Simple Poverty Scorecard® Poverty-Assessment Tool Timor-Leste

Mark Schreiner

12 June 2012

Dokumentu ida ne'e mak disponivel iha lian Tetun iha `SimplePovertyScorecard.com`. This document is in English at `SimplePovertyScorecard.com`.

Abstract

The Simple Poverty Scorecard-brand poverty-assessment tool uses 10 low-cost indicators from Timor-Leste's 2007 Survey of Living Standards to estimate the likelihood that a household has consumption below a given poverty line. Field workers can collect responses in about ten minutes. The scorecard's accuracy is reported for a range of poverty lines. The scorecard is a practical way for pro-poor programs in Timor-Leste to measure poverty rates, to track changes in poverty rates over time, and to segment clients for differentiated treatment.

Acknowledgements

Simple Poverty Scorecard® Poverty-Assessment Tool

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

Indicator	Response	Points	Score
1. How many members does the household have?	A. Nine or more	0	
	B. Eight	6	
	C. Seven	9	
	D. Six	13	
	E. Five	18	
	F. Four	25	
	G. Three	33	
	H. One or two	44	
2. Are all household members ages 8 to 17 currently attending school?	A. No, or no members 8 to 17	0	
	B. Yes	4	
3. What is the highest level and class that the male head/spouse has completed in school?	A. None, pre-school, or primary class 1	0	
	B. Primary class 2 to 5	1	
	C. Primary class 6 to pre-secondary class 2	2	
	D. No male head/spouse	3	
	E. Pre-secondary class 3, or higher	3	
4. What was the main occupation of the male head/spouse in the past 12 months?	A. No male head/spouse	0	
	B. Agriculture and animal husbandry (farming), forestry, fishing, or hunting	0	
	C. Does not work	7	
	D. Others	11	
5. What is the major construction material of the external walls?	A. Mud, wood, bamboo, rattan, tin, or other	0	
	B. Brick, concrete, or unbaked brick	7	
6. What is the primary material of the floor?	A. Earth/clay, wood, bamboo, or other	0	
	B. Concrete/brick, floor tile/cement, marble/ceramic	1	
7. What is the primary material of the roof?	A. Leaves, or other	0	
	B. Metal sheets/zinc, concrete, wood, tile, or sugar palm fibre	9	
8. Does the household own any televisions, tape players/CD players, or radios?	A. No	0	
	B. Yes	5	
9. How many clothes cupboards (*lemari pakaian*) does the household own?	A. None	0	
	B. One	2	
	C. Two or more	4	
10. How many square-meters of land does the household cultivate (or has or controls, even if the land does not belong to the household) that is for annual crops or fallow, tree crops, pasture, plantation, grassland, or garden/garden plot?	A. 1 to 1,499	0	
	B. None	2	
	C. 1,500 to 2,999	3	
	D. 3,000 to 9,999	5	
	E. 10,000 or more	12	

Back-page Worksheet: Household Roster

Complete this roster using the definition of *household member* below and in the
"Guidelines for the Interpretation of Scorecard Indicators".

Name	How old is [name]? (completed years)	Does [name] currently attend school? (ask only for ages 8 to 17)		
1.		No	Yes	Not 8 to 17
2.		No	Yes	Not 8 to 17
3.		No	Yes	Not 8 to 17
4.		No	Yes	Not 8 to 17
5.		No	Yes	Not 8 to 17
6.		No	Yes	Not 8 to 17
7.		No	Yes	Not 8 to 17
8.		No	Yes	Not 8 to 17
9.		No	Yes	Not 8 to 17
10.		No	Yes	Not 8 to 17
11.		No	Yes	Not 8 to 17
12.		No	Yes	Not 8 to 17

Use an additional sheet if there are more than 12 household members.

- Record the number of household members for "household size" at the top of the scorecard
- Mark Question 1 based on the number of household members
- Mark Question 3 based on whether all household members ages 8 to 17 attend school

Household members are the people who normally live in the same house and eat from the same pot (that is, they have a common arrangement for food, such as using a common kitchen or a common food budget), including young children, newborns, old people, sick people, people who are temporarily absent, domestic employees who work indoors, and relatives, pensioners, or guests. The persons may or may not be related to each other, and they may include non-relatives, including servants or employees who stay with the employer. Make a complete list of all the people who normally live and eat their meals together in this dwelling. Everyone who currently lives in the residence count as members. People who normally live in the residence but who are currently absent are also counted as members if the total length of their absence (counting both actual past absence and expected future absence) is six months or less. Newborns, newlyweds, and people who recently joined the household permanently are always counted as members. Please consult the "Guidelines for the Interpretation of Scorecard Indicators".

Look-up table for converting scores to poverty likelihoods, by poverty line

	Poverty likelihood (%)						
	National					Intl. 2005 PPP	
Score	Lower Natl.	Upper Natl.	150% Upper Natl.	200% Upper Natl.	USAID 'Extreme'	$1.25	$2.50
0–4	100.0	100.0	100.0	100.0	100.0	100.0	100.0
5–9	100.0	100.0	100.0	100.0	88.6	100.0	100.0
10–14	70.5	93.1	100.0	100.0	64.3	70.5	100.0
15–19	78.3	93.1	99.5	100.0	57.7	73.8	100.0
20–24	64.7	82.6	98.8	100.0	54.0	61.4	99.9
25–29	57.0	78.5	98.2	99.0	41.9	52.4	98.9
30–34	37.9	61.1	89.7	96.1	25.5	38.0	92.8
35–39	31.1	51.5	84.2	94.2	17.5	25.0	87.2
40–44	9.5	27.4	75.0	92.4	3.8	8.8	80.3
45–49	9.2	23.4	59.7	84.9	4.1	8.1	69.7
50–54	2.9	8.6	52.4	78.0	1.4	2.7	61.6
55–59	5.3	9.5	33.9	57.2	1.9	4.7	43.5
60–64	2.2	4.4	20.6	54.7	1.7	3.9	24.4
65–69	0.0	0.1	17.4	56.2	0.0	0.0	31.6
70–74	0.0	0.0	0.0	19.5	0.0	0.0	0.0
75–79	0.0	0.0	0.0	52.9	0.0	0.0	0.0
80–84	0.0	0.0	0.0	0.0	0.0	0.0	0.0
85–89	0.0	0.0	0.0	0.0	0.0	0.0	0.0
90–94	0.0	0.0	0.0	0.0	0.0	0.0	0.0
95–100	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Simple Poverty Scorecard® Poverty-Assessment Tool Timor-Leste

1. Introduction

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

The direct approach to poverty measurement via surveys asks households about a long list of consumption items and so is difficult and costly. As a case in point, Timor-Leste's 2007 Survey of Living Standards (TLSLS) runs 62 pages. It is conducted over four visits and covers about 300 consumption items. As an example, one such consumption item is: "Has your household consumed local rice in the past seven days? How much local rice did your household consume? What is the value of the local rice consumed? Did you obtain this local rice via purchase, self-production, and/or as a gift? Now then, has your household consumed imported rice the past seven days? . . ."

In contrast, the indirect approach via the scorecard is simple, quick, and inexpensive. It uses ten verifiable indicators (such as "What is the primary material of the floor?" and "How many clothes cupboards (*lemari pakaian*) does the household

own?") to get a score that is highly correlated with poverty status as measured by the exhaustive survey.

The scorecard here differs from "proxy-means tests" (Coady, Grosh, and Hoddinott, 2004) 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 local organizations are typically subjective and relative (such as participatory wealth ranking by skilled field agents) or blunt (such as rules based on land-ownership or housing quality). Measurements from these approaches may not be comparable across organizations, they may be costly, and their accuracy is unknown.

Local, pro-poor organizations in Timor-Leste 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 the scorecard to report how many of their participants are among the poorest half of people below the upper national poverty line. Local organizations can also use the tool to measure movement across a poverty line. For all these purposes, the scorecard provides an consumption-based, objective tool with known accuracy. While consumption surveys are costly even for governments, some small, local organizations may be able to implement an inexpensive poverty-assessment tool that can help with 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 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 non-specialists (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, transparent scorecards are often about as accurate as complex, opaque ones.

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

The scorecard is based on the 2007 TLSLS conducted by Timor-Leste's *Direcção Nacional de Estatística* (DNE). Indicators are selected to be:

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

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

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

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

Third, the scorecard can be used to estimate changes in the poverty rate for a group of households (or for two independent 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) 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 consumption data and Timor-Leste's upper national poverty line. Scores from this one scorecard are calibrated to poverty likelihoods for seven poverty lines.

The scorecard is constructed and calibrated using half the households in the 2007 TLSLS, 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

4

the scorecard was built), they are—like all predictive models—biased to some extent when applied to a different population or at a later point in time.[1]

Thus, while the indirect scoring approach is less costly than the direct survey approach, it is also biased. There is bias because scoring must assume that future relationships between indicators and poverty will be the same as in the data used to build the scorecard and that the scorecard will be applied to nationally representative samples.[2] Of course, these assumptions—inevitable in predictive modeling—hold only partly.

When applied to the validation sample for Timor-Leste 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 for the upper national poverty line is $+2.3$ percentage points. The average difference across all seven lines is $+1.3$ percentage points. These differences are due to sampling variation and not bias; the average of each difference would be zero if the entire 2007 TLSLS were to be repeatedly redrawn and divided into sub-samples before repeating the entire process of construction/calibration and validation.

The 90-percent confidence intervals for these estimates are ± 0.7 percentage points or less. For $n = 1,024$, these intervals are ± 2.6 percentage points or less.

---

[1] Important cases in practice include nationally representative samples at a different point in time or non-nationally representative sub-groups (Tarozzi and Deaton, 2009).
[2] Bias may also result from changes in data quality or data definitions.

Section 2 below describes data and poverty lines. Sections 3 and 4 describe scorecard construction and offer guidelines for use in practice. Sections 5 and 6 detail the estimation of households' poverty likelihoods and the estimation 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 compares the new scorecard here to a similar tool for Timor-Leste, 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 4,477 households in Timor-Leste's nationally representative TLSLS that was conducted from January 2007 through January 2008.

For the purposes of the scorecard, the households in the 2007 TLSLS are randomly divided into two sub-samples (Figure 1):

- *Construction and calibration* for selecting indicators and points and for associating scores with poverty likelihoods
- *Validation* for measuring 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 consumption (divided by the number of household members) is below a given poverty line.

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

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

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

Whether the household-level rate or the person-level rate is relevant depends on the situation. If an organization's "participants" include all the people in a household, then the person-level rate is relevant. Governments, for example, are concerned with the well-being of people, regardless of how 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 1 reports poverty lines and poverty rates for Timor-Leste at both the household-level and the person-level. Figure 2 reports the same information for the six

geographic regions for which poverty lines are defined. The national poverty rates are weighted averages of the regional poverty rates.

The scorecard is constructed using the 2007 TLSLS 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

The derivation of national poverty lines for Timor-Leste is documented in DNE (2008) and in Datt and Cumpa (2008). The lines are defined by urban/rural for East, Central, and West. The food line for a given region is defined as the average food bundle observed for the poor in the 2007 TLSLS, scaled to provide 2,100 Calories and valued at the median food prices paid by the poor.

Two "national" (food + non-food) poverty lines are defined (Ravallion and Bidani, 1994). The lower national line is the food line, plus the average non-food consumption of people whose *total* consumption is within 5 percent of the food line. In contrast, the upper national line is the food line, plus the average non-food consumption of people whose *food* consumption is within 5 percent of the food line.

The "poor" reference group used in this process to define the average food bundle and the prices of its elements is defined iteratively (Pradhan *et al.*, 2001). The first

iteration takes as "poor" as the people in the bottom two quintiles of consumption. A

interim food line is then derived. In the second and subsequent iterations, the "poor" are

taken as those with consumption below the poverty lines from the previous iteration.

This then implies new poverty lines and a different group of "poor" people. Iterations

continue until the absolute change in the poverty lines between iterations is less than

five percent (Datt and Cumpa, 2008).

For Timor-Leste as a whole, this process leads to an upper national line of

USD0.88 per person per day,[3] giving a household-level poverty rate of 41.6 percent and

a person-level poverty rate of 49.9 percent (Figure 1).

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:[4]

- Lower national
- Upper national
- 150% of upper national
- 200% of upper national
- USAID "extreme"
- $1.25/day 2005 PPP
- $2.50/day 2005 PPP

150% (or 200%) of the upper national line is defined as the upper national line

multiplied by 1.5 (or by 2).

---

[3] All monetary figures in this paper are in constant units as of December 2007.

[4] To avoid confusion and contradictions, this paper does not report food lines, nor does
it calibrate scores to them. DNE (2008) and Datt and Cumpa (2008) report food
poverty rates based on comparing food consumption to the food lines. International
practice, however, is to compare total consumption to food lines.

For a given region, the USAID "extreme" line is defined as the median consumption of people (not households) below the national line (U.S. Congress, 2004).

Timor-Leste's regional $1.25/day 2005 PPP poverty lines are derived from:

- 2005 PPP exchange rate for "individual consumption expenditure by households" (World Bank, 2008): USD0.49 per $1.00[5]
- Average monthly country-wide CPI in 2005 for Timor-Leste of 117.02
- Country-wide CPI for December 2007 of 135.2
- Six regional upper national poverty lines (L_i, $i = 1, 2, \ldots 6$) from Figure 2
- Person-weighted average of the six regional upper national lines L_i of USD0.88

Given this, the $1.25/day 2005 PPP line for region r in USD as of December 2007 is (Sillers, 2006):

$$\left(2005 \text{ PPP exchange rate}\right) \cdot \$1.25 \cdot \left(\frac{\text{CPI}_{\text{Dec. 2007}}}{\text{CPI}_{\text{Ave. 2005}}}\right) \cdot \frac{L_r}{\left(\sum_{i=1}^{6} \frac{L_i}{6}\right)}.$$

For the example of the Center Urban region, the upper national line is USD1.15 per person per day (Figure 2), so the $1.25/day 2005 PPP line is:

$$0.49 \cdot \$1.25 \cdot \left(\frac{135.2}{117.02}\right) \cdot \left(\frac{1.15}{0.88}\right) = 0.93.$$

The average $1.25/day line across all six regions in Timor-Leste is USD0.71 per person per day, giving a household-level poverty rate of 25.5 percent and a person-level rate of 31.8 percent (Figure 1).

---

[5] Timor-Leste uses U.S. dollars. The 2005 PPP exchange rate means that, on average in 2005, 0.49 dollars in Timor-Leste has the same purchasing power as 1 dollar in the United States of America. That is, a dollar goes twice as far in Timor-Leste as in the United States of America.

The poverty rates for the $1.25/day 2005 PPP line differ from those of the lower national line, even though both lines have the same average value (USD0.71). The poverty rates differ because the two lines differ slightly at the regional level, as the $1.25/day line is derived using the upper national line, which is based on a different food basket and different prices than the lower national line.

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

3. Scorecard construction

For Timor-Leste, about 100 potential indicators are initially prepared in the areas of:

- Household composition (such as number of members)
- Education (such as school attendance)
- Employment (such as the main occupation of the male head/spouse)
- Housing (such as type of floor, walls, and roof)
- Ownership of durable goods (such as clothes cupboards (*lemari pakaian*))
- Agriculture (such as area of land cultivated, owned, or controlled)

Figure 3 lists all potential indicators. They are ordered 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, the number of clothes cupboards (*lemari pakaian*) is probably more likely to change in response to changes in poverty than is the age of the male head/spouse.

The scorecard itself is built using the upper national poverty line and Logit regression on the construction/calibration 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 power 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). These include 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 criteria. This 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 Timor-Leste. Evidence from India and Mexico (Schreiner, 2006 and 2005a), Sri Lanka (Narayan and Yoshida, 2005), and Jamaica (Grosh and Baker, 1995) suggests that segmenting poverty-assessment tools by urban/rural does not improve targeting accuracy much, although it may improve the accuracy of estimates of poverty rates for sub-groups (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 the scorecard is actually adopted and used (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 predictive power, thanks to the empirical phenomenon known as 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 additional 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. Additional work is minimized; non-specialists can compute scores by hand in the field because the scorecard has:

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

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

- Record participant identifiers at the top of the scorecard
- Record data on individual household members using the household roster page
- Record household size (number of members) at the top of the scorecard
- Mark question 1 based on the number of members in the household
- Mark question 2 based on school attendance as recorded in the household roster
- Read each remaining question one-by-one from the scorecard (except questions 5, 6, and 7, which—if possible—should be recorded as observed without being asked)
- Circle the household's 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

4.1 Data quality

Of course, field agents must be trained. Quality outputs depend on quality inputs. If organizations or field agents 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 (2007a) and Toohig (2008) are useful nuts-and-bolts guides for budgeting, training field agents and supervisors, managing logistics, sampling, interviewing, piloting, recording data, and controlling quality.

In particular, while collecting scorecard indicators is relatively easier than alternatives, it is still absolutely difficult. Training and explicit definitions of terms and concepts in the scorecard is essential. For this reason, field agents should study the "Guidelines for the Interpretation of Scorecard Indicators" because it—as well as the "Back-page Worksheet" are is as integral parts of the Simple Poverty Scorecard tool.

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 car ownership. 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 Mexican social program, 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, self-reports can be checked

---

[6] If an organization does not want field agents to know the points associated with indicators, then they can make a version of the scorecard without points and apply the points later in at a central office. Schreiner (2011) argues, however, that experience in Colombia (Camacho and Conover, 2011) suggests that hiding points does little to deter cheating and that cheating by an organization's central office may be more likely and more damaging than cheating by field agents and respondents.

by field agents who verify responses with a home visit, and this is suggested for Timor-Leste.

4.2   Sample design

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 and from the business questions that the analysis seeks to inform. Determining these goals and questions is the key to the entire process.

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 at the respondent's homestead and then filed at an office
- On paper at the home and then keyed into a database or spreadsheet at an office
- On portable electronic devices at the home and downloaded to a database

Given a population of interest for a given question, the participants to be scored can be:

- All participants in the group of interest
- A representative sample of all participants
- All participants in a representative sample of service points
- A representative sample of all participants in a representative sample of service points

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 time interval (allowing measuring change)
- Each time a field agent 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 sample of participants
- With the same set of participants

An example set of design choices is illustrated by BRAC and ASA, two microlenders in Bangladesh. Each has more than 7 million participants, and each is using the (Schreiner, 2013). Their design is that loan officers in a random sample of branches score all their 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 more than 50,000 participants (far more than most pro-poor organizations would need).

5. Estimates of household poverty likelihoods

The sum of scorecard points for a household is called the *score.* For Timor-Leste, 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 line, the scores themselves have only relative units. For example, doubling the score increases the likelihood of being above a given poverty line, but it does not double the likelihood.

To get absolute units, scores are 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 upper national line, scores of 35–39 have a poverty likelihood of 51.5 percent, and scores of 40–44 have a poverty likelihood of 27.4 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 51.5 percent for the upper national line but 25.0 percent for the $1.25/day 2005 PPP 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] From Figure 4 on, many figures have seven versions, one for each of the seven poverty lines. To keep them straight, they are grouped by poverty line. Single tables pertaining to all poverty lines are placed with the first group of tables for the upper national line.

For the example of the upper national line (Figure 5), there are 13,399 (normalized) households in the calibration sub-sample with a score of 35–39, of whom 6,896 (normalized) are below the poverty line. The estimated poverty likelihood associated with a score of 35–39 is then 51.5 percent, because 6,896 ÷ 13,399 = 51.5 percent.

To illustrate with the upper national line and a score of 40–44, there are 10,108 (normalized) households in the construction/calibration sample, of whom 2,771 (normalized) are below the line (Figure 5). Thus, the poverty likelihood for this score is 2,771 ÷ 10,108 = 27.4 percent.

This method is used to calibrate scores with estimated poverty likelihoods for all seven poverty lines.

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

- 17.5 percent   below the USAID "extreme" line
- 7.5 percent   between the USAID "extreme" line and $1.25/day 2005 PPP
- 6.1 percent   between $1.25/day 2005 PPP and the lower national line
- 20.4 percent   between the lower national line and the upper national line
- 32.7 percent   between the upper national line and 150% of the upper natl. line
- 3.0 percent   between 150% of the upper national line and $2.50/day 2005 PPP
- 7.0 percent   between $2.50/day 2005 PPP and 200% of the upper national line
- 5.8 percent   above 200% of the upper national line

Even though the scorecard is constructed partly based on judgment, the calibration process produces poverty likelihoods that are objective, that is, derived from survey data on consumption and quantitative poverty lines. The poverty likelihoods

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

Although the points in the Timor-Leste 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, going from scores to poverty likelihoods in this way requires no arithmetic at all, just a look-up table. This calibration can also improve accuracy, especially with large samples.

5.2 Accuracy of estimates of households' poverty likelihoods

As long as the relationships between indicators and poverty do not change, and as long as the scorecard is applied to households who are representative of the same population from which the scorecard was constructed, then this calibration process produces unbiased estimates of households' poverty likelihoods. *Unbiased* means that in repeated samples from the same population, the average estimate matches the true value. The scorecard also produces unbiased estimates of groups' poverty rates at a point in time, as well as unbiased estimates of changes in groups' poverty rates between two points in time.[8]

Of course, the relationship between indicators and poverty does change to some unknown extent with time and also across sub-groups in Timor-Leste. Thus, the scorecard will generally be biased when applied after January 2008 (when fieldwork for the 2007 TLSLS ended) or when applied with non-nationally representative sub-groups.

---

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

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

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

For the upper national line, the average poverty likelihood across bootstrap samples for scores of 35–39 in the validation sample is too high by 13.4 percentage points. For scores of 40–44, the estimate is too high by 1.2 percentage points.[9]

The 90-percent confidence interval for the differences for scores of 35–39 is ±1.9 percentage points (Figure 7). This means that in 900 of 1,000 bootstraps, the average

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

difference between the estimate and the true value for households in this score range is between +11.5 and +15.3 percentage points (because +13.4 − 1.9 = +11.5, and +13.4 + 1.9 = +15.3). In 950 of 1,000 bootstraps (95 percent), the difference is +13.4 ± 2.2 percentage points, and in 990 of 1,000 bootstraps (99 percent), the difference is +13.4 ± 2.9 percentage points.

For some scores, 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 Timor-Leste'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. This is generally the case, as discussed in the next section.

Another possible source of differences between estimates and true values is overfitting. By construction, the scorecard here is unbiased, but it may still be *overfit* when applied in practice. That is, it may fit the construction data from the 2007 TLSLS 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 construction/calibration

data. Or the scorecard may be overfit in the sense that its bias is highly sensitive to

changes over time in the relationships between indicators and poverty or to applications

with non-nationally representative sub-groups.

Overfitting can be mitigated by simplifying the scorecard and by not relying only

on data but rather also considering experience, judgment, and theory. Of course, the

scorecard here does this. Combining scorecards can also help, at the cost of complexity.

Most errors in individual households' likelihoods, however, cancel out in the

estimates of groups' poverty rates (see later sections). Furthermore, at least some of the

differences arise from non-scorecard sources (such as sampling variation) that 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, 2012 and that they have scores of 20, 30, and 40, corresponding to poverty likelihoods of 82.6, 61.1, and 27.4 percent (upper national line, Figure 4). The group's estimated poverty rate is the households' average poverty likelihood of $(82.6 + 61.1 + 27.4) \div 3 = 57.0$ percent.[10]

6.1 Accuracy of estimated poverty rates at a point in time

For the Timor-Leste scorecard applied to the validation sample with $n = 16{,}384$, the difference between the estimated poverty rate at a point in time and the true rate for the upper national line is $+2.3$ percentage points (Figure 9, summarizing Figure 8 for all poverty lines). Across all seven lines, all absolute differences are 2.3 percentage points or less, and the average absolute difference is 1.3 percentage points. Part of these differences is due to sampling variation and the division of the 2007 TLSLS into two sub-samples.

---

[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 61.1 percent, which differs from the average of the three poverty likelihoods associated with each of the three scores (57.0 percent).

In terms of precision, the 90-percent confidence interval for a group's estimated poverty rate at a point in time with $n = 16{,}384$ is ± 0.7 percentage points or less (Figure 9). 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 the upper 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 $+2.3 - 0.7 = +1.6$ to $+2.3 + 0.7 = +3.0$ percentage points. This is because $+2.3$ is the average difference, and ± 0.7 is its 90-percent confidence interval. The average difference is $+2.3$ because the average scorecard estimate is too high by 2.3 percentage points; the average estimated poverty rate for the validation sample is 44.5 percent, but the true value is 42.2 percent (Figure 1).

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 poverty-assessment tools (Schreiner, 2008a), note that the textbook formula (Cochran, 1977) that relates confidence intervals with standard errors in the case of direct measurement of a proportion is $c = \pm z \cdot \sigma$, 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, $\sqrt{\dfrac{p \cdot (1 - p)}{n}}$,

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 40.9 percent (the poverty rate in the construction and calibration samples in Figure 1 for the upper national line), the

confidence interval c is $\pm z \cdot \sqrt{\dfrac{p \cdot (1 - p)}{n}} = \pm 1.64 \cdot \sqrt{\dfrac{0.409 \cdot (1 - 0.409)}{16{,}384}} = \pm 0.630$

percentage points.

The scorecard, however, does not measure poverty directly, so this formula is not immediately applicable. To derive a formula for the Timor-Leste scorecard, consider Figure 8, which reports empirical confidence intervals c for the differences for the scorecard applied to 1,000 bootstrap samples of various sizes from the validation

31

sample. For $n = 16{,}384$ and the upper national line, the 90-percent confidence interval is 0.660 percentage points.[11]

Thus, the 90-percent confidence interval with $n = 16{,}384$ is ± 0.660 percentage points for the Timor-Leste scorecard and ± 0.630 percentage points for direct measurement. The ratio of the two intervals is $0.660 \div 0.630 = 1.05$.

Now consider the same case, but with $n = 8{,}192$. The confidence interval under direct measurement is $\pm 1.64 \cdot \sqrt{\dfrac{0.409 \cdot (1 - 0.409)}{8{,}192}} = \pm 0.891$ percentage points. The empirical confidence interval with the Timor-Leste scorecard (Figure 8) is 0.895 percentage points. Thus for $n = 8{,}192$, the ratio of the two intervals is $0.895 \div 0.891 = 1.00$.

This ratio of 1.00 for $n = 8{,}192$ is close to the ratio of 1.05 for $n = 16{,}384$. Across all sample sizes of 256 or more in Figure 8, the average ratio turns out to be 1.03, implying that confidence intervals for indirect estimates of poverty rates via the Timor-Leste scorecard and the upper national poverty line are slightly wider than for direct estimates via the 2007 TLSLS. This 1.03 appears in Figure 9 as the "α factor" because if $\alpha = 1.03$, then the formula relating confidence intervals c and standard errors σ for the Timor-Leste scorecard is $c = \pm z \cdot \alpha \cdot \sigma$. That is, formula for the standard error σ for point-in-time estimates of poverty rates via scoring is $\alpha \cdot \sqrt{\dfrac{p \cdot (1 - p)}{n}}$.

---

[11] Due to rounding, Figure 8 displays 0.7, not 0.660.

In general, α can be more or less than 1.00. When α is less than 1.00, it means that the scorecard is more precise than direct measurement. This occurs for one of seven poverty lines in Figure 9.

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

To illustrate how to use this, suppose $c = 0.05455$ and $z = 1.64$ (90-percent confidence). Then the formula gives $n = \left(\dfrac{1.03 \cdot 1.64}{0.05455}\right)^2 \cdot 0.409 \cdot (1 - 0.409) = 232$, not too far from the sample size of 256 observed for these parameters in Figure 8 for the upper national line.

Of course, the α factors in Figure 9 are specific to Timor-Leste, its poverty lines, its poverty rates, and this scorecard. The derivation of the formulas, however, is valid for any scorecard following the approach in this paper.

---

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

In practice after the end of fieldwork for the TLSLS in January 2008, an organization would select a poverty line (say, the upper 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 41.6 percent national average in the 2007 TLSLS in Figure 1), look up α (here, 1.03), 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, $n = \left(\dfrac{1.03 \cdot 1.64}{0.02}\right)^2 \cdot 0.416 \cdot (1 - 0.416) = 1{,}734$.

---

[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 January 2008 will resemble that in the 2007 TLSLS with deterioration to the extent that the relationships between indicators and poverty status change over time and to the extent that the sub-group being scored is not nationally representative.

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 TLSLS, this paper cannot test estimates of change over time for Timor-Leste, and it can only suggest approximate formulas for standard errors. Nevertheless, the relevant concepts are presented here because, in practice, pro-poor organizations can apply the scorecard to collect their own data and 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 on poverty 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, 2012, a program samples three households who score 20, 30, and 40 and so have poverty likelihoods of 82.6, 61.1, and 27.4 percent (upper national line, Figure 4). The group's baseline estimated poverty rate is the households' average poverty likelihood of (82.6 + 61.1 + 27.4) ÷ 3 = 57.0 percent.

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

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

By way of illustration, suppose that a year later on Jan. 1, 2013, 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 78.5, 51.5, and 23.4 percent, upper national line, Figure 4). Their average poverty likelihood at follow-up is now (78.5 + 51.5 + 23.4) ÷ 3 = 51.1 percent, an improvement of 57.0 − 51.1 = 5.9 percentage points.[14]

This suggests that about one in 17 participants in this hypothetical example crossed the poverty line in 2012.[15] Among those who started below the line, about one in nine (5.9 ÷ 51.1 = 11.5 percent) on net ended up above the line.[16]

---

[14] Of course, such a large reduction in poverty in one year would be miraculous, 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 TLSLS, 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 Timor-Leste 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:

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

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

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

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

For countries for which this α has been measured, the simple average of α across poverty lines and years for a given country and then across countries is 1.19.[18] This is as reasonable a number as any to use for Timor-Leste.

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 upper national line, $\alpha = 1.19$, and $\hat{p} = 0.416$ (from Figure 1). Then the baseline sample size is

$n = 2 \cdot \left(\frac{1.19 \cdot 1.64}{0.02} \right)^2 \cdot 0.416 \cdot (1 - 0.416) = 4{,}627$, and the follow-up sample size is also

4,627.

[18] Schreiner, 2010, 2009a, 2009b, 2009c, 2009d, 2009e, and 2008b; Chen and Schreiner, 2009a and 2009b; and Schreiner and Woller, 2010a and 2010b.

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:[19]

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

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

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

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

---

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

Because $\hat{p}_*$ can range from 0–0.5, more information is needed to apply this formula. In Peru (Schreiner, 2009a), 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 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 Timor-Leste scorecard is applied twice (once after January 2008 and then again later) is

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

Peru is the only country with an estimate of α for this case (Schreiner 2009a). There, 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 upper national line, and the sample will first be scored in 2012 and then again in 2015 ($y = 3$). The before-baseline poverty rate is 41.6 percent ($p_{2007} = 0.416$, Figure 1), and suppose $\alpha = 1.30$. Then the baseline sample size is

$$n = 2 \cdot \left(\frac{1.30 \cdot 1.64}{0.02}\right)^2 \cdot \{-0.02 + 0.016 \cdot 3 + 0.47 \cdot [0.416 \cdot (1 - 0.416)]\} = 3{,}232.$$ The same

group of 3,232 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 consumption below a poverty line). Poverty status is a fact that depends on whether consumption is below a poverty line as directly measured by a survey. In contrast, targeting status is a program's policy choice that depends on a cut-off and on an indirect estimate from a scorecard.

Targeting is successful when households truly below a poverty line are targeted (*inclusion*) and when households truly above a poverty line are not targeted (*exclusion*). Of course, no scorecard is perfect, and targeting is unsuccessful when households truly below a poverty line are not targeted (*undercoverage*) or when households truly above a poverty line are targeted (*leakage*). Figure 10 depicts these four possible targeting outcomes. Targeting accuracy varies with 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. A formal way to do this is to assign net benefits—based on a program's values and mission—to each of

the four possible targeting outcomes and then to choose the cut-off that maximizes total

net benefits (Adams and Hand, 2000; Hoadley and Oliver, 1998).

Figure 11 shows the distribution of households in Timor-Leste by targeting

outcome. For an example cut-off of 35–39, outcomes for the upper national line in the

validation sample are:

- Inclusion: 33.3 percent are below the line and correctly targeted
- Undercoverage: 8.9 percent are below the line and mistakenly not targeted
- Leakage: 16.5 percent are above the line and mistakenly targeted
- Exclusion: 41.4 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: 36.5 percent are below the line and correctly targeted
- Undercoverage: 5.6 percent are below the line and mistakenly not targeted
- Leakage: 23.3 percent are above the line and mistakenly targeted
- Exclusion: 34.5 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 the 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

42

how it values successful inclusion or exclusion versus errors of undercoverage or 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 Timor-Leste scorecard. For the upper national line in the validation sample, total net benefit is greatest (75.7) for a cut-off of 34 or less, with about three in four households in Timor-Leste 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).[20]

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

---

[20] Figure 12 also reports "BPAC", a measure discussed in the next section.

achieve a desired poverty rate among targeted households. The third column of Figure 12 ("% targeted who are poor") shows, for the Timor-Leste scorecard applied to the validation sample, the expected poverty rate among households who score at or below a given cut-off. For the example of the upper national line, targeting households who score 39 or less would target 49.8 percent of all households (second column), and that targeted group would have a poverty rate is 66.9 percent (third column).

Figure 12 also reports two other measures of targeting accuracy. The first is a version of inclusion ("% of poor who are targeted"). For the example of the upper national line in the validation sample and a cut-off of 39 or less, 79.0 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 upper national line in the validation sample and a cut-off of 39 or less, covering 2.0 poor households means leaking to 1 non-poor household.

9. Context of poverty-assessment tools for Timor-Leste

This section discusses an existing poverty-assessment tool for Timor-Leste in terms of its goals, methods, poverty lines, data, indicators, cost, accuracy, and precision. Compared with this other tool, the main advantages of the new scorecard here are its use of the latest nationally representative data, its simplicity, and its reporting of formulas for standard errors.

USAID commissioned IRIS Center (2007c) to build a poverty-assessment tool (PAT) so that USAID's microenterprise partners in Timor-Leste could report on their participants' poverty rates. IRIS uses Timor-Leste's 2001 Living Standards Measurement Survey (IRIS, 2007d). The PAT supports one poverty line ($1.08/day 1993 PPP), giving a household-level poverty rate of 44.7 percent.

IRIS tests a variety of statistical approaches to tool construction (IRIS, 2005), settling on a Probit regression that estimates the likelihood that per-capita household consumption is less than $1.08/day 1993 PPP. This Probit is essentially the same as the Logit used here, except that IRIS uses an automated stepwise procedure to select indicators and that IRIS computes poverty likelihoods directly from the Probit formula rather than using non-parametric calibration and a look-up table. Also, IRIS does not use the poverty likelihoods as likelihoods when computing estimated poverty rates; rather, it compares the estimated likelihoods to an arbitrary cut-off of 50 percent, labeling those below the cut-off as "poor" and others as "non-poor". This happens to

work for Timor-Leste because the poverty rate for $1.08/day 1993 PPP in 2001 is close

to 50 percent. On the whole, even though the indicators and points are reported (IRIS,

2007d), IRIS' tool is more complex and less transparent than the new scorecard here.

IRIS' 16 indicators are similar in their simplicity and ease-of-collection to those

in this paper:

- Demographics:
    — Household size (and its square)
    — Age of the head (and its square)
- Education: Number of members 18-years-old or older who can read
- Characteristics of the residence:
    — Type of external walls
    — Type of roof
    — Number of rooms
    — Source of lighting
    — Whether the source of drinking water is private, public, or shared
    — Type of toilet arrangement
- Ownership of consumer durables:
    — Number of fans
    — Number of baskets
- Ownership of agricultural assets:
    — Farmland
    — Number of adult chickens
    — Number of axes
- Location:
    — Region
    — Urban/rural

The purpose of the IRIS PAT is to estimate poverty rates for USAID. For its

single poverty line, bias is +0.6 percentage points, Total Accuracy is 77.1 percent, and

the "Balanced Poverty Accuracy Criterion" is 73.8 (IRIS, 2010). BPAC is a measure

invented by IRIS (2005) that USAID has adopted as its criterion for certifying PATs. It

considers accuracy in terms of the estimated poverty rate (the purpose of the PAT) and in terms of inclusion (a targeting purpose that IRIS disavows). The formula is:

BPAC = (Inclusion – |Undercoverage – Leakage|) x [100 ÷ (Inclusion+Undercoverage)].

Because the data used by IRIS and in this paper differ both in age and in sample size, accuracy comparisons with the scorecard are not done here. Furthermore, IRIS uses *in-sample* tests; that is, it uses the same data to construct the tool and to test its accuracy. In contrast, this paper uses *out-of-sample* tests that divide data in two parts, one for scorecard construction/calibration and another for validating accuracy. In practice, poverty-assessment tools are used out-of-sample, so out-of-sample tests are more relevant. Furthermore, in-sample tests tend to overstate accuracy relative to out-of-sample standards. Thus, out-of-sample tests are preferred.

IRIS states that the PAT should not be used for targeting,[21] and IRIS doubts that the PAT can be useful for measuring changes in poverty rates, noting that "it is unclear that the tools will be able to identify real changes in poverty over time due to their inherent measurement errors. Unless the changes in the poverty rate are exceptionally large and the tools exceptionally accurate, the changes identified are likely to be contained within the margin of error."[22] In contrast, this paper supports these uses, reporting various aspects of targeting accuracy and margins of error (that is, standard errors and confidence intervals) for measures of change over time so that users can decide for themselves whether accuracy is adequate for their purposes. Finally, IRIS does not report standard errors for its estimator of poverty rates.

---

[21] `povertytools.org/faq/faq.html#11`, retrieved 19 February 2009.
[22] `povertytools.org/faq/faq.html#12`, retrieved 19 February 2009.

10. Conclusion

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

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

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

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

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

When the scorecard is applied to the validation sample with $n = 16,384$, the absolute difference between estimates versus true poverty rates for groups of households at a point in time is always 2.3 percentage points or less and averages—across the seven poverty lines—about 1.3 percentage points. For $n = 16,384$ and 90-percent confidence, the precision of these differences is ± 0.7 percentage points or better.

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 ten indicators that are inexpensive to collect and that are straightforward to verify. Points are all zeroes or positive integers, and scores range from 0 to 100. 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 to 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 Timor-Leste to estimate poverty rates, track changes in poverty rates over time, and target services. The same approach can be applied in any country with similar data.

References

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

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

Caire, Dean. (2004) "Building Credit Scorecards for Small-Business Lending in Developing Markets", `microfinance.com/English/Papers/ Scoring_SMEs_Hybrid.pdf`, retrieved 26 July 2011.

Camacho, Adriana; and Emily Conover. (2011) "Manipulation of Social-Program Eligibility", *American Economic Journal: Economic Policy*, Vol. 3, No. 2, pp. 41–65.

Chen, Shiyuan; and Mark Schreiner. (2009) "Simple Poverty Scorecard Poverty-Assessment Tool: Vietnam", `SimplePovertyScorecard.com/VNM_2006_ENG.pdf`, retrieved 11 July 2016.

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.*

Datt, Gaurav; and Martín Cumpa. (2008) "Timor-Leste: Poverty in a Young Nation", `www.ausaid.gov.au/publications/ pdf/et-timorleste-poverty-young-nation.pdf`, retrieved 26 July 2011.

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

Direcção Nacional de Estatística. (2008) *Final Statistical Abstract: Timor-Leste Survey of Living Standards 2007.*

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

Falkenstein, Eric. (2008) "DefProb™: A Corporate Probability-of-Default Model", `papers.ssrn.com/sol3/papers.cfm?abstract_id=1103404`, retrieved 26 July 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 26 July 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 26 July 2011.

Grosh, Margaret; and Judy L. Baker. (1995) "Proxy Means Tests for Targeting Social Programs: Simulations and Speculation", World Bank LSMS Working Paper No. 118, `go.worldbank.org/W9OWN57PD0`, retrieved 26 July 2011.

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

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

IRIS Center. (2010) "Accuracy Results for 26 Poverty Assessment Tool Countries", `povertytools.org/other_documents/Accuracy%20Notes/PAT_26_country_accuracy_results_Jan2010.pdf`, retrieved 26 July 2011.

_____. (2007a) "Manual for the Implementation of USAID Poverty Assessment Tools", `povertytools.org/training_documents/Manuals/USAID_PAT_Manual_Eng.pdf`, retrieved 26 July 2011.

_____. (2007b) "Introduction to Sampling for the Implementation of PATs", `povertytools.org/training_documents/Sampling/Introduction_Sampling.ppt`, retrieved 26 July 2011.

_____. (2007c) "Client Assessment Survey—Timor-Leste", `povertytools.org/countries/East_Timor/USAID_PAT_Timor_9-2007.xls`, retrieved 30 July 2011.

_____. (2007d) "Poverty Assessment Tool Accuracy Submission: USAID/IRIS Tool for Timor-Leste", `povertytools.org/countries/East_Timor/USAID_PAT_Timor_Leste.pdf`, retrieved 30 July 2011.

_____. (2005) "Notes on Assessment and Improvement of Tool Accuracy", `povertytools.org/other_documents/AssessingImproving_Accuracy.pdf`, retrieved 26 July 2011.

Johnson, Glenn. (2007) "Lesson 3: Two-Way Tables—Dependent Samples", `onlinecourses.science.psu.edu/stat504/node/96`, retrieved 26 July 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.

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", *Journal of the European Economic Association*, Vol. 4, No. 6, pp. 886–908.

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 26 July 2011.

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

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

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

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

Pradhan, Menno; Suryahadi, Asep; Sumarto, Sudarno; and Lant Pritchett. (2001) "Eating Like which 'Joneses'? An Iterative Solution to the Choice of a Poverty Line 'Reference Group'", *Review of Income and Wealth*, Series 47, No. 4, pp. 473–487.

Ravallion, Martin; and Benu Bidani. (1994) "How Robust Is a Poverty Profile?", *World Bank Economic Review*, Vol. 8, No. 1, pp. 75–102.

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 26 July 2011.

Schreiner, Mark. (2011) "Simple Poverty Scorecard Poverty-Assessment Tool: Colombia", `SimplePovertyScorecard.com/COL_2009_ENG.pdf`, retrieved 11 July 2016.

_____. (2010) "Simple Poverty Scorecard Poverty-Assessment Tool: Honduras", `SimplePovertyScorecard.com/HND_2007_ENG.pdf`, retrieved 11 July 2016.

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

_____. (2009b) "Simple Poverty Scorecard Poverty-Assessment Tool: Philippines", `SimplePovertyScorecard.com/PHL_2002_ENG.pdf`, retrieved 11 July 2016.

_____. (2009c) "Simple Poverty Scorecard Poverty-Assessment Tool: Pakistan", `SimplePovertyScorecard.com/PAK_2005_ENG.pdf`, retrieved 11 July 2016.

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

_____. (2009e) "Simple Poverty Scorecard Poverty-Assessment Tool: Mexico", `SimplePovertyScorecard.com/MEX_2008_ENG.pdf`, retrieved 11 July 2016.

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

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

_____. (2006) "Is One Simple Poverty Scorecard Poverty-Assessment Tool Enough for India?", `microfinance.com/English/Papers/Scoring_Poverty_India_Segments.pdf`, retrieved 11 July 2016.

_____. (2005a) "La Herramienta del Índice Calificación de la Pobreza™: México", `SimplePovertyScorecard.com/MEX_2002_SPA.pdf`, retrieved 11 July 2016.

_____. (2005b) "IRIS Questions on the Simple Poverty Scorecard Poverty-Assessment Tool", `microfinance.com/English/Papers/Scoring_Poverty_Response_to_IRIS.pdf`, retrieved 26 July 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 26 July 2011.

_____; and Gary Woller. (2010a) "Simple Poverty Scorecard Poverty-Assessment Tool: Ghana", `SimplePovertyScorecard.com/GHA_2005_ENG.pdf`, retrieved 11 July 2016.

_____; and Gary Woller. (2010b) "Simple Poverty Scorecard Poverty-Assessment Tool: Guatemala", `SimplePovertyScorecard.com/GTM_2006_ENG.pdf`, retrieved 11 July 2016.

Sillers, Don. (2006) "National and International Poverty Lines: An Overview", `pdf.usaid.gov/pdf_docs/Pnadh069.pdf`, retrieved 31 May 2012.

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

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

Toohig, Jeff. (2007) "Progress out of Poverty Index: Training Guide", `progressoutofpoverty.org/toolkit`, retrieved 26 July 2011.

United States Congress. (2004) "Microenterprise Results and Accountability Act of 2004 (HR 3818 RDS)", November 20, `smith4nj.com/laws/108-484.pdf`, retrieved 26 May 2012.

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

World Bank. (2008) "Estimation of PPPs for Non-Benchmark Economies for the 2005 ICP Round", `siteresources.worldbank.org/ICPINT/Resources/non-benhmark.pdf` (sic), retrieved 26 July 2011.

Zeller, Manfred. (2004) "Review of Poverty Assessment Tools", `pdf.usaid.gov/pdf_docs/PNADH120.pdf`, retrieved 26 July 2011.

Guidelines for the Interpretation
of Scorecard Indicators

The following is from the 2007 TLSLS questionnaire and from:

Direcção Nacional de Estatística. (2006) *TLSLS–2006: Interviewer's Manual*, Dili: Ministério do Plano e Finanças. (the *Manual*).

According to both the questionnaire and the *Manual*, the respondent should be the "household head or the most informed household member".

According to p. 13 (twice) in the *Manual*: "You should always use this field manual as a guide."

According to pp. 11–14 of the *Manual*, the following rules for how to conduct an interview should be followed:

- Be presentable
- Be polite
- Introduce yourself and the survey
- Explain the objectives of the survey
- Explain that all responses will be kept confidential
- Do not be in a hurry to conduct the interview
- Be positive
- Try to avoid refusals to cooperate by:
    — Acknowledging that some questions may be embarrassing to some people
    — Creating a relaxed atmosphere of trust so that the respondent can freely speak without feeling embarrassed or shy
- Conduct the interview in a place with few distractions or interruptions
- Always maintain a neutral attitude
- Do not suggest answers
- Always use this field manual as a guide (the *Manual* repeats this injunction twice)
- Probe if an answer is unsatisfactory:
    — Repeat the question
    — Explain the concept if necessary
    — Ask for an estimate, if appropriate
    — Ask a follow-up question (such as "What do you mean by that?")
- Thank the respondent for his/her cooperation

According to pp. 14–15 of the *Manual*, the following rules should be observed when asking questions:

- Ask all the questions exactly as they are worded in the questionnaire
- Ask the questions in the order that they are presented in the questionnaire
- Do not read the coded answers to the respondent
- Never ask a leading question, that is, one that seems to suggest a desired answer

Page 24 of the *Manual* makes the following self-referential note: "An attempt has been made to provide explanations and instructions for the completion of the questionnaire. Explanations are provided for questions that are complicated or could be interpreted in more than one way. Where a question is presumed to be clear, based on pre-testing experience, no explanation is provided. In cases where interviewers need additional clarification of any of the questions, they should ask their supervisors."

1. <u>How many members does the household have?</u>

According to p. 8 of the questionnaire, a *household member* is defined according to the following criteria.

- Anyone who has not been away from the household for more than one month in the last 12 months is a household member
- Anyone who has been away from the household for nine months or more in the last 12 months is not a household member, with the following exceptions:
  — Infants less than three months old
  — Newlywed couples and close relatives who have recently joined the household permanently
  — A bride who has just joined her husband's household
- Servants who live and take their meals with the household are to be counted as household members, even though they may have no blood relationship with the household head

According to p. 21 of the *Manual*, "A *household* is a group of persons (or a single person) who usually live together and have a common arrangement for food, such as using a common kitchen or a common food budget. The persons may be related to each other, or they may be non-relatives, including servants or other employees staying with the employer.

"Students, boarders and employees residing in and having a common food arrangement with the household are considered members of the household if they have been in the household for more than a year or if they have no other place of residence.

"However, if there are five or more boarders/lodgers in a housing unit, they should not be reported as members of the household. They are considered to be living in a dormitory or boarding house operated by the household."

According to p. 25 of the *Manual, household members* are "the people who normally live in the same house and eat from the same pot, including young children, newborns, old people, sick people, people who are temporarily absent, domestic employees (who work indoors), and relatives, pensioners and guests."

According to pp. 26–27 of the *Manual*, the way to list household members is to ask the following questions:

- "I would like to make a complete list of all the people who normally live and eat their meals together in this dwelling.
- "First, I would like to have the names of all the members of your immediate family, who normally live and eat their meals together in this dwelling. Start with the head of the household, wife/husband of household head, and his/her children in order of age.
- "Please give me the names of any other persons related to the head of the household or to his/her wife/husband, together with their families, who normally live and eat their meals here.
- "Please give me the names of any other persons not related to the head of household or to his/her wife/husband but who normally live and eat their meals here. For instance, tenants, lodgers, servants or other persons who are not relatives.
- "Are there any other persons not now present but who normally live and eat their meals here? For example, any person studying somewhere else or who is on vacation or who is visiting other people?"

The discussion above reports what is documented in the questionnaire and the *Manual*. These sources do not address, however, potential household members who are away from the residence for more than one month but less than nine months. When asked about this, DNE said that, in practice, they use the following procedure to determine household membership:

- Ask whether the potential member currently lives in the residence. If yes, then the person is counted as a member
- If no, then ask the following questions to determine the total length of the absence (combining both actual past absence and expected future absence):
  — Why is the potential household member not living there?
  — How long has the potential member been gone?
  — When is the potential member expected to return?

If the total length of the current absence exceeds six months, then the person is not counted as a household member. If it is six months or less, then the person is counted as a household member.

As suggested by the formal documentation above, the "six-month rule" would not apply to newborns, newlyweds, and people who recently joined the household permanently.

2. <u>Are all household members ages 8 to 17 currently attending school?</u>

According to p. 28 of the *Manual*, ages are to be considered in terms of completed years. When asking for ages, probe for the birth certificate or ID. If this is available, use it to determine age. If not, then use the age provided by the respondent.

3. <u>What is the highest level and class that the male head/spouse has completed in school?</u>

According to p. 22 of the *Manual*, "The *head of household* is the adult member of the household who is accepted and recognized by the other household members as head."

The purposes of the scorecard:, the *male head/spouse* is defined as:

- The household head, if the head is a man
- The spouse of the household head, if the head is a woman
- Non existent, if neither of the previous two criteria are met

According to p. 47 of the *Manual*, this question pertains to "the highest grade successfully completed by the person."

4. <u>What was the main occupation of the male head/spouse in the past 12 months?</u>

According to p. 22 of the *Manual*, "*Occupation* refers to the type of work, trade or profession performed by the individual during the reference period. If the person is not at work but with a job, then *occupation* refers to the kind of work that the person will be doing when he reports for work. . . .
　　"*Work* is defined as an economic activity that a person performs for pay, profit or family gain. It includes paid employment; operating a farm or business; working in a household economic activity (like food processing or raising of livestock) without pay; working as an apprentice in order to learn a skill or craft, without necessarily receiving wages; and production of paddy or vegetables, say, solely for home consumption. Also, included is the holding of a job, even if the person is temporarily absent because of vacation, strike, or illness. Production of fixed assets for own household use (such as building or repairing the house) is also considered as work."

According to p. 60 of the *Manual*, all jobs conducted in the past 12 months count here, including jobs that are not currently conducted.

5. <u>What is the major construction material of the external walls?</u>

According to p. 31 of the *Manual*, this question should be answered by observation (not by asking the respondent).

According to p. 10 of the questionnaire, the *residence* is considered to include "all the rooms and all separate buildings used by household members to live in."

According to p. 19 of the *Manual*, "Choose from the list of responses the one that best fits the respondent's answer. . . . The response codes include 'other'. The reason for this is that although the questionnaires include coded responses for what are thought to be the most common answers, there are bound to be cases in which an answer may not clearly fit any of the coded answers. In such cases, write in the numeric code for 'other'."

6. <u>What is the primary material of the floor?</u>

According to p. 31 of the *Manual*, this question should be answered by observation (not by asking the respondent).

According to p. 10 of the questionnaire, the *residence* is considered to include "all the rooms and all separate buildings used by household members to live in."

According to p. 19 of the *Manual*, "Choose from the list of responses the one that best fits the respondent's answer. . . . The response codes include 'other'. The reason for this is that although the questionnaires include coded responses for what are thought to be the most common answers, there are bound to be cases in which an answer may not clearly fit any of the coded answers. In such cases, write in the numeric code for 'other'."

7. <u>What is the primary material of the roof?</u>

According to p. 31 of the *Manual*, this question should be answered by observation (not by asking the respondent).

According to p. 10 of the questionnaire, the *residence* is considered to include "all the rooms and all separate buildings used by household members to live in."

According to p. 19 of the *Manual*, "Choose from the list of responses the one that best fits the respondent's answer. . . . The response codes include 'other'. The reason for this is that although the questionnaires include coded responses for what are thought to be the most common answers, there are bound to be cases in which an answer may not clearly fit any of the coded answers. In such cases, write in the numeric code for 'other'."

8. <u>Does the household own any televisions, tape players/CD players, or radios?</u>

The questionnaire and the *Manual* do not provide any additional information about this indicator.

9. <u>How many clothes cupboards (*lemari pakaian*) does the household own?</u>

The questionnaire and the *Manual* do not provide any additional information about this indicator.

10. <u>How many square-meters of land does the household cultivate (or has or controls, even if the land does not belong to the household) that is for annual crops or fallow, tree crops, pasture, plantation, grassland, or garden/garden plot?</u>

The question should be asked in terms of square meters. Respondents, however, may give answers in terms of hectares. One hectare is 10,000 (ten thousand) square meters. Thus, the response options can be converted as follows:

<u>In square meters</u>	<u>In hectares</u>
A. 1 to 1,499 square meters	0.0001 to 0.1499 hectares
B. 1,500 to 2,999 square meters	0.1500 to 0.2999 hectares
C. 3,000 to 9,999 square meters	0.3000 to 0.9999 hectares
D. None	0 hectares
E. 10,000 or more square meters	1.0000 or more hectares

If the respondent has difficulty answering, ask him or her for an estimate (p. 14 of the *Manual*). Page 20 of the *Manual* states that for some questions, "Some respondents will not be able to answer. There are many possible reasons: either they do not remember very well, or they do not possess the information, or they do not fully understand the question. In some cases, when the question asks the respondent to 'estimate' some quantity, you should encourage the respondent to provide his/her best guess."

According to p. 70 of the *Manual*, "Sometimes it is difficult to get this information [area of agricultural land], but try at least to get an estimate from the respondent. . . . The area should be recorded in square meters."

Figure 1: Sample sizes and poverty rates by construction-calibration/validation sub-sample and by poverty line at both the household level and the person level

		% with per-capita daily household expenditure below a poverty line						
		National					Intl. 2005 PPP	
				150%	200%			
	#	Lower	Upper	Upper	Upper	USAID		
Sub-sample	households	Natl.	Natl.	Natl.	Natl.	'Extreme'	$1.25	$2.50
Poverty line (USD/person/day)	4,477	0.71	0.88	1.32	1.75	0.66	0.71	1.42
<u>**All Timor-Leste poverty rates (%)**</u>								
Household level	4,477	27.0	41.6	70.7	85.0	19.2	25.5	75.0
Person level	4,477	33.2	49.9	79.0	90.4	24.9	31.8	82.6
<u>**Construction and calibration: Selecting indicators and points, and associating scores with likelihoods**</u>								
Household level	2,233	26.6	40.9	69.9	85.1	18.6	24.9	75.1
Person level	2,233	32.9	49.4	78.7	90.7	24.1	31.1	83.1
<u>**Validation: Measuring accuracy**</u>								
Household level	2,244	27.3	42.2	71.4	84.8	19.7	26.1	74.8
Person level	2,244	33.5	50.3	79.3	90.2	25.6	32.5	82.1

Source: 2007 Survey of Living Standards, and Datt and Cumpa (2008)

All monetary units in dollars of the United States of America (USD) in constant terms as of Dec. 2007.

Figure 2: Poverty lines and poverty rates at the household level and person level by region

	Poverty line (USD/person/day) and poverty rate (%)						
	National					Intl. 2005 PPP	
			150%	200%			
	Lower	Upper	Upper	Upper	USAID		
Region	Natl.	Natl.	Natl.	Natl.	'Extreme'	$1.25	$2.50
All Timor-Leste							
Poverty line	0.71	0.88	1.32	1.75	0.66	0.71	1.42
Household-level poverty rate	27.0	41.6	70.7	85.0	19.2	25.5	75.0
Person-level poverty rate	33.2	49.9	79.0	90.4	24.9	31.8	82.6
East rural (Baucau, Lautem, and Viqueque)							
Poverty line	0.50	0.61	0.92	1.23	0.51	0.50	0.99
Household-level poverty rate	8.3	18.5	47.9	70.7	8.6	7.9	55.0
Person-level poverty rate	12.6	26.4	61.1	81.2	13.1	11.9	67.4
East urban (Baucau, Lautem, and Viqueque)							
Poverty line	0.57	0.71	1.07	1.43	0.58	0.58	1.15
Household-level poverty rate	8.6	20.2	48.1	86.7	9.1	9.0	56.4
Person-level poverty rate	13.1	27.7	58.4	79.1	13.7	13.6	66.7
Center rural (Aileu, Ainaro, Dili, Ermera, Liquiçá, Manufahi, and Manututo)							
Poverty line	0.78	0.90	1.34	1.80	0.64	0.73	1.45
Household-level poverty rate	44.1	55.0	83.8	92.9	25.0	37.7	86.8
Person-level poverty rate	52.7	64.2	90.2	96.5	32.0	45.7	92.5
Center urban (Aileu, Ainaro, Dili, Ermera, Liquiçá, Manufahi, and Manututo)							
Poverty line	0.84	1.15	1.73	2.30	0.85	0.93	1.86
Household-level poverty rate	18.4	41.2	71.7	86.0	19.5	24.4	75.0
Person-level poverty rate	22.6	47.7	78.3	89.9	23.8	29.2	81.4
West rural (Bobonaro, Cova Lima, and Oecussi)							
Poverty line	0.68	0.84	1.25	1.67	0.63	0.67	1.35
Household-level poverty rate	30.0	48.5	76.8	89.2	22.5	28.6	80.1
Person-level poverty rate	37.5	57.4	84.0	92.7	28.6	35.9	86.3
West urban (Bobonaro, Cova Lima, and Oecussi)							
Poverty line	0.74	0.89	1.34	1.78	0.70	0.72	1.44
Household-level poverty rate	18.2	33.0	61.7	79.5	15.9	17.6	67.3
Person-level poverty rate	22.0	38.8	68.3	84.4	19.3	21.6	93.0

Source: 2007 Survey of Living Standards, and Datt and Cumpa (2008)

All monetary units in dollars of the United States of America (USD) in constant terms as of Dec. 2007.

Figure 3: Poverty indicators by uncertainty coefficient

Uncertainty coefficient	Indicator (Answers ordered starting with those most strongly linked with higher poverty likelihoods)
1,124	How many members does the household have? (Nine or more; Eight; Seven; Six; Five; Four; Three; One or two)
1,104	How many household members are 18-years-old or younger? (Seven or more; Six; Five; Four; Three; Two; One; None)
1,067	How many household members are 17-years-old or younger? (Six or more; Four or five; Three; Two; One; None)
1,034	How many household members are 16-years-old or younger? (Six or more; Four or five; Three; Two; One; None)
1,015	How many household members are 15-years-old or younger? (Six or more; Four or five; Three; Two; One; None)
1,003	How many household members are 14-years-old or younger? (Five or more; Four; Three; Two; One; None)
925	How many household members are 13-years-old or younger? (Five or more; Four; Three; Two; One; None)
839	What is the mother tongue of the female head/spouse? (Mambae; Tetum Terik, or other; Kemak; Tetum; Bunak; Baequeno; Tokodete; No female head/spouse; Macasae, Galolen, Isni, Kaklun Bikeli, Laklei, Macalero, Mangilih, Midiki, Naueti, Sa Ani, Uaimua, Waweloi, Bahasa Indonesian, Portuguese, or Inggris; Fatalucu)
832	How many household members are 12-years-old or younger? (Five or more; Four; Three; Two; One; None)
791	How many household members are 11-years-old or younger? (Five or more; Four; Three; Two; One; None)
747	In what strata does the household live? (Urban or rural Ainaro, Manufahi, or Manatuto; Urban or rural Oecussi; Rural Aileu, Dili, or Ermera; Rural Bobonaro, Cova Lima, or Liquiçá; Urban Bobonaro, Cova Lima, or Liquiçá; Urban Aileu, Dili, or Ermera; Rural Baucau, Lautem, or Viqueque; Urban Baucau, Lautem, or Viqueque)
698	How many household members are 8-years-old or younger? (Four or more; Three; Two; One; None)
679	In what region does the household live? (Ainaro, Manufahi, or Manatuto; Aileu, Dili, or Ermera; Bobonaro, Cova Lima, Liquiçá, or Oecussi; Baucau, Lautem, or Viqueque)

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

Uncertainty coefficient	Indicator (Answers ordered starting with those most strongly linked with higher poverty likelihoods)
605	Are all household members ages 8 to 14 currently attending school? (No; No members 8 to 14; Yes)
592	Are all household members ages 8 to 15 currently attending school? (No; No members 8 to 15; Yes)
581	What is the mother tongue of the male head/spouse? (Kemak; Other; Mambae; Tetum Terik; Baequeno; Bunak; Tokodete; No male head/spouse; Tetum; Macasae; Galolen, Isni, Kaklun Bikeli, Laklei, Macalero, Mangilih, Midiki, Naueti, Sa Ani, Uaimua, Waweloi, Bahasa Indonesian, Portuguese, or Inggris; Fatalucu)
574	Does the household live in an urban or rural area in the East, Central, or West? (Center rural (Aileu, Ainaro, Dili, Ermera, Liquica, Manufahi, and Manututo); West rural (Bobonaro, Cova Lima, and Oecussi); Center urban (Aileu, Ainaro, Dili, Ermera, Liquica, Manufahi, and Manututo); West urban (Bobonaro, Cova Lima, and Oecussi); East rural (Baucau, Lautem, and Viqueque); East urban (Baucau, Lautem, and Viqueque))
573	Are all household members ages 8 to 16 currently attending school? (No; No members 8 to 16; Yes)
571	Are all household members ages 8 to 13 currently attending school? (No; No members 8 to 13; Yes)
567	Are all household members ages 8 to 18 currently attending school? (No; No members 8 to 18; Yes)
562	In their main occupation in the past 12 months, how many household members were workers in agriculture and animal husbandry (farming), forestry, fishing, or hunting? (Four or more; Three; Two; One; None)
558	Are all household members ages 8 to 17 currently attending school? (No, or no members 8 to 17; Yes)
519	What type of toilet is used by the household? (Other; No toilet, or bowl/bucket; Pit latrine with slab; Ventilated improved pit latrine; Pit latrine without slab/open pit; Flush toilet)
486	Are all household members ages 8 to 12 currently attending school? (No; No members 8 to 12; Yes)
482	Does the household live in the East, Center, or West? (Center (Aileu, Ainaro, Dili, Ermera, Liquica, Manufahi, and Manututo); West (Bobonaro, Cova Lima, and Oecussi); East (Baucau, Lautem, and Viqueque))

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

Uncertainty coefficient	Indicator (Answers ordered starting with those most strongly linked with higher poverty likelihoods)
472	Are all household members ages 8 to 11 currently attending school? (No; No members 8 to 11; Yes)
471	What was the main occupation of the male head/spouse in the past 12 months? (No male head/spouse; Agriculture and animal husbandry (farmer), forestry, fishing, or hunting; Does not work; Other)
341	What is the main occupation of the male head/spouse? (Sharecropper, or farm laborer; Farmer; No male head/spouse; Does not work; Non-farm laborer, fisher, trader, skilled worker, teacher, pensioner, housekeeper, school student, or university student; Civil servant)
330	What is the primary material of the floor? (Earth/clay, wood, bamboo, or other; Concrete/brick, floor tile/cement, marble/ceramic)
313	Where do members of the household bathe? (Other; River, pond, etc.; Outdoor bath/shower; Indoor bath/shower)
307	What is the major construction material of the external walls? (Mud, wood, bamboo, rattan, tin, or other; Brick, concrete, or unbaked brick)
304	Does the household have a mobile telephone? (No; Yes)
292	What is the highest level and class that the male head/spouse has completed in school? (None, pre-school, or primary class 1; Primary class 2 to 5; Primary class 6 to pre-secondary class 2; No male head/spouse; Pre-secondary class 3, or higher)
283	Can either the household head or the spouse write a letter (with or without difficulty)? (Two spouses, neither can write; Two spouses, one can write; Two spouses, both can write; Female head/spouse only or male head/spouse only, and he/she cannot write; Female head/spouse only or male head/spouse only, and he/she can write)

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

Uncertainty coefficient	Indicator (Answers ordered starting with those most strongly linked with higher poverty likelihoods)
280	In their main occupation in the past 12 months, was any household member an unskilled worker (including agricultural laborer)? (No; Yes)
280	Does the household own any bicycles, motorcycle/scooters, cars or trucks, or boats? (No; Bicycle only; Boat only; Motorcycle/scooter, regardless of bicycle, no cars, trucks, or boats; Cars or trucks, regardless of others)
267	In their main occupation in the past 12 months, how many household members were unskilled workers (including agricultural laborers) or were workers in agriculture and animal husbandry (farming), forestry, fishing, or hunting? (Four or more; Three, Two; One; None)
267	What is the major material of the roof? (Leaves, or other; Metal sheets/zinc, concrete, wood, tile, or sugar palm fibre)
260	Does the household own any mobile phones? (No; Yes)
252	What was the main occupation of the female head/spouse in the past 12 months? (Worker in agriculture and animal husbandry (farmer), forestry, fishing, or hunting; Does not work; No female head/spouse; Professional or technical expert, managerial, administrative, or decision-making staff, clerical, sales worker, service worker, manufacturing worker, transportation operator, unskilled worker (including agricultural laborer), or other)
250	What type of dwelling is it? (Bamboo house, traditional house, or small house in compound of main house; Semi-permanent; Permanent house; Emergency/tent, or other)
230	Does the household own any televisions, tape players/CD players, or radios? (No; Yes)
226	Does the male head/spouse speak Portuguese or English? (No; No male head/spouse; Yes)
225	Does the household own any televisions? (No; Yes)

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

Uncertainty coefficient	Indicator (Answers ordered starting with those most strongly linked with higher poverty likelihoods)
224	How many square-meters of land does the household cultivate (or has or controls, even if the land does not belong to the household) that is for annual crops or fallow, tree crops, pasture, plantation, grassland, or garden/garden plot? (1 to 1,499; None; 1,500 to 2,999; 3,000 to 9,999; 10,000 or more)
224	What is the main source of light in your dwelling? (Lamp, or candles or battery flashlights; Other; Electricity, privately generated electricity, or Petromax (kerosene pressure lantern))
207	What is the highest level and class that the female head/spouse has completed in school? (None, pre-school, or primary, class 1 to 5; Primary, class 6, to pre-secondary, class 2; Pre-secondary, class 3, to secondary, class 2; No female head/spouse; Secondary, class 3, or higher)
201	What is the main occupation of the female head/spouse? (Sharecropper, or farm laborer; Housekeeper; Other; Farmer; No female head/spouse; Non-farm laborer, fisher, trader, skilled worker, civil servant, teacher, pensioner, school student, or university student)
186	Can the male head/spouse write a letter? (No; No male head/spouse; Yes (with or without difficulty))
177	How many clothes cupboards (*lemari pakaian*) does the household own? (None; One; Two or more)
167	Does the female head/spouse speak Indonesian? (No; Yes; No female head/spouse)
161	Did the male head/spouse work for a wage in his main occupation in the past 12 months? (No; No male head/spouse; Yes)
154	What is the condition of the dwelling unit? (Severely damaged; Damaged a little; Mediocre; Good)
152	Is this bath/shower used only by your household? (No; Yes)
151	Can the female head/spouse write a letter? (No; Yes, but with difficulty; Yes, without difficulty; No female head/spouse)
146	Does the household own any tape players/CD players? (No; Yes)
146	In their main occupation in the past 12 months, did any household members have a wage job or a non-wage, non-farming job? (No; Yes)

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

Uncertainty coefficient	Indicator (Answers ordered starting with those most strongly linked with higher poverty likelihoods)
140	Does the female head/spouse speak Portuguese or English? (No; No female head/spouse; Yes)
139	Does the male head/spouse speak Indonesian? (No; Yes; No male head/spouse)
134	In their main occupation in the past 12 months, did any household members have a wage job for the government, public sector, army, state-owned enterprise, rural public-works program, or NGO? (No; Yes)
127	How many household members worked one or more hours in the past seven days, or have a permanent job even though they did not work in the past seven days? (Four or more; Three; Two; One; None)
124	In their main occupation in the past 12 months, did any household members have a wage job? (No; Yes)
114	What is the structure of household headship? (Both male and female heads/spouses; Female head/spouse only; Male head/spouse only)
106	What is the final disposal of sewage? (River/lake/ocean; Other; Pond/field; Hole; Shore/open field; No toilet; Septic tank)
101	Does the household own any mosquito nets? (No; Yes)
100	Does the female head/spouse speak Tetun? (Yes; No; No female head/spouse)
99	How many picks does the household own now? (Four or more; Three; Two; One; None)
97	How many rooms do the members of the household occupy, including bedrooms, living rooms, and rooms used for household enterprises? (One; Two; Three; Four; Five or more)
97	Does the household own any radios? (No; Yes)
96	Concerning the household's food consumption over the past month, was it . . .? (Less than adequate; Just adequate, or more than adequate)
96	Does your household treat your drinking water in any way? (No; Yes)

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

Uncertainty coefficient	Indicator (Answers ordered starting with those most strongly linked with higher poverty likelihoods)
81	How many pigs, goats, and sheep does your household own today? (One; None; Two; Three; Four; Five; Six or more)
77	Can either the household head or the spouse write a letter without difficulty? (Two spouses, neither can write without difficulty, or female head/spouse only or male head/spouse only, and he/she cannot write without difficulty; Two spouses, one or both can write without difficulty, or female head/spouse only or male head/spouse only, and he/she can write without difficulty)
77	How many household members can write a letter (with or without difficulty)? (None; One; Two; Three or more)
75	Did the female head/spouse work for a wage in her main occupation in the past 12 months? (No; Yes; No female head/spouse)
74	Did any household members sleep under a treated or untreated mosquito net last night? (No; Some or all treated; Yes, only untreated)
72	Did the female head/spouse work one or more hours in the past seven days, or if not, does she have a permanent job even though she did not work in the past seven days? (No; Yes; No female head/spouse)
71	What is the main source of drinking water for the household? (River, stream, lake, pond, or rainwater; Protected well; Unprotected spring; Protected spring; Unprotected well; Tap water; Pump; Bottled water, or other)
71	Did many household members sleep under a mosquito net last night? (No; Yes)
48	What is the main source of water for bathing and washing for your household? (River, stream, lake, pond, rainwater, or other; Public well; Spring; Public taps/standpipe; Private well; Private connection to pipeline)
47	Concerning the household's housing situation over the past month, was it . . .? (Less than adequate; Just adequate, or more than adequate)

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

Uncertainty coefficient	Indicator (Answers ordered starting with those most strongly linked with higher poverty likelihoods)
44	How many goats and sheep does your household own today? (One; None; Two; Three or more)
43	Does the male head/spouse speak Tetun? (Yes; No; No male head/spouse)
39	How many pigs does your household own today? (One; None; Two; Three or more)
38	Did the male head/spouse work one or more hours in the past seven days, or if not, does he have a permanent job even though he did not work in the past seven days? (Yes; No; No male head/spouse)
24	Is the dwelling owned by a member of the household? (Yes; No)
24	Does the household own any hoes now? (Yes; No)
24	Does your household own any buffalo today? (No; Yes)
18	Does the household live in a rural area? (Yes; No)
15	How many household members can write a letter without difficulty? (None; One; Two; Three or more)
8	Does the household own any sickles/reaping hooks now? (Yes; No)
7	Does the household own any big knives now? (Yes; No)
5	Does your household own any buffalo, Bali cows, cows, horses, pigs, goats, sheep, chickens, or ducks today? (Yes; No)
2	Concerning the household's health-care situation over the past month, was it . . .? (Less than adequate; Just adequate, or more than adequate)
1	Does the household own any shovels now? (No; Yes)
0.9	Did any household members sleep under a treated mosquito net last night? (No; Yes)
0.7	Does your household own any buffalo, Bali cows, cows, or horses today? (No; Yes)

Source: 2007 Survey of Living Standards and the upper national poverty line

Tables for
the Upper National Poverty Line

(and tables pertaining to all nine poverty lines)

Figure 4 (Upper 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	100.0
5–9	100.0
10–14	93.1
15–19	93.1
20–24	82.6
25–29	78.5
30–34	61.1
35–39	51.5
40–44	27.4
45–49	23.4
50–54	8.6
55–59	9.5
60–64	4.4
65–69	0.1
70–74	0.0
75–79	0.0
80–84	0.0
85–89	0.0
90–94	0.0
95–100	0.0

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

Score	Households below poverty line		All households at score		Poverty likelihood (estimated, %)
0–4	348	÷	348	=	100.0
5–9	472	÷	472	=	100.0
10–14	1,591	÷	1,709	=	93.1
15–19	4,474	÷	4,808	=	93.1
20–24	6,048	÷	7,323	=	82.6
25–29	7,931	÷	10,106	=	78.5
30–34	7,087	÷	11,606	=	61.1
35–39	6,896	÷	13,399	=	51.5
40–44	2,771	÷	10,108	=	27.4
45–49	2,721	÷	11,633	=	23.4
50–54	843	÷	9,839	=	8.6
55–59	764	÷	8,014	=	9.5
60–64	213	÷	4,847	=	4.4
65–69	4	÷	3,133	=	0.1
70–74	0	÷	1,604	=	0.0
75–79	0	÷	838	=	0.0
80–84	0	÷	125	=	0.0
85–89	0	÷	87	=	0.0
90–94	0	÷	0	=	#N/A
95–100	0	÷	0	=	0.0

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

Figure 6: Distribution of household poverty likelihoods across consumption ranges demarcated by poverty lines per person per day

	Likelihood of having expenditure in ranges demarcated by national poverty lines per person per day							
	<USAID	=>USAID and <$1.25/day	=>$1.25/day and <Lower natl.	=>Lower natl. and <Upper natl.	=>Upper natl. and <150% upper natl.	=>150% upper natl. and <$2.50/day	=>$2.50/day and <200% upper natl.	=>200% upper natl.
Score	<USD0.66	=>USD0.66 and <USD0.71	=>USD0.71 and <USD0.71	=>USD0.71 and <USD0.88	=>USD0.88 and <USD1.32	=>USD1.32 and <USD1.42	=>USD1.42 and <USD1.75	=>USD1.75
0–4	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5–9	88.6	11.4	0.0	0.0	0.0	0.0	0.0	0.0
10–14	64.3	6.2	0.0	22.6	6.9	0.0	0.0	0.0
15–19	57.7	16.1	4.5	14.8	6.4	0.5	0.0	0.0
20–24	54.0	7.4	3.2	17.9	16.2	1.2	0.1	0.0
25–29	41.9	10.5	4.6	21.5	19.7	0.7	0.1	1.0
30–34	25.5	12.5	0.0	23.1	28.7	3.1	3.3	3.9
35–39	17.5	7.5	6.1	20.4	32.7	3.0	7.0	5.8
40–44	3.8	4.9	0.7	17.9	47.6	5.3	12.1	7.6
45–49	4.1	4.0	1.2	14.2	36.3	10.0	15.2	15.2
50–54	1.4	1.3	0.2	5.7	43.8	9.2	16.4	22.0
55–59	1.9	2.7	0.6	4.2	24.4	9.6	13.7	42.8
60–64	1.7	2.2	0.0	0.5	16.2	3.8	30.3	45.3
65–69	0.0	0.0	0.0	0.1	17.2	14.3	24.6	43.8
70–74	0.0	0.0	0.0	0.0	0.0	0.0	19.5	80.5
75–79	0.0	0.0	0.0	0.0	0.0	0.0	52.9	47.1
80–84	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
85–89	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
90–94	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0
95–100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0

Note: All poverty likelihoods in percentage units.

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

Score	Diff.	Difference between estimate and true value		
		Confidence interval ($+/-$ percentage points)		
		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	−6.5	3.4	3.4	3.5
15–19	+11.2	2.9	3.4	4.3
20–24	−2.3	2.1	2.3	3.0
25–29	+8.7	2.1	2.5	3.4
30–34	+7.9	2.2	2.6	3.4
35–39	+13.4	1.9	2.2	2.9
40–44	+1.2	1.9	2.3	2.9
45–49	−10.8	6.6	6.9	7.4
50–54	−2.8	2.0	2.2	2.3
55–59	+5.2	0.8	1.0	1.3
60–64	−0.9	1.5	1.7	2.2
65–69	+0.1	0.0	0.0	0.0
70–74	+0.0	0.0	0.0	0.0
75–79	+0.0	0.0	0.0	0.0
80–84	+0.0	0.0	0.0	0.0
85–89	+0.0	0.0	0.0	0.0
90–94	+0.0	0.0	0.0	0.0
95–100	+0.0	0.0	0.0	0.0

Figure 8 (Upper national line): Differences and precision of differences for bootstrapped estimates of poverty rates for groups of households at a point in time, by sample size, scorecard applied to the validation sample

Sample Size n	Difference between estimate and true value			
	Diff.	Confidence interval (+/− percentage points)		
		90-percent	95-percent	99-percent
1	−0.1	68.8	77.5	87.0
4	+2.6	38.1	45.8	58.4
8	+2.7	28.4	34.3	45.0
16	+2.7	21.3	24.7	32.8
32	+2.4	14.8	17.6	23.0
64	+2.2	10.7	12.4	17.0
128	+2.4	7.9	9.3	11.3
256	+2.3	5.5	6.5	8.2
512	+2.2	3.6	4.5	6.0
1,024	+2.2	2.6	3.1	4.3
2,048	+2.2	1.9	2.2	3.0
4,096	+2.2	1.3	1.5	2.0
8,192	+2.2	0.9	1.1	1.4
16,384	+2.3	0.7	0.8	1.0

Figure 9 (All poverty lines): Differences, precision of differences, and the α factor for bootstrapped estimates of poverty rates for groups of households at a point in time, scorecard applied to the validation sample

		Poverty line						
		National						Intl. 2005 PPP
				150%	200%			
		Lower	Upper	Upper	Upper	USAID		
		Natl.	Natl.	Natl.	Natl.	'Extreme'	$1.25	$2.50
Estimate minus true value								
Scorecard applied to validation sample		+1.8	+2.3	−0.2	+1.0	+2.0	+1.0	+1.1
Precision of difference								
Scorecard applied to validation sample		0.6	0.7	0.6	0.5	0.4	0.6	0.6
α factor for standard errors								
Scorecard applied to validation sample		1.08	1.03	1.06	1.04	0.85	1.08	1.04
Precision is measured as 90-percent confidence intervals in units of +/− percentage points.								
Differences and precision estimated from 1,000 bootstraps of size $n = 16{,}384$.								
α is estimated from 1,000 bootstrap samples of $n = 256, 512, 1{,}024, 2{,}048, 4{,}096, 8{,}192$, and 16,384.								

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

		Targeting segment	
		Targeted	**Non-targeted**
True poverty status	**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 (Upper national line): Households by targeting classification and score, along with "Total Accuracy" and BPAC, scorecard applied to the validation sample

Score	Inclusion: < poverty line correctly targeted	Undercoverage: < poverty line mistakenly non-targeted	Leakage: => poverty line mistakenly targeted	Exclusion: => poverty line correctly non-targeted	Total Accuracy Inclusion + Exclusion	BPAC See text
0–4	0.3	41.8	0.0	57.8	58.2	–98.3
5–9	0.8	41.3	0.0	57.8	58.7	–96.1
10–14	2.5	39.7	0.0	57.8	60.3	–88.1
15–19	6.6	35.5	0.7	57.1	63.8	–66.8
20–24	13.0	29.2	1.7	56.1	69.1	–34.5
25–29	20.3	21.8	4.4	53.4	73.7	+6.9
30–34	27.1	15.1	9.3	48.6	75.7	+50.5
35–39	33.3	8.9	16.5	41.4	74.6	+60.9
40–44	36.5	5.6	23.3	34.5	71.0	+44.6
45–49	39.8	2.3	31.7	26.2	66.0	+24.9
50–54	41.4	0.7	39.9	17.9	59.3	+5.3
55–59	41.9	0.2	47.4	10.4	52.4	–12.5
60–64	42.2	0.0	52.0	5.8	48.0	–23.4
65–69	42.2	0.0	55.2	2.7	44.8	–30.9
70–74	42.2	0.0	56.8	1.1	43.2	–34.7
75–79	42.2	0.0	57.6	0.2	42.4	–36.7
80–84	42.2	0.0	57.7	0.1	42.3	–37.0
85–89	42.2	0.0	57.8	0.0	42.2	–37.2
90–94	42.2	0.0	57.8	0.0	42.2	–37.2
95–100	42.2	0.0	57.8	0.0	42.2	–37.2

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

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

Targeting cut-off	% all households who are targeted	% targeted who are poor	% of poor who are targeted	Poor households targeted per non-poor household targeted
0–4	0.3	100.0	0.8	Only poor targeted
5–9	0.8	100.0	1.9	Only poor targeted
10–14	2.5	98.9	5.9	87.1:1
15–19	7.3	90.5	15.8	9.6:1
20–24	14.7	88.4	30.7	7.6:1
25–29	24.8	82.1	48.2	4.6:1
30–34	36.4	74.5	64.3	2.9:1
35–39	49.8	66.9	79.0	2.0:1
40–44	59.9	61.0	86.6	1.6:1
45–49	71.5	55.7	94.5	1.3:1
50–54	81.4	50.9	98.3	1.0:1
55–59	89.4	46.9	99.5	0.9:1
60–64	94.2	44.8	100.0	0.8:1
65–69	97.3	43.3	100.0	0.8:1
70–74	98.9	42.6	100.0	0.7:1
75–79	99.8	42.3	100.0	0.7:1
80–84	99.9	42.2	100.0	0.7:1
85–89	100.0	42.2	100.0	0.7:1
90–94	100.0	42.2	100.0	0.7:1
95–100	100.0	42.2	100.0	0.7:1

Tables for
the Lower National Poverty Line

Figure 4 (Lower 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	100.0
5–9	100.0
10–14	70.5
15–19	78.3
20–24	64.7
25–29	57.0
30–34	37.9
35–39	31.1
40–44	9.5
45–49	9.2
50–54	2.9
55–59	5.3
60–64	2.2
65–69	0.0
70–74	0.0
75–79	0.0
80–84	0.0
85–89	0.0
90–94	0.0
95–100	0.0

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

Score	Diff.	Difference between estimate and true value Confidence interval (+/− percentage points) 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	−15.5	9.6	10.0	10.5
15–19	+7.1	3.1	3.7	4.7
20–24	−2.8	2.7	2.9	3.8
25–29	+14.0	2.2	2.7	3.5
30–34	+6.6	2.0	2.3	3.1
35–39	+12.0	1.4	1.6	2.1
40–44	−3.2	2.3	2.4	2.6
45–49	−11.4	6.8	7.0	7.4
50–54	−1.0	0.9	0.9	1.1
55–59	+4.6	0.4	0.4	0.5
60–64	+0.6	0.7	0.9	1.3
65–69	+0.0	0.0	0.0	0.0
70–74	+0.0	0.0	0.0	0.0
75–79	+0.0	0.0	0.0	0.0
80–84	+0.0	0.0	0.0	0.0
85–89	+0.0	0.0	0.0	0.0
90–94	+0.0	0.0	0.0	0.0
95–100	+0.0	0.0	0.0	0.0

Figure 8 (Lower national line): Differences and precision of differences for bootstrapped estimates of poverty rates for groups of households at a point in time, by sample size, scorecard applied to the validation sample

Sample Size n	Diff.	90-percent	95-percent	99-percent
		Difference between estimate and true value		
		Confidence interval (+/− percentage points)		
1	−0.5	63.0	77.7	87.1
4	+2.1	33.9	41.4	55.9
8	+2.1	24.8	30.4	41.5
16	+2.1	17.9	22.5	31.1
32	+1.9	14.0	16.6	22.0
64	+1.7	9.5	11.5	14.5
128	+1.7	7.2	8.3	10.6
256	+1.8	5.1	5.8	7.6
512	+1.7	3.5	4.2	5.4
1,024	+1.7	2.5	3.0	3.9
2,048	+1.7	1.7	2.1	2.6
4,096	+1.8	1.2	1.4	1.9
8,192	+1.8	0.9	1.0	1.3
16,384	+1.8	0.6	0.8	0.9

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

Score	Inclusion: < poverty line correctly targeted	Undercoverage: < poverty line mistakenly non-targeted	Leakage: => poverty line mistakenly targeted	Exclusion: => poverty line correctly non-targeted	Total Accuracy Inclusion + Exclusion	BPAC See text
0–4	0.3	26.9	0.0	72.7	73.1	−97.5
5–9	0.8	26.5	0.0	72.7	73.5	−94.0
10–14	2.3	25.0	0.3	72.5	74.7	−82.4
15–19	5.9	21.4	1.5	71.3	77.1	−51.6
20–24	11.0	16.3	3.7	69.0	80.0	−6.1
25–29	15.9	11.4	8.9	63.9	79.8	+49.1
30–34	20.0	7.3	16.4	56.3	76.3	+39.9
35–39	23.5	3.7	26.2	46.5	70.0	+3.9
40–44	25.0	2.3	34.9	37.9	62.9	−27.8
45–49	26.6	0.7	44.9	27.8	54.4	−64.7
50–54	27.1	0.1	54.2	18.5	45.7	−98.7
55–59	27.2	0.1	62.2	10.6	37.8	−127.8
60–64	27.3	0.0	66.9	5.8	33.1	−145.3
65–69	27.3	0.0	70.1	2.7	29.9	−156.8
70–74	27.3	0.0	71.7	1.1	28.3	−162.7
75–79	27.3	0.0	72.5	0.2	27.5	−165.8
80–84	27.3	0.0	72.6	0.1	27.4	−166.2
85–89	27.3	0.0	72.7	0.0	27.3	−166.5
90–94	27.3	0.0	72.7	0.0	27.3	−166.5
95–100	27.3	0.0	72.7	0.0	27.3	−166.5

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

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

Targeting cut-off	% all households who are targeted	% targeted who are poor	% of poor who are targeted	Poor households targeted per non-poor household targeted
0–4	0.3	100.0	1.3	Only poor targeted
5–9	0.8	100.0	3.0	Only poor targeted
10–14	2.5	89.7	8.3	8.7:1
15–19	7.3	80.2	21.6	4.0:1
20–24	14.7	74.7	40.1	3.0:1
25–29	24.8	64.2	58.3	1.8:1
30–34	36.4	55.0	73.3	1.2:1
35–39	49.8	47.3	86.3	0.9:1
40–44	59.9	41.8	91.7	0.7:1
45–49	71.5	37.2	97.4	0.6:1
50–54	81.4	33.4	99.5	0.5:1
55–59	89.4	30.4	99.7	0.4:1
60–64	94.2	29.0	100.0	0.4:1
65–69	97.3	28.0	100.0	0.4:1
70–74	98.9	27.6	100.0	0.4:1
75–79	99.8	27.3	100.0	0.4:1
80–84	99.9	27.3	100.0	0.4:1
85–89	100.0	27.3	100.0	0.4:1
90–94	100.0	27.3	100.0	0.4:1
95–100	100.0	27.3	100.0	0.4:1

Tables for
150% of the Upper National Poverty Line

Figure 4 (150% of the upper 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	100.0
5–9	100.0
10–14	100.0
15–19	99.5
20–24	98.8
25–29	98.2
30–34	89.7
35–39	84.2
40–44	75.0
45–49	59.7
50–54	52.4
55–59	33.9
60–64	20.6
65–69	17.4
70–74	0.0
75–79	0.0
80–84	0.0
85–89	0.0
90–94	0.0
95–100	0.0

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

| Score | Diff. | Difference between estimate and true value | | |
| | | Confidence interval (+/− percentage points) | | |
		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	+0.0	0.0	0.0	0.0
15–19	+8.8	2.5	3.0	4.0
20–24	+1.4	0.8	0.9	1.1
25–29	+4.1	1.3	1.5	1.9
30–34	−2.4	1.8	1.9	2.1
35–39	−1.1	1.3	1.6	2.1
40–44	+14.1	2.5	3.0	3.9
45–49	−15.6	8.7	8.8	9.2
50–54	+11.3	2.5	3.0	4.0
55–59	−2.5	2.5	2.9	3.8
60–64	−6.7	4.8	5.0	5.6
65–69	+0.7	3.5	4.1	5.7
70–74	−20.3	12.3	12.8	13.6
75–79	+0.0	0.0	0.0	0.0
80–84	+0.0	0.0	0.0	0.0
85–89	+0.0	0.0	0.0	0.0
90–94	+0.0	0.0	0.0	0.0
95–100	+0.0	0.0	0.0	0.0

Figure 8 (150% of the upper national line): Differences and precision of differences for bootstrapped estimates of poverty rates for groups of households at a point in time, by sample size, scorecard applied to the validation sample

| Sample Size | Difference between estimate and true value | | | |
| | | Confidence interval (+/− percentage points) | | |
n	Diff.	90-percent	95-percent	99-percent
1	−1.9	70.5	81.8	90.4
4	+0.8	35.8	42.2	52.8
8	+0.5	26.5	31.7	39.0
16	+0.5	19.3	23.0	29.3
32	+0.1	13.9	16.7	22.7
64	−0.3	9.4	11.3	15.7
128	−0.3	6.7	7.9	10.7
256	−0.3	4.6	5.7	7.6
512	−0.3	3.6	4.3	6.0
1,024	−0.3	2.5	3.0	4.1
2,048	−0.3	1.8	2.1	2.7
4,096	−0.2	1.2	1.5	1.9
8,192	−0.2	0.9	1.0	1.4
16,384	−0.2	0.6	0.7	0.9

Figure 11 (150% of the upper national line): Households by targeting classification and score, along with "Total Accuracy" and BPAC, scorecard applied to the validation sample

Score	Inclusion: < poverty line correctly targeted	Undercoverage: < poverty line mistakenly non-targeted	Leakage: => poverty line mistakenly targeted	Exclusion: => poverty line correctly non-targeted	Total Accuracy Inclusion + Exclusion	BPAC See text
0–4	0.3	71.1	0.0	28.6	28.9	–99.0
5–9	0.8	70.6	0.0	28.6	29.4	–97.7
10–14	2.5	68.9	0.0	28.6	31.1	–92.9
15–19	7.1	64.3	0.2	28.3	35.4	–79.8
20–24	14.2	57.2	0.5	28.1	42.3	–59.6
25–29	23.9	47.5	0.9	27.7	51.6	–31.9
30–34	34.6	36.8	1.7	26.8	61.5	–0.6
35–39	46.0	25.5	3.8	24.8	70.7	+34.0
40–44	53.0	18.4	6.9	21.7	74.7	+58.0
45–49	61.1	10.3	10.4	18.2	79.3	+85.5
50–54	66.0	5.4	15.3	13.2	79.2	+78.5
55–59	69.2	2.3	20.2	8.4	77.6	+71.7
60–64	70.6	0.8	23.6	5.0	75.6	+67.0
65–69	71.1	0.3	26.3	2.3	73.4	+63.2
70–74	71.4	0.0	27.5	1.1	72.5	+61.5
75–79	71.4	0.0	28.4	0.2	71.6	+60.3
80–84	71.4	0.0	28.5	0.1	71.5	+60.1
85–89	71.4	0.0	28.6	0.0	71.4	+60.0
90–94	71.4	0.0	28.6	0.0	71.4	+60.0
95–100	71.4	0.0	28.6	0.0	71.4	+60.0

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

Figure 12 (150% of the upper national line): For a given score cut-off, the percentage of all households who are targeted (that is, have a score equal to or less than the cut-off), the percentage of targeted households who are poor (that is, below the poverty line), the percentage of poor households who are targeted, and the number of poor households who are successful targeted (coverage) per non-poor household mistakenly targeted (leakage), scorecard applied to the validation sample

Targeting cut-off	% all households who are targeted	% targeted who are poor	% of poor who are targeted	Poor households targeted per non-poor household targeted
0–4	0.3	100.0	0.5	Only poor targeted
5–9	0.8	100.0	1.1	Only poor targeted
10–14	2.5	100.0	3.5	Only poor targeted
15–19	7.3	96.8	9.9	30.4:1
20–24	14.7	96.9	19.9	31.5:1
25–29	24.8	96.4	33.4	27.1:1
30–34	36.4	95.2	48.5	19.9:1
35–39	49.8	92.3	64.3	12.1:1
40–44	59.9	88.5	74.2	7.7:1
45–49	71.5	85.5	85.6	5.9:1
50–54	81.4	81.1	92.4	4.3:1
55–59	89.4	77.4	96.8	3.4:1
60–64	94.2	75.0	98.9	3.0:1
65–69	97.3	73.0	99.5	2.7:1
70–74	98.9	72.2	100.0	2.6:1
75–79	99.8	71.6	100.0	2.5:1
80–84	99.9	71.5	100.0	2.5:1
85–89	100.0	71.4	100.0	2.5:1
90–94	100.0	71.4	100.0	2.5:1
95–100	100.0	71.4	100.0	2.5:1

Tables for
200% of the Upper National Poverty Line

Figure 4 (200% of the upper 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	100.0
5–9	100.0
10–14	100.0
15–19	100.0
20–24	100.0
25–29	99.0
30–34	96.1
35–39	94.2
40–44	92.4
45–49	84.9
50–54	78.0
55–59	57.2
60–64	54.7
65–69	56.2
70–74	19.5
75–79	52.9
80–84	0.0
85–89	0.0
90–94	0.0
95–100	0.0

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

| Score | Diff. | Difference between estimate and true value | | |
| | | Confidence interval ($+/-$ percentage points) | | |
		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	+0.0	0.0	0.0	0.0
15–19	+0.0	0.0	0.0	0.0
20–24	+1.3	0.5	0.7	0.8
25–29	−1.0	0.5	0.5	0.5
30–34	−2.4	1.5	1.5	1.6
35–39	+0.2	0.9	1.1	1.5
40–44	+2.5	1.3	1.6	2.1
45–49	−5.3	3.2	3.3	3.4
50–54	+5.4	2.4	2.9	3.7
55–59	−5.2	3.8	4.1	4.7
60–64	+8.4	3.2	3.8	5.1
65–69	+22.1	4.1	5.0	6.8
70–74	−7.8	6.3	6.7	7.8
75–79	+32.8	5.3	6.4	8.9
80–84	+0.0	0.0	0.0	0.0
85–89	+0.0	0.0	0.0	0.0
90–94	+0.0	0.0	0.0	0.0
95–100	+0.0	0.0	0.0	0.0

Figure 8 (200% of the upper national line): Differences and precision of differences for bootstrapped estimates of poverty rates for groups of households at a point in time, by sample size, scorecard applied to the validation sample

Sample Size	Difference between estimate and true value			
		Confidence interval (+/− percentage points)		
n	Diff.	90-percent	95-percent	99-percent
1	−0.9	60.7	65.1	87.8
4	+1.9	28.1	35.2	45.0
8	+1.7	20.6	25.5	33.8
16	+1.7	15.5	18.5	22.8
32	+1.3	10.7	12.7	16.5
64	+1.0	7.8	9.1	12.6
128	+1.0	5.4	6.5	8.0
256	+0.9	3.8	4.4	5.9
512	+1.0	2.8	3.3	4.2
1,024	+0.9	1.9	2.4	3.0
2,048	+0.9	1.3	1.6	2.0
4,096	+1.0	0.9	1.1	1.5
8,192	+1.0	0.7	0.8	1.1
16,384	+1.0	0.5	0.6	0.8

Figure 11 (200% of the upper national line): Households by targeting classification and score, along with "Total Accuracy" and BPAC, scorecard applied to the validation sample

Score	Inclusion: < poverty line correctly targeted	Undercoverage: < poverty line mistakenly non-targeted	Leakage: => poverty line mistakenly targeted	Exclusion: => poverty line correctly non-targeted	Total Accuracy Inclusion + Exclusion	BPAC See text
0–4	0.3	84.5	0.0	15.2	15.5	–99.2
5–9	0.8	84.0	0.0	15.2	16.0	–98.1
10–14	2.5	82.3	0.0	15.2	17.7	–94.0
15–19	7.3	77.5	0.0	15.2	22.5	–82.7
20–24	14.5	70.3	0.1	15.1	29.6	–65.6
25–29	24.6	60.2	0.1	15.1	39.7	–41.7
30–34	36.0	48.8	0.3	14.9	50.9	–14.6
35–39	48.6	36.2	1.1	14.1	62.7	+16.0
40–44	57.8	27.1	2.1	13.1	70.8	+38.7
45–49	67.7	17.1	3.8	11.4	79.1	+64.2
50–54	75.2	9.6	6.2	9.0	84.2	+84.6
55–59	80.2	4.6	9.1	6.1	86.3	+89.2
60–64	82.8	2.0	11.4	3.8	86.6	+86.6
65–69	84.1	0.7	13.3	1.9	86.0	+84.4
70–74	84.6	0.2	14.4	0.8	85.4	+83.1
75–79	84.8	0.0	15.0	0.2	85.0	+82.3
80–84	84.8	0.0	15.1	0.1	84.9	+82.2
85–89	84.8	0.0	15.2	0.0	84.8	+82.1
90–94	84.8	0.0	15.2	0.0	84.8	+82.1
95–100	84.8	0.0	15.2	0.0	84.8	+82.1

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

Figure 12 (200% of the upper national line): For a given score cut-off, the percentage of all households who are targeted (that is, have a score equal to or less than the cut-off), the percentage of targeted households who are poor (that is, below the poverty line), the percentage of poor households who are targeted, and the number of poor households who are successful targeted (coverage) per non-poor household mistakenly targeted (leakage), scorecard applied to the validation sample

Targeting cut-off	% all households who are targeted	% targeted who are poor	% of poor who are targeted	Poor households targeted per non-poor household targeted
0–4	0.3	100.0	0.4	Only poor targeted
5–9	0.8	100.0	1.0	Only poor targeted
10–14	2.5	100.0	3.0	Only poor targeted
15–19	7.3	100.0	8.6	Only poor targeted
20–24	14.7	99.2	17.1	125.4:1
25–29	24.8	99.5	29.1	199.5:1
30–34	36.4	99.1	42.5	112.0:1
35–39	49.8	97.7	57.4	43.0:1
40–44	59.9	96.5	68.1	27.2:1
45–49	71.5	94.7	79.9	17.9:1
50–54	81.4	92.4	88.6	12.2:1
55–59	89.4	89.8	94.6	8.8:1
60–64	94.2	87.9	97.6	7.3:1
65–69	97.3	86.4	99.2	6.3:1
70–74	98.9	85.5	99.7	5.9:1
75–79	99.8	85.0	100.0	5.7:1
80–84	99.9	84.9	100.0	5.6:1
85–89	100.0	84.8	100.0	5.6:1
90–94	100.0	84.8	100.0	5.6:1
95–100	100.0	84.8	100.0	5.6:1

Tables for
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	100.0
5–9	88.6
10–14	64.3
15–19	57.7
20–24	54.0
25–29	41.9
30–34	25.5
35–39	17.5
40–44	3.8
45–49	4.1
50–54	1.4
55–59	1.9
60–64	1.7
65–69	0.0
70–74	0.0
75–79	0.0
80–84	0.0
85–89	0.0
90–94	0.0
95–100	0.0

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

Score	Diff.	Difference between estimate and true value		
		Confidence interval (+/− percentage points)		
		90-percent	95-percent	99-percent
0–4	+0.0	0.0	0.0	0.0
5–9	−3.2	5.0	6.0	7.9
10–14	+2.3	5.4	6.3	8.4
15–19	−6.1	4.7	4.9	5.4
20–24	+2.3	2.6	3.1	4.0
25–29	+11.3	2.2	2.5	3.1
30–34	−2.0	2.0	2.2	2.8
35–39	+9.3	0.9	1.1	1.3
40–44	−5.6	3.4	3.5	3.7
45–49	+2.8	0.3	0.4	0.5
50–54	−0.6	0.6	0.6	0.9
55–59	+1.3	0.4	0.4	0.5
60–64	+1.7	0.0	0.0	0.0
65–69	+0.0	0.0	0.0	0.0
70–74	+0.0	0.0	0.0	0.0
75–79	+0.0	0.0	0.0	0.0
80–84	+0.0	0.0	0.0	0.0
85–89	+0.0	0.0	0.0	0.0
90–94	+0.0	0.0	0.0	0.0
95–100	+0.0	0.0	0.0	0.0

Figure 8 (USAID "extreme" line): Differences and precision of differences for bootstrapped estimates of poverty rates for groups of households at a point in time, by sample size, scorecard applied to the validation sample

| Sample Size n | Difference between estimate and true value | | | |
| | Diff. | Confidence interval (+/− percentage points) | | |
		90-percent	95-percent	99-percent
1	−1.0	61.2	68.2	80.2
4	+1.8	28.5	34.3	43.2
8	+2.0	19.6	23.9	29.9
16	+2.0	13.7	16.4	20.5
32	+1.7	9.7	11.6	16.5
64	+1.7	6.7	8.2	11.0
128	+1.7	5.0	6.0	7.5
256	+1.9	3.4	4.1	5.4
512	+1.9	2.4	2.9	3.6
1,024	+1.9	1.8	2.0	2.7
2,048	+2.0	1.2	1.5	2.1
4,096	+2.0	0.9	1.1	1.4
8,192	+1.9	0.6	0.7	1.0
16,384	+2.0	0.4	0.5	0.7

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

Score	Inclusion: < poverty line correctly targeted	Undercoverage: < poverty line mistakenly non-targeted	Leakage: => poverty line mistakenly targeted	Exclusion: => poverty line correctly non-targeted	Total Accuracy Inclusion + Exclusion	BPAC See text
0–4	0.3	19.4	0.0	80.3	80.6	–96.5
5–9	0.8	19.0	0.0	80.2	81.0	–91.9
10–14	1.9	17.8	0.6	79.7	81.6	–77.5
15–19	5.1	14.6	2.2	78.1	83.2	–36.8
20–24	9.4	10.3	5.3	75.0	84.4	+21.9
25–29	12.9	6.9	11.9	68.4	81.2	+39.7
30–34	16.2	3.6	20.2	60.1	76.2	–2.5
35–39	18.0	1.8	31.8	48.5	66.4	–61.3
40–44	19.1	0.7	40.8	39.4	58.5	–106.9
45–49	19.4	0.4	52.2	28.1	47.5	–164.4
50–54	19.7	0.1	61.7	18.6	38.2	–212.7
55–59	19.7	0.0	69.6	10.6	30.4	–253.0
60–64	19.7	0.0	74.5	5.8	25.5	–277.6
65–69	19.7	0.0	77.6	2.7	22.4	–293.5
70–74	19.7	0.0	79.2	1.1	20.8	–301.6
75–79	19.7	0.0	80.1	0.2	19.9	–305.8
80–84	19.7	0.0	80.2	0.1	19.8	–306.5
85–89	19.7	0.0	80.3	0.0	19.7	–306.9
90–94	19.7	0.0	80.3	0.0	19.7	–306.9
95–100	19.7	0.0	80.3	0.0	19.7	–306.9

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

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

Targeting cut-off	% all households who are targeted	% targeted who are poor	% of poor who are targeted	Poor households targeted per non-poor household targeted
0–4	0.3	100.0	1.8	Only poor targeted
5–9	0.8	94.8	3.9	18.1:1
10–14	2.5	75.6	9.7	3.1:1
15–19	7.3	70.1	26.1	2.3:1
20–24	14.7	64.0	47.6	1.8:1
25–29	24.8	52.0	65.2	1.1:1
30–34	36.4	44.4	81.9	0.8:1
35–39	49.8	36.1	91.0	0.6:1
40–44	59.9	31.8	96.6	0.5:1
45–49	71.5	27.1	98.1	0.4:1
50–54	81.4	24.2	99.7	0.3:1
55–59	89.4	22.1	100.0	0.3:1
60–64	94.2	20.9	100.0	0.3:1
65–69	97.3	20.3	100.0	0.3:1
70–74	98.9	19.9	100.0	0.2:1
75–79	99.8	19.8	100.0	0.2:1
80–84	99.9	19.7	100.0	0.2:1
85–89	100.0	19.7	100.0	0.2:1
90–94	100.0	19.7	100.0	0.2:1
95–100	100.0	19.7	100.0	0.2:1

Tables for
$1.25/day 2005 PPP Poverty Line

Figure 4 ($1.25/day 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	70.5
15–19	73.8
20–24	61.4
25–29	52.4
30–34	38.0
35–39	25.0
40–44	8.8
45–49	8.1
50–54	2.7
55–59	4.7
60–64	3.9
65–69	0.0
70–74	0.0
75–79	0.0
80–84	0.0
85–89	0.0
90–94	0.0
95–100	0.0

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

| Score | Diff. | Difference between estimate and true value | | |
| | | Confidence interval (+/− percentage points) | | |
		90-percent	95-percent	99-percent
0–4	+0.0	0.0	0.0	0.0
5–9	+8.2	5.0	6.0	7.9
10–14	−13.3	8.6	8.9	9.4
15–19	+2.6	3.1	3.7	4.7
20–24	−5.1	3.8	4.1	4.4
25–29	+12.6	2.3	2.7	3.4
30–34	+8.1	2.0	2.2	2.8
35–39	+8.9	1.2	1.5	2.0
40–44	−4.2	2.8	2.9	3.1
45–49	−11.7	7.0	7.1	7.7
50–54	−0.4	0.6	0.7	1.0
55–59	+3.0	0.6	0.7	0.9
60–64	+2.3	0.7	0.9	1.3
65–69	+0.0	0.0	0.0	0.0
70–74	+0.0	0.0	0.0	0.0
75–79	+0.0	0.0	0.0	0.0
80–84	+0.0	0.0	0.0	0.0
85–89	+0.0	0.0	0.0	0.0
90–94	+0.0	0.0	0.0	0.0
95–100	+0.0	0.0	0.0	0.0

Figure 8 ($1.25/day line): Differences and precision of differences for bootstrapped estimates of poverty rates for groups of households at a point in time, by sample size, scorecard applied to the validation sample

| Sample Size | Difference between estimate and true value | | | |
| | | Confidence interval (+/− percentage points) | | |
n	Diff.	90-percent	95-percent	99-percent
1	−1.5	63.7	76.7	84.7
4	+1.3	32.4	41.3	55.5
8	+1.4	24.4	29.5	40.3
16	+1.3	17.9	22.8	30.6
32	+1.0	13.9	16.0	21.3
64	+1.0	9.6	10.9	15.0
128	+1.0	7.0	8.1	10.6
256	+1.0	5.0	5.8	7.3
512	+1.0	3.5	4.1	5.2
1,024	+1.0	2.4	2.9	4.0
2,048	+1.0	1.8	2.1	2.6
4,096	+1.0	1.2	1.4	1.9
8,192	+1.0	0.8	1.0	1.3
16,384	+1.0	0.6	0.7	0.9

Figure 11 ($1.25/day line): Households by targeting classification and score, along with "Total Accuracy" and BPAC, scorecard applied to the validation sample

Score	Inclusion: < poverty line correctly targeted	Undercoverage: < poverty line mistakenly non-targeted	Leakage: => poverty line mistakenly targeted	Exclusion: => poverty line correctly non-targeted	Total Accuracy Inclusion + Exclusion	BPAC See text
0–4	0.3	25.7	0.0	73.9	74.3	−97.3
5–9	0.8	25.3	0.0	73.9	74.6	−93.9
10–14	2.2	23.9	0.4	73.6	75.7	−82.0
15–19	5.8	20.3	1.5	72.4	78.1	−49.7
20–24	10.8	15.3	3.9	70.0	80.8	−2.5
25–29	15.3	10.8	9.4	64.5	79.8	+53.7
30–34	19.3	6.8	17.1	56.8	76.1	+34.4
35–39	22.5	3.6	27.3	46.6	69.1	−4.6
40–44	24.0	2.1	35.9	38.0	62.0	−37.7
45–49	25.4	0.7	46.1	27.8	53.2	−76.7
50–54	25.9	0.2	55.5	18.4	44.3	−112.6
55–59	26.0	0.1	63.3	10.6	36.6	−142.8
60–64	26.1	0.0	68.1	5.8	31.9	−161.1
65–69	26.1	0.0	71.3	2.7	28.7	−173.1
70–74	26.1	0.0	72.9	1.1	27.1	−179.2
75–79	26.1	0.0	73.7	0.2	26.3	−182.4
80–84	26.1	0.0	73.8	0.1	26.2	−182.9
85–89	26.1	0.0	73.9	0.0	26.1	−183.3
90–94	26.1	0.0	73.9	0.0	26.1	−183.3
95–100	26.1	0.0	73.9	0.0	26.1	−183.3

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

Figure 12 ($1.25/day line): For a given score cut-off, the percentage of all households who are targeted (that is, have a score equal to or less than the cut-off), the percentage of targeted households who are poor (that is, below the poverty line), the percentage of poor households who are targeted, and the number of poor households who are successful targeted (coverage) per non-poor household mistakenly targeted (leakage), scorecard applied to the validation sample

Targeting cut-off	% all households who are targeted	% targeted who are poor	% of poor who are targeted	Poor households targeted per non-poor household targeted
0–4	0.3	100.0	1.3	Only poor targeted
5–9	0.8	94.8	3.0	18.1:1
10–14	2.5	86.0	8.3	6.2:1
15–19	7.3	78.9	22.2	3.7:1
20–24	14.7	73.5	41.3	2.8:1
25–29	24.8	61.9	58.8	1.6:1
30–34	36.4	53.0	73.8	1.1:1
35–39	49.8	45.1	86.1	0.8:1
40–44	59.9	40.0	91.8	0.7:1
45–49	71.5	35.5	97.3	0.6:1
50–54	81.4	31.8	99.1	0.5:1
55–59	89.4	29.1	99.7	0.4:1
60–64	94.2	27.7	100.0	0.4:1
65–69	97.3	26.8	100.0	0.4:1
70–74	98.9	26.4	100.0	0.4:1
75–79	99.8	26.1	100.0	0.4:1
80–84	99.9	26.1	100.0	0.4:1
85–89	100.0	26.1	100.0	0.4:1
90–94	100.0	26.1	100.0	0.4:1
95–100	100.0	26.1	100.0	0.4:1

Tables for
$2.50/day 2005 PPP Poverty Line

Figure 4 ($2.50/day 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	100.0
20–24	99.9
25–29	98.9
30–34	92.8
35–39	87.2
40–44	80.3
45–49	69.7
50–54	61.6
55–59	43.5
60–64	24.4
65–69	31.6
70–74	0.0
75–79	0.0
80–84	0.0
85–89	0.0
90–94	0.0
95–100	0.0

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

Score	Diff.	Difference between estimate and true value Confidence interval ($+/-$ percentage points) 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	+0.0	0.0	0.0	0.0
15–19	+1.8	0.8	1.0	1.3
20–24	+2.5	0.8	0.9	1.1
25–29	+3.5	1.2	1.4	1.8
30–34	−1.5	1.2	1.3	1.5
35–39	−0.0	1.3	1.5	1.9
40–44	+16.1	2.5	3.1	3.8
45–49	−13.3	7.4	7.5	7.9
50–54	+16.2	2.4	3.0	4.2
55–59	−3.9	3.2	3.5	4.3
60–64	−5.1	4.0	4.2	4.7
65–69	+13.7	3.5	4.3	5.7
70–74	−23.0	13.7	14.3	15.3
75–79	+0.0	0.0	0.0	0.0
80–84	+0.0	0.0	0.0	0.0
85–89	+0.0	0.0	0.0	0.0
90–94	+0.0	0.0	0.0	0.0
95–100	+0.0	0.0	0.0	0.0

Figure 8 ($2.50/day line): Differences and precision of differences for bootstrapped estimates of poverty rates for groups of households at a point in time, by sample size, scorecard applied to the validation sample

Sample Size	Difference between estimate and true value			
	Diff.	Confidence interval (+/− percentage points)		
n		90-percent	95-percent	99-percent
1	−0.8	68.4	77.8	93.3
4	+3.0	35.2	41.0	51.6
8	+2.1	25.1	29.6	38.3
16	+1.9	18.0	21.1	27.4
32	+1.5	13.1	15.7	20.6
64	+1.2	9.3	10.8	14.6
128	+1.1	6.3	7.5	9.8
256	+1.0	4.4	5.2	7.1
512	+1.1	3.3	4.0	5.4
1,024	+1.1	2.4	2.9	3.9
2,048	+1.1	1.6	2.0	2.5
4,096	+1.1	1.1	1.4	1.8
8,192	+1.1	0.8	1.0	1.3
16,384	+1.1	0.6	0.7	0.9

118

Figure 11 ($2.50/day line): Households by targeting classification and score, along with "Total Accuracy" and BPAC, scorecard applied to the validation sample

Score	Inclusion: < poverty line correctly targeted	Undercoverage: < poverty line mistakenly non-targeted	Leakage: => poverty line mistakenly targeted	Exclusion: => poverty line correctly non-targeted	Total Accuracy Inclusion + Exclusion	BPAC See text
0–4	0.3	74.5	0.0	25.2	25.5	–99.1
5–9	0.8	74.0	0.0	25.2	26.0	–97.8
10–14	2.5	72.3	0.0	25.2	27.7	–93.2
15–19	7.3	67.6	0.1	25.1	32.3	–80.5
20–24	14.4	60.5	0.3	24.9	39.2	–61.2
25–29	24.2	50.7	0.6	24.6	48.7	–34.6
30–34	35.1	39.7	1.2	23.9	59.1	–4.5
35–39	46.8	28.0	3.0	22.2	69.0	+29.0
40–44	54.3	20.6	5.6	19.5	73.8	+52.5
45–49	63.0	11.8	8.5	16.7	79.7	+79.8
50–54	68.4	6.4	12.9	12.2	80.7	+82.7
55–59	72.3	2.6	17.1	8.1	80.4	+77.2
60–64	73.9	0.9	20.3	4.9	78.8	+72.9
65–69	74.5	0.4	22.9	2.3	76.7	+69.4
70–74	74.8	0.0	24.1	1.1	75.9	+67.8
75–79	74.8	0.0	24.9	0.2	75.1	+66.7
80–84	74.8	0.0	25.1	0.1	74.9	+66.5
85–89	74.8	0.0	25.2	0.0	74.8	+66.4
90–94	74.8	0.0	25.2	0.0	74.8	+66.4
95–100	74.8	0.0	25.2	0.0	74.8	+66.4

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

Figure 12 ($2.50/day line): For a given score cut-off, the percentage of all households who are targeted (that is, have a score equal to or less than the cut-off), the percentage of targeted households who are poor (that is, below the poverty line), the percentage of poor households who are targeted, and the number of poor households who are successful targeted (coverage) per non-poor household mistakenly targeted (leakage), scorecard applied to the validation sample

Targeting cut-off	% all households who are targeted	% targeted who are poor	% of poor who are targeted	Poor households targeted per non-poor household targeted
0–4	0.3	100.0	0.5	Only poor targeted
5–9	0.8	100.0	1.1	Only poor targeted
10–14	2.5	100.0	3.4	Only poor targeted
15–19	7.3	98.9	9.7	93.7:1
20–24	14.7	98.0	19.2	48.7:1
25–29	24.8	97.6	32.3	41.2:1
30–34	36.4	96.6	47.0	28.6:1
35–39	49.8	94.1	62.5	15.8:1
40–44	59.9	90.6	72.5	9.7:1
45–49	71.5	88.2	84.2	7.4:1
50–54	81.4	84.1	91.4	5.3:1
55–59	89.4	80.9	96.6	4.2:1
60–64	94.2	78.5	98.8	3.6:1
65–69	97.3	76.5	99.5	3.3:1
70–74	98.9	75.6	100.0	3.1:1
75–79	99.8	75.0	100.0	3.0:1
80–84	99.9	74.9	100.0	3.0:1
85–89	100.0	74.8	100.0	3.0:1
90–94	100.0	74.8	100.0	3.0:1
95–100	100.0	74.8	100.0	3.0:1