.

3. Context of poverty-assessment tools for Yemen

This section discusses three existing poverty-assessment tools for Yemen in terms of goals, methods, poverty lines, indicators, accuracy, precision, and cost. Compared with these tools, the main strengths of the scorecard are its simplicity, its out-of-sample tests, and its formulas for standard errors.

3.1 Gwatkin *et al.*

Gwatkin *et al.* (2007) apply to Yemen an approach used in 56 countries with Demographic and Health Surveys (Rutstein and Johnson, 2004). They use Principal Components Analysis to make a "wealth index" from simple, low-cost indicators available for the 10,701 households in Yemen's 1997 DHS. The index is like the scorecard except that, because it is based on a relative definition of poverty, its accuracy is unknown, and it can only be assumed to be a proxy for long-term wealth/economic status.[6] Other examples of the PCA-index approach are Stifel and Christiaensen (2007), Zeller *et al.* (2006), Sahn and Stifle (2003 and 2000), and Filmer and Pritchett (2001).

[6] Still, because the indicators are similar and because the "flat max" is important, carefully built PCA indices and consumption-based poverty-assessment tools probably pick up the same underlying construct (such as "permanent income", see Bollen, Glanville, and Stecklov, 2007), and they probably rank households much the same. Tests of how well PCA indices predict consumption include Filmer and Scott (2008), Lindelow (2006), Wagstaff and Watanabe (2003), and Montgomery *et al.* (2000).

The 41 indicators in Gwatkin *et al.* are similar in their simplicity, inexpensiveness, and verifiability to those in the new scorecard here:

- Characteristics of the residence:
  - Presence of electricity
  - Source of lighting
  - Source of drinking water (and location)
  - Type of fuel for cooking
  - Type of toilet arrangement
  - Type of floors
  - Type of structure
  - Tenancy status
  - Location of kitchen
  - Number of rooms
  - Number of people per sleeping room
- Agriculture:
  - Whether any household member works agricultural land
  - Whether the household keeps any animals

- Ownership of consumer durables:
  - — Presence (and number) of radios
  - — Televisions:
    - Presence of any kind of television
    - Number of black-and-white televisions
    - Number of color televisions
  - — Presence (and number) of refrigerators
  - — Presence (and number) of bicycles
  - — Presence (and number) of motorcycles
  - — Cars:
    - Presence of any kind of car
    - Number of private cars
    - Number of taxi cars
  - — Presence (and number) of telephones
  - — Number of radio cassettes
  - — Number of videos
  - — Number of gas ranges
  - — Number of water heaters
  - — Number of sewing machines
  - — Number of electric fans
  - — Number of washing machines
  - — Number of air conditioners
  - — Number of vacuum cleaners
  - — Number of mixers
  - — Number of dishes

Gwatkin *et al.* has three basic goals for their PCA-based wealth index:

- Segment people by quintiles in order to see how health, population, and nutrition vary with socio-economic status
- Monitor (via exit surveys) how well health-service points reach the poor
- Measure coverage of services via small-scale local surveys

These last two goals resemble the monitoring goals here, and the first goal of ranking households by quintiles is akin to targeting.

Overall, Gwatkin *et al.*'s index is more difficult to use than the scorecard here. Beyond the need to collect 41 indicators, the points have at least four decimal places and are sometimes negative. Furthermore, the points for most indicators require taking

a difference, performing division, and then performing multiplication. Even though Gwatkin *et al.* provide their tool in a three-page format that could be photocopied and taken to the field, the index cannot be computed by hand.

Finally, the scorecard here—unlike the PCA index—is linked directly to an absolute, consumption-based poverty line. Thus, while both approaches can rank households, only the scorecard can estimate quantitative, consumption-based poverty status. Furthermore, relative accuracy (that is, ability to rank or target) is tested here more completely here than in Gwatkin *et al.*; generally, discussion of the accuracy of PCA indices rests on how well they correlate with health, education, or self-assessed poverty, even though their construction does not take any such correlation into account.

3.2 El-Kogali and El-Daw

Like Gwatkin *et al.*, El-Kogali and El-Daw (2001) construct a PCA-based index using Yemen's 1997 DHS. Their goal, however, is not to provide a tool to help local, pro-poor organizations to manage their social performance but rather to create a proxy for standard-of-living that researchers can use to measure the association between poverty status and outcomes in health and education.

To this end, El-Kogali and El-Daw build a 7-indicator poverty-assessment tool:

- Presence of electricity
- Presence of a radio
- Presence of a television
- Presence of a refrigerator
- Presence of a bicycle
- Presence of a car or motorcycle
- Type of floor material

While these indicators are few, simple, inexpensive, and verifiable, the tool would be difficult to use in the field, as it uses three principal components (Gwatkin *et al.* use one) and is presented in a way that makes more sense to researchers than to program managers. El-Kogali and El-Daw do not report targeting accuracy because they seek to relate living standards with health and education outcomes, not to provide a way to help target services to households at-risk for poor health and low education.

El-Kogali and El-Daw apply their index to Yemen's DHS households, rank them by the index, and divide them into three equal-sized groups labeled "poor", "middle", and "rich". They then create corresponding dummy variables and enter them in regressions with some other controls. Unsurprisingly, they find that:

> Children from poor households are less likely to attend school, and, if they attend, they are more likely to drop out and engage in child labor. With regards to health, children from poor households are more likely to be undernourished, more susceptible to disease incidence, and less likely to receive medical treatment and immunization. In general, girls are more disadvantaged than boys, particularly in poor households.

In the end, El-Kogali and El-Daw is part of an unfortunate genre in poverty analysis: run regressions, rediscover well-known associations between outcomes and poverty, and—without offering any new motivations or tools—exhort governments to do

a better job at what governments already know they should be doing. Of course, governments struggle with poverty alleviation not because they do not know, for example, that the poor have worse health and lower education, but rather because they face technical, financial, and—most important—organizational/institutional constraints. The goal of the scorecard is to weaken some of these constraints.

3.3 World Bank

World Bank (2007) uses the 2005/6 HBS to make a "poverty map" (Elbers, Lanjouw, and Lanjouw, 2003) that estimates poverty rates for Yemen's 313 urban/rural districts. It constructs 38 poverty-assessment tools (by urban/rural and by governorate) using generalized least squares, estimating the logarithm of consumption for households in the 2005/6 HBS based only on indicators also found in Yemen's December 2004 census.

The resulting 38 tools are then applied to the 2.8 million households in the census to estimate poverty rates based on the national poverty line for smaller areas than would be possible with only the HBS. Finally, World Bank makes "poverty maps" that quickly show how estimated poverty rates vary across areas in a way that makes sense to lay people.

Poverty mapping and the scorecard are similar in that they both:

- Build poverty-assessment tools with nationally representative survey data and then apply them to other data on sub-groups that may not be nationally representative
- Use simple, verifiable indicators that are quick and inexpensive to collect
- Provide unbiased estimates
- Estimate poverty rates for groups
- Seek to be useful in practice and so aim to be understood by non-specialists

Strengths of poverty mapping include that it:

- Has formally established theoretical properties
- Can be applied straightforwardly to measures of well-being beyond poverty rates
- Requires less data for construction and calibration
- Uses only indicators that appear in a census

Strengths of the scorecard include that it:

- Is simpler in terms of both construction and application
- Tests accuracy empirically and out-of-sample
- Associates poverty likelihoods with scores non-parametrically
- Estimates poverty likelihoods for individual households
- Reports simple formulas for standard errors

The basic difference between the two approaches is that poverty mapping seeks to help governments design pro-poor policies, while the scorecard seeks to help small, local pro-poor organizations to manage their outreach when implementing policies.[7]

The World Bank's 38 tools use various indicators, all of which appear in both the 2005/6 HBS and the 2004 census. The tools usually include 10 to 20 indicators at both the household level and from district-level census means. All indicators are objective, verifiable, simple, and quick-to-collect. Of course, the ease of collecting

[7] Another apparent difference is that the developers of poverty mapping say that it is inappropriate for targeting individual households or persons, while this paper supports such targeting as a legitimate, potentially useful application (Schreiner, 2008a).

indicators (or managing 38 tools) is not an issue in poverty mapping, as researchers apply the tools only after data has been collected for other purposes. Also, some of the 38 tools may be overfit, as few households are available for tool construction in some areas from in the 2005/6 HBS.

Because the census does not measure consumption, World Bank cannot test accuracy out-of-sample (that is, using data that was not already used to construct the scorecard). The World Bank does not report precision in terms of formula for standard errors, so precision cannot be compared with that of the scorecard here.

4. Scorecard construction

For Yemen, about 100 potential indicators are initially prepared in the areas of:

- Family composition (such as household size)
- Education (such as the highest education level attained by the female head/spouse)
- Housing (such as the main flooring material)
- Ownership of durable goods (such as TVs and washing machines)

Each indicator is first screened with the entropy-based "uncertainty coefficient" (Goodman and Kruskal, 1979) that measures how well the indicator predicts poverty on its own. Figure 3 lists the best candidate indicators, ranked by uncertainty coefficient. Responses for each indicator in Figure 3 are ordered starting with those most strongly linked with higher poverty likelihoods.

The scorecard also aims to measure *changes* in poverty through time. This means that, when selecting indicators and holding other considerations constant, preference is given to more sensitive indicators. For example, ownership of a radio/cassette recorder is probably more likely to change in response to changes in poverty than is the education of the female head/spouse.

The scorecard itself is built using the national poverty line and Logit regression on the construction sub-sample (Figure 2). Indicator selection uses both judgment and statistics (forward stepwise, based on "c"). The first step is to use Logit to build one scorecard for each candidate indicator. Each scorecard's accuracy is taken as "c", a measure of ability to rank by poverty status (SAS Institute Inc., 2004).

One of these one-indicator scorecards is then selected based on several factors (Schreiner *et al.*, 2004; Zeller, 2004), including improvement in accuracy, likelihood of acceptance by users (determined by simplicity, cost of collection, and "face validity" in terms of experience, theory, and common sense), sensitivity to changes in poverty status, variety among indicators, and verifiability.

A series of two-indicator scorecards are then built, each based on the one-indicator scorecard selected from the first step, with a second candidate indicator added. The best two-indicator scorecard is then selected, again based on "c" and judgment. These steps are repeated until the scorecard has 10 indicators.

The final step is to transform the Logit coefficients into non-negative integers such that total scores range from 0 (most likely below a poverty line) to 100 (least likely below a poverty line).

This algorithm is the Logit analogue to the common R^2-based stepwise with least-squares regression. It differs from naïve stepwise in that the criteria for selecting indicators include not only statistical accuracy but also judgment and non-statistical factors. The use of non-statistical criteria can improve robustness through time and helps ensure that indicators are simple and make sense to users.

The single scorecard here applies to all of Yemen. Evidence from India and Mexico (Schreiner, 2006 and 2005a), Sri Lanka (Narayan and Yoshida, 2005), and Jamaica (Grosh and Baker, 1995) suggests that segmenting poverty-assessment tools by urban/rural does not improve targeting accuracy much.

5. Practical guidelines for scorecard use

The main challenge of scorecard design is not to maximize statistical accuracy but rather to improve the chances that scoring is actually used in practice (Schreiner, 2005b). When scoring projects fail, the reason is not usually statistical inaccuracy but rather the failure of an organization to decide to do what is needed to integrate scoring in its processes and to learn to use it properly (Schreiner, 2002). After all, most reasonable scorecards predict tolerably well, thanks to the empirical phenomenon known as the "flat max" (Hand, 2006; Baesens *et al.*, 2003; Lovie and Lovie, 1986; Kolesar and Showers, 1985; Stillwell, Barron, and Edwards, 1983; Dawes, 1979; Wainer, 1976; Myers and Forgy, 1963). The bottleneck is less technical and more human, not statistics but organizational change management. Accuracy is easier to achieve than adoption.

The scorecard here is designed to encourage understanding and trust so that users will adopt it and use it properly. Of course, accuracy matters, but it is balanced against simplicity, ease-of-use, and "face validity". Programs are more likely to collect data, compute scores, and pay attention to the results if, in their view, scoring does not make a lot of "extra" work and if the whole process generally seems to make sense.

To this end, the scorecard here fits on one page. The construction process, indicators, and points are simple and transparent. "Extra" work is minimized; non-specialists can compute scores by hand in the field because the scorecard has:

- Only 10 indicators
- Only categorical indicators
- Simple weights (non-negative integers, no arithmetic beyond addition)

The scorecard in is ready to be photocopied. A field worker using the paper scorecard would:

- Record participant identifiers and household size
- Read each question from the scorecard
- Circle the response and its points
- Write the points in the far-right column
- Add up the points to get the total score
- Implement targeting policy (if any)
- Deliver the paper scorecard to a central office for filing or data entry

Of course, field workers must be trained. Quality outputs depend on quality inputs. If organizations or field workers gather their own data and have an incentive to exaggerate poverty rates (for example, if funders reward them for higher poverty rates), then it is wise to do on-going quality control via data review and random audits (Matul and Kline, 2003).[8] IRIS Center (2007a) and Toohig (2008) are useful nuts-and-bolts guides for budgeting, training field workers and supervisors, logistics, sampling, interviewing, piloting, recording data, and controlling quality.

In particular, while collecting scorecard indicators is relatively easier than alternatives, it is still absolutely difficult. Training and explicit definitions of terms and concepts in the scorecard is essential. For the case of Nigeria, Onwujekwe, Hanson, and Fox-Rushby (2006) found distressingly low inter-rater and test-retest correlations for indicators as seemingly simple and obvious as whether the household owns an automobile. In Mexico, however, Martinelli and Parker (2007) find that errors by

[8] If an organization does not want field workers to know the points associated with indicators, then they can use the version of Figure 1 without points and apply the points later in a spreadsheet or database at the central office.

interviewers and lies by respondents have negligible effects on targeting accuracy. For now, it is unknown whether these results are universal or country-specific.

In terms of sampling design, an organization must make choices about:

- Who will do the scoring
- How scores will be recorded
- What participants will be scored
- How many participants will be scored
- How frequently participants will be scored
- Whether scoring will be applied at more than one point in time
- Whether the same participants will be scored at more than one point in time

The non-specialists who apply the scorecard with participants in the field can be:

- Employees of the organization
- Third-party contractors

Responses, scores, and poverty likelihoods can be recorded:

- On paper in the field and then filed at an office
- On paper in the field and then keyed into a database or spreadsheet at an office
- On portable electronic devices in the field and downloaded to a database

The subjects to be scored can be:

- All participants (or all new participants)
- A representative sample of all participants (or of all new participants)
- All participants (or all new participants) in a representative sample of branches
- A representative sample of all participants (or of all new participants) in a representative sample of branches

If not determined by other factors, the number of participants to be scored can be derived from sample-size formulas (presented later) for a desired level of confidence and a desired confidence interval.

Frequency of application can be:

- At in-take of new clients only (precluding measuring change in poverty rates)
- As a once-off project for current participants (precluding measuring change)
- Once a year (or at some other fixed time interval, allowing measuring change)
- Each time a field worker visits a participant at home (allowing measuring change)

When the scorecard is applied more than once in order to measure change in poverty rates, it can be applied:

- With a different set of participants
- With the same set of participants

An example set of choices were made by BRAC and ASA, two microlenders in Bangladesh (each with more than 7 million participants) who are applying the Simple Poverty Scorecard tool for Bangladesh (Schreiner, 2013). Their design is that loan officers in a random sample of branches will score all participants each time they visit a homestead (about once a year) as part of their standard due diligence prior to loan disbursement. Responses are recorded on paper in the field before being sent to a central office to be entered into a database. ASA's and BRAC's sampling plans cover 50,000–100,000 participants each.

6. Estimates of household poverty likelihoods

The sum of scorecard points for a household is called the *score*. For Yemen, scores range from 0 (most likely below a poverty line) to 100 (least likely below a poverty line). While higher scores indicate less likelihood of being below a poverty line, the scores themselves have only relative units. For example, doubling the score does not necessarily double the likelihood of being above a poverty line.

To get absolute units, scores must be converted to *poverty likelihoods*, that is, probabilities of being below a poverty line. This is done via simple look-up tables. For the example of the national line, scores of 10–14 have a poverty likelihood of 59.4 percent, and scores of 40–44 have a poverty likelihood of 10.8 percent (Figure 4).

The poverty likelihood associated with a score varies by poverty line. For example, scores of 40–44 are associated with a poverty likelihood of 10.8 percent for the national line but 3.0 percent for the food line.[9]

[9] Starting with Figure 4, most figures have seven versions, one for each of seven poverty lines. To keep them straight, they are grouped by poverty line. Single tables that pertain to all poverty lines are placed with the first group of tables for the national line.

6.1 Calibrating scores with poverty likelihoods

A given score is non-parametrically associated ("calibrated") with a poverty likelihood by defining the poverty likelihood as the share of households in the calibration sub-sample who have the score and who are below a given poverty line.

For the example of the national line (Figure 5), there are 11,977 (normalized) households in the calibration sub-sample with a score of 20–24, of whom 4,346 (normalized) are below the poverty line. The estimated poverty likelihood associated with a score of 20–24 is then 36.3 percent, because 4,346 ÷ 11,977 = 36.3 percent.

To illustrate with the national line and a score of 40–44, there are 6,868 (normalized) households in the calibration sample, of whom 743 (normalized) are below the line (Figure 5). Thus, the poverty likelihood for this score is 743 ÷ 6,868 = 10.8 percent.

The same method is used to calibrate scores with estimated poverty likelihoods for the other six poverty lines.

Figure 6 shows, for all scores, the likelihood that consumption falls in a range demarcated by two adjacent poverty lines. For example, the daily consumption of someone with a score of 35–39 falls in the following ranges with probability:

- 5.1 percent below the USD1.25/day 2005 PPP line
- 0.3 percent between the USD1.25/day 2005 PPP and food lines
- 3.8 percent between the food and USAID lines
- 10.4 percent between the USAID and national lines
- 17.8 percent between the national and USD2.50/day 2005 PPP lines
- 15.2 percent between the USD2.50/day and USD3.00/day 2005 PPP lines
- 21.2 percent between the USD3.00/day and USD4.00/day 2005 PPP lines
- 26.3 percent above the USD4.00/day 2005 PPP line

Even though the scorecard is constructed partly based on judgment, the calibration process produces poverty likelihoods that are objective, that is, derived from survey data on consumption and quantitative poverty lines. The poverty likelihoods would be objective even if indicators and/or points were selected without any data at all. In fact, objective scorecards of proven accuracy are often constructed using only expert judgment (Fuller, 2006; Caire, 2004; Schreiner *et al.*, 2004). Of course, the scorecard here is constructed with both data and judgment. The fact that this paper acknowledges that some choices in scorecard construction—as in any statistical analysis—are informed by judgment in no way impugns the objectivity of the poverty likelihoods, as this depends on using data in score calibration, not on using data (and nothing else) in scorecard construction.

Although the points in the Yemen scorecard are transformed coefficients from a Logit regression, scores are not converted to poverty likelihoods via the Logit formula of $2.718281828^{\text{score}} \times (1+ 2.718281828^{\text{score}})^{-1}$. This is because the Logit formula is esoteric and

difficult to compute by hand. Non-specialists find it more intuitive to define the poverty likelihood as the share of households with a given score in the calibration sample who are below a poverty line. In the field, converting scores to poverty likelihoods requires no arithmetic at all, just a look-up table. This non-parametric calibration can also improve accuracy, especially with large samples.

6.2 Accuracy of estimates of households' poverty likelihoods

As long as the relationship between indicators and poverty does not change and as long as the scorecard is applied to households from the same population from which it was constructed, then this calibration process produces unbiased estimates of poverty likelihoods. *Unbiased* means that in repeated samples from the same population, the average estimate matches the true poverty likelihood. The scorecard also produces unbiased estimates of poverty rates at a point in time, as well as unbiased estimates of changes in poverty rates between two points in time.[10]

Of course, the relationship between indicators and poverty does change to some unknown extent with time and also across sub-groups in Yemen's population, so the scorecard will generally be biased when applied after March 2006 (the end date of fieldwork for the 2005/6 HBS) or when applied with non-nationally representative groups.

[10] This follows because these estimates of groups' poverty rates are linear functions of the unbiased estimates of households' poverty likelihoods.

How accurate are estimates of households' poverty likelihoods? To measure, the scorecard is applied to 1,000 bootstrap samples of size $n = 16{,}384$ from the validation sub-sample. Bootstrapping entails (Efron and Tibshirani, 1993):

- Score each household in the validation sample
- Draw a new bootstrap sample *with replacement* from the validation sample
- For each score, compute the true poverty likelihood in the bootstrap sample, that is, the share of households with the score and consumption below a poverty line
- For each score, record the difference between the estimated poverty likelihood (Figure 4) and the true poverty likelihood in the bootstrap sample
- Repeat the previous three steps 1,000 times
- For each score, report the average difference between estimated and true poverty likelihoods across the 1,000 bootstrap samples
- For each score, report the two-sided interval containing the central 900, 950, or 990 differences between estimated and true poverty likelihoods

For each score range and for $n = 16{,}384$, Figure 7 shows the average difference between estimated and true poverty likelihoods, as well as confidence intervals for the differences.

For the national line, the average poverty likelihood across bootstrap samples for scores of 20–24 in the validation sample is too low by 4.0 percentage points (Figure 7). For scores of 15–19, the estimate is too high by 2.0 percentage points.[11]

The 90-percent confidence interval for the differences for scores of 20–24 is ±3.1 percentage points (Figure 7). This means that in 900 of 1,000 bootstraps, the difference between the estimate and the true value is between –7.1 and –0.9 percentage points

[11] These differences are not zero, in spite of the estimator's unbiasedness, because the scorecard comes from a single sample. The average difference by score would be zero if samples were repeatedly drawn from the population and split into sub-samples before repeating the entire process of scorecard building and calibration.

(because –4.0 – 3.1 = –7.1, and –4.0 + 3.1 = –0.9). In 950 of 1,000 bootstraps (95 percent), the difference is –4.0 ±3.3 percentage points, and in 990 of 1,000 bootstraps (99 percent), the difference is –4.0 ±3.7 percentage points.

For scores below 70, Figure 7 shows differences—sometimes large ones—between estimated poverty likelihoods and true values. This is because the validation sub-sample is a single sample that—thanks to sampling variation—differs in distribution from the construction/calibration sub-samples and from Yemen's population. For targeting, however, what matters is less the difference in all score ranges and more the difference in score ranges just above and below the targeting cut-off. This mitigates the effects of bias and sampling variation on targeting (Friedman, 1997). Section 9 below looks at targeting accuracy in detail.

Of course, if estimates of groups' poverty rates are to be usefully accurate, then errors for individual households must largely cancel out. This is generally the case, as discussed in the next section.

Another possible source of differences between estimates and true values is overfitting. By construction, the scorecard here is unbiased, but it may still be *overfit* when applied after the end of the HBS fieldwork in March 2006. That is, it may fit the 2005/6 HBS data so closely that it captures not only some timeless patterns but also some random patterns that, due to sampling variation, show up only in the 2005/6 HBS. Or the scorecard may be overfit in the sense that it is not robust to changes in

the relationships between indicators and poverty over time or when it is applied to non-nationally representative samples.

Overfitting can be mitigated by simplifying the scorecard and by not relying only on data but rather also considering experience, judgment, and theory. Of course, the scorecard here does this. Combining scorecards can also help, at the cost of greater complexity.

Most errors in individual households' likelihoods, however, cancel out in the estimates of groups' poverty rates (see later sections). Furthermore, at least some of the differences come from non-scorecard sources such as changes in the relationship between indicators and poverty, sampling variation, changes in poverty lines, inconsistencies in data quality across time, and imperfections in cost-of-living adjustments across time and space. These factors can be addressed only by improving data quantity and quality (which is beyond the scope of the scorecard) or by reducing overfitting (which likely has limited returns, given the scorecard's parsimony).

7. Estimates of a group's poverty rate at a point in time

A group's estimated poverty rate at a point in time is the average of the estimated poverty likelihoods of the individual households in the group.

To illustrate, suppose a program samples three households on Jan. 1, 2009 and that they have scores of 20, 30, and 40, corresponding to poverty likelihoods of 36.3, 21.6, and 10.8 percent (national line, Figure 4). The group's estimated poverty rate is the households' average poverty likelihood of $(36.3 + 21.6 + 10.8) \div 3 = 22.9$ percent.[12]

7.1 Accuracy of estimated poverty rates at a point in time

For the Yemen scorecard applied to the validation sample with $n = 16{,}384$, the absolute differences between the estimated poverty rate at a point in time and the true rate are 2.0 percentage points or less (Figure 8, summarizing Figure 9 across poverty lines). The average absolute difference across the seven poverty lines is 1.0 percentage points. At least part of these differences is due to sampling variation in the validation sample and in the random division of the 2005/6 HBS into three sub-samples.

In terms of precision, the 90-percent confidence interval for a group's estimated poverty rate at a point in time with $n = 16{,}384$ is ± 0.7 percentage points or less (Figure 8). This means that in 900 of 1,000 bootstraps of this size, the difference between the

[12] The group's poverty rate is *not* the poverty likelihood associated with the average score. Here, the average score is $(20 + 30 + 40) \div 3 = 30$, and the poverty likelihood associated with the average score is 21.6 percent. This is not the 22.9 percent found as the average of the three poverty likelihoods associated with each of the three scores.

estimate and the true value is within 0.7 percentage points of the average difference. In the specific case of the national line and the validation sample, 90 percent of all samples of n = 16,384 produce estimates that differ from the true value in the range of 1.5 – 0.6 = 0.9 to 1.5 + 0.6 = 2.1 percentage points. This is because 1.5 is the average difference, and ±0.6 is its 90-percent confidence interval. The average difference is 1.5 because the average scorecard estimate is too high by 1.5 percentage points; it estimates a poverty rate of 30.7 percent for the validation sample, but the true value is 29.2 percent (Figure 2).

7.2 Formula for standard errors for estimates of poverty rates

How precise are the point-in-time estimates? Because they are averages of binary (0/1, or poor/non-poor) variables, the estimates (in "large" samples) have a Normal distribution and can be characterized by their average difference vis-à-vis true values together with the standard error of the average difference.

To derive a formula for the standard errors of estimated poverty rates at a point in time from indirect measurement via poverty-assessment tools (Schreiner, 2008a), note that the textbook formula (Cochran, 1977) that relates confidence intervals with standard errors in the case of direct measurement of poverty status is $c = +/- z \cdot \sigma$, where:

c is the confidence interval as a proportion
(for example, 0.2 for ±2 percentage points),

z is from the Normal distribution and is $\begin{cases} 1.64 \text{ for confidence levels of 90 percent} \\ 1.96 \text{ for confidence levels of 95 percent} \\ 2.58 \text{ for confidence levels of 99 percent} \end{cases}$,

σ is the standard error of the estimated poverty rate, that is, $\sqrt{\frac{p \cdot (1-p)}{n}}$,

p is the proportion of households below the poverty line in the sample, and

n is the sample size.

For example, this implies that for a sample n of 16,384 with 90-percent confidence ($z = 1.64$) and a poverty rate p of 28.8 percent (the average poverty rate in the construction and calibration samples in Figure 2 for the national line), the confidence interval c is $+/- z \cdot \sqrt{\frac{p \cdot (1-p)}{n}} = +/- 1.64 \cdot \sqrt{\frac{0.288 \cdot (1-0.288)}{16{,}384}} = \pm 0.0058$, or 0.58 percentage points.

The scorecard, however, does not measure poverty directly, so this formula is not immediately applicable. To derive a formula for the Yemen scorecard, consider Figure 9, which reports empirical confidence intervals c for the differences for the scorecard

applied to 1,000 bootstrap samples of various sample sizes from the validation sample. For $n = 16{,}384$ and the national line, the 90-percent confidence interval is 0.63 percentage points.[13]

Thus, the 90-percent confidence interval with $n = 16{,}384$ is 0.63 percentage points for Yemen's scorecard and 0.58 percentage points for direct measurement. The ratio of the two intervals is $0.63 \div 0.58 = 1.09$.

Now consider the same case, but with $n = 8{,}192$. The confidence interval under direct measurement is $+/-1.64 \cdot \sqrt{\frac{0.288 \cdot (1-0.288)}{8{,}192}} = \pm 0.008205$, or about 0.8205 percentage points. The empirical confidence interval with the Yemen scorecard (Figure 9) is 0.00900, or about 0.9 percentage points. Thus for $n = 8{,}192$, the ratio of the two intervals is $0.9 \div 0.8205 = 1.10$.

This ratio of 1.10 for $n = 8{,}182$ is not far from the ratio of 1.09 for $n = 16{,}384$. Across all sample sizes of 256 or more in Figure 9, the average ratio turns out to be 1.07, implying that confidence intervals for indirect estimates of poverty rates via the Yemen scorecard and this poverty line are about 1.07 times as wide as confidence intervals for direct estimates via the 2005/6 HBS. This 1.07 appears in Figure 8 as the "α factor" because if $\alpha = 1.07$, then the formula relating confidence intervals c and standard errors σ for the Yemen scorecard is $c = +/- z \cdot \alpha \cdot \sigma$. That is, formula for the

[13] Due to rounding, Figure 9 displays 0.6, not 0.63.

standard error σ for point-in-time estimates of poverty rates via scoring is

$$\alpha \cdot \sqrt{\frac{p \cdot (1-p)}{n}}.$$

In general, α can be more or less than 1.00. When α is less than 1.00, it means that the scorecard is more precise than direct measurement. This occurs for one of seven poverty lines in Figure 8.

The formula relating confidence intervals to standard errors for the scorecard can be rearranged to give a formula for determining sample size before measurement.[14] If $\hat{p}$ is the expected poverty rate before measurement, then the formula for sample size n based on the desired confidence level that corresponds to z and the desired confidence interval $\pm c$ is $n = \left(\frac{\alpha \cdot z}{c}\right)^2 \cdot \hat{p} \cdot (1-\hat{p})$.

To illustrate how to use this, suppose $c = 0.0502$ and $z = 1.64$ (90-percent confidence). Then the formula gives $n = \left(\frac{1.07 \cdot 1.64}{0.0502}\right)^2 \cdot 0.288 \cdot (1-0.288) = 251$, close to the sample size of 256 observed for these parameters in Figure 9.

[14] IRIS Center (2007a and 2007b) says that a sample size of $n = 300$ is sufficient for USAID reporting. If a poverty-assessment tool is as precise as direct measurement, if the expected (before measurement) poverty rate is 50 percent, and if the confidence level is 90 percent, then $n = 300$ implies a confidence interval of ±2.2 percentage points. In fact, USAID has not specified confidence levels or intervals. Furthermore, the expected poverty rate may not be 50 percent, and the poverty-assessment tool could be more or less precise than direct measurement.

Of course, the α factors in Figure 8 are specific to Yemen, its poverty lines, its poverty rates, and this scorecard. The derivation of the formulas, however, is valid for any poverty-assessment tool following the approach in this paper.

In practice after the end of fieldwork for the HBS in March 2006, an organization would select a poverty line (say, the national line), select a desired confidence level (say, 90 percent, or $z = 1.64$), select a desired confidence interval (say, ±2.0 percentage points, or $c = 0.02$), make an assumption about $\hat{p}$ (perhaps based on a previous measurement such as the 28.9 percent national average in the 2005/6 HBS in Figure 2), look up α (here, 1.07), assume that the scorecard will still work in the future and/or for non-nationally representative sub-groups,[15] and then compute the required sample size.

In this illustration, $n = \left(\frac{1.07 \cdot 1.64}{0.02} \right)^2 \cdot 0.289 \cdot (1 - 0.289) = 1{,}582$.

[15] This paper reports accuracy for the scorecard applied to the validation sample, but it cannot test accuracy for later years or for other groups. Still, performance after March 2006 will probably resemble that in the 2005/6 HBS, with some deterioration over time.

8. Estimates of changes in group poverty rates over time

The change in a group's poverty rate between two points in time is estimated as the change in the average poverty likelihood of the households in the group. With data for 2005/6 HBS only, this paper cannot estimate changes over time, nor can it present formula for standard errors. Nevertheless, the relevant concepts are presented here because, in practice, pro-poor organizations can apply the scorecard to collect their own data and measure change through time.

8.1 Warning: Change is not impact

Scoring can estimate change. Of course, poverty could get better or worse, and scoring does not indicate what caused change. This point is often forgotten or confused, so it bears repeating: the scorecard simply estimates change, and it does not, in and of itself, indicate the reason for the change. In particular, estimating the impact of program participation requires knowing what would have happened to participants if they had not been participants (Moffitt, 1991). Knowing this requires either strong assumptions or a control group that resembles participants in all ways except participation. To belabor the point, the scorecard can help estimate program impact only if there is some way to know what would have happened in the absence of the program. And that information must come from somewhere beyond the scorecard. Even measuring simple change usually requires assuming that the population is constant over time and that program drop-outs do not differ from non-drop-outs.

8.2 Calculating estimated changes in poverty rates over time

Consider the illustration begun in the previous section. On Jan. 1, 2009, a program samples three households who score 20, 30, and 40 and so have poverty likelihoods of 36.3, 21.6, and 10.8 percent (national line, Figure 4). The group's baseline estimated poverty rate is the households' average poverty likelihood of (36.3 + 21.6 + 10.8) ÷ 3 = 22.9 percent.

After baseline, two sampling approaches are possible for the follow-up round:

- Score a new, independent sample, measuring change by cohort across samples
- Score the same sample at follow-up as at baseline

By way of illustration, suppose that a year later on Jan. 1, 2010, the program samples three additional households who are in the same cohort as the three households originally sampled (or suppose that the program scores the same three original households a second time) and finds that their scores are 25, 35, and 45 (poverty likelihoods of 32.8, 19.5, and 6.8 percent, national line, Figure 4). Their average poverty likelihood at follow-up is now (32.8 + 19.5 + 6.8) ÷ 3 = 19.7 percent, an improvement of 22.9 – 19.7 = 3.2 percentage points.

This suggests that about one of thirty participants in this hypothetical example crossed the poverty line in 2009.[16] Among those who started below the line, one in seven (3.2 ÷ 22.9 = 14.0 percent) on net ended up above the line.[17]

[16] This is a net figure; some people start above the line and end below it, and vice versa.
[17] The scorecard does not reveal the reasons for this change.

8.3 Accuracy for estimated change in two independent samples

With only the 2005/6 HBS, it is not possible to measure the accuracy of scorecard estimates of changes in groups' poverty rates over time. In practice, of course, local pro-poor organizations can still apply Yemen's scorecard to estimate change. The rest of this section suggests approximate formulas for standard errors and sample sizes that may be used until there is additional data.

For two equal-sized independent samples, the same logic as above can be used to derive a formula relating the confidence interval c with the standard error σ of a scorecard's estimate of the change in poverty rates over time:

$$c = +/- z \cdot \sigma = +/- z \cdot \alpha \cdot \sqrt{\frac{2 \cdot p \cdot (1-p)}{n}} \, .$$

z, c, and p are defined as above, n is the sample size at both baseline and follow-up,[18] and α is the average (across a range of bootstrapped sample sizes) of the ratio of the observed confidence interval from a scorecard and the theoretical confidence interval under direct measurement.

[18] This means that, for a given precision and with direct measurement, estimating the change in a poverty rate between two points in time requires four times as many measurements (not twice as many) as does estimating a poverty rate at a point in time.

As before, the formula for standard errors can be rearranged to give a formula for sample sizes before indirect measurement via a scorecard, where $\hat{p}$ is based on previous measurements and is assumed equal at both baseline and follow-up:

$$n = 2 \cdot \left(\frac{\alpha \cdot z}{c} \right)^2 \cdot \hat{p} \cdot (1 - \hat{p}) .$$

For the countries for which this α has been measured (Peru, the Philippines, India, Vietnam, and Bangladesh, see Schreiner, 2009a, 2009b, and 2008b and Chen and Schreiner, 2009), the average α across poverty lines is 0.77, 0.77, 1.40, 0.68, and 1.03. The average across countries (0.93) may be reasonable for Yemen.

To illustrate the use of the formula above to determine sample size for estimating changes in poverty rates across two independent samples, suppose the desired confidence level is 90 percent ($z = 1.64$), the desired confidence interval is 2 percentage points ($c = 0.02$), the poverty line is the national line, $\alpha = 0.93$, and $\hat{p} = 0.289$ (from Figure 2). Then the baseline sample size is $n = 2 \cdot \left(\frac{0.93 \cdot 1.64}{0.02} \right)^2 \cdot 0.289 \cdot (1 - 0.289) =$ 2,390, and the follow-up sample size is also 2,390.

8.4 Accuracy for estimated change for one sample, scored twice

Analogous to previous derivations, the general formula relating the confidence interval c to the standard error σ when using a scorecard to estimate change for a single group of households, all of whom are scored at two points in time, is:[19]

$$c = +/- z \cdot \sigma = +/- z \cdot \alpha \cdot \sqrt{\frac{p_{12} \cdot (1 - p_{12}) + p_{21} \cdot (1 - p_{21}) + 2 \cdot p_{12} \cdot p_{21}}{n}},$$

where z, c, and α are defined as usual, p_{12} is the share of all sampled households that move from below the poverty line to above it, and p_{21} is the share of all sampled households that move from above the line to below it.

The formula for standard errors can be rearranged to give a formula for sample size before measurement. This requires an estimate (based on information available before measurement) of the expected shares of all households who cross the poverty line $\hat{p}_{12}$ and $\hat{p}_{21}$. Before measurement, it is reasonable to assume that the change in the poverty rate will be zero, which implies $\hat{p}_{12} = \hat{p}_{21} = \hat{p}_*$, giving:

$$n = 2 \cdot \left(\frac{\alpha \cdot z}{c} \right)^2 \cdot \hat{p}_* .$$

[19] See McNemar (1947) and Johnson (2007). John Pezzullo helped find this formula.

Because $\hat{p}_*$ could be anything between 0–1, more information is needed to apply this formula. Suppose that the observed relationship between $\hat{p}_*$, the number of years y between baseline and follow-up, and $p_{\text{baseline}} \cdot (1 - p_{\text{baseline}})$ is—as in Peru—close to:

$$\hat{p}_* = -0.02 + 0.016 \cdot y + 0.47 \cdot [p_{\text{baseline}} \cdot (1 - p_{\text{baseline}})].$$

Given this, a sample-size formula for a group of households to whom the Yemen scorecard is applied twice (once after March 2006 and then again later) is:

$$n = 2 \cdot \left(\frac{\alpha \cdot z}{c} \right)^2 \cdot \{-0.02 + 0.016 \cdot y + 0.47 \cdot [p_{\text{baseline}} \cdot (1 - p_{\text{baseline}})]\}.$$

In Peru (the only other country for which there is an estimate, Schreiner 2009a), the average α across years and poverty lines is about 1.3.

To illustrate the use of this formula, suppose the desired confidence level is 90 percent ($z = 1.64$), the desired confidence interval is 2.0 percentage points ($c = 0.02$), the poverty line is the national line, and the sample will first be scored in 2009 and then again in 2012 ($y = 3$). The before-baseline poverty rate is 28.9 percent ($p_{2005} = 0.289$, Figure 2), and suppose $\alpha = 1.3$. Then the baseline sample size is

$n = 2 \cdot \left(\frac{1.3 \cdot 1.64}{0.02} \right)^2 \cdot \{-0.02 + 0.016 \cdot 3 + 0.47 \cdot [0.289 \cdot (1 - 0.289)]\} = 2{,}832$. The same group of 2,832 households is scored at follow-up as well.

9. Targeting

When a program uses the scorecard for targeting, households with scores at or below a cut-off are labeled *targeted* and treated—for program purposes—as if they are below a given poverty line. Households with scores above a cut-off are labeled *non-targeted* and treated—for program purposes—as if they are above a given poverty line.

There is a distinction between *targeting status* (scoring at or below a targeting cut-off) and *poverty status* (consumption below a poverty line). Poverty status is a fact that depends on whether consumption is below a poverty line as directly measured by a survey. In contrast, targeting status is a program's policy choice that depends on a cut-off and on an indirect estimate from a scorecard.

Targeting is successful when households truly below a poverty line are targeted (*inclusion*) and when households truly above a poverty line are not targeted (*exclusion*). Of course, no scorecard is perfect, and targeting is unsuccessful when households truly below a poverty line are not targeted (*undercoverage*) or when households truly above a poverty line are targeted (*leakage*). Figure 10 depicts these four possible targeting outcomes. Targeting accuracy varies by cut-off; a higher cut-off has better inclusion (but greater leakage), while a lower cut-off has better exclusion (but higher undercoverage).

It makes sense for a program to weigh these trade-offs when setting a cut-off. A formal way to do this is to assign net benefits—based on a program's values and

mission—to each of the four possible targeting outcomes and then to choose the cut-off that maximizes total net benefits (Adams and Hand, 2000; Hoadley and Oliver, 1998).

Figure 11 shows the distribution of households by targeting outcome. For an example cut-off of 35–39, outcomes for the national line in the validation sample are:

- Inclusion: 27.1 percent are below the line and correctly targeted
- Undercoverage: 2.0 percent are below the line and mistakenly not targeted
- Leakage: 44.7 percent are above the line and mistakenly targeted
- Exclusion: 26.1 percent are above the line and correctly not targeted

Increasing the cut-off to 40–44 improves inclusion and undercoverage but worsens leakage and exclusion:

- Inclusion: 28.3 percent are below the line and correctly targeted
- Undercoverage: 0.9 percent are below the line and mistakenly not targeted
- Leakage: 50.5 percent are above the line and mistakenly targeted
- Exclusion: 20.4 percent are above the line and correctly not targeted

Which cut-off is preferred depends on total net benefit. If each targeting outcome has a per-household benefit or cost, then total net benefit for a given cut-off is:

Benefit per household correctly included	x	Households correctly included	–
Cost per household mistakenly not covered	x	Households mistakenly not covered	–
Cost per household mistakenly leaked	x	Households mistakenly leaked	+
Benefit per household correctly excluded	x	Households correctly excluded.	

To set an optimal cut-off, a program would:

- Assign benefits and costs to possible outcomes, based on its values and mission
- Tally total net benefits for each cut-off using Figure 11 for a given poverty line
- Select the cut-off with the highest total net benefit

The most difficult step is assigning benefits and costs to targeting outcomes. It makes sense that any program that uses targeting—with or without scoring—would thoughtfully consider how it values successful inclusion or exclusion versus errors of

undercoverage and leakage. It is healthy to go through a process of thinking explicitly and intentionally about how possible targeting outcomes are valued.

A common choice of benefits and costs is "Total Accuracy" (IRIS Center, 2005; Grootaert and Braithwaite, 1998). With "Total Accuracy", total net benefit is the number of households correctly included or correctly excluded:

Total Accuracy =	1	x	Households correctly included	–
	0	x	Households mistakenly undercovered	–
	0	x	Households mistakenly leaked	+
	1	x	Households correctly excluded.	

Figure 11 shows "Total Accuracy" for all cut-offs for the Yemen scorecard. For the national line in the validation sample, total net benefit is greatest (74.4) for a cut-off of 10–14, with about three in four Yemeni households correctly classified.

"Total Accuracy" weighs successful inclusion of households below the line the same as successful exclusion of households above the line. If a program valued inclusion more (say, twice as much) than exclusion, it could reflect this by setting the benefit for inclusion to 2 and the benefit for exclusion to 1. Then the chosen cut-off would maximize (2 x Households correctly included) + (1 x Households correctly excluded).

As an alternative to assigning benefits and costs to targeting outcomes and then choosing a cut-off to maximize total net benefit, a program could set a cut-off to achieve a desired poverty rate among targeted households. The third column of Figure 12 ("% targeted who are poor") shows, for the Yemen scorecard applied to the validation sample, the expected poverty rate among households who score at or below a given cut-off. For the example of the national line in the validation sample, targeting

households who score 35–39 or less would target 71.9 percent of all households (second column) and produce a poverty rate among those targeted of 37.8 percent (third column).

Figure 12 also reports two other measures of targeting accuracy. The first is a version of coverage ("% of poor who are targeted"). For the example of the national line in the validation sample and a cut-off of 35–39, 93.1 percent of all poor households are covered.

The final targeting measure in Figure 12 is the number of successfully targeted poor households for each non-poor household mistakenly targeted (right-most column). For the national line in the validation sample and a cut-off of 35–39, covering 6 poor households means leaking to 10 non-poor household.

10. Conclusion

Pro-poor programs in Yemen can use the scorecard to segment clients for differentiated treatment as well as to estimate:

- The likelihood that a household has consumption below a given poverty line
- The poverty rate of a population at a point in time
- The change in the poverty rate of a population over time

The scorecard is inexpensive to use and can be understood by non-specialists. It is designed to be practical for pro-poor organizations in Yemen that want to improve how they monitor and manage their social performance.

The scorecard is built with a sub-sample of data from the 2005/6 HBS, tested on a different sub-sample from the 2005/6 HBS, and calibrated to seven poverty lines (national, food, USAID "extreme", USD1.25/day 2005 PPP, USD2.50/day 2005 PPP, USD3.00/day 2005 PPP, and USD4.00/day 2005 PPP).

Accuracy and precision are reported for estimates of households' poverty likelihoods, groups' poverty rates at a point in time, and changes in groups' poverty rates over time. Of course, the scorecard's estimates of changes in poverty rates are not the same as estimates of program impact. Targeting accuracy is also reported.

When the scorecard is applied to the validation sample with $n = 16{,}384$, the absolute difference between estimates versus true poverty rates for groups of households at a point in time is always less than 2.0 percentage points and averages—across the seven poverty lines—about 1.0 percentage points. For $n = 16{,}384$ and 90-percent confidence, the precision of these differences is ± 0.7 percentage points or better, and for

$n = 1{,}024$, precision is ± 2.7 percentage points or better. For one poverty line, the scorecard is more precise than direct measurement.

If a program wants to use the scorecard for targeting, then the results here provide the information needed to select a cut-off that fits its values and mission.

Although the statistical technique is innovative, and although technical accuracy is important, the design of the scorecard here focuses on transparency and ease-of-use. After all, a perfectly accurate scorecard is worthless if programs feel so daunted by its complexity or its cost that they do not even try to use it. For this reason, the scorecard is kept simple, using ten indicators that are inexpensive to collect and that are straightforward to verify. Points are all zeros or positive integers, and scores range from 0 (most likely below a poverty line) to 100 (least likely below a poverty line). Scores are related to poverty likelihoods via simple look-up tables, and targeting cut-offs are likewise simple to apply. The design attempts to facilitate adoption by helping managers understand and trust scoring and by allowing non-specialists to generate scores quickly in the field.

In sum, the scorecard is a practical, objective way for pro-poor programs in Yemen to monitor poverty rates, track changes in poverty rates over time, and target services. The same approach can be applied to any country with similar data from a national consumption survey.

References

Adams, Niall M.; and David J. Hand. (2000) "Improving the Practice of Classifier Performance Assessment", *Neural Computation*, Vol. 12, pp. 305–311.

Baesens, Bart; Van Gestel, Tony; Viaene, Stijn; Stepanova, Maria; Suykens, Johan A. K.; and Jan Vanthienen. (2003) "Benchmarking State-of-the-Art Classification Algorithms for Credit Scoring", *Journal of the Operational Research Society*, Vol. 54, pp. 627–635.

Bollen, Kenneth A.; Glanville, Jennifer L.; and Guy Stecklov. (2007) "Socio-Economic Status, Permanent Income, and Fertility: A Latent-Variable Approach", *Population Studies*, Vol. 61, No. 1, pp. 15–34.

Caire, Dean. (2004) "Building Credit Scorecards for Small Business Lending in Developing Markets", `microfinance.com/English/Papers/Scoring_SMEs_Hybrid.pdf`, accessed 19 May 2009.

Chen, Shiyuan; and Mark Schreiner. (2009) "Simple Poverty Scorecard Poverty-Assessment Tool: Vietnam", `SimplePovertyScorecard.com/VNM_2006_ENG.pdf`, accessed 19 May 2009.

Coady, David; Grosh, Margaret; and John Hoddinott. (2004) *Targeting of Transfers in Developing Countries*, `hdl.handle.net/10986/14902`, retrieved 13 May 2016.

Cochran, William G. (1977) *Sampling Techniques, Third Edition*.

Dawes, Robyn M. (1979) "The Robust Beauty of Improper Linear Models in Decision Making", *American Psychologist*, Vol. 34, No. 7, pp. 571–582.

Efron, Bradley; and Robert J. Tibshirani. (1993) *An Introduction to the Bootstrap*.

El-Kogali, Safaa; and Suliman El-Daw. (2001) "Poverty, Human Capital, and Gender: A Comparative Study of Yemen and Egypt", ERF Working Paper No. 0123, `www.erf.org.eg/CMS/getFile.php?id=86`, accessed 20 May 2009.

Elbers, Chris; Lanjouw, Jean O.; and Peter Lanjouw. (2003) "Micro-Level Estimation of Poverty and Inequality", *Econometrica*, Vol. 71, No. 1, pp. 355–364.

Filmer, Deon; and Lant Pritchett. (2001) "Estimating Wealth Effects without Expenditure Data—or Tears: An Application to Educational Enrollments in States of India", *Demography*, Vol. 38, No. 1, pp. 115–132.

_____; and Kinnon Scott. (2008) "Assessing Asset Indices", World Bank Policy Research Working Paper No. 4605, papers.ssrn.com/sol3/papers.cfm?abstract_id=1149108, accessed 21 March 2009.

Friedman, Jerome H. (1997) "On Bias, Variance, 0–1 Loss, and the Curse-of-Dimensionality", *Data Mining and Knowledge Discovery*, Vol. 1, pp. 55–77.

Fuller, Rob. (2006) "Measuring the Poverty of Microfinance Clients in Haiti", microfinance.com/English/Papers/Scoring_Poverty_Haiti_Fuller.pdf, accessed 19 May 2009.

Goodman, Leo A.; and Kruskal, William H. (1979) *Measures of Association for Cross Classification*.

Grootaert, Christiaan; and Jeanine Braithwaite. (1998) "Poverty Correlates and Indicator-Based Targeting in Eastern Europe and the Former Soviet Union", World Bank Policy Research Working Paper No. 1942, dx.doi.org/10.1596/1813-9450-1942, retrieved 15 May 2016.

Grosh, Margaret; and Judy L. Baker. (1995) "Proxy-Means Tests for Targeting Social Programs: Simulations and Speculation", World Bank LSMS Working Paper No. 118, go.worldbank.org/W9OWN57PD0, retrieved 13 May 2016.

Gwatkin, Davidson R.; Rutstein, Shea; Johnson, Kiersten; Suliman, Eldaw; Wagstaff, Adam; and Agbessi Amouzou. (2007) "Socio-Economic Differences in Health, Nutrition, and Population: Yemen", Country Reports on HNP and Poverty, go.worldbank.org/T6LCN5A340, retrieved 18 March 2017.

Hand, David J. (2006) "Classifier Technology and the Illusion of Progress", *Statistical Science*, Vol. 22, No. 1, pp. 1–15.

Hoadley, Bruce; and Robert M. Oliver. (1998) "Business Measures of Scorecard Benefit", *IMA Journal of Mathematics Applied in Business and Industry*, Vol. 9, pp. 55–64.

International Comparison Project. (2008) "Tables of Results", siteresources.worldbank.org/ICPINT/Resources/icp-final-tables.pdf, accessed 19 May 2009.

IRIS Center. (2007a) "Manual for the Implementation of USAID Poverty Assessment Tools", povertytools.org/training_documents/Manuals/USAID_PAT_Manual_Eng.pdf, accessed 19 May 2009.

_____. (2007b) "Introduction to Sampling for the Implementation of PATs", povertytools.org/training_documents/Sampling/Introduction_Sampling.ppt, accessed 19 May 2009.

_____. (2005) "Notes on Assessment and Improvement of Tool Accuracy", povertytools.org/other_documents/AssessingImproving_Accuracy.pdf, accessed 19 May 2009.

Johnson, Glenn. (2007) "Lesson 3: Two-Way Tables—Dependent Samples", www.stat.psu.edu/online/development/stat504/03_2way/53_2way_compare.htm, accessed 19 May 2009.

Kolesar, Peter; and Janet L. Showers. (1985) "A Robust Credit Screening Model Using Categorical Data", *Management Science*, Vol. 31, No. 2, pp. 124–133.

Lindelow, Magnus. (2006) "Sometimes More Equal Than Others: How Health Inequalities Depend on the Choice of Welfare Indicator", *Health Economics*, Vol. 15, pp. 263–279.

Lovie, Alexander D.; and Patricia Lovie. (1986) "The Flat Maximum Effect and Linear Scoring Models for Prediction", *Journal of Forecasting*, Vol. 5, pp. 159–168.

Martinelli, César; and Susan W. Parker. (2007) "Deception and Misreporting in a Social Program", ciep.itam.mx/~martinel/lies4.pdf, accessed 19 May 2009.

Matul, Michal; and Sean Kline. (2003) "Scoring Change: Prizma's Approach to Assessing Poverty", Microfinance Centre for Central and Eastern Europe and the New Independent States Spotlight Note No. 4, www.mfc.org.pl/doc/Research/ImpAct/SN/MFC_SN04_eng.pdf, accessed 19 May 2009.

McNemar, Quinn. (1947) "Note on the Sampling Error of the Difference between Correlated Proportions or Percentages", *Psychometrika*, Vol. 17, pp. 153–157.

Moffitt, Robert. (1991) "Program Evaluation with Non-experimental Data", *Evaluation Review*, Vol. 15, No. 3, pp. 291–314.

Montgomery, Mark; Gragnolati, Michele; Burke, Kathleen A.; and Edmundo Paredes. (2000) "Measuring Living Standards with Proxy Variables", *Demography*, Vol. 37, No. 2, pp. 155–174.

Myers, James H.; and Edward W. Forgy. (1963) "The Development of Numerical Credit Evaluation Systems", *Journal of the American Statistical Association*, Vol. 58, No. 303, pp. 779–806.

Narayan, Ambar; and Nobuo Yoshida. (2005) "Proxy Means Tests for Targeting Welfare Benefits in Sri Lanka", World Bank Report No. SASPR–7, documents.worldbank.org/curated/en/2005/07/6209268/proxy-means-test-targeting-welfare-benefits-sri-lanka, retrieved 5 May 2016.

Onwujekwe, Obinna; Hanson, Kara; and Julia Fox-Rushby. (2006) "Some Indicators of Socio-Economic Status May Not Be Reliable and Use of Indices with These Data Could Worsen Equity", *Health Economics*, Vol. 15, pp. 639–644.

Rutstein, Shea Oscar; and Kiersten Johnson. (2004) "The DHS Wealth Index", DHS Comparative Reports No. 6, measuredhs.com/pubs/pdf/CR6/CR6.pdf, accessed 2 March 2009.

Sahn, David E.; and David Stifel. (2003) "Exploring Alternative Measures of Welfare in the Absence of Expenditure Data", *Review of Income and Wealth*, Series 49, No. 4, pp. 463–489.

_____. (2000) "Poverty Comparisons Over Time and Across Countries in Africa", *World Development*, Vol. 28, No. 12, pp. 2123–2155.

SAS Institute Inc. (2004) "The LOGISTIC Procedure: Rank Correlation of Observed Responses and Predicted Probabilities", in *SAS/STAT User's Guide, Version 9*, support.sas.com/documentation/cdl/en/statug/59654/HTML/default/statug_logistic_sect035.htm, accessed 19 May 2009.

Schreiner, Mark. (2013) "Simple Poverty Scorecard Poverty-Assessment Tool: Bangladesh", SimplePovertyScorecard.com/BGD_2010_ENG.pdf, accessed 9 July 2016.

_____. (2009a) "Simple Poverty Scorecard Poverty-Assessment Tool: Peru", SimplePovertyScorecard.com/PER_2007_ENG.pdf, accessed 9 July 2016.

_____. (2009b) "Simple Poverty Scorecard Poverty-Assessment Tool: Philippines", SimplePovertyScorecard.com/PHL_2002_ENG.pdf, accessed 9 July 2016.

_____. (2008a) "Simple Poverty Scorecard Poverty-Assessment Tool: Peru", SimplePovertyScorecard.com/PER_2003_ENG.pdf, accessed 9 July 2016.

_____. (2008b) "Simple Poverty Scorecard Poverty-Assessment Tool: India", SimplePovertyScorecard.com/IND_2005_ENG.pdf, accessed 9 July 2016.

_____. (2006) "Is One Simple Poverty Scorecard Poverty-Assessment Tool Enough for India?", microfinance.com/English/Papers/Scoring_Poverty_India_Segments.pdf, accessed 9 July 2016.

_____. (2005a) "La Herramienta del Índice de Calificación de la Pobreza™: México", SimplePovertyScorecard.com/MEX_2002_SPA.pdf, accessed 9 July 2016.

_____. (2005b) "IRIS Questions on the Simple Poverty Scorecard Poverty-Assessment Tool", microfinance.com/English/Papers/Scoring_Poverty_Response_to_IRIS.pdf, accessed 9 July 2016.

_____. (2002) *Scoring: The Next Breakthrough in Microfinance?* CGAP Occasional Paper No. 7, microfinance.com/English/Papers/Scoring_Breakthrough_CGAP.pdf, retrieved 13 May 2016.

_____; Matul, Michal; Pawlak, Ewa; and Sean Kline. (2004) "Poverty Scoring: Lessons from a Microlender in Bosnia-Herzegovina", microfinance.com/English/Papers/Scoring_Poverty_in_BiH_Short.pdf, accessed 19 May 2009.

Sillers, Don. (2006) "National and International Poverty Lines: An Overview", pdf.usaid.gov/pdf_docs/Pnadh069.pdf, retrieved 13 May 2016.

Stifel, David; and Luc Christiaensen. (2007) "Tracking Poverty over Time in the Absence of Comparable Consumption Data", *World Bank Economic Review*, Vol. 21, No. 2, pp. 317–341.

Stillwell, William G.; Barron, F. Hutton; and Ward Edwards. (1983) "Evaluating Credit Applications: A Validation of Multi-Attribute Utility Weight Elicitation Techniques", *Organizational Behavior and Human Performance*, Vol. 32, pp. 87–108.

Tarozzi, Alessandro; and Angus Deaton. (2009) "Using Census and Survey Data to Estimate Poverty and Inequality for Small Areas", *Review of Economics and Statistics*, Vol. 91, No. 4, pp. 773–792.

Toohig, Jeff. (2008) "PPI Pilot Training Guide", progressoutofpoverty.org/toolkit, accessed 19 May 2009.

United States Congress. (2004) "Microenterprise Results and Accountability Act of 2004 (HR 3818 RDS)", November 20, `smith4nj.com/laws/108-484.pdf`, retrieved 13 May 2016.

Wagstaff, Adam; and Naoko Watanabe. (2003) "What Difference Does the Choice of SES Make in Health Inequality Measurement?", *Health Economics*, Vol. 12, No. 10, pp. 885–890.

Wainer, Howard. (1976) "Estimating Coefficients in Linear Models: It Don't Make No Nevermind", *Psychological Bulletin*, Vol. 83, pp. 223–227.

World Bank. (2007) "Yemen Poverty Assessment", `go.worldbank.org/DZXB1DU520`, accessed 21 May 2009.

Zeller, Manfred. (2004) "Review of Poverty Assessment Tools", `pdf.usaid.gov/pdf_docs/PNADH120.pdf`, retrieved 13 May 2016.

_____; Sharma, Manohar; Henry, Carla; and Cécile Lapenu. (2006) "An Operational Method for Assessing the Poverty Outreach Performance of Development Policies and Projects: Results of Case Studies in Africa, Asia, and Latin America", *World Development*, Vol. 34, No. 3, pp. 446–464.

Figure 2: Sample sizes and household poverty rates by sub-sample and poverty line

		% with expenditure below a poverty line						
			National	USAID	International (2005 PPP)			
Sub-sample	**Households**	**National**	**Food**	**'Extreme'**	**$1.25/day**	**$2.50/day**	**$3.00/day**	**$4.00/day**
All Yemen	13,136	28.9	10.1	14.3	8.3	47.6	60.6	76.8
Construction								
Selecting indicators and weights	4,379	28.9	10.0	14.2	8.2	47.3	60.0	76.5
Calibration								
Associating scores with likelihoods	4,329	28.7	10.1	14.5	8.3	47.5	59.9	76.5
Validation								
Measuring accuracy	4,428	29.2	10.1	14.2	8.5	47.9	61.9	77.6
Change in poverty rate (percentage points)								
From construction/calibration to validation		-0.4	-0.1	0.1	-0.2	-0.5	-1.9	-1.1

Source: 2005/6 HBS

Figure 3: Poverty indicators by uncertainty coefficient

Uncertainty coefficient	Indicator (Answers ordered starting with those most strongly linked with higher poverty likelihoods)
86	How many members does the household have? (Ten or more; Eight or nine; Seven; Six; Four or five; Three; One or two)
60	How many household members aged 12 to 18 currently attend school? (Not all; All; No children aged 12 to 18)
44	What is the main material used for the floor? (Concrete, mud/soil, stone, or other; Floor tiles or marble)
36	What is the highest education level attained by the female head/spouse? (None, or no female head; Primary school or higher)
36	Does this household or any of its members own an iron? (No; Yes)
35	What is the highest education level attained by the male head/spouse? (None; Primary school; No male head, or pre-high school or higher)
35	Does the household or any of its members own a washing machine? (No; Yes)
34	Does this household or any of its members own a refrigerator? (No; Yes)
34	Does this household or any of its members own a color TV? (No; Yes)
33	What is the type of sewage disposal system? (None, open pit, or other; Public network or closed pit)
33	What type of toilet does the household have? (Non-flush toilet, other, or no toilet; flush toilet)
32	What is the main source of lighting? (Kerosene lantern or gas lamp; Public network, cooperative network, house generator, or other; Private network)
32	Does this household or any of its members own a telephone? (No; Yes)
32	Does this household or any of its members own a satellite dish? (No; Yes)
32	Does this household or any of its members own a mobile phone? (No; Yes)

Figure 3 (cont.): Poverty indicators by uncertainty coefficient

Uncertainty coefficient	Indicator (Answers ordered starting with those most strongly linked with higher poverty likelihoods)
29	Does this household or any of its members own a mixer? (No; Yes)
26	What is the main source of water supply? (On foot or riding animals; Tap outside the house or water tank vehicle; Public/cooperative/private network, private car, or other)
25	What is the main source of water to the household? (Normal well, stream, covered/open tank, dam, traditional rain collection, or other; Artesian well; Public/cooperative/private network)
23	Does the household or any of its members own a TV? (No; Yes)
21	What is the main material used for the ceiling? (Wood and mud, wood, hay, mud and hay, metal sheets and mud, or other; Reinforced concrete, wood and concrete, or metal sheets)
19	What are the main sources of energy for cooking? (Wood, coal, kerosene, garbage, animal dung, or other; Gas or electricity)
17	What is the main material used for the external walls? (Regular stone, mud, hay, or fabric; Cement block, sun-dried brick, cooked/burned brick, or other; Cut stone)
16	Did the household own any sheep or goats during the past 12 months? (Yes; No)
12	Does this household or any of its members own an electrical water heater? (No; Yes)
12	Did the household own any donkeys during the past 12 months? (Yes; No)
11	Does this household or any of its members own a car? (No; Yes)
9	What type of kitchen does the household have? (Outdoors private, outdoors shared, or no kitchen; Indoors private or indoors shared)
6	Does this household or any of its members own an air conditioner? (No; Yes)
6	Does this household or any of its members own a gas stove? (No; Yes)
5	Does the household own any cultivated land? (Yes; No)

Figure 3 (cont.): Poverty indicators by uncertainty coefficient

Uncertainty coefficient	Indicator (Answers ordered starting with those most strongly linked with higher poverty likelihoods)
4	How many rooms are there in the house, excluding bathrooms and kitchens? (One; Two; Three; Four; Five; Six, or more)
3	Does this household or any of its members own an electric fan? (No; Yes)
3	Does this household or any of its members own a motorbike? (No; Yes)
3	Does this household or any of its members own a black-and-white TV? (No; Yes)
2	How many of cows did the household own during the past 12 months? (None; One; Two or more)
2	Can anyone in the household read and write in at least one language? (No; Yes)
1	Does the household or any of its members own a gas cylinder? (No; Yes)
1	Does the household or any of its members own a radio/cassette recorder? (No; Yes)
1	Does this household or any of its members own a sewing machine? (No; Yes)
1	Does this household or any of its members own a bicycle? (No; Yes)

National Poverty Line Tables

(and Tables Pertaining to All Seven Poverty Lines)

Figure 4 (National poverty line): Estimated poverty likelihoods associated with scores

If a household's score is . . .	. . . then the likelihood (%) of being below the poverty line is:
0–4	86.4
5–9	60.8
10–14	59.4
15–19	47.6
20–24	36.3
25–29	32.8
30–34	21.6
35–39	19.5
40–44	10.8
45–49	6.8
50–54	3.9
55–59	4.4
60–64	0.8
65–69	0.1
70–74	0.0
75–79	0.0
80–84	0.0
85–89	0.0
90–94	0.0
95–100	0.0

Based on the 2005/06 HBS

Figure 5 (National poverty line): Derivation of estimated poverty likelihoods associated with scores

Score	Households at score and below poverty line		All households at score		Poverty likelihood (estimated, %)
0–4	1,245	÷	1,442	=	86.4
5–9	3,339	÷	5,487	=	60.8
10–14	4,570	÷	7,688	=	59.4
15–19	5,609	÷	11,779	=	47.6
20–24	4,346	÷	11,977	=	36.3
25–29	4,231	÷	12,891	=	32.8
30–34	2,595	÷	12,032	=	21.6
35–39	1,674	÷	8,581	=	19.5
40–44	743	÷	6,868	=	10.8
45–49	477	÷	6,968	=	6.8
50–54	199	÷	5,113	=	3.9
55–59	138	÷	3,162	=	4.4
60–64	20	÷	2,449	=	0.8
65–69	2	÷	1,854	=	0.1
70–74	0	÷	716	=	0.0
75–79	0	÷	487	=	0.0
80–84	0	÷	328	=	0.0
85–89	0	÷	152	=	0.0
90–94	0	÷	28	=	0.0
95–100	0	÷	0	=	0.0

Number of all households normalized to sum to 100,000.

Based on the 2005/06 HBS.

Figure 6 (All poverty lines): Distribution of household poverty likelihoods across consumption ranges demarcated by poverty lines

	Likelihood of having expenditure in range demarcated by poverty lines per day per capita							
	<$1.25/day	=>$1.25/day and <Food	=>Food and <USAID	=>USAID and <National	=>National and <$2.50/day	=>$2.50/day and <$3.00/day	=>$3.00/day and <$4.00/day	=>$4.00/day
Score	<YR117	=>YR117 and <YR124	=>YR124 and <YR137	=>YR137 and <YR179	=>YR179 and <YR235	=>YR235 and <YR282	=>YR282 and <YR376	=>YR376
0–4	39.8	4.1	3.5	38.9	5.9	0.0	7.8	0.0
5–9	25.3	4.1	5.7	25.7	20.3	9.4	5.2	4.2
10–14	19.2	4.1	6.2	29.9	22.5	6.3	9.2	2.7
15–19	16.4	5.2	7.6	18.4	23.0	13.4	10.5	5.5
20–24	8.6	2.3	6.0	19.5	24.7	14.4	17.3	7.3
25–29	7.0	1.0	7.0	17.7	26.6	15.0	12.8	12.8
30–34	4.2	1.0	5.2	11.2	21.2	14.1	21.8	21.3
35–39	5.1	0.3	3.8	10.4	17.8	15.2	21.2	26.3
40–44	1.9	1.2	0.3	7.5	14.4	18.1	25.9	30.8
45–49	1.0	0.0	0.5	5.4	13.2	13.0	19.6	47.4
50–54	0.3	0.0	0.3	3.3	8.6	10.0	23.6	54.0
55–59	1.0	0.0	0.6	2.8	13.6	8.9	19.0	54.2
60–64	0.0	0.0	0.0	0.8	2.7	4.7	20.8	71.0
65–69	0.0	0.0	0.1	0.0	2.2	3.4	14.4	79.9
70–74	0.0	0.0	0.0	0.0	1.5	1.2	2.6	94.7
75–79	0.0	0.0	0.0	0.0	0.0	0.0	2.8	97.2
80–84	0.0	0.0	0.0	0.0	0.0	0.0	6.5	93.5
85–89	0.0	0.0	0.0	0.0	0.0	4.0	7.5	88.5
90–94	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
95–100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0

All poverty likelihoods are in percentage units.

Figure 7 (National poverty line): Bootstrapped differences between estimated and true poverty likelihoods for households in a large sample (n = 16,384) with confidence intervals, scorecard applied to the validation sample

	Difference between estimate and true value			
		Confidence interval (+/− percentage points)		
Score	**Diff.**	**90-percent**	**95-percent**	**99-percent**
0–4	+9.1	4.9	5.7	7.6
5–9	–2.2	2.7	3.4	4.5
10–14	+1.1	2.5	2.9	3.6
15–19	+2.0	2.1	2.5	3.2
20–24	–4.0	3.1	3.3	3.7
25–29	+9.2	1.7	2.1	2.7
30–34	–0.8	1.7	2.0	2.5
35–39	+9.9	1.3	1.5	2.0
40–44	–9.0	5.7	6.0	6.6
45–49	+2.4	1.0	1.2	1.7
50–54	–2.4	2.0	2.2	2.5
55–59	+4.0	0.3	0.4	0.5
60–64	–7.9	5.8	6.3	7.0
65–69	–0.9	0.9	1.0	1.3
70–74	+0.0	0.0	0.0	0.0
75–79	+0.0	0.0	0.0	0.0
80–84	+0.0	0.0	0.0	0.0
85–89	+0.0	0.0	0.0	0.0
90–94	+0.0	0.0	0.0	0.0
95–100	+0.0	0.0	0.0	0.0

Figure 8 (All poverty lines): Differences, precision of differences, and the α factor for bootstrapped estimates of poverty rates for groups of households at a point in time, scorecard applied to the validation sample

	Poverty line						
		National	USAID	International (2005 PPP)			
	National	Food	'Extreme'	$1.25/day	$2.50/day	$3.00/day	$4.00/day
Estimate minus true value							
Scorecard applied to the validation sample	+1.5	+0.6	+1.3	+0.5	+2.0	–0.6	–0.6
Precision of difference							
Scorecard applied to the validation sample	0.6	0.4	0.5	0.4	0.7	0.7	0.5
α for sample size							
Scorecard applied to the validation sample	1.07	1.10	1.10	1.11	1.06	1.02	0.94

Precision is measured as 90-percent confidence intervals in units of +/– percentage points.
Differences and precision estimated from 1,000 bootstraps of size n = 16,384.
α is estimated from 1,000 bootstrap samples of n = 256, 512, 1,024, 2,048, 4,096, 8,192, and 16,384.

Figure 9 (National poverty line): Differences and precision of differences for bootstrapped estimates of poverty rates for groups of households at a point in time, by sample size, scorecard applied to the validation sample

Sample Size	Difference between estimate and true value			
		Confidence interval (+/− percentage points)		
n	Diff.	90-percent	95-percent	99-percent
1	−1.3	68.9	70.0	78.5
4	+1.2	38.3	43.4	55.3
8	+1.2	28.5	33.1	39.2
16	+1.2	20.2	23.8	29.6
32	+1.5	14.5	17.5	22.1
64	+1.5	10.4	12.3	15.6
128	+1.3	7.5	8.6	10.2
256	+1.4	5.0	6.0	8.0
512	+1.5	3.4	4.0	5.8
1,024	+1.5	2.4	2.9	3.8
2,048	+1.5	1.7	2.0	2.9
4,096	+1.5	1.3	1.5	2.1
8,192	+1.5	0.9	1.0	1.4
16,384	+1.5	0.6	0.7	1.0

Figure 10 (All poverty lines): Possible types of outcomes from targeting by poverty score

True poverty status	Targeting segment: Targeted	Targeting segment: Non-targeted
Below poverty line	**Inclusion** Under poverty line Correctly Targeted	**Undercoverage** Under poverty line Mistakenly Non-targeted
Above poverty line	**Leakage** Above poverty line Mistakenly Targeted	**Exclusion** Above poverty line Correctly Non-targeted

Figure 11 (National poverty line): Households by targeting classification and score, along with "Total Accuracy" and BPAC, scorecard applied to the validation sample

Score	Inclusion: < poverty line correctly targeted	Undercoverage: < poverty line mistakenly non-targeted	Leakage: => poverty line mistakenly targeted	Exclusion: => poverty line correctly non-targeted	Total Accuracy Inclusion + Exclusion	BPAC See text
0–4	1.1	28.0	0.3	70.5	71.7	–91.2
5–9	4.7	24.5	2.3	68.6	73.3	–60.2
10–14	9.1	20.1	5.5	65.3	74.4	–18.6
15–19	14.6	14.5	11.8	59.1	73.7	+40.8
20–24	19.6	9.6	18.8	52.0	71.6	+35.5
25–29	23.2	6.0	28.1	42.8	66.0	+3.7
30–34	26.0	3.2	37.3	33.5	59.5	–28.0
35–39	27.1	2.0	44.7	26.1	53.3	–53.4
40–44	28.3	0.9	50.5	20.4	48.7	–73.1
45–49	28.7	0.5	57.0	13.8	42.5	–95.6
50–54	29.0	0.2	61.8	9.0	38.0	–112.0
55–59	29.0	0.1	65.0	5.9	34.9	–122.8
60–64	29.1	0.0	67.3	3.5	32.7	–130.8
65–69	29.2	0.0	69.1	1.7	30.9	–137.1
70–74	29.2	0.0	69.8	1.0	30.2	–139.6
75–79	29.2	0.0	70.3	0.5	29.7	–141.2
80–84	29.2	0.0	70.7	0.2	29.3	–142.4
85–89	29.2	0.0	70.8	0.0	29.2	–142.9
90–94	29.2	0.0	70.8	0.0	29.2	–143.0
95–100	29.2	0.0	70.8	0.0	29.2	–143.0

Inclusion, undercoverage, leakage, and exclusion normalized to sum to 100.

Figure 12 (National poverty line): For a given score cut-off, the percentage of all households who are targeted (that is, have a score equal to or less than the cut-off), the percentage of targeted households who are poor (that is, below the poverty line), the percentage of poor households who are targeted, and the number of poor households who are successful targeted (coverage) per non-poor household mistakenly targeted (leakage), scorecard applied to the validation sample

Targeting cut-off	% all households who are targeted	% targeted who are poor	% of poor who are targeted	Poor households targeted per non-poor household targeted
0–4	1.4	78.9	3.9	3.7:1
5–9	6.9	67.4	16.0	2.1:1
10–14	14.6	62.3	31.2	1.7:1
15–19	26.4	55.5	50.2	1.2:1
20–24	38.4	51.0	67.1	1.0:1
25–29	51.3	45.2	79.5	0.8:1
30–34	63.3	41.0	89.1	0.7:1
35–39	71.9	37.8	93.1	0.6:1
40–44	78.7	35.9	97.0	0.6:1
45–49	85.7	33.5	98.4	0.5:1
50–54	90.8	31.9	99.5	0.5:1
55–59	94.0	30.9	99.6	0.4:1
60–64	96.4	30.2	99.9	0.4:1
65–69	98.3	29.7	100.0	0.4:1
70–74	99.0	29.5	100.0	0.4:1
75–79	99.5	29.3	100.0	0.4:1
80–84	99.8	29.2	100.0	0.4:1
85–89	100.0	29.2	100.0	0.4:1
90–94	100.0	29.2	100.0	0.4:1
95–100	100.0	29.2	100.0	0.4:1

National Food Poverty Line Tables

Figure 4 (National food poverty line): Estimated poverty likelihoods associated with scores

If a household's score is . . .	. . . then the likelihood (%) of being below the poverty line is:
0–4	44.0
5–9	29.4
10–14	23.3
15–19	21.6
20–24	10.8
25–29	8.0
30–34	5.2
35–39	5.3
40–44	3.0
45–49	1.0
50–54	0.3
55–59	1.0
60–64	0.0
65–69	0.0
70–74	0.0
75–79	0.0
80–84	0.0
85–89	0.0
90–94	0.0
95–100	0.0

Based on the 2005/06 HBS

Figure 7 (National food poverty line): Bootstrapped differences between estimated and true poverty likelihoods for households in a large sample (n = 16,384) with confidence intervals, scorecard applied to the validation sample

	Difference between estimate and true value			
		Confidence interval (+/− percentage points)		
Score	**Diff.**	**90-percent**	**95-percent**	**99-percent**
0–4	–7.9	6.8	7.5	8.7
5–9	+4.7	2.5	3.0	3.9
10–14	–8.1	5.2	5.5	5.9
15–19	+6.0	1.5	1.8	2.2
20–24	–0.8	1.3	1.6	2.2
25–29	+2.5	0.9	1.0	1.3
30–34	+1.9	0.6	0.7	1.0
35–39	+3.2	0.6	0.7	0.9
40–44	–2.1	1.8	1.9	2.1
45–49	–0.2	0.5	0.6	0.8
50–54	–1.6	1.3	1.4	1.5
55–59	+1.0	0.0	0.0	0.0
60–64	–7.9	5.8	6.2	7.1
65–69	–0.2	0.3	0.3	0.4
70–74	+0.0	0.0	0.0	0.0
75–79	+0.0	0.0	0.0	0.0
80–84	+0.0	0.0	0.0	0.0
85–89	+0.0	0.0	0.0	0.0
90–94	+0.0	0.0	0.0	0.0
95–100	+0.0	0.0	0.0	0.0

Figure 9 (National food poverty line): Differences and precision of differences for bootstrapped estimates of poverty rates for groups of households at a point in time, by sample size, scorecard applied to the validation sample

Sample Size	Difference between estimate and true value			
		Confidence interval (+/− percentage points)		
n	Diff.	90-percent	95-percent	99-percent
1	+0.6	50.9	59.3	62.1
4	+0.5	26.5	32.5	41.8
8	+0.7	18.6	22.4	28.4
16	+0.7	13.3	16.1	21.2
32	+0.7	9.8	11.7	13.8
64	+0.6	6.9	8.2	11.0
128	+0.5	4.7	5.8	7.2
256	+0.6	3.4	4.0	5.1
512	+0.6	2.4	2.9	3.7
1,024	+0.6	1.7	2.0	2.6
2,048	+0.6	1.1	1.4	2.0
4,096	+0.6	0.9	1.0	1.6
8,192	+0.6	0.6	0.7	0.9
16,384	+0.6	0.4	0.5	0.7

Figure 11 (National food poverty line): Households by targeting classification and score, along with "Total Accuracy" and BPAC, scorecard applied to the validation sample

Score	Inclusion: < poverty line correctly targeted	Undercoverage: < poverty line mistakenly non-targeted	Leakage: => poverty line mistakenly targeted	Exclusion: => poverty line correctly non-targeted	Total Accuracy Inclusion + Exclusion	BPAC See text
0–4	0.7	9.4	0.7	89.2	89.9	–78.7
5–9	2.2	7.9	4.8	85.1	87.3	–10.1
10–14	4.5	5.6	10.1	79.8	84.3	+0.4
15–19	6.4	3.7	19.9	69.9	76.4	–97.2
20–24	7.9	2.2	30.5	59.4	67.3	–201.1
25–29	8.8	1.3	42.5	47.4	56.2	–319.7
30–34	9.4	0.8	53.9	36.0	45.3	–433.0
35–39	9.6	0.5	62.3	27.6	37.3	–515.3
40–44	9.9	0.3	68.9	21.0	30.9	–580.8
45–49	10.0	0.1	75.7	14.1	24.1	–648.6
50–54	10.0	0.1	80.8	9.1	19.1	–698.5
55–59	10.0	0.1	83.9	5.9	16.0	–729.7
60–64	10.1	0.0	86.3	3.6	13.7	–753.2
65–69	10.1	0.0	88.2	1.7	11.8	–771.5
70–74	10.1	0.0	88.9	1.0	11.1	–778.6
75–79	10.1	0.0	89.4	0.5	10.6	–783.4
80–84	10.1	0.0	89.7	0.2	10.3	–786.6
85–89	10.1	0.0	89.9	0.0	10.1	–788.1
90–94	10.1	0.0	89.9	0.0	10.1	–788.4
95–100	10.1	0.0	89.9	0.0	10.1	–788.4

Inclusion, undercoverage, leakage, and exclusion normalized to sum to 100.

Figure 12 (National food poverty line): For a given score cut-off, the percentage of all households who are targeted (that is, have a score equal to or less than the cut-off), the percentage of targeted households who are poor (that is, below the poverty line), the percentage of poor households who are targeted, and the number of poor households who are successful targeted (coverage) per non-poor household mistakenly targeted (leakage), scorecard applied to the validation sample

Targeting cut-off	% all households who are targeted	% targeted who are poor	% of poor who are targeted	Poor households targeted per non-poor household targeted
0–4	1.4	49.5	7.1	1.0:1
5–9	6.9	31.3	21.4	0.5:1
10–14	14.6	31.1	44.9	0.5:1
15–19	26.4	24.4	63.7	0.3:1
20–24	38.4	20.6	78.2	0.3:1
25–29	51.3	17.2	87.0	0.2:1
30–34	63.3	14.8	92.6	0.2:1
35–39	71.9	13.4	95.1	0.2:1
40–44	78.7	12.5	97.5	0.1:1
45–49	85.7	11.6	98.6	0.1:1
50–54	90.8	11.1	99.2	0.1:1
55–59	94.0	10.7	99.2	0.1:1
60–64	96.4	10.5	99.9	0.1:1
65–69	98.3	10.3	100.0	0.1:1
70–74	99.0	10.2	100.0	0.1:1
75–79	99.5	10.2	100.0	0.1:1
80–84	99.8	10.1	100.0	0.1:1
85–89	100.0	10.1	100.0	0.1:1
90–94	100.0	10.1	100.0	0.1:1
95–100	100.0	10.1	100.0	0.1:1

USAID "Extreme" Poverty Line Tables

Figure 4 (USAID "Extreme" poverty line): Estimated poverty likelihoods associated with scores

If a household's score is . . .	. . . then the likelihood (%) of being below the poverty line is:
0–4	47.4
5–9	35.1
10–14	29.6
15–19	29.2
20–24	16.8
25–29	15.1
30–34	10.4
35–39	9.1
40–44	3.4
45–49	1.5
50–54	0.6
55–59	1.6
60–64	0.0
65–69	0.1
70–74	0.0
75–79	0.0
80–84	0.0
85–89	0.0
90–94	0.0
95–100	0.0

Based on the 2005/06 HBS

Figure 7 (USAID "Extreme" poverty line): Bootstrapped differences between estimated and true poverty likelihoods for households in a large sample (n = 16,384) with confidence intervals, scorecard applied to the validation sample

	Difference between estimate and true value			
		Confidence interval (+/− percentage points)		
Score	Diff.	90-percent	95-percent	99-percent
0–4	–13.1	9.4	9.9	10.9
5–9	+5.3	2.7	3.1	4.1
10–14	–8.2	5.4	5.6	6.3
15–19	+7.3	1.8	2.0	2.7
20–24	–0.4	1.5	1.8	2.4
25–29	+5.8	1.1	1.3	1.7
30–34	+0.3	1.3	1.5	1.9
35–39	+5.1	0.8	1.0	1.2
40–44	–2.4	1.9	2.1	2.3
45–49	+0.3	0.5	0.6	0.8
50–54	–1.3	1.1	1.2	1.4
55–59	+1.6	0.0	0.0	0.0
60–64	–7.9	5.8	6.2	7.1
65–69	–0.1	0.3	0.3	0.4
70–74	+0.0	0.0	0.0	0.0
75–79	+0.0	0.0	0.0	0.0
80–84	+0.0	0.0	0.0	0.0
85–89	+0.0	0.0	0.0	0.0
90–94	+0.0	0.0	0.0	0.0
95–100	+0.0	0.0	0.0	0.0

Figure 9 (USAID "Extreme" poverty line): Differences and precision of differences for bootstrapped estimates of poverty rates for groups of households at a point in time, by sample size, scorecard applied to the validation sample

Sample Size	Difference between estimate and true value			
		Confidence interval (+/− percentage points)		
n	Diff.	90-percent	95-percent	99-percent
1	+0.6	56.4	60.0	65.9
4	+1.4	30.5	35.5	45.9
8	+1.0	21.7	26.2	33.4
16	+1.1	16.1	19.5	27.0
32	+1.3	11.7	14.1	18.3
64	+1.2	8.1	9.9	12.9
128	+1.2	5.9	6.8	8.8
256	+1.3	3.9	4.6	6.2
512	+1.2	2.8	3.4	4.8
1,024	+1.3	2.0	2.3	3.2
2,048	+1.3	1.4	1.7	2.4
4,096	+1.3	1.0	1.2	1.6
8,192	+1.3	0.7	0.8	1.1
16,384	+1.3	0.5	0.6	0.8

Figure 11 (USAID "Extreme" poverty line): Households by targeting classification and score, along with "Total Accuracy" and BPAC, scorecard applied to the validation sample

Score	Inclusion: < poverty line correctly targeted	Undercoverage: < poverty line mistakenly non-targeted	Leakage: => poverty line mistakenly targeted	Exclusion: => poverty line correctly non-targeted	Total Accuracy Inclusion + Exclusion	BPAC See text
0–4	0.8	13.4	0.6	85.2	86.0	–83.9
5–9	2.6	11.6	4.3	81.4	84.0	–33.2
10–14	5.5	8.7	9.1	76.7	82.2	+36.0
15–19	8.2	6.1	18.2	67.5	75.7	–28.1
20–24	10.4	3.9	28.0	57.8	68.1	–96.8
25–29	11.9	2.4	39.4	46.4	58.3	–176.8
30–34	13.1	1.1	50.2	35.6	48.7	–252.6
35–39	13.6	0.6	58.2	27.5	41.2	–309.2
40–44	14.0	0.3	64.8	21.0	35.0	–355.1
45–49	14.1	0.1	71.6	14.1	28.2	–403.2
50–54	14.2	0.1	76.7	9.1	23.3	–438.7
55–59	14.2	0.1	79.8	5.9	20.1	–460.9
60–64	14.2	0.0	82.2	3.6	17.8	–477.6
65–69	14.2	0.0	84.1	1.7	15.9	–490.6
70–74	14.2	0.0	84.8	1.0	15.2	–495.6
75–79	14.2	0.0	85.3	0.5	14.7	–499.1
80–84	14.2	0.0	85.6	0.2	14.4	–501.4
85–89	14.2	0.0	85.7	0.0	14.3	–502.4
90–94	14.2	0.0	85.8	0.0	14.2	–502.6
95–100	14.2	0.0	85.8	0.0	14.2	–502.6

Inclusion, undercoverage, leakage, and exclusion normalized to sum to 100.

Figure 12 (USAID "Extreme" poverty line): For a given score cut-off, the percentage of all households who are targeted (that is, have a score equal to or less than the cut-off), the percentage of targeted households who are poor (that is, below the poverty line), the percentage of poor households who are targeted, and the number of poor households who are successful targeted (coverage) per non-poor household mistakenly targeted (leakage), scorecard applied to the validation sample

Targeting cut-off	% all households who are targeted	% targeted who are poor	% of poor who are targeted	Poor households targeted per non-poor household targeted
0–4	1.4	58.8	6.0	1.4:1
5–9	6.9	37.3	18.2	0.6:1
10–14	14.6	37.7	38.7	0.6:1
15–19	26.4	30.9	57.3	0.4:1
20–24	38.4	27.0	72.8	0.4:1
25–29	51.3	23.2	83.4	0.3:1
30–34	63.3	20.7	92.2	0.3:1
35–39	71.9	19.0	95.8	0.2:1
40–44	78.7	17.7	98.2	0.2:1
45–49	85.7	16.4	99.0	0.2:1
50–54	90.8	15.6	99.5	0.2:1
55–59	94.0	15.1	99.5	0.2:1
60–64	96.4	14.7	99.9	0.2:1
65–69	98.3	14.5	100.0	0.2:1
70–74	99.0	14.4	100.0	0.2:1
75–79	99.5	14.3	100.0	0.2:1
80–84	99.8	14.3	100.0	0.2:1
85–89	100.0	14.2	100.0	0.2:1
90–94	100.0	14.2	100.0	0.2:1
95–100	100.0	14.2	100.0	0.2:1

USD1.25/Day 2005 PPP Poverty Line Tables

Figure 4 (USD1.25/day 2005 PPP line): Estimated poverty likelihoods associated with scores

If a household's score is . . .	. . . then the likelihood (%) of being below the poverty line is:
0–4	39.8
5–9	25.3
10–14	19.2
15–19	16.4
20–24	8.6
25–29	7.0
30–34	4.2
35–39	5.1
40–44	1.9
45–49	1.0
50–54	0.3
55–59	1.0
60–64	0.0
65–69	0.0
70–74	0.0
75–79	0.0
80–84	0.0
85–89	0.0
90–94	0.0
95–100	0.0

Based on the 2005/06 HBS

Figure 7 (USD1.25/day 2005 PPP line): Bootstrapped differences between estimated and true poverty likelihoods for households in a large sample (n = 16,384) with confidence intervals, scorecard applied to the validation sample

	Difference between estimate and true value			
		Confidence interval (+/− percentage points)		
Score	Diff.	90-percent	95-percent	99-percent
0–4	–12.1	8.9	9.6	10.5
5–9	+4.6	2.4	2.9	3.9
10–14	–5.8	4.0	4.2	4.6
15–19	+3.1	1.4	1.6	2.2
20–24	–0.5	1.2	1.4	1.8
25–29	+2.1	0.8	1.0	1.3
30–34	+2.1	0.5	0.5	0.7
35–39	+3.0	0.6	0.7	0.9
40–44	–3.3	2.4	2.5	2.7
45–49	+0.2	0.5	0.5	0.7
50–54	+0.3	0.0	0.0	0.0
55–59	+1.0	0.0	0.0	0.0
60–64	+0.0	0.0	0.0	0.0
65–69	+0.0	0.0	0.0	0.0
70–74	+0.0	0.0	0.0	0.0
75–79	+0.0	0.0	0.0	0.0
80–84	+0.0	0.0	0.0	0.0
85–89	+0.0	0.0	0.0	0.0
90–94	+0.0	0.0	0.0	0.0
95–100	+0.0	0.0	0.0	0.0

Figure 9 (USD1.25/day 2005 PPP line): Differences and precision of differences for bootstrapped estimates of poverty rates for groups of households at a point in time, by sample size, scorecard applied to the validation sample

Sample Size	Difference between estimate and true value			
		Confidence interval (+/− percentage points)		
n	Diff.	90-percent	95-percent	99-percent
1	+0.0	50.0	56.4	59.6
4	+0.4	24.7	28.6	40.4
8	+0.7	16.5	19.7	26.2
16	+0.6	12.0	14.3	19.8
32	+0.5	8.9	10.4	14.0
64	+0.4	6.2	7.3	10.0
128	+0.4	4.2	5.1	6.8
256	+0.5	3.1	3.6	4.9
512	+0.5	2.3	2.7	3.5
1,024	+0.5	1.6	1.8	2.5
2,048	+0.5	1.1	1.3	1.7
4,096	+0.5	0.8	1.0	1.3
8,192	+0.5	0.5	0.6	0.9
16,384	+0.5	0.4	0.5	0.7

Figure 11 (USD1.25/day 2005 PPP line): Households by targeting classification and score, along with "Total Accuracy" and BPAC, scorecard applied to the validation sample

Score	Inclusion: < poverty line correctly targeted	Undercoverage: < poverty line mistakenly non-targeted	Leakage: => poverty line mistakenly targeted	Exclusion: => poverty line correctly non-targeted	Total Accuracy Inclusion + Exclusion	BPAC See text
0–4	0.7	7.8	0.7	90.8	91.5	–74.6
5–9	2.0	6.5	4.9	86.6	88.5	+5.0
10–14	3.9	4.5	10.7	80.8	84.8	–25.8
15–19	5.6	2.9	20.8	70.7	76.2	–145.5
20–24	6.8	1.7	31.6	59.9	66.7	–272.5
25–29	7.5	1.0	43.7	47.8	55.3	–415.4
30–34	7.9	0.6	55.4	36.2	44.1	–552.3
35–39	8.2	0.3	63.7	27.8	36.0	–650.6
40–44	8.4	0.1	70.3	21.2	29.6	–728.6
45–49	8.5	0.0	77.2	14.3	22.8	–809.9
50–54	8.5	0.0	82.3	9.2	17.7	–870.2
55–59	8.5	0.0	85.5	6.0	14.5	–907.4
60–64	8.5	0.0	87.9	3.6	12.1	–936.3
65–69	8.5	0.0	89.8	1.7	10.2	–958.1
70–74	8.5	0.0	90.5	1.0	9.5	–966.5
75–79	8.5	0.0	91.0	0.5	9.0	–972.3
80–84	8.5	0.0	91.3	0.2	8.7	–976.1
85–89	8.5	0.0	91.5	0.0	8.5	–977.9
90–94	8.5	0.0	91.5	0.0	8.5	–978.3
95–100	8.5	0.0	91.5	0.0	8.5	–978.3

Inclusion, undercoverage, leakage, and exclusion normalized to sum to 100.

Figure 12 (USD1.25/day 2005 PPP line): For a given score cut-off, the percentage of all households who are targeted (that is, have a score equal to or less than the cut-off), the percentage of targeted households who are poor (that is, below the poverty line), the percentage of poor households who are targeted, and the number of poor households who are successful targeted (coverage) per non-poor household mistakenly targeted (leakage), scorecard applied to the validation sample

Targeting cut-off	% all households who are targeted	% targeted who are poor	% of poor who are targeted	Poor households targeted per non-poor household targeted
0–4	1.4	49.5	8.4	1.0:1
5–9	6.9	28.6	23.3	0.4:1
10–14	14.6	27.0	46.4	0.4:1
15–19	26.4	21.1	65.5	0.3:1
20–24	38.4	17.6	79.6	0.2:1
25–29	51.3	14.7	88.6	0.2:1
30–34	63.3	12.5	93.5	0.1:1
35–39	71.9	11.4	96.3	0.1:1
40–44	78.7	10.7	99.2	0.1:1
45–49	85.7	9.9	100.0	0.1:1
50–54	90.8	9.3	100.0	0.1:1
55–59	94.0	9.0	100.0	0.1:1
60–64	96.4	8.8	100.0	0.1:1
65–69	98.3	8.6	100.0	0.1:1
70–74	99.0	8.6	100.0	0.1:1
75–79	99.5	8.5	100.0	0.1:1
80–84	99.8	8.5	100.0	0.1:1
85–89	100.0	8.5	100.0	0.1:1
90–94	100.0	8.5	100.0	0.1:1
95–100	100.0	8.5	100.0	0.1:1

USD2.50/Day 2005 PPP Poverty Line Tables

Figure 4 (USD2.50/day 2005 PPP line): Estimated poverty likelihoods associated with scores

If a household's score is . . .	. . . then the likelihood (%) of being below the poverty line is:
0–4	92.2
5–9	81.1
10–14	81.9
15–19	70.6
20–24	61.0
25–29	59.5
30–34	42.8
35–39	37.3
40–44	25.2
45–49	20.1
50–54	12.5
55–59	17.9
60–64	3.5
65–69	2.3
70–74	1.5
75–79	0.0
80–84	0.0
85–89	0.0
90–94	0.0
95–100	0.0

Based on the 2005/06 HBS

Figure 7 (USD2.50/day 2005 PPP line): Bootstrapped differences between estimated and true poverty likelihoods for households in a large sample (n = 16,384) with confidence intervals, scorecard applied to the validation sample

	Difference between estimate and true value			
		Confidence interval (+/– percentage points)		
Score	Diff.	90-percent	95-percent	99-percent
0–4	+0.6	3.1	3.6	4.9
5–9	–4.6	3.3	3.5	3.8
10–14	+11.4	2.2	2.7	3.2
15–19	+0.5	2.0	2.4	3.2
20–24	–3.2	2.5	2.8	3.1
25–29	+9.2	2.3	2.7	3.5
30–34	–5.7	3.8	4.1	4.5
35–39	+14.1	2.0	2.4	3.2
40–44	–8.2	5.6	5.8	6.2
45–49	+1.1	2.3	2.7	3.9
50–54	+2.2	1.9	2.2	3.1
55–59	+2.8	3.8	4.6	6.2
60–64	–9.1	6.3	6.9	7.7
65–69	+1.1	0.9	1.0	1.3
70–74	+1.5	0.0	0.0	0.0
75–79	+0.0	0.0	0.0	0.0
80–84	+0.0	0.0	0.0	0.0
85–89	+0.0	0.0	0.0	0.0
90–94	+0.0	0.0	0.0	0.0
95–100	+0.0	0.0	0.0	0.0

Figure 9 (USD2.50/day 2005 PPP line): Differences and precision of differences for bootstrapped estimates of poverty rates for groups of households at a point in time, by sample size, scorecard applied to the validation sample

Sample Size	Difference between estimate and true value			
		Confidence interval (+/– percentage points)		
n	Diff.	90-percent	95-percent	99-percent
1	–0.4	66.7	80.5	84.7
4	+1.0	43.2	47.3	59.7
8	+1.4	31.4	37.2	45.9
16	+1.6	21.6	25.5	33.5
32	+1.8	16.3	19.0	25.1
64	+1.8	11.0	13.0	16.6
128	+1.8	7.5	8.9	11.6
256	+1.9	5.4	6.1	8.8
512	+1.9	3.8	4.5	5.7
1,024	+2.0	2.7	3.2	4.4
2,048	+2.0	1.9	2.2	2.9
4,096	+2.0	1.4	1.6	2.1
8,192	+2.0	1.0	1.1	1.5
16,384	+2.0	0.7	0.8	1.1

Figure 11 (USD2.50/day 2005 PPP line): Households by targeting classification and score, along with "Total Accuracy" and BPAC, scorecard applied to the validation sample

Score	Inclusion: < poverty line correctly targeted	Undercoverage: < poverty line mistakenly non-targeted	Leakage: => poverty line mistakenly targeted	Exclusion: => poverty line correctly non-targeted	Total Accuracy Inclusion + Exclusion	BPAC See text
0–4	1.3	46.6	0.1	51.9	53.3	–94.2
5–9	6.0	41.9	0.9	51.2	57.2	–73.0
10–14	11.5	36.5	3.2	48.9	60.4	–45.6
15–19	20.0	27.9	6.4	45.7	65.7	–3.2
20–24	27.7	20.2	10.6	41.4	69.2	+37.9
25–29	34.9	13.0	16.4	35.7	70.6	+65.8
30–34	40.6	7.3	22.7	29.4	70.0	+52.6
35–39	43.2	4.8	28.7	23.4	66.5	+40.1
40–44	45.4	2.5	33.3	18.8	64.2	+30.5
45–49	46.7	1.2	39.0	13.1	59.8	+18.6
50–54	47.3	0.6	43.5	8.6	55.9	+9.3
55–59	47.7	0.3	46.3	5.7	53.4	+3.3
60–64	47.9	0.0	48.5	3.5	51.4	–1.3
65–69	47.9	0.0	50.4	1.7	49.6	–5.1
70–74	47.9	0.0	51.1	1.0	48.9	–6.6
75–79	47.9	0.0	51.6	0.5	48.4	–7.6
80–84	47.9	0.0	51.9	0.2	48.1	–8.3
85–89	47.9	0.0	52.0	0.0	48.0	–8.6
90–94	47.9	0.0	52.1	0.0	47.9	–8.6
95–100	47.9	0.0	52.1	0.0	47.9	–8.6

Inclusion, undercoverage, leakage, and exclusion normalized to sum to 100.

Figure 12 (USD2.50/day 2005 PPP line): For a given score cut-off, the percentage of all households who are targeted (that is, have a score equal to or less than the cut-off), the percentage of targeted households who are poor (that is, below the poverty line), the percentage of poor households who are targeted, and the number of poor households who are successful targeted (coverage) per non-poor household mistakenly targeted (leakage), scorecard applied to the validation sample

Targeting cut-off	% all households who are targeted	% targeted who are poor	% of poor who are targeted	Poor households targeted per non-poor household targeted
0–4	1.4	91.3	2.7	10.5:1
5–9	6.9	86.9	12.6	6.6:1
10–14	14.6	78.4	23.9	3.6:1
15–19	26.4	75.9	41.8	3.1:1
20–24	38.4	72.3	57.9	2.6:1
25–29	51.3	68.1	72.8	2.1:1
30–34	63.3	64.1	84.7	1.8:1
35–39	71.9	60.0	90.1	1.5:1
40–44	78.7	57.7	94.8	1.4:1
45–49	85.7	54.5	97.4	1.2:1
50–54	90.8	52.1	98.8	1.1:1
55–59	94.0	50.7	99.4	1.0:1
60–64	96.4	49.7	99.9	1.0:1
65–69	98.3	48.8	100.0	1.0:1
70–74	99.0	48.4	100.0	0.9:1
75–79	99.5	48.2	100.0	0.9:1
80–84	99.8	48.0	100.0	0.9:1
85–89	100.0	47.9	100.0	0.9:1
90–94	100.0	47.9	100.0	0.9:1
95–100	100.0	47.9	100.0	0.9:1

USD3.00/Day 2005 PPP Poverty Line Tables

Figure 4 (USD3.00/day 2005 PPP line): Estimated poverty likelihoods associated with scores

If a household's score is . . .	. . . then the likelihood (%) of being below the poverty line is:
0–4	92.2
5–9	90.5
10–14	88.2
15–19	84.0
20–24	75.5
25–29	74.4
30–34	56.9
35–39	52.5
40–44	43.3
45–49	33.1
50–54	22.4
55–59	26.8
60–64	8.3
65–69	5.7
70–74	2.8
75–79	0.0
80–84	0.0
85–89	4.0
90–94	0.0
95–100	0.0

Based on the 2005/06 HBS

Figure 7 (USD3.00/day 2005 PPP line): Bootstrapped differences between estimated and true poverty likelihoods for households in a large sample (n = 16,384) with confidence intervals, scorecard applied to the validation sample

	Difference between estimate and true value			
		Confidence interval (+/– percentage points)		
Score	Diff.	90-percent	95-percent	99-percent
0–4	–3.8	2.9	3.0	3.3
5–9	–2.1	1.8	1.9	2.3
10–14	+1.8	1.7	1.9	2.5
15–19	+0.3	1.6	1.9	2.6
20–24	–0.9	1.7	2.1	2.8
25–29	+3.4	2.1	2.5	3.3
30–34	–10.4	6.2	6.3	6.7
35–39	+2.4	2.7	3.1	4.4
40–44	–1.9	3.1	3.6	4.6
45–49	–1.4	2.7	3.2	4.6
50–54	+4.0	2.6	3.1	4.1
55–59	+6.2	4.1	5.0	6.4
60–64	–5.9	4.9	5.2	6.5
65–69	+2.7	1.4	1.6	2.1
70–74	–2.4	3.5	4.3	5.7
75–79	+0.0	0.0	0.0	0.0
80–84	+0.0	0.0	0.0	0.0
85–89	+4.0	0.0	0.0	0.0
90–94	+0.0	0.0	0.0	0.0
95–100	+0.0	0.0	0.0	0.0

Figure 9 (USD3.00/day 2005 PPP line): Differences and precision of differences for bootstrapped estimates of poverty rates for groups of households at a point in time, by sample size, scorecard applied to the validation sample

Sample Size	Difference between estimate and true value			
		Confidence interval (+/– percentage points)		
n	Diff.	90-percent	95-percent	99-percent
1	–1.3	71.2	75.5	91.1
4	–1.9	38.4	47.7	58.8
8	–1.1	29.3	34.9	46.8
16	–0.7	20.2	24.5	33.8
32	–0.6	14.5	18.0	23.2
64	–0.6	10.1	12.4	15.9
128	–0.6	7.2	8.7	11.8
256	–0.7	5.3	6.1	7.7
512	–0.6	3.5	4.3	5.5
1,024	–0.6	2.5	3.0	3.8
2,048	–0.6	1.8	2.1	2.6
4,096	–0.6	1.3	1.5	2.1
8,192	–0.6	0.9	1.1	1.5
16,384	–0.6	0.7	0.8	1.0

Figure 11 (USD3.00/day 2005 PPP line): Households by targeting classification and score, along with "Total Accuracy" and BPAC, scorecard applied to the validation sample

Score	Inclusion: < poverty line correctly targeted	Undercoverage: < poverty line mistakenly non-targeted	Leakage: => poverty line mistakenly targeted	Exclusion: => poverty line correctly non-targeted	Total Accuracy Inclusion + Exclusion	BPAC See text
0–4	1.4	60.5	0.1	38.0	39.4	–95.4
5–9	6.4	55.5	0.5	37.6	44.1	–78.4
10–14	13.0	48.9	1.6	36.5	49.5	–55.4
15–19	23.1	38.8	3.3	34.8	58.0	–20.0
20–24	32.3	29.6	6.1	32.0	64.3	+14.1
25–29	41.9	20.0	9.4	28.7	70.6	+50.5
30–34	49.9	12.0	13.4	24.7	74.7	+78.4
35–39	54.5	7.4	17.4	20.7	75.2	+71.9
40–44	57.6	4.3	21.2	16.9	74.5	+65.8
45–49	59.8	2.1	25.9	12.2	72.0	+58.2
50–54	60.9	1.0	29.9	8.2	69.1	+51.7
55–59	61.5	0.4	32.5	5.6	67.1	+47.5
60–64	61.8	0.1	34.7	3.4	65.2	+44.0
65–69	61.9	0.0	36.4	1.7	63.5	+41.2
70–74	61.9	0.0	37.1	1.0	62.9	+40.1
75–79	61.9	0.0	37.6	0.5	62.4	+39.3
80–84	61.9	0.0	37.9	0.2	62.1	+38.7
85–89	61.9	0.0	38.1	0.0	61.9	+38.5
90–94	61.9	0.0	38.1	0.0	61.9	+38.5
95–100	61.9	0.0	38.1	0.0	61.9	+38.5

Inclusion, undercoverage, leakage, and exclusion normalized to sum to 100.

Figure 12 (USD3.00/day 2005 PPP line): For a given score cut-off, the percentage of all households who are targeted (that is, have a score equal to or less than the cut-off), the percentage of targeted households who are poor (that is, below the poverty line), the percentage of poor households who are targeted, and the number of poor households who are successful targeted (coverage) per non-poor household mistakenly targeted (leakage), scorecard applied to the validation sample

Targeting cut-off	% all households who are targeted	% targeted who are poor	% of poor who are targeted	Poor households targeted per non-poor household targeted
0–4	1.4	95.8	2.2	22.6:1
5–9	6.9	93.0	10.4	13.2:1
10–14	14.6	89.0	21.0	8.1:1
15–19	26.4	87.6	37.4	7.1:1
20–24	38.4	84.1	52.1	5.3:1
25–29	51.3	81.7	67.7	4.5:1
30–34	63.3	78.9	80.7	3.7:1
35–39	71.9	75.8	88.0	3.1:1
40–44	78.7	73.1	93.0	2.7:1
45–49	85.7	69.8	96.7	2.3:1
50–54	90.8	67.1	98.4	2.0:1
55–59	94.0	65.4	99.3	1.9:1
60–64	96.4	64.1	99.8	1.8:1
65–69	98.3	62.9	99.9	1.7:1
70–74	99.0	62.5	100.0	1.7:1
75–79	99.5	62.2	100.0	1.6:1
80–84	99.8	62.0	100.0	1.6:1
85–89	100.0	61.9	100.0	1.6:1
90–94	100.0	61.9	100.0	1.6:1
95–100	100.0	61.9	100.0	1.6:1

USD4.00/Day 2005 PPP Poverty Line Tables

Figure 4 (USD4.00/day 2005 PPP line): Estimated poverty likelihoods associated with scores

If a household's score is . . .	. . . then the likelihood (%) of being below the poverty line is:
0–4	100.0
5–9	95.8
10–14	97.3
15–19	94.5
20–24	92.7
25–29	87.2
30–34	78.7
35–39	73.7
40–44	69.2
45–49	52.6
50–54	46.0
55–59	45.8
60–64	29.0
65–69	20.1
70–74	5.3
75–79	2.8
80–84	6.5
85–89	11.5
90–94	0.0
95–100	0.0

Based on the 2005/06 HBS

Figure 7 (USD4.00/day 2005 PPP line): Bootstrapped differences between estimated and true poverty likelihoods for households in a large sample (n = 16,384) with confidence intervals, scorecard applied to the validation sample

	Difference between estimate and true value			
		Confidence interval (+/– percentage points)		
Score	Diff.	90-percent	95-percent	99-percent
0–4	+0.0	0.0	0.0	0.0
5–9	–1.5	1.2	1.3	1.5
10–14	+0.6	0.8	0.9	1.2
15–19	–0.5	1.0	1.1	1.4
20–24	–0.3	1.1	1.3	1.6
25–29	–0.6	1.4	1.6	2.1
30–34	–6.5	4.0	4.1	4.4
35–39	–2.8	2.4	2.7	3.3
40–44	+10.8	3.2	3.7	4.6
45–49	–6.0	4.4	4.6	4.9
50–54	+5.9	3.6	4.3	5.4
55–59	+7.1	4.9	5.8	8.0
60–64	+5.0	4.2	5.0	7.4
65–69	–11.7	8.8	9.4	10.6
70–74	–4.5	4.8	5.5	6.8
75–79	–2.8	4.0	5.0	6.6
80–84	+5.7	1.0	1.2	1.7
85–89	+3.1	9.2	10.2	13.1
90–94	+0.0	0.0	0.0	0.0
95–100	+0.0	0.0	0.0	0.0

Figure 9 (USD4.00/day 2005 PPP line): Differences and precision of differences for bootstrapped estimates of poverty rates for groups of households at a point in time, by sample size, scorecard applied to the validation sample

Sample Size	Difference between estimate and true value			
		Confidence interval (+/– percentage points)		
n	Diff.	90-percent	95-percent	99-percent
1	–1.0	64.7	70.6	83.4
4	–1.7	31.2	37.6	48.5
8	–1.1	23.5	28.6	35.2
16	–0.9	15.8	19.0	24.4
32	–0.7	11.6	12.8	17.7
64	–0.7	8.1	9.9	12.8
128	–0.7	5.8	6.9	9.4
256	–0.7	4.2	4.9	6.8
512	–0.6	2.9	3.3	4.5
1,024	–0.6	2.0	2.4	3.3
2,048	–0.6	1.5	1.7	2.3
4,096	–0.6	1.1	1.2	1.5
8,192	–0.6	0.7	0.8	1.1
16,384	–0.6	0.5	0.6	0.8

Figure 11 (USD4.00/day 2005 PPP line): Households by targeting classification and score, along with "Total Accuracy" and BPAC, scorecard applied to the validation sample

Score	Inclusion: < poverty line correctly targeted	Undercoverage: < poverty line mistakenly non-targeted	Leakage: => poverty line mistakenly targeted	Exclusion: => poverty line correctly non-targeted	Total Accuracy Inclusion + Exclusion	BPAC See text
0–4	1.4	76.1	0.0	22.4	23.9	–96.3
5–9	6.7	70.8	0.2	22.3	29.0	–82.4
10–14	14.1	63.4	0.5	21.9	36.1	–62.9
15–19	25.4	52.2	1.0	21.4	46.8	–33.2
20–24	36.5	41.1	1.9	20.5	57.0	–3.5
25–29	48.0	29.6	3.3	19.1	67.1	+27.9
30–34	58.2	19.4	5.1	17.3	75.5	+56.6
35–39	64.7	12.9	7.2	15.2	79.9	+76.0
40–44	68.9	8.6	9.8	12.6	81.6	+87.3
45–49	73.0	4.6	12.7	9.7	82.7	+83.6
50–54	75.2	2.4	15.6	6.8	82.0	+79.8
55–59	76.3	1.3	17.7	4.7	81.0	+77.2
60–64	77.0	0.6	19.5	3.0	80.0	+74.9
65–69	77.4	0.1	20.8	1.6	79.0	+73.1
70–74	77.5	0.1	21.5	0.9	78.5	+72.3
75–79	77.5	0.0	21.9	0.5	78.0	+71.7
80–84	77.6	0.0	22.3	0.2	77.7	+71.3
85–89	77.6	0.0	22.4	0.0	77.6	+71.1
90–94	77.6	0.0	22.4	0.0	77.6	+71.1
95–100	77.6	0.0	22.4	0.0	77.6	+71.1

Inclusion, undercoverage, leakage, and exclusion normalized to sum to 100.

Figure 12 (USD4.00/day 2005 PPP line): For a given score cut-off, the percentage of all households who are targeted (that is, have a score equal to or less than the cut-off), the percentage of targeted households who are poor (that is, below the poverty line), the percentage of poor households who are targeted, and the number of poor households who are successful targeted (coverage) per non-poor household mistakenly targeted (leakage), scorecard applied to the validation sample

Targeting cut-off	% all households who are targeted	% targeted who are poor	% of poor who are targeted	Poor households targeted per non-poor household targeted
0–4	1.4	100.0	1.9	Only poor targeted
5–9	6.9	97.4	8.7	37.3:1
10–14	14.6	96.6	18.2	28.7:1
15–19	26.4	96.2	32.7	25.5:1
20–24	38.4	95.0	47.0	19.1:1
25–29	51.3	93.5	61.8	14.5:1
30–34	63.3	91.9	75.0	11.4:1
35–39	71.9	90.0	83.4	9.0:1
40–44	78.7	87.5	88.9	7.0:1
45–49	85.7	85.1	94.1	5.7:1
50–54	90.8	82.8	96.9	4.8:1
55–59	94.0	81.2	98.4	4.3:1
60–64	96.4	79.8	99.2	4.0:1
65–69	98.3	78.8	99.8	3.7:1
70–74	99.0	78.3	99.9	3.6:1
75–79	99.5	77.9	100.0	3.5:1
80–84	99.8	77.7	100.0	3.5:1
85–89	100.0	77.6	100.0	3.5:1
90–94	100.0	77.6	100.0	3.5:1
95–100	100.0	77.6	100.0	3.5:1

Poverty Lines and Poverty Rates, by Governorate, by Urban/Rural, and by Household-Level/Person-Level

Figure A1: Poverty lines and poverty rates, by urban/rural and by governorate, at the household level, 2005/6

		Poverty line (per capita per day) and poverty rate (%)																					
		National			National Food			USAID 'Extreme'			International (2005 PPP) $1.25/day			$2.50/day			$3.00/day			$4.00/day		
Province	Line or rate	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total
Ibb	Line	168	166		112	118		134	126		110	109		220	218		264	261		353	348	
	Rate	13.8	29.3		2.8	8.7		6.6	14.1		2.6	7.2		26.4	50.6		39.9	64.8		55.9	82.8	
Abyan	Line	195	192		124	129		152	133		128	126		255	251		306	302		409	402	
	Rate	26.3	44.5		8.1	20.0		13.2	22.2		8.1	18.5		48.5	68.6		60.6	81.2		75.3	93.3	
Sana'a City	Line	198	200		129	131		154	154		129	131		259	261		310	314		414	418	
	Rate	11.7	0.0		2.3	0.0		5.7	0.0		2.3	0.0		22.1	33.3		31.8	33.3		48.0	55.6	
Al-Baida	Line	181	179		118	123		136	115		119	117		237	235		285	282		380	376	
	Rate	14.4	56.7		2.9	29.5		8.2	26.2		3.5	26.3		24.3	74.5		37.4	83.3		61.0	94.0	
Taiz	Line	185	172		119	120		147	130		121	113		242	225		290	270		387	360	
	Rate	17.5	35.3		4.1	13.3		9.2	18.1		4.3	11.0		30.3	52.3		42.4	67.4		60.2	81.6	
Al-Jawf	Line	164	177		107	122		133	136		107	116		215	232		258	278		344	371	
	Rate	25.9	44.9		2.9	16.4		12.5	20.0		3.4	14.4		46.6	69.5		59.7	82.8		79.3	93.5	
Hajja	Line	184	177		126	128		151	128		121	116		241	231		290	278		386	370	
	Rate	17.0	39.7		4.5	19.2		9.3	20.3		4.2	15.9		33.5	55.8		44.6	66.9		62.0	81.8	
Al-Hodeida	Line	166	168		119	120		138	118		109	110		218	220		261	264		349	352	
	Rate	16.9	29.0		5.1	10.0		8.8	13.8		3.5	7.7		31.8	49.4		44.9	65.9		62.7	83.3	
Hadramout	Line	189	186		127	130		158	151		124	122		248	244		298	293		397	390	
	Rate	23.6	28.9		2.1	6.0		11.1	12.3		1.9	4.1		44.1	53.9		59.7	69.9		79.0	87.9	
Dhamar	Line	188	187		126	137		142	151		123	122		246	245		295	294		394	391	
	Rate	24.5	21.5		7.0	7.3		12.9	11.1		6.7	5.3		40.2	46.0		51.1	58.2		69.0	77.9	
Shabwah	Line	188	213		122	143		154	145		123	140		246	279		295	335		393	446	
	Rate	33.9	52.7		7.7	24.9		17.3	27.1		8.3	23.3		49.4	71.9		59.6	81.1		74.6	88.3	

Figure A1 (cont.): Poverty lines and poverty rates, by urban/rural and by governorate, at the household level, 2005/6

		Poverty line (per capita per day) and poverty rate (%)																				
		National			National Food			USAID 'Extreme'			International (2005 PPP) $1.25/day			$2.50/day			$3.00/day			$4.00/day		
Province	Line or rate	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total
Sa'adah	Line	170	172		118	123		134	129		112	112		223	225		268	270		357	359	
	Rate	17.0	14.6		4.2	5.0		8.2	7.0		3.2	2.5		30.9	45.3		44.0	56.5		68.2	77.5	
Sana'a Region	Line	-	174		-	131		-	138		-	114		-	227		-	273		-	364	
	Rate	-	24.4		-	10.8		-	12.4		-	5.6		-	40.5		-	58.9		-	79.4	
Aden	Line	202	-		130	-		164	-		132	-		264	0%		317	-		422	-	
	Rate	11.1	-		1.5	-		6.2	-		1.5	-		23.4	0%		35.7	-		56.4	-	
Laheg	Line	181	179		118	126		147	128		118	117		236	234		284	281		378	375	
	Rate	18.5	44.9		2.7	17.5		9.2	20.7		2.7	15.6		29.9	67.2		43.7	80.3		63.2	89.9	
Mareb	Line	178	190		119	128		138	113		117	125		233	249		280	299		373	399	
	Rate	13.3	42.7		3.8	25.0		7.2	19.0		3.4	25.9		26.3	56.3		35.8	63.2		53.9	77.6	
Al-Mahweet	Line	175	172		117	124		146	134		115	113		230	225		276	270		367	360	
	Rate	18.3	26.6		3.5	4.8		9.0	12.5		3.2	3.7		40.7	54.1		55.8	70.6		73.8	87.2	
Al-Maharh	Line	179	210		122	136		153	183		117	137		234	275		281	330		375	439	
	Rate	7.1	4.0		1.6	0.8		4.0	2.5		1.6	0.8		20.8	15.7		28.5	25.9		45.0	52.3	
Amran	Line	192	193		125	132		143	145		126	126		251	252		301	303		402	404	
	Rate	33.7	63.3		9.3	21.6		16.1	31.4		10.0	18.9		62.0	81.3		72.7	88.4		85.8	94.9	
Al-Dhale	Line	183	181		123	121		148	151		120	119		240	238		288	285		384	380	
	Rate	23.2	39.8		5.7	8.6		10.9	23.4		5.7	6.1		39.2	65.1		53.6	75.8		70.2	89.7	
Remah	Line	183	180		128	127		105	139		120	118		240	235		288	282		383	377	
	Rate	8.3	29.4		8.3	10.4		8.3	15.6		8.3	8.3		16.7	52.5		25.0	67.3		33.3	84.7	
Yemen	Line	187	177	180	124	125	125	149	132	137	122	116	118	245	231	235	294	278	282	392	370	376
	Rate	16.0	34.1	28.9	3.4	12.7	10.1	8.1	16.8	14.3	3.2	10.4	8.3	29.8	54.8	47.6	41.6	68.3	60.6	59.2	84.0	76.8

Figure A2: Poverty lines and poverty rates, by urban/rural and by governorate, at the person level, 2005/6

		Poverty line (per capita per day) and poverty rate (%)																					
		National			National Food			USAID 'Extreme'			International (2005 PPP) $1.25/day			$2.50/day			$3.00/day			$4.00/day		
Province	Line or rate	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total
Ibb	Line	170	167		112	117		134	126		111	109		223	218		268	262		357	350	
	Rate	16.4	32.8		3.7	10.6		8.2	16.6		3.5	8.9		30.8	54.4		44.8	69.6		60.8	85.4	
Abyan	Line	192	191		123	128		152	133		126	125		251	250		301	300		402	399	
	Rate	31.4	50.4		8.8	23.0		15.2	25.4		8.8	21.4		55.0	74.7		67.4	85.2		82.1	95.8	
Sana'a City	Line	198	204		128	132		154	154		129	133		259	267		310	320		414	427	
	Rate	15.0	0.0		2.9	0.0		7.5	0.0		3.1	0.0		27.8	29.8		39.6	29.8		56.1	68.4	
Al-Baida	Line	182	180		117	122		136	115		119	118		238	235		285	282		380	376	
	Rate	16.7	59.8		3.1	34.0		9.1	30.7		4.1	31.1		28.3	76.8		44.7	84.9		70.0	95.1	
Taiz	Line	184	173		118	119		147	130		120	113		241	227		289	272		385	363	
	Rate	23.7	41.5		5.5	16.7		12.1	21.6		5.8	14.4		38.2	60.3		51.8	74.4		70.7	87.2	
Al-Jawf	Line	165	178		107	121		133	136		108	117		217	234		260	280		347	374	
	Rate	32.6	52.6		4.2	22.0		16.3	25.5		4.9	19.9		52.2	76.5		65.5	88.0		82.9	96.3	
Hajja	Line	185	178		125	126		151	128		121	116		242	233		291	279		388	372	
	Rate	20.9	50.0		5.8	24.3		10.6	25.3		5.3	20.2		41.7	67.3		54.8	77.2		72.8	90.0	
Al-Hodeida	Line	165	164		117	117		138	118		108	107		216	214		259	257		346	343	
	Rate	21.6	36.4		6.7	12.9		11.5	17.9		4.7	10.3		39.6	57.5		54.4	73.4		72.3	88.5	
Hadramout	Line	186	184		124	128		158	151		122	120		244	241		293	289		390	385	
	Rate	31.4	39.2		2.0	9.1		15.7	19.8		1.9	6.7		55.8	63.4		71.3	77.7		87.2	92.1	
Dhamar	Line	189	186		125	135		142	151		123	122		247	243		296	292		395	389	
	Rate	29.7	25.3		9.2	9.3		15.1	13.3		8.9	7.0		46.7	51.0		57.6	63.7		74.7	83.9	
Shabwah	Line	189	209		122	141		154	145		124	137		247	274		297	329		396	439	
	Rate	39.4	56.8		9.3	25.8		20.2	29.5		9.9	24.4		57.3	76.8		68.4	84.3		81.7	89.9	

Figure A2 (cont.): Poverty lines and poverty rates, by urban/rural and by governorate, at the person level, 2005/6

		Poverty line (per capita per day) and poverty rate (%)																				
		National			National Food			USAID 'Extreme'			International (2005 PPP) $1.25/day			$2.50/day			$3.00/day			$4.00/day		
Province	Line or rate	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total	Urban	Rural	Total
Sa'adah	Line	169	168		117	121		134	129		111	110		221	220		265	264		354	352	
	Rate	18.2	16.2		4.1	5.3		8.8	7.7		3.0	3.0		34.5	46.6		46.5	58.1		70.7	79.2	
Sana'a Region	Line	-	176		-	130		-	138		-	115		-	230		-	276		-	368	
	Rate	-	28.1		-	11.8		-	13.8		-	6.9		-	44.5		-	64.8		-	85.3	
Aden	Line	200	-		128	-		164	-		131	-		262	-		314	-		419	-	
	Rate	16.9	-		2.1	-		9.0	-		2.1	-		31.6	-		44.8	-		67.1	-	
Laheg	Line	178	178		117	124		147	128		117	116		233	233		280	279		373	372	
	Rate	22.9	49.5		3.2	19.6		11.8	24.3		3.2	17.8		36.5	71.5		49.9	82.0		71.2	91.9	
Mareb	Line	180	192		119	127		138	113		118	126		235	251		282	302		376	402	
	Rate	18.0	50.1		4.6	31.5		9.1	25.1		4.1	33.0		32.4	65.6		41.7	73.7		59.3	87.4	
Al-Mahweet	Line	176	171		117	122		146	134		115	112		231	224		277	269		369	358	
	Rate	21.9	31.5		4.5	6.5		11.2	16.1		4.3	5.0		48.1	60.5		64.8	77.0		82.9	90.8	
Al-Maharh	Line	178	207		122	133		153	183		117	136		233	271		280	326		373	434	
	Rate	11.4	6.3		3.3	1.2		7.1	3.3		3.3	1.2		30.0	22.9		38.7	33.8		54.9	59.0	
Amran	Line	191	191		124	131		143	145		125	125		250	250		300	300		400	401	
	Rate	33.9	70.6		9.8	23.8		17.3	35.1		10.3	20.1		64.1	88.0		73.4	93.3		84.9	97.6	
Al-Dhale	Line	183	182		122	121		148	151		120	119		240	238		288	286		384	381	
	Rate	28.2	46.4		6.9	9.0		13.5	23.8		6.9	6.4		45.5	70.0		60.4	79.9		77.7	90.6	
Remah	Line	181	177		126	124		105	139		119	116		238	232		285	278		380	371	
	Rate	5.4	35.3		5.4	12.2		5.4	18.3		5.4	9.0		8.6	62.6		26.9	76.9		41.9	90.2	
Yemen	Line	186	177	179	123	124	124	149	133	137	122	116	117	244	231	235	293	278	282	390	370	376
	Rate	20.7	40.1	34.8	4.3	15.5	12.5	10.6	20.3	17.6	4.1	12.9	10.5	37.1	61.4	54.7	50.1	74.4	67.7	67.8	88.5	82.8