Simple Poverty Scorecard® Poverty-Assessment Tool Tanzania

Mark Schreiner

7 December 2013

This document is at `SimplePovertyScorecard.com`.

Abstract

The Simple Poverty Scorecard-brand poverty-assessment tool uses ten low-cost indicators from Tanzania's 2007 Household Budget Survey to estimate the likelihood that a household has expenditure 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 Tanzania to measure poverty rates, to track changes in poverty rates over time, and to segment clients for targeted services.

Version note

This paper corrects a problem with incomplete data for household members that is in the 8 March 2011 version. This affects the points for most indicators as well as all the poverty likelihoods and accuracy results. It also corrects errors in the 2005 PPP poverty lines that are in the 24 July 2012 version and in the 7 July 2013 version. These corrections affect all text and figures related to those poverty lines.

Acknowledgements

This paper was funded by the Ford Foundation via a grant to Grameen Foundation (GF). The data come from the National Bureau of Statistics. Thanks go to Marc Bavois, Sharlene Brown, Novati Buberwa, Lula Chen, Albina Chuwa, Clara Galeazzi, Lucas Kilyapeka, Ahmed Makbel, Rose Mungai, Sydney Neuschel, Trudy Owens, Mohamed Rajab, Jochen Schloesser, Dane Shikman, Jeff Toohig, Patrick Ward, and Cindy Yang. This scorecard was re-branded by GF as a Progress out of Poverty Index® tool. The PPI® is a performance-management tool that GF promotes to help organizations achieve their social objectives more effectively. "Progress out of Poverty Index" and "PPI" are Registered Trademarks of Innovations for Poverty Action. "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:	TZA	Field agent:		
Scorecard:	001	Service point:		
Sampling wgt.:		Number of household members:		

Indicator	Value	Points	Score
1. How many household members are 17-years-old or younger?	A. Four or more	0	
	B. Three	10	
	C. Two	15	
	D. One	20	
	E. None	30	
2. Do all children ages 6 to 17 attend school?	A. No	0	
	B. Yes, or no children ages 6 to 17	3	
3. Can the female head/spouse read and write?	A. No	0	
	B. Yes, but not in Kiswahili nor English	0	
	C. No female head/spouse	0	
	D. Yes, only in Kiswahili	6	
	E. Yes, in English (regardless of others)	13	
4. What is the main building material of the floor of the main dwelling?	A. Earth	0	
	B. Concrete, cement, tiles, timber, or other	11	
5. What is the main building material of the roof of the main dwelling?	A. Mud and grass	0	
	B. Grass, leaves, bamboo	8	
	C. Concrete, cement, metal sheets (GCI), asbestos sheets, tiles, or other	9	
6. How many bicycles, mopeds, motorcycles, tractors, or motor vehicles does your household own?	A. None	0	
	B. One	3	
	C. Two or more	11	
7. Does your household own any radios or radio cassettes?	A. No	0	
	B. Yes	6	
8. Does your household own any lanterns?	A. No	0	
	B. Yes	6	
9. Does your household own any irons (charcoal or electric)?	A. No	0	
	B. Yes	5	
10. How many tables does your household own?	A. None	0	
	B. One	2	
	C. Two	4	
	D. Three or more	6	

Score:

Look-up table to convert scores to poverty likelihoods

	Poverty likelihood (%) by poverty line						
	National				USAID	Intl. 2005 PPP	
Score	Food	100%	150%	200%	'Extreme'	$1.25	$2.50
0–4	55.2	81.3	95.7	98.6	70.2	97.8	100.0
5–9	45.9	70.8	93.3	97.9	50.0	96.5	100.0
10–14	33.8	64.8	88.4	97.9	37.3	96.5	100.0
15–19	31.2	57.2	82.1	93.6	35.2	92.7	99.5
20–24	30.9	53.5	81.5	92.2	33.9	89.5	99.5
25–29	26.1	48.4	81.5	92.2	25.9	89.5	99.2
30–34	17.6	38.7	73.1	89.6	16.6	85.6	99.0
35–39	13.2	29.6	57.9	81.0	12.9	73.9	97.9
40–44	7.7	22.8	54.3	75.1	7.3	66.8	95.1
45–49	7.4	21.2	50.8	70.8	7.3	60.0	93.5
50–54	7.4	17.0	40.8	62.7	6.0	51.9	88.9
55–59	5.4	12.0	31.5	54.2	4.0	38.0	87.6
60–64	3.5	7.8	27.1	45.8	2.8	34.6	81.0
65–69	0.7	7.0	19.4	37.7	0.6	25.4	75.1
70–74	0.7	3.2	12.8	34.0	0.6	17.2	63.4
75–79	0.7	2.0	6.9	22.0	0.6	8.4	56.3
80–84	0.6	2.0	6.8	19.1	0.5	6.7	51.9
85–89	0.0	0.0	1.7	11.5	0.0	0.5	32.1
90–94	0.0	0.0	0.0	5.9	0.0	0.0	20.1
95–100	0.0	0.0	0.0	0.0	0.0	0.0	2.3

Simple Poverty Scorecard® Poverty-Assessment Tool Tanzania

1. Introduction

Pro-poor programs in Tanzania can use the Simple Poverty Scorecard poverty-assessment tool to estimate the likelihood that a household has expenditure below a given poverty line, to measure groups' poverty rates at a point in time, to track changes in groups' poverty rates over time, and to segment clients for targeted services.

The direct approach to poverty measurement via surveys is difficult and costly. As a case in point, the 2007 Tanzania Household Budget Survey (HBS) asks more than 20 pages of questions over the course of two visits to a household, following that with visits every 2–3 days for a month to record the entries in a household diary of the purchase and consumption of about 400 possible income and expenditure items.

In contrast, the indirect approach via the scorecard is simple, quick, and inexpensive. It uses 10 verifiable indicators (such as "What is the main building material of the floor of the main dwelling?" or "Does your household own any irons (charcoal or electric)?") 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). Results from these approaches are not comparable across organizations, they may be costly, and their accuracy is unknown.

Pro-poor organizations can use the scorecard to measure the share of their participants who are below a given poverty line, such as the Millennium Development Goals' $1.25/day line at 2005 purchase-power parity. USAID microenterprise partners can use it to report how many of their participants are among the poorest half of people below the national poverty line. Organizations can also use it to measure movement across a poverty line. In all these cases, the scorecard provides an expenditure-based, objective tool with known accuracy. While expenditure surveys are costly even for governments, many small, local organizations may be able to implement an inexpensive scorecard that can serve for monitoring 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 maximum", simple scorecards are often about as accurate as complex ones.

The technical approach here is 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 accuracy tests are simple and commonplace in statistical practice and in the for-profit field of credit-risk scoring, they have rarely been applied to poverty-assessment tools.

The scorecard is based on the 2007 HBS conducted by Tanzania's National Bureau of Statistics (NBS) from 1 January 2007 to 31 December 2007.[1] 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 about 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 or per-adult-equivalent expenditure below a given poverty line.

[1] The survey—and thus the scorecard here—applies only to mainland Tanzania. Data is available at `http://www.nbs.go.tz/nada3/index.php/catalog`.

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

Third, the scorecard can estimate changes in the poverty rate for a group of households (or for two independent samples of households that are representative of 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 an 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 expenditure data and Tanzania's national ("basic needs") poverty line. Scores from this one scorecard are calibrated to poverty likelihoods for seven poverty lines.

The scorecard is constructed and calibrated using half of the data from the 2007 HBS, and its accuracy is validated on the other half.

While all three scoring estimators are *unbiased* (that is, they match the true value on average in repeated samples when applied to 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.[2]

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 poverty will be the same as in the data used to build the scorecard. 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 –1.3 percentage points for the national line, and the average absolute difference across all seven lines is 0.7 percentage points. These differences are due to sampling variation and not bias; the average of each difference would be zero if the whole 2007 HBS were to be repeatedly redrawn and divided into sub-samples before repeating the entire process of building and calibrating scorecards.[3]

The 90-percent confidence intervals for these estimates are +/–0.7 percentage points or less. For n = 1,024, the 90-percent intervals are +/–2.5 percentage points or less.

[2] Important examples include nationally representative samples at a different point in time or non-nationally representative sub-groups (Tarozzi and Deaton, 2009).
[3] Most differences for Tanzania seems to be due to constructing the scorecard with an adult-equivalent poverty line and then applying it to per-capita poverty lines.

Section 2 below describes data and poverty lines. Sections 3 and 4 describe scorecard construction and offer guidelines for use in the field. Sections 5 and 6 detail the estimation of households' poverty likelihoods and of groups' poverty rates at a point in time. Section 7 discusses estimating changes in poverty rates over time, and Section 8 covers targeting. Section 9 places the new scorecard here in the context of existing exercises for Tanzania, and Section 10 is a summary.

2. Data and poverty lines

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

2.1 Data

The scorecard is based on data from the 2007 HBS. Households are randomly divided into two sub-samples (Figure 2):

- *Construction* and *calibration* for selecting indicators and points and for associating scores with poverty likelihoods
- *Validation* for testing accuracy on 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 group who live in households whose total household expenditure (divided by the number of adult equivalents or 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 expenditure (or per-adult-equivalent expenditure) above a poverty line (it is "non-poor") and that the second household has per-capita expenditure (or per-adult-equivalent expenditure) 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 more 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 may be relevant. For example, if a microlender has only one borrower in a household, then it might prefer to report household-level poverty rates.

Figure 2 reports poverty lines and poverty rates at both the household-level and the person-level for the three regions of Dar es Salaam, Other Urban, and Rural, as well as for mainland Tanzania. The scorecard is constructed using the 2007 HBS and household-level lines, scores are calibrated to household-level poverty likelihoods, and accuracy is measured for household-level rates.

Person-level poverty rates can be estimated by taking a household-size-weighted average of the household-level poverty likelihoods. It is possible to construct, calibrate, and validate a scorecard based on person-level weights, but it is not done here.

2.2.2 Poverty lines

Tanzania's food poverty line is TZS359 per adult equivalent per day.[4] This is defined as the cost of 2,200 calories using food items consumed by the poorest half of the population in the 2000/1 HBS. This is adjusted to 2007 by a factor of 1.93 to account for price changes, and it is also adjusted for price differences across the three regions in 2007 (NBS, 2009 and 2002b).

Tanzania's national ("basic needs") poverty line of TZS492 per adult equivalent per day is defined as the food line, divided by the share of expenditure among the poorest quartile of households that goes to food (NBS, 2002b). This was 73 percent in the 2000/1 HBS, giving a national line of TZS359/0.73 = TZS492. This adjustment accounts for the fact that households below the 2,200-calorie benchmark nevertheless have some expenditure on (apparently necessary) non-food items (Ravaillon, 1998).

The scorecard here is constructed using the national poverty line (sometimes called "100% of the national line"). For mainland Tanzania, the national line implies a household-level poverty rate of 26.6 percent and a person-level poverty rate of 33.6

[4] The adult-equivalence scale appears in NBS (2002a).

percent. For the food poverty line, the household-level poverty rate for mainland Tanzania is 12.6 percent, and the person-level rate is 16.6 percent.

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:

- Food
- National
- 150% of national
- 200% of national
- USAID "extreme"
- $1.25/day 2005 PPP
- $2.50/day 2005 PPP

The USAID "extreme" line (U.S. Congress, 2002) is defined as the median per-capita expenditure of people (not households) below the national line in a given poverty-line region.

The $1.25/day 2005 PPP line is derived from:

- 2005 PPP exchange rate for "individual consumption expenditure by households" of TZS482.45 per $1 (World Bank, 2008)
- Average mainland Tanzania Consumer Price Index (CPI) for 2005 of 120.9[5]
- Average CPI for 2007 of 138.7

[5] http://www.nbs.go.tz/index.php?option=com_phocadownload&view=category&id=1 38:summary-cpi&download=411:cpi-summary-2002-2010-releases&Itemid=106, retrieved 28 January 2011.

The $1.25/day 2005 PPP line for mainland Tanzania applied to the 2007 HBS is (Sillers, 2006):

$$(\text{2005 PPP exchange rate}) \cdot \$1.25 \cdot \left(\frac{\text{CPI}_{\text{Ave. 2007}}}{\text{CPI}_{\text{Ave. 2005}}} \right) =$$

$$\left(\frac{\text{TZS}482.45}{\$1.00} \right) \cdot \$1.25 \cdot \left(\frac{138.7}{120.9} \right) = 692\text{TZS}.$$

The $2.50/day 2005 PPP line is twice the $1.25/day 2005 PPP line.

The mainland-Tanzania $1.25/day line of TZS692 is adjusted for cost-of-living difference across Tanzania's three poverty-line regions. A given poverty-line region's $1.25/day line is the mainland-Tanzania line of TZS692, divided by the value of the national poverty line applicable in that region, and then divided by the average person-weighted national line across all regions.

3. Scorecard construction

About 75 potential indicators are initially prepared in the areas of:

- Family composition (such as household size)
- Education (such as literacy of the female head/spouse)
- Housing (such as the type of floor or roof)
- Ownership of durable goods (such as lanterns or irons)

Figure 3 lists all potential indicators, ranked by the entropy-based "uncertainty coefficient" that measures how well a given indicator predicts poverty on its own (Goodman and Kruskal, 1979).

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 lantern or an iron is probably more likely to change in response to changes in poverty than is the marital status of the male head/spouse.

The scorecard itself is built using the national poverty line and Logit regression on the construction sub-sample. Indicator selection uses both judgment and statistics. 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 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 mainland Tanzania. Evidence from India and Mexico (Schreiner, 2006 and 2005a), Sri Lanka (Narayan and Yoshida, 2005), and Jamaica (Grosh and Baker, 1995) suggests that segmenting scorecards by urban/rural does not improve targeting accuracy much, although it may improve the accuracy of estimates of poverty rates (Tarozzi and Deaton, 2009).

4. 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 have similar targeting accuracy, thanks to the empirical phenomenon of the "flat maximum" (Falkenstein, 2008; 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, and no arithmetic beyond addition)

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 point value
- Write the point value 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 and analysis

Of course, field workers must be trained. The quality of outputs depends on the quality of inputs. If organizations or field workers gather their own data and believe that they 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 audits (Matul and Kline, 2003).[6] IRIS Center (2007) 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

[6] If an organization does not want field workers to know the points associated with indicators, then they can use the version of the scorecard without points and apply the points later at the central office.

concepts in the scorecard is essential. For the example 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. At the same time, Grosh and Baker (1995) find that gross underreporting of assets does not affect targeting. For the first stage of targeting in a conditional cash-transfer program in Mexico, Martinelli and Parker (2007) find that "underreporting [of asset ownership] is widespread but not overwhelming, except for a few goods . . . [and] overreporting is common for a few goods, which implies that self-reporting may lead to the exclusion of deserving households" (pp. 24–25). Still, as is done in Mexico in the second stage of its targeting process, most false self-reports can be corrected by field agents who verify responses with a home visit, and this is the suggested procedure for the scorecard in Tanzania.

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

In general, the sampling design should follow from the organization's goals for the exercise.

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

Given a population of interest, the subjects to be scored can be:

- All participants
- A representative sample of all participants
- All participants in a representative sample of branches
- A representative sample of all 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 are illustrated by BRAC and ASA, two microlenders in Bangladesh who each have more than 7 million participants and who are applying the Simple Poverty Scorecard tool for Bangladesh (Chen and Schreiner, 2009b). Their design is that loan officers in a random sample of branches 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.

5. Estimates of household poverty likelihoods

The sum of scorecard points for a household is called the *score*. For Tanzania, 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 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 35–39 have a poverty likelihood of 29.6 percent, and scores of 40–44 have a poverty likelihood of 22.8 percent (Figure 4).

The poverty likelihood associated with a score varies by poverty line. For example, scores of 35–39 are associated with a poverty likelihood of 29.6 percent for the national line but 13.2 percent for the food line.[7]

5.1 Calibrating scores with poverty likelihoods

A given score is 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.

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

For the example of the national line (Figure 5), there are 8,395 households in the calibration sub-sample with a score of 35–39, of whom 2,485 are below the poverty line. The estimated poverty likelihood associated with a score of 35–39 is then 29.6 percent, because 2,485 ÷ 8,395 = 29.6 percent.

To illustrate with the national line and a score of 40–44, there are 10,897 households in the calibration sample, of whom 2,482 are below the line (Figure 5). Thus, the poverty likelihood for this score is 2,482 ÷ 10,897 = 22.8 percent.

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

Figure 6 shows, for all scores, the likelihood that expenditure falls in a range demarcated by two adjacent poverty lines. There are two versions of Figure 6, one for the national poverty lines in units of adult equivalents per day, and another for the USAID "extreme" line and the 2005 PPP lines in units of people per day.

For the per-adult-equivalent national lines, the daily expenditure of someone with a score of 35–39 falls in the following ranges with probability:

- 13.2 percent below the food line
- 16.4 percent between the food line and 100% of the national line
- 28.3 percent between 100% and 150% of the national line
- 23.1 percent between 150% and 200% of the national line
- 19.0 percent above 200% of the national line

For the example of the per-capita lines, the daily expenditure of someone with a score of 35–39 falls in the following ranges with probability:

- 12.9 percent below the USAID "extreme" line
- 61.0 percent between the USAID "extreme" line and $1.25/day 2005 PPP line
- 24.0 percent between the $1.25/day and $2.50/day 2005 PPP lines
- 2.1 percent above the $2.50/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 expenditure 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 based only on 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 Tanzania's scorecard are transformed coefficients from a Logit regression, scores are not converted to poverty likelihoods via the Logit formula of $2.718281828^{score} \times (1+ 2.718281828^{score})^{-1}$. 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 calibration samples.

5.2 Accuracy of estimates of poverty likelihoods

As long as the relationship between indicators and poverty does not change, 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 and of changes in poverty rates between two points in time.[8]

Of course, the relationship between indicators and poverty changes with time, so the scorecard applied after 2007 (as it must be in practice) will generally be biased.

[8] 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 poverty likelihoods given the assumption of a sample that is representative of mainland Tanzania in 2007? 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 expenditure 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, 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 35–39 in the validation sample is too high by 0.9 percentage points (Figure 7). For scores of 40–44, the estimate is too low by 2.9 percentage points.[9]

For the validation sample, the 90-percent confidence interval for the differences for scores of 35–39 is +/–2.2 percentage points (Figure 7). This means that in 900 of 1,000 bootstraps, the difference between the estimate and the true value is between –1.3

[9] These differences are not zero, despite the estimator's unbiasedness, because the bootstrap estimates come from a single sample (the 2007 HBS). Their average difference would be zero if samples were repeatedly drawn from the population and split into sub-samples before repeating the entire scorecard-building process.

and +3.1 percentage points (because +0.9 – 2.2 = –1.3, and +0.9 + 2.2 = +3.1). In 950 of 1,000 bootstraps (95 percent), the difference is +0.9 +/–2.7 percentage points, and in 990 of 1,000 bootstraps (99 percent), the difference is +0.9 +/–3.3 percentage points.

For many score ranges, Figure 7 shows differences (some of them large) 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 mainland Tanzania'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 8 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. As discussed later, this is generally the case.

By construction, the scorecard here is unbiased. It may still, however, be *overfit* when applied after the end of the HBS fieldwork in December 2007. That is, it may fit the 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 HBS. Or the scorecard may be overfit in the sense that its bias is highly sensitive to changes over time in the relationship between indicators and poverty.

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. Bootstrapping can also mitigate overfitting by reducing (but not eliminating) dependence on a single sampling instance. 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, much of the differences may come from non-scorecard sources such as changes in the relationship between indicators and poverty, sampling variation, and inconsistencies in data quality. 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).

6. 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, 2011 and that they have scores of 20, 30, and 40, corresponding to poverty likelihoods of 53.5, 38.7, and 22.8 percent (national line, Figure 4). The group's estimated poverty rate is the households' average poverty likelihood of (53.5 + 38.7 + 22.8) ÷ 3 = 38.3 percent.[10]

6.1 Accuracy of estimated poverty rates at a point in time

For the Tanzania 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 for the seven poverty lines are 1.3 percentage points or less (Figure 8, summarizing Figure 9 across poverty lines). The average absolute difference across the seven poverty lines is 0.7 percentage points. At least part of these differences is due to sampling variation in the validation sample and in the division of the 2007 HBS into two 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

[10] The group's poverty rate is *not* the poverty likelihood associated with the average score. Here, the poverty likelihood associated with the average score of 30 is 46.1 percent, which differs from the average of the three poverty likelihoods associated with each of the three scores (42.9 percent).

(Figure 8). This means that in 900 of 1,000 bootstraps of this size, the difference between the estimate and the true value is within 0.7 percentage points of the average difference. In the specific case of 100% 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 –1.3 – 0.6 = –1.9 to –1.3 + 0.6 = –0.7 percentage points. This is because –1.3 is the average difference, and +/–0.6 is its 90-percent confidence interval. The average difference is –1.3 because the average scorecard estimate is too low by 1.3 percentage points; the scorecard tends to estimate a poverty rate of 25.2 percent for the validation sample, but the true value is 26.5 percent (Figure 2).

6.2 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 scorecards (Schreiner, 2008a), note that the textbook formula (Cochran, 1977) that relates confidence intervals with standard errors in the case of direct measurement of rates is $c = +/- z \cdot \sigma$, where:

c is a confidence interval as a proportion (*e.g.*, 0.02 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 26.7 percent (the average poverty rate in the construction and calibration samples in Figure 2 for the national line), the confidence interval c is $+/- z \cdot \sqrt{\frac{p \cdot (1-p)}{n}} = +/- 1.64 \cdot \sqrt{\frac{0.267 \cdot (1-0.267)}{16{,}384}} = +/–0.567$ percentage points.

The scorecard, however, does not measure poverty directly, so this formula is not immediately applicable. To derive a formula for the Tanzania 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 0.645 percentage points.[11]

Thus, the 90-percent confidence interval with $n = 16{,}384$ is 0.645 percentage points for the Tanzania scorecard and 0.567 percentage points for direct measurement. The ratio of the two intervals is $0.645 \div 0.567 = 1.13$.

Now consider the same case, but with $n = 8{,}192$. The confidence interval under direct measurement is $+/-1.64 \cdot \sqrt{\frac{0.267 \cdot (1-0.267)}{8{,}192}} = +/-0.802$ percentage points. The empirical confidence interval with the Tanzania scorecard (Figure 9) is 0.920 percentage points. Thus for $n = 8{,}192$, the ratio of the two intervals is $0.920 \div 0.802 = 1.15$.

This ratio is 1.15 for both $n = 8{,}182$ and 1.13 for $n = 16{,}384$. Across all sample sizes of 256 or more in Figure 9, the average ratio turns out to be 1.12, implying that confidence intervals for indirect estimates of poverty rates via the Tanzania scorecard and this poverty line are about 12 percent wider than confidence intervals for direct estimates via the 2007 HBS. This 1.12 appears in Figure 8 as the "α factor" because if $\alpha = 1.12$, then the formula relating confidence intervals c and standard errors σ for the Tanzania scorecard is $c = +/- z \cdot \alpha \cdot \sigma$. In turn, the formula for the standard error σ for point-in-time estimates of poverty rates via scoring is $\alpha \cdot \sqrt{\frac{p \cdot (1-p)}{n}}$.

[11] Due to rounding, Figure 9 displays 0.6, not 0.620.

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 is the case for four of the seven poverty lines in Figure 8.

The formula relating confidence intervals with standard errors for the scorecard can be rearranged to give a formula for determining sample size before measurement.[12] 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 $+/-c$ is $n = \left(\frac{\alpha \cdot z}{c}\right)^2 \cdot \hat{p} \cdot (1 - \hat{p})$.

To illustrate how to use this, suppose $c = 0.04925$ and $z = 1.64$ (90-percent confidence). Then the formula gives $n = \left(\frac{1.12 \cdot 1.64}{0.04925}\right)^2 \cdot 0.267 \cdot (1 - 0.267) = 273$, not too far from the sample size of 256 observed for these parameters in Figure 9 for the national line.

Of course, the α factors in Figure 8 are specific to Tanzania, its poverty lines, its poverty rates, and this scorecard. The approach to deriving the formulas, however, is valid for any scorecard following the approach in this paper.

---

[12] Although USAID has not specified required confidence levels or intervals, IRIS Center (2007a and 2007b) says that a sample size of $n = 300$ is sufficient for USAID reporting. Given $\alpha = 0.95$ for the \$1.25/day 2005 PPP line to be used for reporting by USAID microenterprise partners in Tanzania, an expected before-measurement poverty rate of 61.3 percent (the all-Tanzania \$1.25/day rate for 2007), and a confidence level of 90 percent, then $n = 300$ implies a confidence interval of $\pm 0.95 \cdot 1.64 \cdot \sqrt{\frac{0.613 \cdot (1 - 0.613)}{300}}$ $= \pm 4.4$ percentage points.

In practice after the end of fieldwork for the HBS in December 2007, 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 26.6 percent national average in the 2007 HBS in Figure 2), look up α (here, 1.12), assume that the scorecard will still work in the future and/or for non-nationally representative sub-groups,[13] and then compute the required sample size. In this illustration, $n = \left(\frac{1.12 \cdot 1.64}{0.02}\right)^2 \cdot 0.266 \cdot (1 - 0.266) = 1{,}649$.

[13] This paper reports accuracy for the scorecard applied to the validation sample, but it cannot test accuracy for later years or for other groups. Performance after 2007 will resemble that in the 2007 HBS with deterioration to the extent that the relationships between indicators and poverty status change over time.

7. 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 only from the 2007 HBS, this paper cannot test estimates of change over time for Tanzania, and it can only suggest approximate formulas for standard errors. Nevertheless, the relevant concepts are presented here because, in practice, pro-poor organizations can use the scorecard to get data to measure change through time.

7.1 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. 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.

7.2 Calculating estimated changes in poverty rates over time

Consider the illustration begun in the previous section. On Jan. 1, 2011, a program samples three households who score 20, 30, and 40 and so have poverty likelihoods of 53.5, 38.7, and 22.8 percent (national line, Figure 4). The group's baseline estimated poverty rate is the households' average poverty likelihood of (53.5 + 38.7 + 22.8) ÷ 3 = 38.3 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, 2012, 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 48.4, 29.6, and 21.2 percent, national line, Figure 4). Their average poverty likelihood at follow-up is now (48.4 + 29.6 + 21.2) ÷ 3 = 33.1 percent, an improvement of 38.3 – 33.1 = 5.2 percentage points.[14]

This suggests that about one in twenty participants in this hypothetical example crossed the poverty line in 2011.[15] Among those who started below the line, about one in seven (5.2 ÷ 38.3 = 13.6 percent) on net ended up above the line.[16]

[14] Of course, such a huge reduction in poverty in one year is unlikely, but this is just an example to show how the scorecard can be used to estimate change.
[15] This is a net figure; some people start above the line and end below it, and vice versa.

7.3 Accuracy for estimated change in two independent samples

With only the 2007 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 the Tanzania 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,[17] 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.

[16] The scorecard does not reveal the reasons for this change.

[17] 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 standard-error formula can be rearranged to give a sample-size formula 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 countries for which this α has been measured (Schreiner, 2010, 2009a, 2009b, 2009c, 2009d, 2009e, and 2008b; Schreiner and Woller, 2010a and 2010b; and Chen and Schreiner, 2009a and 2009b), the simple average of α across poverty lines and years for a given country and then across countries is 1.19. This is as reasonable a figure as any to use for Tanzania.

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, $\alpha = 1.19$, and $\hat{p} = 0.266$ (from Figure 2). Then the baseline sample size is $n = 2 \cdot \left(\frac{1.19 \cdot 1.64}{0.02} \right)^2 \cdot 0.266 \cdot (1 - 0.266) =$ 3,719, and the follow-up sample size is also 3,719.

7.4 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:[18]

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

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:

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

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

Because $\hat{p}_*$ could be anything between 0–0.5, 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 (Schreiner, 2009a)—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 Tanzania scorecard is applied twice (once after December 2007 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 a data-based estimate, Schreiner 2009a), the average α across years and poverty lines is about 1.30.

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 2011 and then again in 2014 ($y = 3$). The before-baseline poverty rate is 26.6 percent ($p_{2007} = 0.266$, Figure 2), and suppose $\alpha = 1.30$. Then the baseline sample size is $n = 2 \cdot \left(\frac{1.30 \cdot 1.64}{0.02} \right)^2 \cdot \{-0.02 + 0.016 \cdot 3 + 0.47 \cdot [0.266 \cdot (1 - 0.266)]\} = 2{,}722$. The same group of 2,722 households is scored at follow-up as well.

8. 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* (having expenditure below a poverty line). Poverty status is a fact that depends on whether expenditure 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 the cut-off score; a higher cut-off has better inclusion (but greater leakage), while a lower cut-off has better exclusion (but higher undercoverage).

Programs should weigh these trade-offs when setting a cut-off. An formal way to do this is to assign explicit 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: 18.0 percent are below the line and correctly targeted
- Undercoverage: 8.6 percent are below the line and mistakenly not targeted
- Leakage: 20.8 percent are above the line and mistakenly targeted
- Exclusion: 52.6 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: 20.7 percent are below the line and correctly targeted
- Undercoverage: 5.9 percent are below the line and mistakenly not targeted
- Leakage: 29.0 percent are above the line and mistakenly targeted
- Exclusion: 44.4 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. A program that uses targeting—with or without scoring—should 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 Tanzania scorecard. For the national line in the validation sample, total net benefit is greatest (76.6) for a cut-off of 20–24, with about three in four households in mainland Tanzania 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).[19]

[19] Figure 11 also reports "BPAC", the Balanced Poverty Accuracy Criteria adopted by USAID for certifying poverty-assessment tools. IRIS Center (2005) says that BPAC considers accuracy in terms of estimated poverty rates and in terms of targeting inclusion. BPAC = (Inclusion – |Undercoverage – Leakage|) x [100 ÷ (Inclusion + Undercoverage)].

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 Tanzania scorecard applied to the validation sample, the expected poverty rate among households who score at or below a given cut-off. For the national line, targeting households who score 39 or less would target 38.8 percent of all households (second column) and produce a poverty rate among those targeted of 46.3 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 39 or less, 67.7 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 39 or less, covering 0.9 poor households means leaking to 1 non-poor household.

9. Context of poverty-assessment tools in Tanzania

This section discusses four existing poverty-assessment tools for Tanzania in terms of their goals, methods, data, poverty lines, indicators, accuracy, standard-error formula, and costs.

The comparative strengths of the scorecard here is that it is based on more recent data, it covers more poverty lines, and it reports measures of accuracy and formulas for standard errors.

9.1 Gwatkin *et al.*

Gwatkin *et al.* (2007) apply to Tanzania an approach used in 56 countries with Demographic and Health Surveys (Rutstein and Johnson, 2004). They use Principal Components Analysis to make an asset index from simple, low-cost indicators available for the 9,735 households in the Tanzania 2004 DHS. The PCA index is like the scorecard here except that, because the DHS does not collect data on income or expenditure, it is based on a different conception of poverty, its accuracy vis-à-vis expenditure-based poverty is unknown, and it can only be assumed to be a proxy for long-term wealth/economic status.[20] Well-known examples of the PCA asset-index

[20] Still, because the indicators are similar and because the "flat maximum" is important, carefully built PCA indices and expenditure-based poverty-assessment tools may pick up the same underlying construct (perhaps "permanent income", see Bollen, Glanville, and Stecklov, 2007), and they rank households much the same. Tests of how well rankings by PCA indices correspond with rankings by expenditure-based scorecards

approach include Ferguson *et al.* (2003), Sahn and Stifel (2003), and Filmer and Pritchett (2001).

Gwatkin *et al.* discuss three basic uses for their index:

- Segmenting households by quintiles to see how health, population, and nutrition vary with socio-economic status
- Monitoring (via exit surveys) how well local health-service posts reach the poor
- Measuring coverage of health services via local, small-scale surveys

The first goal is akin to targeting, and the last two goals resemble the monitoring goals here, so the uses of the index are similar to those of the scorecard here.

Most of the 20 indicators in Gwatkin *et al.* are similar to those in the scorecard here in terms of their simplicity, low cost, and verifiability:

- Characteristics of the residence:
  - Type of floor
  - Type of wall
  - Type of roof
  - Type of fuel for cooking
  - Type of lighting
  - Source of drinking water
  - Type of toilet arrangement
  - Number of people per sleeping room
- Ownership of consumer durables:
  - Radio
  - Television
  - Refrigerator
  - Bicycle
  - Motorcycles or scooter
  - Car or truck
  - Telephone
  - Iron

---

include Howe *et al.* (2009), Filmer and Scott (2008), Lindelow (2006), Wagstaff and Watanabe (2003), and Montgomery *et al.* (2000).

- Ownership of bank account
- Presence of a domestic worker
- Land ownership:
    - — Acres for farming
    - — Acres for grazing

A couple of these indicators are complex (number of people per sleeping room) or are non-verifiable or falsifiable (presence of bank account).

The Gwatkin *et al.* index is also more difficult and costly because it cannot be computed by hand in the field, as it has 114 point values, half of them negative, and all with five decimal places.

Unlike the PCA index, the scorecard here is linked directly to an absolute, expenditure-based poverty line. Thus, while both approaches can rank households, only the scorecard can estimate expenditure-based poverty status.

In essence, Gwatkin *et al.*—like all PCA asset indices—define poverty in terms of the indicators in their index. Thus, the index can be seen not as a proxy standing in for something else (such as expenditure) but rather as a direct measure of a non-expenditure-based definition of poverty. There is nothing wrong—and a lot right—about defining poverty in this way, but it is not as common or as well-understood as an expenditure-based definition.

9.2 Sahn and Stifel (2000)

Sahn and Stifel (2000) use factor analysis (a close relative of PCA that gives similar results) to construct an asset index meant to measure poverty in terms of long-

term wealth. They construct their index by pooling Tanzania's 1991 and 1996 DHS. Defining poverty status according to lines set at the 25th and 40th percentiles of the asset index, they then compare the distribution of the index and poverty rates over time (within Tanzania) and across countries (Tanzania and 10 other sub-Saharan countries).

For the cross-country analysis, Sahn and Stifel construct a single cross-country index from pooled DHS data for the 11 countries (plus five others for which only a single DHS round is available). This is possible because the DHS in all rounds and countries uses a common set of simple, inexpensive, and verifiable indicators.

The nine indicators in Sahn and Stifel are similar to those in the scorecard here in terms of their simplicity, low cost, and verifiability:

- Education of the head
- Characteristics of the residence:
  - Type of floor
  - Source of drinking water
  - Type of toilet arrangement
- Ownership of consumer durables:
  - Radio
  - Television
  - Refrigerator
  - Bicycle
  - Motorized transport

Like Gwatkin *et al.*, this approach shares many of the strengths of the approach here in that it can be used for targeting and in that it is flexible, low-cost, and adaptable to diverse contexts. Because it does not require price adjustments over time or between countries—or even expenditure data—it is more adaptable in those dimensions than the scorecard here.

Sahn and Stifel also share with Gwatkin *et al.* the disadvantages of using a less-common definition of poverty and of not reporting formula for standard errors. Also, their purpose is to inform governments and donors about the broad progress of poverty-reduction efforts in Africa, not to provide a tool to help local, pro-poor organizations in their poverty-alleviation efforts.[21]

9.3 Minot *et al.* (2006)

Minot *et al.* (2006) construct a PCA-based asset index using only indicators in both the 1991/2 HBS and Tanzania's DHS of 1991/2, 1996, 1999, and 2003. They seek to estimate trends in poverty from 1991/2 through 2003. Minot *et al.* say that this is the first application of such an asset index based on a household expenditure survey to survey data that does not include a measure of expenditure. Other examples include Mathiassen (2007), Stifel and Christiaensen (2007), Azzarri *et al.* (2005), and Simler, Harrower, and Massingarela (2003). This is "poverty mapping" applied to DHS data instead of census data (Elbers, Lanjouw, and Lanjouw, 2003; Hentschel *et al.*, 2000).

Minot *et al.* build four regional poverty-assessment tools (Dar es Salaam, large towns, small towns, and rural) using least-squares regression on the logarithm of per-capita expenditure from the 1991/2 HBS. Indicators must match across surveys and be

[21] Booysen *et al.* (2008) covers Tanzania in a way like Sahn and Stifel except that they use Multiple Correspondence Analysis instead of factor analysis, they look at both poverty rates and inequality measures, and they use three DHS rounds rather than two.

statistically significant at $p < 0.20$. The 25 indicators closely resemble those in Gwatkin *et al.* and in Sahn and Stifel, except for the detailed age/sex household breakdown:

- Household demographics:
  - — Number of members (and its square)
  - — Share of males younger than 5
  - — Share of females younger than 5
  - — Share of males aged 5 to 15
  - — Share of females aged 5 to 15
  - — Share of males aged 16 to 30
  - — Share of females aged 16 to 30
  - — Share of males aged 31 to 55
  - — Share of females aged 31 to 55
  - — Share of males 56 or older
  - — Share of females 56 or older
  - — Sex of the head
  - — Age of the head
- Education:
  - — Head
  - — Spouse of head
- Characteristics of the residence:
  - — Type of floor
  - — Source of drinking water
  - — Type of toilet arrangement
  - — Presence of electricity
- Ownership of consumer durables:
  - — Radio
  - — Television
  - — Refrigerator
  - — Motorbike
  - — Car
- Region of residence

The tools are applied to DHS households to estimate poverty rates for the national poverty line developed for the 1991/2 HBS (which differs—even after adjusting for inflation—from the line developed for the 2000/1 HBS that is used here). Given the assumption that the relationship between indicators and poverty is constant from 1991

to 2003,[22] Minot *et al.* find that poverty fell slightly more from 1991/2 to 2000/1 than suggested by the HBS surveys and that poverty fell rapidly from 2000/1 to 2003 (a period which lacks direct expenditure-based measurements of poverty).

The main difference between the scorecard here and the tool of Minot *et al.* is focus. The scorecard is meant as a tool that local pro-poor organizations might be willing and able to pick up on their own to help manage their social performance. In contrast, Minot *et al.* is meant to track poverty over time at the country level for periods that are not covered by household expenditure surveys.

Minot *et al.* report 95-percent confidence intervals, but they do not report regional sample sizes, and they cannot compare the accuracy of their estimates with direct measurements of expenditure-based poverty. Thus, precision and accuracy cannot be compared with that of the scorecard here.

9.4 Setel *et al.* (2005)

Setel *et al.* (2005) use the 2000/1 HBS to build a tool to estimate consumption poverty. As in this paper, they estimate poverty likelihoods with a focus on practicality in a real-world institutional setting, seeking a tool that is:

- Rapid and inexpensive to implement
- Easily integrated into routine surveys
- Able to produce estimates of expenditure-based poverty status
- Simple enough for local professionals to update

[22] For a scorecard based on the 2000/1 HBS, Minot *et al.* find a smaller decrease in poverty. This suggests a changing relationship between indicators and poverty.

Setel *et al.* use stepwise regression based on R^2 to select indicators related to the food poverty line in tools for the three regions of Dar es Salaam, Kilimanjaro, and Morogoro. Indicators in one or more of the regional tools include:

- Household demographics:
  - Number of members
  - Sex of head
  - Age of head
  - Dependency ratio
- Education of the head
- Number of household members employed
- Characteristics of the residence:
  - Type of wall
  - Source of drinking water
  - Type of toilet arrangement
  - Type of lighting
  - Number of people per sleeping room
- Ownership of consumer durables:
  - Bed net
  - Electric/gas stove
  - Iron
  - Lamp
  - Sofa
  - Bicycle
  - Automobile
- Area of land owned for farming/pastoralism
- Main source of cash income

- Past events:
  - — Number of days in past week in which meat was eaten
  - — Number of days in past week in which dairy products were consumed
  - — Whether in the past month the household paid to purchase wheat flour
  - — Whether in the past month the household paid to purchase cooking bananas
  - — Whether in the past month the household paid to purchase potatoes
  - — Whether in the past month the household paid to purchase fresh fish
  - — Whether in the past month the household paid to purchase beer
  - — Whether in the past month the household paid to purchase newspapers
  - — Whether in the past month the household paid to purchase poultry products
  - — Whether in the past month the household paid to purchase eggs
  - — Whether in the past month the household paid to purchase a snack or beverage outside of the household
  - — Whether in the past year the household paid to purchase seeds
  - — Whether in the past year the household paid to purchase fertilizer or manure

Several of these indicators—especially those asking about past events—are complex and/or non-verifiable.

In a validation sample comprised of households that were not used to construct the tools, Setel *et al.* find differences between estimated at true values of +0.4, –0.8, and +1.6 percentage points, which is generally better than for the scorecard here.

10. Conclusion

Pro-poor programs in Tanzania can use the scorecard to segment clients for targeted services as well as to estimate:

- The likelihood that a household has expenditure 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 between two points in time

The scorecard is inexpensive to use and can be understood by non-specialists. It is designed to be practical for local pro-poor organizations who want to improve how they monitor and manage their social performance in order to speed up their participants' progress out of poverty.

The scorecard is built with a sub-sample of data from the 2007 HBS, tested with a different sub-sample, and calibrated to seven poverty lines.

Accuracy and formulas for standard errors 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, the absolute difference between estimates versus true poverty rates for groups of households at a point in time is less than 1.3 percentage points, and the average absolute difference across the seven poverty lines is 0.7 percentage points. For $n = 16{,}384$ and 90-percent confidence, the

precision of these differences is +/–0.7 percentage points or less, and for $n = 1{,}024$, precision is +/–2.5 percentage points or less.

For targeting, programs can use the results reported here to select a cut-off that fits their 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 10 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 Tanzania 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 income or expenditure survey.

References

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

Azzarri, Carlo; Carletto, Gero; Davis, Benjamin; and Alberto Zezza. (2005) "Monitoring Poverty without Consumption Data: An Application Using the Albania Panel Survey", *Eastern European Economics*, Vol. 44, No. 1, pp. 59–82.

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.

Booysen, Frikkie; van der Berg, Servaas; Burger, Ronelle; von Maltitz, Michael; and Gideon du Rand. (2008) "Using an Asset Index to Assess Trends in Poverty in Seven Sub-Saharan African Countries", *World Development*, Vol. 36, No. 6, pp. 1113–1130.

Caire, Dean. (2004) "Building Credit Scorecards for Small Business Lending in Developing Markets", `microfinance.com/English/Papers/Scoring_SMEs_Hybrid.pdf`, retrieved 31 January 2011.

Chen, Shiyuan; and Mark Schreiner. (2009a) "Simple Poverty Scorecard Poverty-Assessment Tool: Bangladesh", `SimplePovertyScorecard.com/BGD_2005_ENG.pdf`, retrieved 11 January 2016.

_____. (2009b) "Simple Poverty Scorecard Poverty-Assessment Tool: Vietnam", `SimplePovertyScorecard.com/VNM_2006_ENG.pdf`, retrieved 31 January 2011.

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*.

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

Falkenstein, Eric. (2008) "DefProb™: A Corporate Probability-of-Default Model", `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1103404`, retrieved 31 January 2011.

Ferguson, Brodie D.; Tandon, Ajay; Gakidou, Emmanuela; and Christopher J.L. Murray. (2003) "Estimating Permanent Income Using Indicator Variables", pp. 747–760 in Christopher J.L. Murray and David B. Evans (eds) *Health Systems Performance Assessment: Debates, Methods, and Empiricism*.

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.

Filmer, Deon; and Kinnon Scott. (2008) "Assessing Asset Indices", World Bank Policy Research Working Paper No. 4605, `papers.ssrn.com/sol3/papers.cfm?abstract_id=1149108`, retrieved 31 January 2011.

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`, retrieved 31 January 2011.

Goodman, L.A.; and Kruskal, W.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, `go.worldbank.org/VPMWVLU8E0`, retrieved 31 January 2011.

Grosh, Margaret; and Judy L. Baker. (1995) "Proxy Means Tests for Targeting Social Programs: Simulations and Speculation", LSMS Working Paper No. 118, `go.worldbank.org/W9OWN57PD0`, retrieved 31 January 2011.

Gwatkin, Davidson R.; Rutstein, Shea; Johnson, Kiersten; Suliman, Eldaw; Wagstaff, Adam; and Agbessi Amouzou. (2007) "Socio-Economic Differences in Health, Nutrition, and Population: Tanzania", 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.

Hentschel, Jesko; Lanjouw, Jean Olson; Lanjouw, Peter; and Javier Poggi. (2000) "Combining Census and Survey Data to Trace the Spatial Dimensions of Poverty: A Case Study of Ecuador", *World Bank Economic Review*, Vol. 14, No. 1, pp. 147–165.

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.

Howe, Laura D.; Hargreaves, James R.; Gabrysch, Sabine; and Sharon R.A. Huttly. (2009) "Is the Wealth Index a Proxy for Consumption Expenditure? A Systematic Review", *Journal of Epidemiology and Community Health*, Vol. 63, pp. 871–880.

IRIS Center. (2007) "Manual for the Implementation of USAID Poverty Assessment Tools", povertytools.org/training_documents/Manuals/USAID_PAT_Manual_Eng.pdf, retrieved 31 January 2011.

_____. (2005) "Notes on Assessment and Improvement of Tool Accuracy", povertytools.org/other_documents/AssessingImproving_Accuracy.pdf, retrieved 31 January 2011.

Johnson, Glenn. (2007) "Lesson 3: Two-Way Tables—Dependent Samples", www.stat.psu.edu/online/courses/stat504/03_2way/53_2way_compare.htm, retrieved 31 January 2011.

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`, retrieved 31 January 2011.

Mathiassen, Astrid. (2007) "A Model-Based Approach for Predicting Annual Poverty Rates without Expenditure Data", *Journal of Economic Inequality*, `DOI: 10.1007/s10888007-9059-7`.

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, `mfc.org.pl/doc/Research/ImpAct/SN/MFC_SN04_eng.pdf`, retrieved 31 January 2011.

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

Minot, Nicholas; Simler, Ken; Benson, Todd; Kilama, Blandina; Luvanda, Eliab; and Ahmed Makbel. (2006) "Poverty and Malnutrition in Tanzania: New Approaches for Examining Trends and Spatial Patterns", `www.repoa.or.tz/documents_storage/Research%20and%20Analysis/IFPRI%20Report.pdf`, retrieved 31 January 2011.

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.

National Bureau of Statistics (2009) *Main Report: Household Budget Survey 2007—Mainland Tanzania*, `www.nbs.go.tz/nada3/index.php/ddibrowser/2/download/65`, retrieved 29 January 2011.

_____. (2002a) *Household Budget Survey 2000/1*, www.tanzania.go.tz/hbs/HomePage_HBS.html, retrieved 29 January 2011.

_____. (2002b) "Key Findings: Household Budget Survey 2000/1", www.nbs.go.tz/nada3/index.php/ddibrowser/4/download/54, retrieved 29 January 2011.

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.

Ravallion, Martin. (1998) "Poverty Lines in Theory and Practice", World Bank LSMS Working Paper No. 133, go.worldbank.org/8P3IBJPQS1, retrieved 29 January 2011.

Rutstein, Shea Oscar; and Kiersten Johnson. (2004) "The DHS Wealth Index", DHS Comparative Reports No. 6, measuredhs.com/pubs/pdf/CR6/CR6.pdf, retrieved 29 January 2011.

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/63033/HTML/default/viewer.htm#statug_logistic_sect035.htm, retrieved 31 January 2011.

Schreiner, Mark. (2010) "Simple Poverty Scorecard Poverty-Assessment Tool: Honduras", SimplePovertyScorecard.com/HND_2007_ENG.pdf, retrieved 31 January 2011.

_____. (2009a) "Simple Poverty Scorecard Poverty-Assessment Tool Peru", SimplePovertyScorecard.com/PER_2007_ENG.pdf, retrieved 11 January 2016.

_____. (2009b) "Simple Poverty Scorecard Poverty-Assessment Tool: Philippines", SimplePovertyScorecard.com/PHL_2002_ENG.pdf, retrieved 31 January 2011.

_____. (2009c) "Simple Poverty Scorecard Poverty-Assessment Tool: Pakistan", SimplePovertyScorecard.com/PAK_2005_ENG.pdf, retrieved 31 January 2011.

_____. (2009d) "Simple Poverty Scorecard Poverty-Assessment Tool: Bolivia", SimplePovertyScorecard.com/BOL_2007_ENG.pdf, retrieved 11 January 2016.

_____. (2009e) "Simple Poverty Scorecard Poverty-Assessment Tool Mexico", SimplePovertyScorecard.com/MEX_2008_ENG.pdf, retrieved 31 January 2011.

_____. (2008a) "Simple Poverty Scorecard Poverty-Assessment Tool: Peru", SimplePovertyScorecard.com/PER_2003_ENG.pdf, retrieved 11 January 2016.

_____. (2008b) "Simple Poverty Scorecard Poverty-Assessment Tool: India", SimplePovertyScorecard.com/IND_2005_ENG.pdf, retrieved 11 January 2016.

_____. (2006) "Is One Simple Poverty Scorecard Poverty-Assessment Tool Enough for India?", microfinance.com/English/Papers/Scoring_Poverty_India_Segments.pdf, retrieved 31 January 2011.

_____. (2005a) "La Herramienta del Índice de Calificación de Pobreza™: México", SimplePovertyScorecard.com/MEX_2002_SPA.pdf, retrieved 31 January 2011.

_____. (2005b) "IRIS Questions on the Simple Poverty Scorecard Poverty-Assessment Tool", microfinance.com/English/Papers/Scoring_Poverty_Response_to_IRIS.pdf, retrieved 31 January 2011.

_____. (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, retrieved 31 January 2011.

_____; and Gary Woller. (2010a) "Simple Poverty Scorecard Poverty-Assessment Tool: Ghana", SimplePovertyScorecard.com/GHA_2005_ENG.pdf, retrieved 31 January 2011.

_____. (2010b) "Simple Poverty Scorecard Poverty-Assessment Tool: Guatemala", SimplePovertyScorecard.com/GTM_2006_ENG.pdf, retrieved 31 January 2011.

Setel, Philip; Abeyasekera, Savitri; Ward, Patrick; Hemed, Yusuf; Whiting, David; Mswia, Robert; and Manos Antoninis. (2005) "Development, Validation, and Performance of a Rapid Consumption Expenditure Proxy for Measuring Income Poverty in Tanzania: Experience from AMMP Demographic Surveillance Sites", pp. 169–183 in de Savigny, Razzaque, Debpuur, *et al.* (eds.) *Measuring Health Equity in Small Areas: Findings from Demographic Surveillance Systems.*

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

Simler, Kenneth R.; Harrower, Sarah; and Claudio Massingarela. (2003) "Estimating Poverty Indices from Simple Indicator Surveys", `www.csae.ox.ac.uk/conferences/2004-GPRaHDiA/papers/3p-SimlerMassingerela-CSAE2004.pdf`, retrieved 31 January 2011.

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, Pages 773–792.

Toohig, Jeff. (2007) "PPI Training Guide", `progressoutofpoverty.org/toolkit`, retrieved 31 January 2011.

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. (2008) "International Comparison Project: Tables of Results", siteresources.worldbank.org/ICPINT/Resources/icp-final-tables.pdf, retrieved 31 January 2011.

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

Appendix:
Guide to Interpretation of Scorecard Indicators

The following information comes from:

National Bureau of Statistics. (2006) "Instruction Manual", Dar es Salaam.

1. How many household members are 17-years-old or younger?

According to p. 37 of the "Instruction Manual", a *household* "may be one-person or multi-person-household.

"One-person household is a person who lives alone in whole or part of a housing unit and has independent consumption.

"Multi-person household is a group of two or more persons who occupy the whole or part of a housing unit and share their consumption. Usual households of this type contain husband, wife and children. Other relatives, boarders, visitors and their persons are included as members of the household if they pool their resources, share their consumption and have been living with the household for at least two weeks.

"Household servants will be counted as member of household if and only if they are taking their meals in that household and recognize the head of household as their head.

"Children who are at boarding school will be counted as household members. . . .

"A person who share residential and meals by paying will be considered as household member. But if s/he does not contribute and share meals with his/her resident s/he will counted as different household. . . .

"A husband with more than one wife and spend his time in more than one household will be counted as household member if he spent at least more than half of his time in that household.

"Therefore, following the above given definitions i.e. item 1 to 7, it can be stated that a household is formed when the members of the household share the consumption by pooling their resources together.

2. Do all children ages 6 to 17 attend school?

The "Instruction Manual" provides no additional information about this indicator.

3. Can the female head/spouse read and write?

The *female head/spouse* is:

- The head of the household, if the head is female
- The spouse of the head of the household, if the head is male and has a spouse
- Non-existent, if the head is male and does not have a spouse

According to p. 37 of the "Instruction Manual", the *household head* is "a person recognized as such by the rest of the members of the household to be the head. Often s/he is responsible for the financial support and welfare of the household members"

4. What is the main building material of the floor of the main dwelling?

According to p. 37 of the "Instruction Manual", the *dwelling* is "all the living space occupied by one household regardless of the physical arrangement of facilities available. It may be one room occupied by lodgers or it may be one, two or more housing units occupied by an extended family or household."

5. What is the main building material of the roof of the main dwelling?

According to p. 37 of the "Instruction Manual", the *dwelling* is "all the living space occupied by one household regardless of the physical arrangement of facilities available. It may be one room occupied by lodgers or it may be one, two or more housing units occupied by an extended family or household."

6. How many bicycles, mopeds, motorcycles, tractors, or motor vehicles does your household own?

According to p. 60 of the "Instruction Manual", "assets may be owned privately or jointly".

7. Does your household own any radios or radio cassettes?

According to p. 60 of the "Instruction Manual", "assets may be owned privately or jointly".

8. Does your household own any lanterns?

According to p. 60 of the "Instruction Manual", "assets may be owned privately or jointly".

9. Does your household own any irons (charcoal or electric)?

According to p. 60 of the "Instruction Manual", "assets may be owned privately or jointly".

10. How many tables does your household own?

According to p. 60 of the "Instruction Manual", "assets may be owned privately or jointly".

Figure 2: Sample sizes and poverty rates at the household level and the person level for mainland Tanzania and by region, sub-sample, and poverty line

			Poverty rates (% with expenditure below a poverty line) and poverty lines (TZS/person/day)						
				National ("Basic Needs") Lines				Intl. 2005 PPP	
	Level	Sample size	Natl. Food	100%	150%	200%	USAID 'Extreme'	$1.25/day	$2.50/day
Poverty lines:									
Mainland Tanzania	N/A		359	492	739	985	282	692	1,384
Dar es Salaam	N/A		468	641	961	1,281	404	900	1,800
Other urban	N/A		388	532	798	1,064	296	747	1,494
Rural	N/A		342	468	703	937	266	658	1,316
Poverty Rates:									
Mainland Tanzania	Households	10,405	12.6	26.6	51.8	69.6	12.6	61.3	90.0
	People	N/A	16.6	33.6	61.6	78.7	16.8	71.3	94.9
Dar es Salaam	Households	3,435	5.2	11.6	30.0	48.9	5.6	36.2	79.5
	People	N/A	7.4	16.4	38.1	57.4	8.2	46.0	86.7
Other urban	Households	3,707	9.5	18.8	39.5	58.7	8.8	49.4	85.2
	People	N/A	12.9	24.1	48.4	68.5	12.0	59.1	92.2
Rural	Households	3,263	14.4	30.8	58.1	75.3	14.6	68.0	92.7
	People	N/A	18.4	37.6	67.0	83.2	18.8	76.8	96.4
Construction and calibration samples									
Selecting indicators and points, and	Households	5,188	12.6	26.7	51.8	69.7	12.6	61.1	90.1
associating scores with likelihoods	People	N/A	16.2	33.4	61.2	78.5	16.5	71.5	95.1
Validation sample									
Measuring accuracy	Households	5,217	12.6	26.5	51.9	69.5	12.6	61.2	89.9
	People	N/A	17.0	33.9	62.0	78.8	17.1	71.1	94.7
Change in poverty rate (percentage points)									
Construction/calibration to validation	Households		+0.0	+0.2	–0.1	+0.1	–0.0	–0.1	+0.2

The national lines are in per-adult-equivalent units. The USAID "extreme" line and the $1.25/day and $2.50/day 2005 PPP lines are in per-person units.

Figure 3: Poverty indicators by uncertainty coefficient

Uncertainty coefficient	Indicator (Responses ordered starting with those most strongly indicative of poverty)
535	What is the main building material of the floor of the main dwelling? (Earth; Concrete, cement, tiles, timber, or other)
523	Does the household own any telephones (landline or mobile)? (No; Yes)
513	How many members does the household have? (Eight or more; Seven; Six; Five; Four; Three; Two; One)
494	How many household members are 18-years-old or younger? (Four or more; Three; Two; One; None)
464	How many household members are 17-years-old or younger? (Four or more; Three; Two; One; None)
447	How many household members are 16-years-old or younger? (Four or more; Three; Two; One; None)
415	How many household members are 15-years-old or younger? (Four or more; Three; Two; One; None)
405	What is the main building material of the roof of the main dwelling? (Mud and grass; Grass, leaves, bamboo; Concrete, cement, metal sheets (GCI), asbestos sheets, tiles, or other)
379	What is the major fuel used for cooking by the household? (Firewood, animal residuals, solar, gas (biogas), wood/farm residuals, or other; Charcoal, electricity, generator/private sources, gas (industrial), paraffin, or coal)
374	Does your household own any stoves (electric/gas or non-electric gas)? (No; Yes)
364	How many household members are 14-years-old or younger? (Four or more; Three; Two; One; None)
351	What is the main building material of the walls of the main dwelling? (Poles (including bamboo), branches, or grass; Mud only; Mud bricks; Poles and mud/mud and stones; Baked/burnt bricks; Concrete, cement, stones, or other)
337	Does your household own any sofas? (No; Yes)
331	What is the main building material of the foundation of the main dwelling? (No foundation, or other; Stones in mud-mortar; Concrete, cement, soil cement, baked/burnt bricks, stones in cement or in lime-mortar; Stones loosely laid)
331	What is the highest grade that the female head/spouse has completed? (No education, adult education only, pre-school, or Standard 1; Standards 2 to 6; Standard 7; No female head/spouse; Standard 8, course after primary education, Form I to VI, course after secondary education, course after form VI, diploma course, other certificate, or university degree)

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

Uncertainty coefficient	Indicator (Respsonses ordered starting with those most strongly associated with poverty)
318	How many household members are 13-years-old or younger? (Three or more; Two; One; None)
305	Does your household own any lanterns? (No; Yes)
299	What is the major fuel used for lighting by the household? (Not electricity; Electricity)
293	How many tables does your household own? (None; One; Two; Three or more)
284	What is the main building material of the roof-frame of the main dwelling? (Poles (including bamboo), or other; Sawn timber, or iron bars)
282	Does the household have electricity from the public utility? (No; Yes)
275	How many watches does your household own? (None; One; Two or more)
272	What is the highest grade that the male head/spouse has completed? (No education, adult education only, pre-school, or Standard 1; Standards 2 to 3; No male head/spouse; Standards 5 to 8; Course after primary education, or Forms I to II; Forms III to VI, course after secondary education, course after form VI, diploma course, other certificate, university degree)
269	What is the main drinking-water supply of the household? (Public well (unprotected), or other; Public well (protected); Private well (unprotected), or spring (unprotected); Piped water in community supply; Spring (protected), river, dam, lake, etc.; Private piped water outside housing unit, private well (protected), or rain catchment tank; Piped water in neighboring housing unit; Private piped water in housing unit, water vendors)
265	Does your household own any irons (charcoal or electric)? (No; Yes)
258	Can the male head/spouse read and write? (No, or yes (but not in English or Kiswahili); No male head/spouse; Yes, in Kiswahili; Yes, in English (regardless of others))
258	What toilet facilities does the household use? (No toilet, or other; Private pit latrine; Shared pit latrine; Private flush toilet, shared flush toilet, private VIP, or shared VIP)
242	How many household members are 12-years-old or younger? (Three or more; Two; One; None)
235	How many household members are 11-years-old or younger? (Three or more; Two; One; None)
221	How many cupboards, chests-of-drawers, boxes, wardrobes, and bookcases does your household have? (None; One; Two or more)

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

Uncertainty coefficient	Indicator (Respsonses ordered starting with those most strongly associated with poverty)
221	Do all children ages 6 to 17 attend school? (No; Yes, or no children ages 6 to 17)
206	Does your household own any radios or radio cassettes? (No; Yes)
201	Does the household own any televisions, videos, or dish antennas/decoders? (No; Yes)
195	How many acres of land for farming/pastoralism are owned by the household this year? (None; >0 to 1; >1 to 2; >2 to 3; >3 to 5; >5)
191	How many acres of land for farming/pastoralism are owned or used (that are not owned) by the household this year? (None; >0 to 1; >1 to 2; >2 to 3; >3 to 5; >5)
186	What is the tenure status of the household in its main dwelling? (Owned by household; Not owned by the household)
180	Do all children ages 6 to 18 attend school? (No; Yes; No children ages 6 to 18)
177	Do all children ages 6 to 16 attend school? (No; Yes; No children ages 6 to 16)
177	Can the female head/spouse read and write? (No; Yes, but not in Kiswahili nor English; No female head/spouse; Yes, only in Kiswahili; Yes, in English (regardless of others))
167	Do all children ages 6 to 15 attend school? (No; Yes; No children ages 6 to 15)
161	How many chairs does your household own? (None; One; Two; Three: Four; Five or more)
152	Do all children ages 6 to 14 attend school? (No; Yes; No children ages 6 to 14)
152	Does the household own any refrigerators or freezers? (No; Yes)
146	Does the household have any cooling facilities? (No; Yes)
137	Do all children ages 6 to 13 attend school? (No; Yes; No children ages 6 to 13)
128	How many mosquito nets does your household own? (None; One; Two or more)
127	How many household members are 11-years-old or younger? (One or more; None)
116	How many books (not school books) does your household own? (None; One; Two; Three: Four or more)
116	Do all children ages 6 to 12 attend school? (No; Yes; No children ages 6 to 12)
113	What is the marital status of the female head/spouse? (Widow; Separated; Married; Living together; Divorced; No female head/spouse; Never married)

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

113	Do all children ages 6 to 11 attend school? (No; Yes; No children ages 6 to 11)
106	How many houses does your household own? (None; One; Two or more)
102	How many fields/land does your household own? (Four or more; Three; Two; One; None)
98	How many doors from the outside lead into the main dwelling used by the household? (None; One; Two or more)
88	Does your household own any fans/air conditioners? (No; Yes)
82	How do you dispose of your garbage? (Thrown inside compound; Thrown outside compound; Rubbish pit inside compound; Rubbish pit outside compound; Rubbish bin, or other)
78	How many rooms in the main dwelling used by the household? (One; Two; Three; Four or more)
78	Does the household own any sewing machines? (No; Yes)
73	How many beds does your household own? (None; One; Two; Three; Four or more)
69	How many hoes does your household own? (Five or more; Four; Three; Two; One; None)
59	What is the marital status of the male head/spouse? (No male head/spouse; Married or widowed; Living together, separated, divorced, or never married)
54	How many windows are in the main dwelling used by the household? (None; One; Two; Three; Four; Five or more)
46	What is the structure of household headship? (Female head/spouse only; Both male and female heads/spouses; Male head/spouse only)
38	How many bicycles, mopeds, motorcycles, tractors, or motor vehicles does your household own? (None; One; Two or more)
35	How many rooms in the main dwelling are used by the household for sleeping? (One; Two; Three or more)
20	How many bicycles does your household own? (None; One; Two or more)
14	How many cattle, other large livestock, sheep, goats, or other medium-sized animals are owned by the household this year? (One or more; None)
14	Does your household own any ploughs? (Yes; No)
14	Does your household own any poultry? (Yes; No)
12	Can all household members read and write? (No; Yes)

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

Uncertainty coefficient	Indicator (Responses ordered starting with those most strongly associated with poverty)
9	Does the household own any cattle and other large livestock? (Yes; No)
5	How many acres of land does the household use that it does not own this year? (None; >0 to 1; >1)
4	Does your household own any livestock? (Yes; No)
3	How many sheep, goats, or other medium-sized animals are owned by the household? (No; Yes)
0	Does your household own any water heaters? (No; Yes)

Source: 2004/5 IHS, national poverty line.

Tables for the National Poverty Line

(and tables pertaining to all seven poverty lines)

Figure 4 (National 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	81.3
5–9	70.8
10–14	64.8
15–19	57.2
20–24	53.5
25–29	48.4
30–34	38.7
35–39	29.6
40–44	22.8
45–49	21.2
50–54	17.0
55–59	12.0
60–64	7.8
65–69	7.0
70–74	3.2
75–79	2.0
80–84	2.0
85–89	0.0
90–94	0.0
95–100	0.0

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

Score	Households below poverty line		All households at score		Poverty likelihood (estimated, %)
0–4	429	÷	527	=	81.3
5–9	868	÷	1,226	=	70.8
10–14	1,522	÷	2,350	=	64.8
15–19	2,057	÷	3,599	=	57.2
20–24	2,728	÷	5,094	=	53.5
25–29	4,150	÷	8,576	=	48.4
30–34	3,494	÷	9,033	=	38.7
35–39	2,485	÷	8,395	=	29.6
40–44	2,482	÷	10,897	=	22.8
45–49	2,355	÷	11,125	=	21.2
50–54	1,463	÷	8,622	=	17.0
55–59	880	÷	7,365	=	12.0
60–64	624	÷	8,015	=	7.8
65–69	442	÷	6,353	=	7.0
70–74	142	÷	4,407	=	3.2
75–79	40	÷	2,035	=	2.0
80–84	25	÷	1,295	=	2.0
85–89	0	÷	673	=	0.0
90–94	0	÷	342	=	0.0
95–100	0	÷	70	=	0.0

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

Figure 6 (National poverty lines): Distribution of household poverty likelihoods across ranges demarcated by national poverty lines in units of adult equivalents

	Likelihood (%) of having expenditure in range demarcated by adjacent poverty lines				
	<Food	=>Food and <100% Natl.	=>100% Natl. and <150% Natl.	=>150% Natl. and <200% Natl.	=>200% Natl.
Score	<TZS359	=>TZS359 and <TZS492	=>TZS492 and <TZS739	=>TZS739 and <TZS985	=>TZS985
0–4	55.2	26.2	14.4	2.9	1.4
5–9	45.9	24.9	22.5	4.6	2.1
10–14	33.8	31.0	23.6	9.5	2.1
15–19	31.2	26.0	24.9	11.5	6.4
20–24	30.9	22.7	28.0	10.7	7.8
25–29	26.1	22.3	33.1	10.7	7.8
30–34	17.6	21.1	34.5	16.5	10.4
35–39	13.2	16.4	28.3	23.1	19.0
40–44	7.7	15.1	31.6	20.8	24.9
45–49	7.4	13.8	29.6	20.1	29.2
50–54	7.4	9.6	23.9	21.8	37.4
55–59	5.4	6.5	19.5	22.7	45.8
60–64	3.5	4.3	19.3	18.7	54.2
65–69	0.7	6.2	12.4	18.3	62.3
70–74	0.7	2.5	9.6	21.2	66.0
75–79	0.7	1.3	4.9	15.2	78.0
80–84	0.6	1.4	4.9	12.3	80.9
85–89	0.0	0.0	1.7	9.9	88.5
90–94	0.0	0.0	0.0	5.9	94.2
95–100	0.0	0.0	0.0	0.0	100.0

Figure 6 (USAID "extreme" and 2005 PPP poverty lines): Distribution of household poverty likelihoods across ranges demarcated by USAID "extreme" and 2005 PPP poverty lines in units of people

	Likelihood (%) of having expenditure in range demarcated by adjacent poverty			
Score	<USAID	=>USAID and <$1.25/day	=>$1.25/day and <$2.50/day	=>$2.50/day
	<TSZ282	=>TSZ282 and <TSZ692	=>TSZ692 and <TSZ1,384	=>TSZ1,384
0–4	70.2	27.6	2.2	0.0
5–9	50.0	46.5	3.5	0.0
10–14	37.3	59.2	3.5	0.0
15–19	35.2	57.4	6.8	0.5
20–24	33.9	55.6	10.0	0.5
25–29	25.9	63.6	9.7	0.8
30–34	16.6	69.0	13.4	1.1
35–39	12.9	61.0	24.0	2.1
40–44	7.3	59.5	28.3	5.0
45–49	7.3	52.7	33.5	6.6
50–54	6.0	45.9	36.9	11.1
55–59	4.0	34.1	49.6	12.4
60–64	2.8	31.9	46.4	19.0
65–69	0.6	24.8	49.7	24.9
70–74	0.6	16.5	46.2	36.6
75–79	0.6	7.8	47.9	43.7
80–84	0.5	6.2	45.2	48.1
85–89	0.0	0.5	31.6	67.9
90–94	0.0	0.0	20.1	79.9
95–100	0.0	0.0	2.3	97.7

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

	Difference between estimate and true value			
		Confidence interval (+/– percentage points)		
Score	**Diff.**	**90-percent**	**95-percent**	**99-percent**
0–4	–11.4	7.7	8.1	8.4
5–9	–15.4	9.7	10.0	10.8
10–14	–0.2	4.4	5.2	6.8
15–19	–6.3	4.9	5.2	5.8
20–24	–0.5	2.9	3.4	4.6
25–29	+3.9	2.3	2.8	3.5
30–34	–4.1	3.2	3.4	4.0
35–39	+0.9	2.2	2.7	3.3
40–44	–2.9	2.4	2.6	3.1
45–49	+1.2	1.8	2.2	3.0
50–54	–4.2	3.2	3.5	3.9
55–59	–1.6	2.0	2.5	3.3
60–64	+1.6	1.3	1.5	1.9
65–69	+2.8	1.2	1.5	1.8
70–74	–0.1	1.4	1.6	2.2
75–79	+1.6	0.3	0.4	0.5
80–84	–24.9	16.5	17.1	18.6
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 α factor for bootstrapped estimates of poverty rates for groups of households at a point in time for the scorecard applied to the validation sample

	Poverty line				
	Natl.	National ("Basic Needs") Lines			USAID	Intl. 2005 PPP	
	Food	100%	150%	200%	'Extreme'	$1.25/day	$2.50/day
Estimate minus true value							
Scorecard applied to validation sample	–0.8	–1.3	–1.1	–0.1	–0.7	–0.9	+0.1
Precision of difference							
Scorecard applied to validation sample	0.5	0.6	0.7	0.6	0.5	0.6	0.4
α factor							
Scorecard applied to validation sample	1.23	1.12	0.99	0.97	1.21	0.95	0.96
Precision is measured as 90-percent confidence intervals in units of +/– percentage points.							
Differences and precision estimated from 500 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.							
The four national lines are in per-adult-equivalent units.							
The USAID "extreme" line and the $1.25/day and $2.50/day 2005 PPP lines are in per-person units.							

Figure 9 (National line): Differences and precision of differences for bootstrapped estimates of households' poverty likelihoods, by sample size, scorecard applied to validation sample

Sample Size	Difference between estimate and true value			
		Confidence interval (+/– percentage points)		
n	Diff.	90-percent	95-percent	99-percent
1	–1.2	63.6	68.3	80.8
4	–0.4	38.0	43.5	55.7
8	–0.4	27.2	30.9	43.0
16	–1.1	20.0	23.7	30.6
32	–1.2	14.5	17.2	22.5
64	–1.3	10.8	12.9	15.6
128	–1.2	7.2	8.9	11.6
256	–1.3	4.925	5.7	7.3
512	–1.3	3.5	4.1	5.4
1,024	–1.3	2.5	2.9	4.2
2,048	–1.2	1.8	2.1	2.7
4,096	–1.3	1.3	1.5	2.0
8,192	–1.3	0.920	1.1	1.4
16,384	–1.3	0.645	0.8	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 line): Households by targeting classification and score, along with "Total Accuracy" and BPAC, scorecard applied to 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	0.5	26.0	0.0	73.4	73.9	–96.2
5–9	1.5	25.0	0.2	73.3	74.8	–87.6
10–14	3.0	23.5	1.1	72.4	75.4	–73.1
15–19	5.2	21.4	2.5	70.9	76.1	–51.5
20–24	8.0	18.6	4.8	68.6	76.6	–21.7
25–29	11.7	14.8	9.7	63.8	75.5	+24.7
30–34	15.5	11.0	14.9	58.6	74.0	+43.7
35–39	18.0	8.6	20.8	52.6	70.6	+21.4
40–44	20.7	5.9	29.0	44.4	65.1	–9.5
45–49	22.8	3.8	38.1	35.4	58.2	–43.6
50–54	24.4	2.1	45.0	28.4	52.8	–69.9
55–59	25.4	1.1	51.4	22.1	47.4	–93.9
60–64	25.9	0.6	58.9	14.6	40.5	–122.2
65–69	26.2	0.3	65.0	8.5	34.7	–145.0
70–74	26.4	0.2	69.2	4.3	30.6	–161.0
75–79	26.4	0.1	71.2	2.3	28.6	–168.6
80–84	26.5	0.0	72.4	1.1	27.6	–173.0
85–89	26.5	0.0	73.1	0.4	26.9	–175.5
90–94	26.5	0.0	73.4	0.1	26.6	–176.8
95–100	26.5	0.0	73.5	0.0	26.5	–177.1

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

Figure 12 (National line): Households below the poverty line and all households at a given score or at or below a given score cut-off, scorecard applied to 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	0.5	92.1	1.8	11.7:1
5–9	1.8	87.3	5.8	6.9:1
10–14	4.1	73.9	11.4	2.8:1
15–19	7.7	67.0	19.5	2.0:1
20–24	12.8	62.2	30.0	1.6:1
25–29	21.4	54.7	44.1	1.2:1
30–34	30.4	50.9	58.4	1.0:1
35–39	38.8	46.3	67.7	0.9:1
40–44	49.7	41.6	77.9	0.7:1
45–49	60.8	37.4	85.8	0.6:1
50–54	69.4	35.1	92.0	0.5:1
55–59	76.8	33.0	95.7	0.5:1
60–64	84.8	30.5	97.7	0.4:1
65–69	91.2	28.8	98.9	0.4:1
70–74	95.6	27.6	99.4	0.4:1
75–79	97.6	27.0	99.5	0.4:1
80–84	98.9	26.8	100.0	0.4:1
85–89	99.6	26.6	100.0	0.4:1
90–94	99.9	26.5	100.0	0.4:1
95–100	100.0	26.5	100.0	0.4:1

Tables for the Food Poverty Line

Figure 4 (Food 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	55.2
5–9	45.9
10–14	33.8
15–19	31.2
20–24	30.9
25–29	26.1
30–34	17.6
35–39	13.2
40–44	7.7
45–49	7.4
50–54	7.4
55–59	5.4
60–64	3.5
65–69	0.7
70–74	0.7
75–79	0.7
80–84	0.6
85–89	0.0
90–94	0.0
95–100	0.0

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

	Difference between estimate and true value			
		Confidence interval (+/− percentage points)		
Score	Diff.	90-percent	95-percent	99-percent
0–4	–2.2	8.7	10.5	13.7
5–9	+0.6	6.2	7.3	9.7
10–14	+1.8	4.3	5.1	6.4
15–19	–5.5	4.6	5.0	5.8
20–24	+7.8	2.5	2.9	3.7
25–29	+0.7	2.1	2.6	3.5
30–34	–7.0	4.6	4.7	5.1
35–39	+1.5	1.6	1.9	2.5
40–44	–2.3	1.8	1.9	2.2
45–49	–2.4	1.9	2.0	2.3
50–54	+1.8	1.2	1.4	2.1
55–59	–0.8	1.4	1.6	2.2
60–64	+2.5	0.4	0.4	0.6
65–69	–2.6	1.9	2.0	2.3
70–74	–2.2	1.8	2.0	2.3
75–79	+0.5	0.3	0.3	0.4
80–84	+0.6	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 (Food line): Differences and precision of differences for bootstrapped estimates of households' poverty likelihoods, by sample size, scorecard applied to validation sample

Sample Size	Difference between estimate and true value			
		Confidence interval (+/– percentage points)		
n	Diff.	90-percent	95-percent	99-percent
1	–0.1	56.6	63.0	68.9
4	+0.5	29.0	35.2	44.2
8	–0.1	21.3	25.1	34.9
16	–0.7	16.3	18.5	24.1
32	–0.8	10.8	12.9	16.7
64	–0.6	8.3	9.8	11.9
128	–0.6	6.1	7.1	8.7
256	–0.7	4.0	4.8	6.6
512	–0.7	2.9	3.4	4.8
1,024	–0.8	2.1	2.5	3.3
2,048	–0.8	1.5	1.8	2.5
4,096	–0.8	1.1	1.3	1.7
8,192	–0.8	0.8	0.9	1.2
16,384	–0.8	0.5	0.6	0.9

Figure 11 (Food line): Households by targeting classification and score, along with "Total Accuracy" and BPAC, scorecard applied to 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	0.3	12.3	0.2	87.2	87.5	–93.6
5–9	0.8	11.8	0.9	86.5	87.3	–79.5
10–14	1.7	10.9	2.4	85.0	86.7	–54.1
15–19	2.8	9.8	4.9	82.5	85.4	–16.4
20–24	4.1	8.5	8.7	78.7	82.8	+30.8
25–29	6.1	6.4	15.2	72.2	78.3	–21.2
30–34	8.2	4.4	22.2	65.2	73.4	–76.5
35–39	9.1	3.4	29.7	57.8	66.9	–135.8
40–44	10.3	2.3	39.4	48.0	58.3	–213.5
45–49	11.2	1.4	49.6	37.8	49.0	–294.4
50–54	11.7	0.9	57.8	29.6	41.3	–359.3
55–59	12.1	0.5	64.7	22.7	34.9	–414.3
60–64	12.3	0.3	72.5	14.9	27.2	–476.6
65–69	12.5	0.1	78.7	8.7	21.2	–525.7
70–74	12.6	0.0	83.0	4.4	17.0	–560.0
75–79	12.6	0.0	85.0	2.4	15.0	–576.1
80–84	12.6	0.0	86.3	1.1	13.7	–586.4
85–89	12.6	0.0	87.0	0.4	13.0	–591.7
90–94	12.6	0.0	87.4	0.1	12.6	–594.4
95–100	12.6	0.0	87.4	0.0	12.6	–595.0

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

Figure 12 (Food line): Households below the poverty line and all households at a given score or at or below a given score cut-off, scorecard applied to 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	0.5	53.6	2.2	1.2:1
5–9	1.8	46.7	6.5	0.9:1
10–14	4.1	40.8	13.3	0.7:1
15–19	7.7	36.6	22.4	0.6:1
20–24	12.8	32.0	32.6	0.5:1
25–29	21.4	28.7	48.8	0.4:1
30–34	30.4	27.0	65.2	0.4:1
35–39	38.8	23.6	72.7	0.3:1
40–44	49.7	20.7	81.6	0.3:1
45–49	60.8	18.4	89.1	0.2:1
50–54	69.4	16.8	92.7	0.2:1
55–59	76.8	15.8	96.4	0.2:1
60–64	84.8	14.5	97.8	0.2:1
65–69	91.2	13.7	99.2	0.2:1
70–74	95.6	13.1	99.9	0.2:1
75–79	97.6	12.9	100.0	0.1:1
80–84	98.9	12.7	100.0	0.1:1
85–89	99.6	12.6	100.0	0.1:1
90–94	99.9	12.6	100.0	0.1:1
95–100	100.0	12.6	100.0	0.1:1

Tables for
150% of the National ("Basic Needs") Poverty Line

Figure 4 (150% of national 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	95.7
5–9	93.3
10–14	88.4
15–19	82.1
20–24	81.5
25–29	81.5
30–34	73.1
35–39	57.9
40–44	54.3
45–49	50.8
50–54	40.8
55–59	31.5
60–64	27.1
65–69	19.4
70–74	12.8
75–79	6.9
80–84	6.8
85–89	1.7
90–94	0.0
95–100	0.0

Figure 7 (150% of national line): Bootstrapped differences between estimated and true poverty likelihoods for households in a large sample (n = 16,384) from the validation sample, with confidence intervals

	Difference between estimate and true value			
		Confidence interval (+/– percentage points)		
Score	Diff.	90-percent	95-percent	99-percent
0–4	+3.1	4.3	5.3	7.5
5–9	–6.7	3.3	3.3	3.3
10–14	–5.0	3.4	3.5	3.8
15–19	–4.1	3.2	3.5	3.8
20–24	–6.7	4.3	4.5	4.8
25–29	+5.8	2.1	2.5	3.1
30–34	+0.2	2.0	2.4	3.0
35–39	–6.1	4.2	4.5	4.8
40–44	–1.1	2.1	2.6	3.3
45–49	+3.0	2.1	2.5	3.5
50–54	–5.5	4.1	4.3	4.9
55–59	+0.5	2.7	3.2	4.3
60–64	–2.0	2.6	3.1	4.0
65–69	+3.7	2.3	2.8	3.6
70–74	+3.0	2.1	2.5	3.2
75–79	–0.7	3.3	3.9	5.0
80–84	–24.9	16.4	17.2	18.8
85–89	+0.5	1.0	1.2	1.5
90–94	+0.0	0.0	0.0	0.0
95–100	+0.0	0.0	0.0	0.0

Figure 9 (150% of national line): Differences and precision of differences for bootstrapped estimates of households' poverty likelihoods, by sample size, scorecard applied to validation sample

Sample Size	Difference between estimate and true value			
		Confidence interval (+/– percentage points)		
n	Diff.	90-percent	95-percent	99-percent
1	–1.8	73.0	81.1	87.5
4	–2.0	42.5	51.5	64.5
8	–0.8	30.2	35.0	46.4
16	–1.2	21.2	25.4	32.3
32	–1.1	15.1	18.0	23.4
64	–1.0	10.4	12.7	16.6
128	–1.1	7.2	8.8	11.9
256	–1.1	5.3	6.1	7.7
512	–1.1	3.7	4.4	6.0
1,024	–1.1	2.6	3.0	4.0
2,048	–1.1	1.8	2.1	2.8
4,096	–1.1	1.2	1.5	2.0
8,192	–1.1	0.9	1.0	1.3
16,384	–1.1	0.7	0.8	1.0

Figure 11 (150% of national line): Households by targeting classification and score, along with "Total Accuracy" and BPAC, scorecard applied to 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	0.5	51.4	0.0	48.1	48.5	–98.0
5–9	1.7	50.2	0.0	48.1	49.8	–93.3
10–14	3.9	48.0	0.2	47.9	51.7	–84.6
15–19	6.9	45.0	0.8	47.3	54.2	–71.8
20–24	11.4	40.5	1.4	46.7	58.1	–53.4
25–29	18.0	33.9	3.4	44.7	62.7	–24.1
30–34	24.6	27.3	5.8	42.3	66.8	+5.9
35–39	30.0	21.9	8.8	39.3	69.3	+32.6
40–44	36.1	15.8	13.6	34.5	70.6	+65.3
45–49	41.4	10.5	19.4	28.7	70.1	+62.6
50–54	45.2	6.7	24.2	23.9	69.1	+53.3
55–59	47.7	4.2	29.1	19.0	66.7	+43.9
60–64	49.8	2.1	35.0	13.1	62.9	+32.6
65–69	50.9	1.0	40.3	7.8	58.8	+22.4
70–74	51.5	0.4	44.1	4.0	55.5	+15.0
75–79	51.7	0.2	46.0	2.1	53.8	+11.4
80–84	51.9	0.0	47.0	1.1	52.9	+9.3
85–89	51.9	0.0	47.7	0.4	52.3	+8.1
90–94	51.9	0.0	48.0	0.1	52.0	+7.4
95–100	51.9	0.0	48.1	0.0	51.9	+7.3

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

Figure 12 (150% of national line): Households below the poverty line and all households at a given score or at or below a given score cut-off, scorecard applied to 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	0.5	92.1	0.9	11.7:1
5–9	1.8	97.6	3.3	41.2:1
10–14	4.1	94.2	7.4	16.1:1
15–19	7.7	89.9	13.3	8.9:1
20–24	12.8	89.1	22.0	8.2:1
25–29	21.4	84.2	34.7	5.3:1
30–34	30.4	80.8	47.3	4.2:1
35–39	38.8	77.3	57.8	3.4:1
40–44	49.7	72.6	69.5	2.7:1
45–49	60.8	68.1	79.8	2.1:1
50–54	69.4	65.1	87.1	1.9:1
55–59	76.8	62.1	91.9	1.6:1
60–64	84.8	58.7	96.0	1.4:1
65–69	91.2	55.8	98.1	1.3:1
70–74	95.6	53.9	99.2	1.2:1
75–79	97.6	52.9	99.6	1.1:1
80–84	98.9	52.4	100.0	1.1:1
85–89	99.6	52.1	100.0	1.1:1
90–94	99.9	51.9	100.0	1.1:1
95–100	100.0	51.9	100.0	1.1:1

Tables for
200% of the National ("Basic Needs") Poverty Line

Figure 4 (200% of national 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	98.6
5–9	97.9
10–14	97.9
15–19	93.6
20–24	92.2
25–29	92.2
30–34	89.6
35–39	81.0
40–44	75.1
45–49	70.8
50–54	62.7
55–59	54.2
60–64	45.8
65–69	37.7
70–74	34.0
75–79	22.0
80–84	19.1
85–89	11.5
90–94	5.9
95–100	0.0

Figure 7 (200% of national line): Bootstrapped differences between estimated and true poverty likelihoods for households in a large sample (n = 16,384) from the validation sample, with confidence intervals

	Difference between estimate and true value			
		Confidence interval (+/− percentage points)		
Score	Diff.	90-percent	95-percent	99-percent
0–4	–1.4	0.7	0.7	0.7
5–9	–2.1	1.1	1.1	1.1
10–14	–0.0	1.0	1.1	1.5
15–19	–4.2	2.5	2.6	2.7
20–24	–3.8	2.5	2.6	2.8
25–29	+3.4	1.5	1.8	2.2
30–34	–1.6	1.4	1.6	2.1
35–39	–1.3	2.1	2.4	3.2
40–44	–0.6	1.9	2.3	2.9
45–49	+3.5	2.1	2.5	3.2
50–54	–4.0	3.2	3.5	3.9
55–59	+10.5	2.9	3.4	4.9
60–64	–5.7	4.3	4.5	5.1
65–69	+0.9	3.2	3.8	5.1
70–74	+9.3	3.1	3.9	5.1
75–79	–5.9	5.8	6.7	8.9
80–84	–20.8	14.0	15.3	16.2
85–89	+10.3	1.0	1.2	1.5
90–94	–24.1	18.5	19.7	22.5
95–100	–4.4	6.8	8.3	10.7

Figure 9 (200% of national line): Differences and precision of differences for bootstrapped estimates of households' poverty likelihoods, by sample size, scorecard applied to validation sample

Sample Size	Difference between estimate and true value			
		Confidence interval (+/– percentage points)		
n	Diff.	90-percent	95-percent	99-percent
1	–1.2	68.7	73.5	86.6
4	–0.9	37.8	44.1	61.9
8	+0.0	26.3	30.9	42.2
16	–0.5	18.1	21.8	26.6
32	–0.3	13.2	15.3	18.9
64	–0.2	9.3	10.6	13.8
128	–0.3	6.3	7.5	9.8
256	–0.3	4.6	5.5	7.2
512	–0.2	3.2	3.9	5.2
1,024	–0.1	2.3	2.6	3.7
2,048	–0.1	1.6	2.0	2.6
4,096	–0.1	1.2	1.4	1.7
8,192	–0.1	0.8	0.9	1.2
16,384	–0.1	0.6	0.7	0.9

Figure 11 (200% of national line): Households by targeting classification and score, along with "Total Accuracy" and BPAC, scorecard applied to 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	0.5	69.0	0.0	30.5	31.0	–98.5
5–9	1.8	67.8	0.0	30.5	32.2	–95.0
10–14	4.0	65.5	0.1	30.4	34.4	–88.3
15–19	7.5	62.0	0.2	30.3	37.8	–78.1
20–24	12.4	57.1	0.4	30.1	42.5	–63.7
25–29	20.1	49.4	1.3	29.2	49.3	–40.3
30–34	28.3	41.2	2.1	28.4	56.7	–15.5
35–39	35.4	34.1	3.4	27.1	62.5	+6.8
40–44	43.8	25.7	5.9	24.6	68.4	+34.5
45–49	51.5	18.0	9.3	21.2	72.7	+61.6
50–54	57.3	12.3	12.2	18.3	75.6	+82.3
55–59	61.0	8.6	15.9	14.6	75.6	+77.2
60–64	64.9	4.7	20.0	10.5	75.4	+71.3
65–69	67.3	2.3	23.9	6.6	73.8	+65.6
70–74	68.5	1.0	27.1	3.4	71.9	+61.0
75–79	69.1	0.5	28.6	1.9	71.0	+58.9
80–84	69.4	0.1	29.5	1.0	70.4	+57.6
85–89	69.4	0.1	30.2	0.3	69.8	+56.6
90–94	69.5	0.0	30.4	0.1	69.6	+56.3
95–100	69.5	0.0	30.5	0.0	69.5	+56.2

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

Figure 12 (200% of national line): Households below the poverty line and all households at a given score or at or below a given score cut-off, scorecard applied to 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	0.5	100.0	0.8	Only poor targeted
5–9	1.8	100.0	2.5	Only poor targeted
10–14	4.1	98.4	5.8	60.9:1
15–19	7.7	97.7	10.8	43.4:1
20–24	12.8	97.1	17.9	33.2:1
25–29	21.4	94.1	28.9	15.9:1
30–34	30.4	93.2	40.8	13.7:1
35–39	38.8	91.3	51.0	10.5:1
40–44	49.7	88.1	63.0	7.4:1
45–49	60.8	84.7	74.1	5.5:1
50–54	69.4	82.5	82.4	4.7:1
55–59	76.8	79.4	87.7	3.8:1
60–64	84.8	76.5	93.3	3.2:1
65–69	91.2	73.8	96.7	2.8:1
70–74	95.6	71.7	98.5	2.5:1
75–79	97.6	70.7	99.3	2.4:1
80–84	98.9	70.2	99.8	2.4:1
85–89	99.6	69.7	99.9	2.3:1
90–94	99.9	69.6	100.0	2.3:1
95–100	100.0	69.5	100.0	2.3:1

Tables for the USAID "Extreme" Poverty Line

Figure 4 (USAID "extreme" 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	70.2
5–9	50.0
10–14	37.3
15–19	35.2
20–24	33.9
25–29	25.9
30–34	16.6
35–39	12.9
40–44	7.3
45–49	7.3
50–54	6.0
55–59	4.0
60–64	2.8
65–69	0.6
70–74	0.6
75–79	0.6
80–84	0.5
85–89	0.0
90–94	0.0
95–100	0.0

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

	Difference between estimate and true value			
		Confidence interval (+/– percentage points)		
Score	Diff.	90-percent	95-percent	99-percent
0–4	+7.0	8.5	10.3	13.8
5–9	+1.6	6.3	7.4	9.1
10–14	+1.5	4.4	5.1	7.0
15–19	–6.8	5.2	5.6	6.2
20–24	+8.1	2.6	3.1	4.3
25–29	–0.8	2.1	2.6	3.3
30–34	–10.5	6.3	6.6	7.0
35–39	+1.2	1.6	1.9	2.4
40–44	–1.1	1.2	1.5	2.0
45–49	+0.4	1.2	1.4	2.0
50–54	+3.3	0.8	1.0	1.2
55–59	+0.1	1.1	1.3	1.6
60–64	+1.8	0.4	0.4	0.5
65–69	–2.5	1.9	2.0	2.2
70–74	–2.3	1.9	2.0	2.3
75–79	+0.5	0.2	0.2	0.2
80–84	+0.5	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" line): Differences and precision of differences for bootstrapped estimates of households' poverty likelihoods, by sample size, scorecard applied to 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	58.6	61.2	67.8
4	+0.5	27.7	33.7	43.3
8	–0.1	21.4	25.5	35.3
16	–0.7	16.3	19.4	24.5
32	–0.7	11.2	13.5	16.9
64	–0.5	8.2	9.5	12.3
128	–0.6	5.9	6.8	8.8
256	–0.6	4.0	4.8	6.4
512	–0.6	2.8	3.4	4.6
1,024	–0.7	2.1	2.5	3.1
2,048	–0.7	1.5	1.7	2.3
4,096	–0.7	1.1	1.2	1.6
8,192	–0.7	0.7	0.9	1.2
16,384	–0.7	0.5	0.6	0.8

Figure 11 (USAID "extreme" line): Households by targeting classification and score, along with "Total Accuracy" and BPAC, scorecard applied to 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	0.3	12.3	0.2	87.2	87.5	–93.3
5–9	0.9	11.7	0.9	86.5	87.4	–79.0
10–14	1.8	10.8	2.3	85.1	86.9	–53.1
15–19	3.2	9.5	4.5	82.8	86.0	–14.0
20–24	4.5	8.1	8.3	79.1	83.6	+34.6
25–29	6.7	5.9	14.7	72.7	79.4	–16.2
30–34	9.0	3.6	21.4	66.0	75.0	–69.4
35–39	10.0	2.7	28.8	58.5	68.5	–128.2
40–44	10.9	1.7	38.8	48.6	59.5	–206.8
45–49	11.6	1.0	49.2	38.2	49.8	–289.3
50–54	11.9	0.7	57.5	29.9	41.8	–355.0
55–59	12.2	0.4	64.6	22.8	35.0	–410.9
60–64	12.4	0.2	72.4	14.9	27.3	–473.2
65–69	12.5	0.1	78.6	8.7	21.3	–522.3
70–74	12.6	0.0	83.0	4.4	17.0	–556.4
75–79	12.6	0.0	85.0	2.4	15.0	–572.5
80–84	12.6	0.0	86.3	1.1	13.7	–582.7
85–89	12.6	0.0	87.0	0.4	13.0	–588.0
90–94	12.6	0.0	87.3	0.1	12.7	–590.7
95–100	12.6	0.0	87.4	0.0	12.6	–591.3

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

Figure 12 (USAID "extreme" line): Households below the poverty line and all households at a given score or at or below a given score cut-off, scorecard applied to 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	0.5	60.6	2.5	1.5:1
5–9	1.8	51.3	7.1	1.1:1
10–14	4.1	44.3	14.4	0.8:1
15–19	7.7	41.1	25.1	0.7:1
20–24	12.8	35.4	35.9	0.5:1
25–29	21.4	31.3	53.0	0.5:1
30–34	30.4	29.6	71.2	0.4:1
35–39	38.8	25.7	78.8	0.3:1
40–44	49.7	22.0	86.4	0.3:1
45–49	60.8	19.1	92.0	0.2:1
50–54	69.4	17.2	94.5	0.2:1
55–59	76.8	15.9	96.9	0.2:1
60–64	84.8	14.6	98.0	0.2:1
65–69	91.2	13.8	99.2	0.2:1
70–74	95.6	13.2	99.9	0.2:1
75–79	97.6	12.9	100.0	0.1:1
80–84	98.9	12.8	100.0	0.1:1
85–89	99.6	12.7	100.0	0.1:1
90–94	99.9	12.6	100.0	0.1:1
95–100	100.0	12.6	100.0	0.1:1

Tables for the $1.25/day 2005 PPP Poverty Line

Figure 4 ($1.25/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	97.8
5–9	96.5
10–14	96.5
15–19	92.7
20–24	89.5
25–29	89.5
30–34	85.6
35–39	73.9
40–44	66.8
45–49	60.0
50–54	51.9
55–59	38.0
60–64	34.6
65–69	25.4
70–74	17.2
75–79	8.4
80–84	6.7
85–89	0.5
90–94	0.0
95–100	0.0

Figure 7 ($1.25/day 2005 PPP line): Bootstrapped differences between estimated and true poverty likelihoods for households in a large sample (n = 16,384) from the validation sample, with confidence intervals

	Difference between estimate and true value			
		Confidence interval (+/– percentage points)		
Score	Diff.	90-percent	95-percent	99-percent
0–4	–2.2	1.1	1.1	1.1
5–9	–3.5	1.7	1.7	1.7
10–14	–1.4	1.2	1.2	1.5
15–19	–3.4	2.3	2.3	2.5
20–24	–4.6	3.0	3.1	3.3
25–29	+4.2	1.8	2.1	2.6
30–34	–0.3	1.6	1.9	2.4
35–39	–3.0	2.6	2.8	3.4
40–44	–1.2	2.0	2.3	3.1
45–49	+1.9	2.2	2.5	3.5
50–54	–4.8	3.7	3.9	4.4
55–59	–1.7	2.7	3.2	4.7
60–64	–2.3	2.7	3.2	4.1
65–69	+4.0	2.8	3.3	4.1
70–74	+4.3	2.5	2.9	3.9
75–79	–0.4	3.4	4.0	5.1
80–84	–24.8	16.4	17.1	18.7
85–89	–0.6	1.0	1.2	1.5
90–94	+0.0	0.0	0.0	0.0
95–100	+0.0	0.0	0.0	0.0

Figure 9 ($1.25/day 2005 PPP line): Differences and precision of differences for bootstrapped estimates of households' poverty likelihoods, by sample size, scorecard applied to validation sample

Sample Size	Difference between estimate and true value			
		Confidence interval (+/– percentage points)		
n	Diff.	90-percent	95-percent	99-percent
1	–1.6	69.6	80.1	91.4
4	–1.9	39.8	48.1	64.2
8	–1.0	26.8	32.6	45.4
16	–1.2	19.4	23.4	31.0
32	–1.2	14.3	17.2	20.6
64	–1.1	9.9	11.6	14.9
128	–1.0	6.5	7.7	10.8
256	–1.0	4.9	5.9	7.8
512	–1.0	3.4	4.2	5.4
1,024	–0.9	2.4	2.9	3.8
2,048	–0.9	1.6	2.0	2.6
4,096	–0.9	1.2	1.4	1.8
8,192	–0.9	0.8	1.0	1.3
16,384	–0.9	0.6	0.7	1.0

Figure 11 ($1.25/day 2005 PPP line): Households by targeting classification and score, along with "Total Accuracy" and BPAC, scorecard applied to 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	0.5	60.9	0.0	38.5	39.1	–98.3
5–9	1.8	59.7	0.0	38.5	40.3	–94.3
10–14	4.0	57.4	0.1	38.5	42.5	–86.8
15–19	7.4	54.0	0.3	38.3	45.7	–75.4
20–24	12.2	49.3	0.6	37.9	50.1	–59.3
25–29	19.7	41.8	1.7	36.8	56.5	–33.3
30–34	27.4	34.1	3.0	35.5	62.9	–5.9
35–39	34.1	27.4	4.7	33.8	67.9	+18.5
40–44	41.5	19.9	8.2	30.4	71.9	+48.4
45–49	48.2	13.3	12.7	25.9	74.0	+77.3
50–54	53.0	8.5	16.5	22.0	75.0	+73.2
55–59	56.1	5.4	20.7	17.8	73.9	+66.3
60–64	58.8	2.7	26.0	12.5	71.3	+57.7
65–69	60.3	1.2	30.9	7.6	67.9	+49.8
70–74	61.0	0.5	34.6	3.9	65.0	+43.8
75–79	61.3	0.2	36.4	2.2	63.4	+40.9
80–84	61.5	0.0	37.5	1.1	62.5	+39.1
85–89	61.5	0.0	38.1	0.4	61.9	+38.0
90–94	61.5	0.0	38.5	0.1	61.5	+37.4
95–100	61.5	0.0	38.5	0.0	61.5	+37.3

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

Figure 12 ($1.25/day 2005 PPP line): Households below the poverty line and all households at a given score or at or below a given score cut-off, scorecard applied to 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	0.5	100.0	0.9	Only poor targeted
5–9	1.8	100.0	2.9	Only poor targeted
10–14	4.1	98.4	6.6	60.9:1
15–19	7.7	96.5	12.1	28.0:1
20–24	12.8	95.4	19.9	20.7:1
25–29	21.4	91.9	32.0	11.4:1
30–34	30.4	90.2	44.6	9.2:1
35–39	38.8	87.8	55.4	7.2:1
40–44	49.7	83.6	67.6	5.1:1
45–49	60.8	79.2	78.3	3.8:1
50–54	69.4	76.3	86.2	3.2:1
55–59	76.8	73.0	91.2	2.7:1
60–64	84.8	69.3	95.6	2.3:1
65–69	91.2	66.1	98.1	2.0:1
70–74	95.6	63.8	99.2	1.8:1
75–79	97.6	62.8	99.6	1.7:1
80–84	98.9	62.1	100.0	1.6:1
85–89	99.6	61.7	100.0	1.6:1
90–94	99.9	61.5	100.0	1.6:1
95–100	100.0	61.5	100.0	1.6:1

Tables for the $2.50/day 2005 PPP Poverty Line

Figure 4 ($2.50/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	100.0
5–9	100.0
10–14	100.0
15–19	99.5
20–24	99.5
25–29	99.2
30–34	99.0
35–39	97.9
40–44	95.1
45–49	93.5
50–54	88.9
55–59	87.6
60–64	81.0
65–69	75.1
70–74	63.4
75–79	56.3
80–84	51.9
85–89	32.1
90–94	20.1
95–100	2.3

Figure 7 ($2.50/day 2005 PPP line): Bootstrapped differences between estimated and true poverty likelihoods for households in a large sample (n = 16,384) from the validation sample, with confidence intervals

	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	+0.0	0.0	0.0	0.0
10–14	+2.1	1.0	1.1	1.5
15–19	+0.1	0.4	0.6	0.7
20–24	–0.5	0.3	0.3	0.3
25–29	+0.1	0.4	0.5	0.6
30–34	+2.0	0.8	1.0	1.4
35–39	–0.9	0.7	0.7	0.8
40–44	–2.6	1.6	1.6	1.7
45–49	–0.6	1.0	1.2	1.6
50–54	–3.1	2.2	2.3	2.5
55–59	+14.5	2.7	3.2	4.2
60–64	–5.8	3.8	3.9	4.2
65–69	–1.4	2.7	3.1	4.2
70–74	+3.7	4.0	4.6	5.7
75–79	+5.4	6.6	7.8	10.6
80–84	–6.5	7.2	8.3	11.6
85–89	+18.6	5.7	7.1	9.3
90–94	–31.5	22.0	23.4	27.0
95–100	–62.2	42.7	44.5	49.0

Figure 9 ($2.50/day 2005 PPP line): Differences and precision of differences for bootstrapped estimates of households' poverty likelihoods, by sample size, scorecard applied to validation sample

Sample Size	Difference between estimate and true value			
		Confidence interval (+/– percentage points)		
n	Diff.	90-percent	95-percent	99-percent
1	+0.1	52.9	62.1	72.3
4	+0.1	22.3	28.8	43.3
8	–0.1	15.4	19.3	27.1
16	–0.2	11.7	13.8	18.6
32	–0.3	8.0	9.6	12.8
64	+0.0	5.8	6.9	8.9
128	+0.1	4.0	4.8	6.3
256	+0.0	2.9	3.5	4.5
512	+0.1	2.1	2.5	3.4
1,024	+0.1	1.5	1.8	2.2
2,048	+0.1	1.0	1.3	1.7
4,096	+0.1	0.7	0.9	1.2
8,192	+0.1	0.5	0.6	0.8
16,384	+0.1	0.4	0.4	0.5

Figure 11 ($2.50/day 2005 PPP line): Households by targeting classification and score, along with "Total Accuracy" and BPAC, scorecard applied to 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	0.5	89.4	0.0	10.1	10.6	–98.8
5–9	1.8	88.1	0.0	10.1	11.9	–96.1
10–14	4.0	85.9	0.1	10.0	14.1	–90.9
15–19	7.6	82.3	0.1	10.0	17.6	–83.0
20–24	12.7	77.2	0.1	10.0	22.7	–71.6
25–29	21.2	68.7	0.2	9.9	31.1	–52.7
30–34	30.0	59.9	0.4	9.7	39.7	–32.8
35–39	38.3	51.6	0.5	9.6	47.9	–14.3
40–44	48.9	41.0	0.8	9.3	58.3	+9.7
45–49	59.4	30.5	1.4	8.7	68.2	+33.8
50–54	67.4	22.5	2.1	8.0	75.4	+52.2
55–59	73.4	16.5	3.4	6.7	80.0	+67.1
60–64	80.3	9.6	4.5	5.6	85.9	+83.7
65–69	85.0	4.9	6.2	3.9	88.8	+93.1
70–74	87.7	2.2	7.9	2.2	90.0	+91.3
75–79	88.9	1.0	8.7	1.4	90.3	+90.3
80–84	89.6	0.3	9.3	0.8	90.3	+89.6
85–89	89.7	0.2	9.9	0.2	89.9	+89.0
90–94	89.9	0.0	10.1	0.0	89.9	+88.8
95–100	89.9	0.0	10.1	0.0	89.9	+88.8

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

Figure 12 ($2.50/day 2005 PPP line): Households below the poverty line and all households at a given score or at or below a given score cut-off, scorecard applied to 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	0.5	100.0	0.6	Only poor targeted
5–9	1.8	100.0	2.0	Only poor targeted
10–14	4.1	98.4	4.5	60.9:1
15–19	7.7	98.7	8.5	76.4:1
20–24	12.8	99.2	14.1	127.6:1
25–29	21.4	99.2	23.6	117.1:1
30–34	30.4	98.7	33.4	75.4:1
35–39	38.8	98.7	42.6	73.6:1
40–44	49.7	98.5	54.4	64.1:1
45–49	60.8	97.7	66.1	43.0:1
50–54	69.4	97.0	75.0	32.7:1
55–59	76.8	95.5	81.6	21.4:1
60–64	84.8	94.7	89.3	17.7:1
65–69	91.2	93.2	94.5	13.6:1
70–74	95.6	91.8	97.6	11.2:1
75–79	97.6	91.1	98.9	10.2:1
80–84	98.9	90.6	99.6	9.6:1
85–89	99.6	90.1	99.8	9.1:1
90–94	99.9	89.9	100.0	8.9:1
95–100	100.0	89.9	100.0	8.9:1