

Simple Poverty Scorecard[®] Poverty-Assessment Tool Palestine (West Bank and Gaza Strip)

Mark Schreiner

8 July 2010

This document is at SimplePovertyScorecard.com.

Abstract

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

Acknowledgements

This paper is funded by the United Nations Relief and Works Agency for Palestine Refugees in the Near East via a contract with Microfinanza SRL. It updates earlier work that used data from 2005. Data come from the Palestine Central Bureau of Statistics. Thanks go to Malika Anand, Nigel Biggar, Frank DeGiovanni, Micol Guarneri, Fabio Malanchini, Alex Pollock, Tony Sheldon, Lucia Spaggiari, Jeff Toohig, and Elsa Valli. "Simple Poverty Scorecard" is a Registered Trademark of Microfinance Risk Management, L.L.C. for its brand of poverty-assessment tools. Copyright © 2010 Microfinance Risk Management.

Simple Poverty Scorecard® Poverty-Assessment Tool

Interview ID: _____	<u>Name</u>	<u>Identifier</u>
Interview date: _____	Participant: _____	_____
Country: _____ PSE	Field agent: _____	_____
Scorecard: _____ 002	Service point: _____	_____
Sampling wgt.: _____	Number of household members: _____	

Indicator	Value	Points	Score
1. How many members does the household have?	A. Ten or more	0	
	B. Nine	1	
	C. Eight	6	
	D. Five, six, or seven	9	
	E. Four	21	
	F. One, two, or three	32	
2. What is the main building material used in the exterior walls of the residence?	A. Cement cob, mud, or other	0	
	B. Cleaned stone, stone and cement, old stone, or concrete	8	
3. Does the household have a solar water heater?	A. No	0	
	B. Yes	2	
4. Does the household have a vacuum cleaner?	A. No	0	
	B. Yes	10	
5. Does the household have a land-line and/or cellular telephone?	A. None	0	
	B. Only cellular	7	
	C. Land-line (regardless of cellular)	9	
6. Does the household have a T.V. and a VCR and/or DVD?	A. None	0	
	B. Only T.V.	2	
	C. T.V. and VCR and/or DVD	6	
7. Does the household have a satellite dish?	A. No	0	
	B. Yes	5	
8. Does the household have a computer?	A. No	0	
	B. Yes	10	
9. Does the household have a bookcase?	A. No	0	
	B. Yes	7	
10. How many household members are employed?	A. None	0	
	B. One	5	
	C. Two or more	11	

SimplePovertyScorecard.com

Score:

Simple Poverty Scorecard[®] Poverty-Assessment Tool Palestine (West Bank and Gaza Strip)

1. Introduction

Pro-poor programs in Palestine¹ (West Bank and Gaza Strip) can use the Simple Poverty Scorecard poverty-assessment tool to estimate the likelihood that a household has consumption below a given poverty line, to estimate a population's poverty rate at a point in time, to track changes in a population's poverty rate over time, and to segment participants for differentiated treatment.

The direct approach to poverty measurement via surveys is difficult and costly. For example, the 2007 Palestine Expenditure and Consumption Survey (PECS) runs more than 40 pages. Enumerators visit households weekly for four weeks, applying a consumption module with hundreds of questions, such as: “In the past week, how much long-grain rice did the household eat? How much was this rice worth? Now then, in the past week, how much short-grain rice did the household eat? . . .”

In contrast, the indirect approach via the scorecard is simple, quick, and inexpensive. It uses ten verifiable indicators (such as “What is the main building material used in the exterior walls of the residence?” and “Does the household have a solar water heater?”) to get a score that is highly correlated with poverty status as measured by consumption from the lengthy survey.

¹ *Palestine* is used here as shorthand for the “occupied Palestinian territories” or the “West Bank and Gaza Strip”.

The scorecard differs from “proxy means tests” (Coady, Grosh, and Hoddinott, 2002) in that it is tailored to the capabilities and purposes not of national governments but rather of local, pro-poor organizations. The feasible poverty-measurement options for these local organizations are typically subjective and relative (such as participatory wealth ranking by skilled field workers) or blunt (such as rules based on land-ownership or housing quality). These approaches may be costly, their results are not comparable across organizations nor across countries, and their accuracy and precision are unknown.

Suppose an organization wants to know what share of its participants are below a poverty line, perhaps because it wants to relate their poverty status to the Millennium Development Goals’ \$1.25/day poverty line at 2005 purchase-power parity (PPP). Or an organization might want to report (as required of USAID microenterprise partners) how many of its participants are among the poorest half of people below the national poverty line. Or an organization might want to measure movement across a poverty line (see, for example, Daley-Harris, 2009). In these cases, what is needed is a consumption-based, objective tool with known accuracy that can serve for monitoring, management, and/or targeting. While consumption surveys are costly even for governments, many small, local organizations can afford to implement a simple, inexpensive poverty-assessment tool.

The statistical approach here aims to be understood by non-specialists. After all, if managers are to adopt the scorecard on their own and apply it to inform their

decisions, they must first trust that it works. Transparency and simplicity build trust. Getting “buy-in” matters; proxy means tests and regressions on the “determinants of poverty” have been around for three decades, but they are rarely used to inform decisions by local pro-poor organizations. This is not because these tools do not work, but because they often have complex indicators and are presented (when they are presented at all) as tables of regression coefficients incomprehensible to non-specialists (with indicator names such as “LGHHSZ_2”, negative points, and points with 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 the accuracy tests are simple and standard in statistical practice and in the for-profit field of credit-risk scoring, they have rarely been applied to poverty-assessment tools.

The scorecard is based on the 2007 PECS conducted by the Palestine Central Bureau of Statistics (PCBS). Indicators for the scorecard 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 zeroes or positive 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 estimate the poverty rate of a group of households at a point in time. This is simply the average poverty likelihood among the households in the group.

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

The scorecard can also be used for targeting services to poorer households. To help managers select an appropriate targeting cut-off, 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 Palestine's national poverty line. Scores from this scorecard are calibrated to poverty likelihoods for seven poverty lines.

The scorecard is constructed and calibrated using two sub-samples from the 2007 PECS, and its accuracy is validated on a third sub-sample. While all three scoring

estimators are unbiased when applied to the population from which they were derived (that is, they match the true value on average in repeated samples from the same population from which the scorecard is built), they are—like all predictive models—biased to some extent when applied to a different population.²

Thus, while the indirect scoring approach is less costly than the direct survey approach, it is also biased in practice. (The direct survey approach is unbiased by definition.) There is bias because scoring must assume that the future relationships between indicators and poverty will be the same as in the data used to build the scorecard. It must also assume that these relationships will be the same in all sub-groups as in the population as a whole.³ Of course, these assumptions—ubiquitous and inevitable in predictive modeling—hold only partly. In particular, Palestine is regularly subject to large political and economic shocks. The scorecard cannot predict changes due to these shocks; it must assume that the situation in Palestine and the relationships between indicators and poverty after 2007 are the same as in 2007.

When applied to the validation sample for Palestine with the national poverty line and $n = 16,384$, the difference between scorecard estimates of groups' poverty rates and the true rates at a point in time is +0.4 percentage points. Across all seven lines,

² Examples of “different populations” include nationally representative samples at another point in time or non-representative sub-groups (Tarozzi and Deaton, 2007).

³ Bias may also result from changes over time in the quality of data, from changes in the real value of poverty lines, from imperfect adjustment of poverty lines to account for differences in cost-of-living, or from sampling variation across surveys.

the average absolute difference is 1.1 percentage points, and the maximum absolute difference is 3.0 percentage points.

Because the validation sample is representative of the same population as the data that is used to construct the scorecard and because all the data come from the same time period, the scorecard estimators are unbiased and these observed differences are due to sampling variation; the average difference would be zero if the 2007 PECS were to be repeatedly redrawn and divided into sub-samples before repeating the entire scorecard-building and accuracy-testing process.

For $n = 16,384$, the 90-percent confidence intervals for these estimates are ± 0.5 percentage points or less. For $n = 1,024$, these intervals are ± 2.2 percentage points or less.

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

2. Data and poverty lines

This section discusses the data used to construct and validate the scorecard. It also defines the measure of consumption, as well as the poverty lines to which scores are calibrated.

2.1 Data

The scorecard is based on data from the 1,231 households in the 2007 PECS. This is the best, most recent national consumption survey available for Palestine. For scoring, the data are further divided into three sub-samples (Figure 2):

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

Among all the surveys that have been used with the poverty-scoring approach in this paper, the 2007 PECS has the fewest households. This means that the scorecard may not generalize as well in future applications. Beyond using a different approach, the only remedy is to gather more data or pool the 2006 and 2007 PECS.

2.2 Household consumption and expenditure

Household consumption in the PECS is the sum of (PCBS, 2006):

Bread and cereals	Furniture and utensils
Meat and poultry	Household operations
Fish and seafood	Medical care
Dairy products and eggs	Transport and communication
Oils and fats	Education
Fruits and nuts	Recreation
Vegetables, legumes, and tubers	Personal care
Sugar and confectionery	Tobacco
Non-alcoholic beverages	Alcoholic beverages
Salt, spices, and other food	Other non-food consumption
Take-away and restaurant meals	Imputed rent
Self-produced food in-kind	
Clothing and footwear	
Housing	

PCBS also defines *household expenditure* as household consumption above, plus:

Remittances in cash
Taxes in cash
Other cash non-consumption expenditure
Social security

minus

Own produced food in kind
Other than food
Imputed rent

Which definition should be used to measure poverty? According to Deaton and

Zaidi (2002), a measure of household economic well-being should include:

- Food consumption
- Non-food consumption
- Imputed housing rent
- Imputed use-value of durable goods
- Education expenditures
- Expenditures on water and electricity

This ideal measure should exclude:

- Taxes (unless they are “fees for services” that directly benefit the household)
- Debt repayment
- Interest payments on debt
- Large and infrequent expenses such as marriages and dowries
- Health expenditures
- Gifts, transfers, and remittances set out

Neither consumption nor expenditure in the 2007 PECS matches the guidelines in Deaton and Zaidi (2002). Both measures exclude the imputed use-value of consumer durables, both exclude expenditure on water and electricity, and both include health expenditures. In addition, the PECS expenditure measure excludes imputed rent and the value of own-produced food in-kind but includes taxes and remittances sent out.

On net, the consumption measure in the 2007 PECS is closer to the Deaton/Zaidi standard than the expenditure measure. Furthermore, other analyses of poverty in Palestine are based on consumption (UNRWA, 2009 and 2008; PCBS, 2008; World Bank, 2004). Thus, this paper uses the PCBS measure of consumption.

2.3 Poverty rates and poverty lines

2.3.1 Rates

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

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 for the group 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 for the group of $2 \div (1 + 2) = 67$ percent.

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

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

The scorecard is constructed using Palestine’s 2007 PECS and household-level lines. Scores are calibrated to household-level poverty likelihoods, and accuracy is measured for household-level rates. This use of the household level reflects the belief that it is the most relevant for most pro-poor organizations.

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

2.3.2 Poverty lines

Figure 2 shows seven poverty lines for Palestine. It also reports poverty rates (for both households and people) for all of Palestine and separately for the West Bank and the Gaza Strip. Finally, Figure 2 shows household-level poverty rates for the three subsamples used in scorecard construction, calibration, and validation.

Palestine’s subsistence poverty line is defined as the 30th percentile of household expenditure on food, clothing, and housing in 1995, adjusted for changes in the prices since then (World Bank, 2004). In 2007, this is 10.33 New Israeli Shekels per person per day.⁴ This official subsistence line applies to the whole country and leads to poverty rates of 19.7 percent (households) and 25.9 percent (people). Poverty rates are much higher in the Gaza Strip than in the West Bank.

⁴ All sources present monthly poverty lines for a household of six. The daily per-capita lines here are divided by six and by the average days in a month.

Palestine’s national poverty line (sometimes referred to in the figures as “100% of the national line”) is defined like the subsistence line, but it considers nine categories of expenditure instead of three. The national line for Palestine is NIS12.92 per person per day, giving poverty rates of 30.0 percent (household) and 37.4 percent (people). Again, the poverty rate is much higher in the Gaza Strip than in the West Bank.

The national line is used here to construct the scorecard. Because pro-poor organizations may want to use different or various poverty lines, this paper calibrates scores from its single scorecard to poverty likelihoods for seven lines:

- National
- Subsistence
- 150% of national
- 200% of national
- USAID “extreme”
- USD3.75/day 2005 PPP
- USD5.00/day 2005 PPP

The USAID “extreme” line is defined as the median consumption of people (not households) below the national line (U.S. Congress, 2002). For Palestine, it is NIS8.91 per person per day.

The \$3.75/day 2005 PPP line is derived using the 2005 PPP exchange rate for “individual consumption expenditure by households” (NIS2.31 per USD1.00, see World Bank, 2008) and the average all-Palestine Consumer Price Index for 2005 and 2007 (146.79 and 156.41).⁵ The \$3.75/day 2005 PPP line in 2007 is then (Sillers, 2006):

$$3 \cdot (\text{2005 PPP exchange rate}) \cdot \text{USD1.25} \cdot \frac{\text{CPI}_{\text{Ave. 2007}}}{\text{CPI}_{\text{Ave. 2005}}} =$$

$$3 \cdot \left(\frac{\text{NIS2.31}}{\text{USD1.00}} \right) \cdot \text{USD1.25} \cdot \frac{156.41}{146.79} = \text{NIS9.23}.$$

The \$5.00/day 2005 PPP line is a multiple of the \$3.75/day 2005 PPP line. This paper does not present lower 2005 PPP poverty lines (such as \$1.25/day and \$2.50/day) because very few households in Palestine are below them.

⁵ http://www.pcbs.gov.ps/Portals/_pcbs/cpi/bfed52be-8c22-47f7-b3d9-d79ba134d37c.htm and http://www.pcbs.gov.ps/Portals/_pcbs/cpi/551ee164-91d4-436e-b789-1535eca17163.htm, retrieved 31 December 2009

3. Scorecard construction

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

- Family composition (such as household size)
- Education (such as school attendance of children)
- Employment (such as the number of household members who are employed)
- Housing (such as the main material of the exterior walls)
- Ownership of durable goods (such as bookcases and computers)

Figure 3 lists all the candidate indicators, ranked by the entropy-based “uncertainty coefficient” that is a measure of how well the indicator predicts poverty on its own (Goodman and Kruskal, 1979). Responses for each indicator are ordered starting with those most strongly linked with higher poverty likelihoods.

The scorecard also aims to measure *changes* in poverty through time. This means that, when selecting indicators and holding other considerations constant, preference is given to more sensitive indicators. For example, ownership of a computer 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 national poverty line and Logit regression on the construction sub-sample. Indicator selection uses both judgment and statistics (forward stepwise, based on “c”). The first step is to use Logit to build one scorecard for each candidate indicator. Each scorecard’s accuracy is taken as “c”, a measure of its ability to rank by poverty status (SAS Institute Inc., 2004).

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

A series of two-indicator scorecards are then built, each based on the one-indicator scorecard selected from the first step, now 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.

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

The 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).

The single scorecard here applies to all of Palestine. Tests for Mexico and India (Schreiner, 2006a and 2006b), Sri Lanka (Narayan and Yoshida, 2005), and Jamaica (Grosh and Baker, 1995) suggest that segmenting poverty-assessment tools by urban/rural does not improve targeting much, although such segmentation may improve the accuracy of estimated poverty rates (Tarozzi and Deaton, 2007).

4. Practical guidelines for scorecard use

The main challenge of scorecard design is not to squeeze out the last drops of accuracy but rather to improve the chances that scoring is actually used (Schreiner, 2005). When scoring projects fail, the reason is not usually technical inaccuracy but rather the failure of an organization to decide to do what is needed to integrate scoring in its processes and to learn to use it properly (Schreiner, 2002). After all, most reasonable scorecards predict tolerably well, thanks to the empirical phenomenon known as the “flat 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 want to adopt it and use it properly. Of course, accuracy is important, but so are simplicity, ease-of-use, and “face validity”. Programs are more likely to collect data, compute scores, and pay attention to the results if, in their view, scoring does not make a lot of “extra” work and if the whole process generally seems to make sense.

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

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

A field worker using the paper scorecard would:

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

4.1 Quality control

Of course, field workers must be trained. High-quality outputs require high-quality inputs. If organizations or field workers gather their own data and if they 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 random audits (Matul and Kline, 2003).⁶ IRIS Center (2007a) and Toohig (2008) are useful nuts-and-bolts guides for planning, budgeting, training field

⁶ If an organization does not want field workers to know the points associated with indicators, then they can use the version of Figure 1 without points and apply them later at the central office.

workers and supervisors, logistics, sampling, interviewing, piloting, recording data, and controlling quality.

In particular, while collecting scorecard indicators is relatively easier than most alternatives, it is still absolutely difficult. Training and explicit definitions of the terms and concepts in the scorecard is essential. For the example of Nigeria, one study finds distressingly low inter-rater and test-retest correlations for indicators as seemingly simple and obvious as whether the household owns an automobile (Onwujekwe, Hanson, and Fox-Rushby, 2006).

As an example from Mexico, Martinelli and Parker (2007) find that in the first stage of targeting a conditional cash-transfer program, “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 Mexico does in the second stage of its targeting process—field agents can verify responses with a home visit and correct false reports, and this same procedure is suggested for the scorecard as well.

4.2 Implementation and sampling

In terms of implementation and sample design, an organization must make choices about:

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

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

- Employees of the organization
- Third-party contractors

Responses, scores, and poverty likelihoods can be recorded:

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

The subjects to be scored can be:

- All participants
- A representative sample of all participants
- All participants in a representative sample of branches
- A representative sample of all participants in a representative sample of branches
- A non-representative sub-group of interest to management for a particular question

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 confidence level and a desired confidence interval.

Frequency of application can be:

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

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

- With different sets of participants, with each set representative of all participants
- With a single set of participants

An example set of choices for implementation and design is provided by BRAC and ASA, two microlenders in Bangladesh (each with more than 7 million participants) who are applying the scorecard (Schreiner, 2013). Their design is that loan officers in a random sample of branches apply the scorecard to their clients 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 and scored. The sampling plans of ASA and BRAC cover 50,000–100,000 participants each, which is far more than would be required to inform most relevant questions at a typical pro-poor organization.

5. Estimates of household poverty likelihoods

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

To get absolute units, scores must be converted to *poverty likelihoods*, that is, probabilities of being below a poverty line. This is done via simple look-up tables. For the example of the national line with the 2007 PECS, scores of 30–34 have a poverty likelihood of 43.1 percent, and scores of 35–39 have a poverty likelihood of 41.6 percent (Figure 4).

Naturally, the poverty likelihood associated with a score varies by poverty line. For example, scores of 30–34 are associated with a poverty likelihood of 43.1 percent for the national line but 28.5 percent for the subsistence line.⁷

5.1 Calibrating scores with poverty likelihoods

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

⁷ Starting with Figure 4, many figures have seven versions, one for each of the seven poverty lines. The tables are grouped by poverty line. Single tables that pertain to all poverty lines are placed with the tables for the national line.

For the example of the national line (Figure 5), there are 11,615 (normalized) households in the calibration sub-sample with a score of 30–34, of whom 5,002 (normalized) are below the poverty line. The estimated poverty likelihood associated with a score of 30–34 is then 43.1 percent, as $5,002 \div 11,615 = 0.431$.

To illustrate further with the national line and a score of 35–39, there are 13,022 (normalized) households in the calibration sample, of whom 5,417 (normalized) are below the line (Figure 5). Thus, the poverty likelihood for this score is $5,417 \div 13,022 = 41.6$ percent.

The same 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 daily consumption of someone with a score of 30–34 falls in the following ranges with probability:

- 22.7 percent below the USAID “extreme” line
- 0.0 percent between the USAID “extreme” and the \$3.75/day 2005 PPP lines
- 5.9 percent between the \$3.75/day 2005 PPP and the subsistence lines
- 11.1 percent between the subsistence and the \$5.00/day 2005 PPP lines
- 3.4 percent between the \$5.00/day 2005 PPP and the national lines
- 38.9 percent between the national and 150% of the national lines
- 8.3 percent between 150% of the national and 200% of the national lines
- 9.7 percent above 200% of the national line

Even though the scorecard is constructed partly based on judgment, this calibration process produces poverty likelihoods that are objective, that is, derived from quantitative poverty lines and survey data on consumption. The poverty likelihoods would be objective even if indicators and/or points were selected without any data at

all. In fact, scorecards with objective poverty likelihoods of proven accuracy are often constructed using only 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 Palestine’s scorecard are transformed coefficients from a Logit regression, scores are not converted to poverty likelihoods via the Logit formula of $2.718281828^{\text{score}} \times (1 + 2.718281828^{\text{score}})^{-1}$. This is because the Logit formula is esoteric and difficult to compute by hand. Non-specialists find it more intuitive to define the poverty likelihood as the share of households with a given score in the calibration sample who are below a poverty line. In the field, converting scores to poverty likelihoods requires no arithmetic at all, just a look-up table. This non-parametric calibration can also improve accuracy, especially with large calibration samples.

5.2 Accuracy of estimates of households’ poverty likelihoods

As long as the relationship between indicators and poverty does not change and as long as the scorecard is applied to households who are representative of the same population from which the scorecard was constructed, this calibration process produces unbiased estimates of poverty likelihoods. *Unbiased* means that in repeated samples

from the same population, the average estimate matches the true poverty likelihood.

The scorecard also produces unbiased estimates of poverty rates at a point in time, as well as unbiased estimates of changes in poverty rates between two points in time.⁸

But the relationship between indicators and poverty does change with time and also across sub-groups in Palestine's population, so the scorecard is generally biased when applied after the end date of fieldwork for the 2007 PECS (as it must necessarily be applied in practice) or when applied with non-nationally representative groups (as it probably would be applied by local, pro-poor organizations).

How accurate are estimates of households' poverty likelihoods, given the assumption of representativeness? To check, 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 who have consumption below a poverty line
- For each score range, 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 range, report the average difference between estimated and true poverty likelihoods across the 1,000 bootstrap samples
- For each score range, report the two-sided interval containing the central 900, 950, or 990 differences between estimated and true poverty likelihoods

⁸ This follows because these estimates of groups' poverty rates are linear functions of the unbiased estimates of households' poverty likelihoods.

For each score range and for $n = 16,384$, Figure 7 shows the average difference between estimated and true poverty likelihoods as well as confidence intervals for the differences. For the national line in the validation sample, the average poverty likelihood across bootstrap samples for scores of 30–34 is too low by 6.0 percentage points. For scores of 35–39, the estimate is too high by 10.9 percentage points.⁹

The 90-percent confidence interval for the differences for scores of 30–34 is ± 4.0 percentage points (Figure 7). This means that in 900 of 1,000 bootstraps, the difference between the estimate and the true value is between -10.0 and -2.0 percentage points (because $-6.0 - 4.0 = -10.0$, and $-6.0 + 4.0 = -2.0$). In 950 of 1,000 bootstraps (95 percent), the difference is -6.0 ± 4.2 percentage points, and in 990 of 1,000 bootstraps (99 percent), the difference is -6.0 ± 4.4 percentage points.

For many 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 Palestine’s population. For targeting, however, what matters is less the differences across all score ranges and more the differences in score ranges just above and below the targeting cut-off. This mitigates the

⁹ These differences are not zero, despite 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 construction and calibration process.

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 balance out. This is generally the case, as discussed in the next section.

Another possible source of bias is overfitting. By construction, the scorecard here is unbiased, but it may still be *overfit* when applied after the 14 January 2008 end of field work for the 2007 PECS. That is, the scorecard may fit the data from the 2007 PECS so closely that it captures not only real patterns but also some random patterns that, due to sampling variation, show up only in the 2007 PECS.¹⁰ Or the scorecard may be overfit in the sense that it is not robust to changes through time in the relationships between indicators and poverty. Finally, the scorecard could also be overfit when it is applied to samples from non-nationally representative sub-groups.

Overfitting can be mitigated by simplifying the scorecard and by not relying only on the 2007 PECS data but rather also considering experience, judgment, and theory. Of course, the scorecard here does just that. Bootstrapping scorecard construction—which is not done here—can also mitigate overfitting by reducing (but not eliminating) dependence on a single sampling instance. Combining scorecards can also help, at the cost of complexity.

¹⁰ Because the 2007 PECS covers so few households, sampling variation is higher than is typical, so overfitting is more likely.

In any case, most errors in individual households' likelihoods balance out in the estimates of groups' poverty rates (see later sections). Furthermore, much of the differences between scorecard estimates and true values may come from non-scorecard sources such as changes in the relationship between indicators and poverty, sampling variation, changes in poverty lines, inconsistencies in data quality across time, and inconsistencies/imperfections in cost-of-living adjustments. These factors can be addressed only by improving data quantity and quality (which is beyond the scope of the scorecard), by updating data, 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, 2010 and that they have scores of 20, 30, and 40, corresponding to poverty likelihoods of 82.2, 43.1, and 30.2 percent (national line, Figure 4). The group's estimated poverty rate is the households' average poverty likelihood of $(82.2 + 43.1 + 30.2) \div 3 = 51.8$ percent.¹¹

6.1 Accuracy of estimated poverty rates at a point in time

How accurate is this estimate? For a range of sample sizes, Figure 9 reports average differences between estimated and true poverty rates as well as precision (confidence intervals for the differences) for the Palestine scorecard applied to 1,000 bootstrap samples from the validation sample.

Summarizing Figure 9 across poverty lines and years for $n = 16,384$, Figure 8 shows that the absolute differences between the estimated poverty rate and the true rate for the scorecard applied to the validation sample are +3.0 percentage points or less. The average absolute difference across the seven poverty lines is 1.1 percentage points.

¹¹ The group's poverty rate is *not* the poverty likelihood associated with the average score. Here, the average score is $(20 + 30 + 40) \div 3 = 30$, and the poverty likelihood associated with the average score is 43.1 percent. This is not the 51.8 percent found as the average of the three poverty likelihoods associated with each of the three scores.

In terms of precision, the 90-percent confidence interval for a group's estimated poverty rate at a point in time in 2007 with $n = 16,384$ and for all poverty lines is ± 0.5 percentage points or less (Figure 8). This means that in 900 of 1,000 bootstraps of this size, the absolute difference between the estimate and the average estimate is 0.5 percentage points or less.

In the specific case of the national line, 90 percent of all samples of $n = 16,384$ produce estimates that differ from the true value in the range of $+0.4 - 0.5 = -0.1$ to $+0.4 + 0.5 = +0.9$ percentage points. This is because $+0.4$ is the average difference and ± 0.5 is its 90-percent confidence interval. The average difference is $+0.4$ because the average scorecard estimate is too high by 0.4 percentage points; the scorecard tends to estimate a poverty rate of 30.5 percent for the validation sample, but the true value is 30.1 percent (Figure 2). Future accuracy will depend on how closely the period of application resembles 2007.

6.2 Standard-error formula for estimates of poverty rates at a point in time

How precise are the point-in-time estimates? Because they are averages, the estimates have a Normal distribution and can be characterized by their average difference vis-à-vis true values, along 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 for indirect measurement via poverty-assessment tools (Schreiner, 2008a), note that the textbook formula (Cochran, 1977) that relates confidence intervals with standard errors in the case of direct measurement of poverty rates is $c = +/- z \cdot \sigma$, where:

c is a confidence interval as a proportion (*e.g.*, 0.02 for ± 2 percentage points),

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

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

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

n is the sample size.

For example, with a sample $n = 16,384$, 90-percent confidence ($z = 1.64$), and a poverty rate p of 30.1 percent (the true rate in the validation sample for the national line in Figure 2), the confidence interval c is

$$+/- z \cdot \sqrt{\frac{p \cdot (1 - p)}{n}} = +/- 1.64 \cdot \sqrt{\frac{0.301 \cdot (1 - 0.301)}{16,384}} = \pm 0.588 \text{ percentage points.}$$

The scorecard, however, does not measure poverty directly, so this formula is not applicable. To derive a formula for the Palestine scorecard, consider Figure 9, which reports empirical confidence intervals c for the differences for the scorecard applied to 1,000 bootstrap samples of various sample sizes from the validation sample. For $n =$

16,384, the national line, and the validation sub-sample, the 90-percent confidence interval is ± 0.530 percentage points.¹² Thus, the ratio of confidence intervals with the scorecard versus direct measurement is $0.530 \div 0.588 = 0.90$.

Now consider the same case, but with $n = 8,192$. The confidence interval under direct measurement is $+/- 1.64 \cdot \sqrt{\frac{0.301 \cdot (1 - 0.301)}{8,192}} = \pm 0.831$ percentage points. The empirical confidence interval with the Palestine scorecard for the national line (Figure 9) is ± 0.735 percentage points. Thus for $n = 8,192$, the ratio is $0.735 \div 0.831 = 0.88$.

This ratio of 0.88 for $n = 8,192$ is close to the ratio of 0.90 for $n = 16,384$.

Indeed, across all sample sizes of 256 or more in Figure 9, the average ratio turns out to be 0.88, implying that confidence intervals for indirect estimates of poverty rates via the Palestine scorecard and the national poverty line are about 12-percent narrower than those for direct estimates. This 0.88 appears in Figure 8 as the “ α factor” because if $\alpha = 0.88$, then the formula relating confidence intervals c and standard errors for the Palestine scorecard is $c = +/- z \cdot \alpha \cdot \sigma$. The standard error for point-in-time estimates of

poverty rates via scoring is $\alpha \cdot \sqrt{\frac{p \cdot (1 - p)}{n}}$.

In general, α could 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 in all seven cases in Figure 8.

¹² Due to rounding, Figure 9 displays 0.5, not 0.530.

The formula relating confidence intervals to standard errors for the scorecard can be rearranged to give a formula for determining sample size n before measurement.¹³ If \hat{p} is the expected poverty rate before measurement, then the formula for n based on the desired confidence level that corresponds to z and the desired confidence interval $\pm c$ under the scorecard is $n = \left(\frac{\alpha \cdot z}{c}\right)^2 \cdot \hat{p} \cdot (1 - \hat{p})$.

To illustrate how to use this, suppose $c = 0.04000$ and $z = 1.64$ (90-percent confidence), and $\hat{p} = 0.2985$ (the average poverty rate for the national line in the construction and calibration sub-samples, Figure 2). Then the formula gives

$$n = \left(\frac{0.88 \cdot 1.64}{0.04000}\right)^2 \cdot 0.2985 \cdot (1 - 0.2985) = 273, \text{ not far from the sample size of 256}$$

observed for these parameters in Figure 9.

Of course, the α factors in Figure 8 are specific to Palestine, its poverty lines, its poverty rates, and this scorecard. The method for deriving standard errors, however, is valid for any poverty-assessment tool following the approach in this paper.

In practice after the end of the PECS field work in mid-January 2008, an organization would select a poverty line (say, the national line), select a desired confidence level (say, 90 percent, or $z = 1.64$), select a desired confidence interval (say,

¹³ IRIS Center (2007a and 2007b) says that a sample size of $n = 300$ is sufficient for reporting estimated poverty rates to USAID. If a poverty-assessment tool is as precise as direct measurement, if the expected (before measurement) poverty rate is 50 percent, and if the confidence level is 90 percent, then $n = 300$ implies a confidence interval of ± 2.2 percentage points. In fact, USAID has not specified confidence levels or intervals. Furthermore, the expected poverty rate may not be 50 percent, and the poverty-assessment tool could be more or less precise than direct measurement.

± 2.0 percentage points, or $c = 0.02$), make an assumption about \hat{p} (perhaps based on a previous measurement such as the 30.0-percent average for the national line in the 2007 PECS in Figure 2), look up α (here, 0.88), assume that the scorecard is still valid in the future and/or for non-nationally representative sub-groups,¹⁴ and then compute the required sample size. In this illustration, $n = \left(\frac{0.88 \cdot 1.64}{0.02} \right)^2 \cdot 0.300 \cdot (1 - 0.300) = 1,094$.

¹⁴ This paper reports accuracy for the scorecard applied to the validation sample, but it cannot test accuracy for later years or other groups. Performance will deteriorate with time to the extent that the relationship between indicators and poverty changes.

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 PECS, this paper cannot test estimates of change over time for Palestine, 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 measure change through time.

7.1 Warning: Change is not impact

Scoring can estimate change. Of course, change could be for the better or for the worse, and scoring does not indicate what caused change. This point is often forgotten, confused, or ignored, 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 status 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, 2010, a program samples three households who score 20, 30, and 40 and so have poverty likelihoods of 82.2, 43.1, and 30.2 percent (national line, Figure 4). The group's baseline estimated poverty rate is the households' average poverty likelihood of $(82.2 + 43.1 + 30.2) \div 3 = 51.8$ percent.

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

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

By way of illustration, suppose that a year later on Jan. 1, 2011, 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 now 25, 35, and 45 (poverty likelihoods of 63.1, 41.6, and 8.3 percent, national line, Figure 4). Their average poverty likelihood at follow-up is $(63.1 + 41.6 + 8.3) \div 3 = 37.7$ percent, an improvement of $51.8 - 37.7 = 14.1$ percentage points.¹⁵

This suggests that about one of seven participants crossed above the poverty line in 2010. (This is a net figure; some people start above the line and end below it, and vice versa.) Among those who started below the line, about one in four ($14.1 \div 51.8 =$

¹⁵ Of course, such a huge reduction in poverty is unlikely in a year's time, but this is just an example to show how the scorecard can be used to estimate change.

27.2 percent) ended up above the line. Of course, the scorecard does not reveal the reasons for this change.

7.3 Estimated changes in poverty rates in Palestine

With only the 2007 PECS, 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 Palestine 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.

7.4 Accuracy for estimated change in two independent samples

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

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

z , c , and p are defined as above, n is the sample size at both baseline and follow-up,¹⁶ and α is the average (across a range of sample sizes) of the ratio of the observed

¹⁶ This means that, for a given precision and with direct measurement, estimating the change in a poverty rate over time requires four times as many measurements (not twice as many) as does estimating a poverty rate at a point in time.

bootstrap confidence intervals from a poverty-assessment tool and the theoretical confidence intervals from the textbook formula for direct measurement.

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

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

For the countries for which this α has been measured (Schreiner, 2009a, 2009b, 2009c, 2009d, 2009e, and 2008b; and Chen and Schreiner, 2009), the simple average α across poverty lines, years, and countries is (1.11). This is as reasonable a figure as any to use for Palestine.

To illustrate the use of the formula above to determine sample size for estimating changes in poverty rates across two independent samples, suppose the desired confidence level is 90 percent ($z = 1.64$), the desired confidence interval is 2 percentage points ($c = 0.02$), the poverty line is the national line, $\alpha = 1.11$, and $\hat{p} = 0.300$ (from

Figure 2). Then the baseline sample size is $n = 2 \cdot \left(\frac{1.11 \cdot 1.64}{0.02} \right)^2 \cdot 0.300 \cdot (1 - 0.300) =$

3,480, and the follow-up sample size is also 3,480.

7.5 Accuracy for estimated change for one sample, scored twice

The general formula relating the confidence interval c to the standard error σ when using scoring to estimate change for a single group of households, all of whom are scored at two points in time, is:¹⁷

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

z , c , and α are defined as before, 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.

As usual, the formula for σ can be rearranged to give a formula for sample size n 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 overall change in the poverty rate will be zero, which implies $\hat{p}_{12} = \hat{p}_{21} = \hat{p}_*$, giving:

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

\hat{p}_* could be anything between 0–1, so more information is needed before applying this formula. Suppose that the observed relationship between \hat{p}_* , the number of years y between baseline and follow-up, and $p_{\text{baseline}} \cdot (1 - p_{\text{baseline}})$ is—as in Peru (Schreiner, 2009a)—close to:

¹⁷ See McNemar (1947) and Johnson (2007). John Pezzullo helped find this formula.

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

Given this, a sample-size formula for a group of households to whom the Palestine scorecard is applied twice (once after the end of field work for the 2007 PECS and then again later) is:

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

In Peru (the only other country for which there is a data-based estimate, Schreiner 2009a), the average α across years and poverty lines is about 1.30.

To illustrate the use of this formula, suppose the desired confidence level is 90 percent ($z = 1.64$), the desired confidence interval is 2.0 percentage points ($c = 0.02$), the poverty line is the national line, and the sample will be scored first in 2010 and then again in 2013 ($y = 3$). The before-baseline poverty rate is 30.0 percent ($p_{2007} = 0.300$, Figure 2), and suppose $\alpha = 1.30$. Then the baseline sample size is

$$n = 2 \cdot \left(\frac{1.30 \cdot 1.64}{0.02} \right)^2 \cdot \{ -0.02 + 0.016 \cdot 3 + 0.47 \cdot [0.300 \cdot (1 - 0.300)] \} = 2,880. \text{ The same}$$

group of 2,880 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 by cut-off; a higher cut-off has better inclusion (but worse leakage), while a lower cut-off has better exclusion (but worse undercoverage).

A program 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 by targeting outcome. For an example cut-off of 34 or less and the scorecard applied to the validation sample, outcomes for the national line are:

- Inclusion: 19.9 percent are below the line and correctly targeted
- Undercoverage: 10.2 percent are below the line and mistakenly not targeted
- Leakage: 11.0 percent are above the line and mistakenly targeted
- Exclusion: 58.9 percent are above the line and correctly not targeted

Increasing the cut-off to 39 or less improves inclusion and undercoverage but worsens leakage and exclusion:

- Inclusion: 23.6 percent are below the line and correctly targeted
- Undercoverage: 6.5 percent are below the line and mistakenly not targeted
- Leakage: 20.4 percent are above the line and mistakenly targeted
- Exclusion: 49.6 percent are above the line and correctly not targeted

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

$$\begin{array}{rcl}
 (\text{Benefit per household correctly included} & \times & \text{Households correctly included}) & - \\
 (\text{Cost per household mistakenly not covered} & \times & \text{Households mistakenly not covered}) & - \\
 (\text{Cost per household mistakenly leaked} & \times & \text{Households mistakenly leaked}) & + \\
 (\text{Benefit per household correctly excluded} & \times & \text{Households correctly excluded}). &
 \end{array}$$

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. Any program that uses targeting—with or without scoring—should thoughtfully consider

how it values successful inclusion or exclusion versus errors of undercoverage and leakage. It is healthy to go through a process of thinking explicitly and intentionally about how possible targeting outcomes are valued.

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

$$\begin{array}{rclcl}
 \text{Total Accuracy} = & 1 & \times & \text{Households correctly included} & - \\
 & 0 & \times & \text{Households mistakenly undercovered} & - \\
 & 0 & \times & \text{Households mistakenly leaked} & + \\
 & 1 & \times & \text{Households correctly excluded.} &
 \end{array}$$

Figure 11 shows “Total Accuracy” for all cut-offs for Palestine’s scorecard. For the national line in the validation sample, total net benefit is greatest (78.8) for a cut-off of 34 or less, with about four in five Palestinian households correctly classified.

“Total Accuracy” weighs successful inclusion of households below the line the same as successful exclusion of households above the line. If a program valued inclusion more (say, twice as much) than exclusion, it could reflect this by setting the benefit for inclusion to 2 and the benefit for exclusion to 1. Then the chosen cut-off would maximize $(2 \times \text{Households correctly included}) + (1 \times \text{Households correctly excluded})$.¹⁸

As an alternative to assigning benefits and costs to targeting outcomes and then choosing a cut-off to maximize total net benefit, a program could set a cut-off to achieve a desired poverty rate among targeted households. The third column of Figure

¹⁸ Figure 11 also reports the Balanced Poverty Accuracy Criteria, discussed later.

12 (“% targeted who are poor”) shows the expected poverty rate among Palestinian households who score at or below a given cut-off. For the example of the national line and the validation sample, targeting households who score 34 or less would target 30.9 percent of all households (second column) and lead to a poverty rate among those targeted of 64.3 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 national line and the validation sample with a cut-off of 34 or less, 66.6 percent of all poor households are covered.

The final targeting measure in Figure 12 is the number of successfully targeted poor households for each non-poor household mistakenly targeted (right-most column). For the national line, the validation sample, and a cut-off of 34 or less, covering 1.8 poor households means leaking to 1 non-poor household.

9. Context of poverty-assessment tools in Palestine

This section discusses two existing Palestine poverty-assessment tools in terms of their goals, data, methods, poverty lines, cost, indicators, accuracy, and precision. The relative strengths of the new scorecard here are that it uses the latest nationally representative data, it uses simpler and fewer indicators, it costs less to implement, it reports accuracy and precision out-of-sample, and it provides formulas for standard errors.

9.1 Astrup and Dessus

Astrup and Dessus (2005) use Palestine’s 1997 census and pooled data from the 1996, 1997, and 1998 PECS to construct poverty-assessment tools that feed into a “poverty map” (Elbers, Lanjouw, and Lanjouw, 2003; Hentschel *et al.*, 2000) that estimates poverty rates for small areas (down to 5,000 people) more precisely than direct estimates based on the PECS.

Astrup and Dessus use Logit to build two poverty-assessment tools (one for West Bank, one for Gaza Strip) to predict whether a household in the 1996–8 PECS has consumption below the national poverty line, using only indicators found both in the PECS and the census. Due to data constraints, Astrup and Dessus then apply the tools not to household-level census data, but rather to locality-level census averages. In effect, they treat the locality as if it were a household whose characteristics are the averages of

the actual characteristics of its particular households.¹⁹ Estimates from the tools are then used to make poverty maps that quickly show—in a way that is clear to non-specialists—how poverty rates vary across localities.

Poverty mapping in Astrup and Dessus and the scorecard in this paper are similar in that they both:

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

Strengths of poverty mapping include that it:

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

Strengths of the scorecard include that it:

- Is simpler in terms of both construction and application
- Tests accuracy empirically
- Associates poverty likelihoods with scores non-parametrically
- Uses judgment and theory in scorecard construction to reduce overfitting
- Estimates poverty likelihoods for individual households
- Reports straightforward formulas for standard errors

The basic difference between the two approaches is that poverty mapping seeks to help governments design and target pro-poor policies, while the scorecard seeks to

¹⁹ Other poverty maps using locality averages are Benson (2002), Bigman and Srinivasan (2002), and Minot (2000).

help small, local pro-poor organizations to manage their outreach when implementing policies.²⁰

Astrup and Dessus use the following indicators for Palestine, expressed as averages at the level of the locality:

- Demographics:
 - Logarithm of number of adult-equivalent household members
 - Marital status of head
- Employment status
- Education of head
- Characteristics of the residence:
 - Type of structure
 - Tenancy status
 - Fuel for heating
 - Fuel for cooking
- Asset ownership:
 - Solar water heater
 - Stove
 - Television
 - VCR
 - Refrigerator
 - Washing machine
 - Telephone
- Characteristics of the locality:
 - Identity as city or village
 - Identity as a region

These 17 indicators are simple and quick to collect (like those in this paper), although a field agent cannot calculate by hand the logarithm of the number of adult

²⁰ Another apparent difference is that the developers of the poverty-mapping approach (Elbers, Lanjouw, and Lanjouw, 2003; Demombynes *et al.*, 2002) say that it is too inaccurate to be used for targeting individual households, while Schreiner (2008c and in this paper) supports such targeting as a legitimate, potentially useful application of the scorecard. Recently, the developers of poverty mapping seem to have taken a small step away from their original position (Elbers *et al.*, 2007).

equivalents. Astrup and Dessus do not report bias or precision—even though a central feature of the poverty-mapping approach is the ability to generate standard errors—so its estimates cannot be compared with those in this paper.

9.2 IRIS Center

USAID commissioned IRIS Center (“IRIS”, 2009) to build a poverty-assessment tool (PAT) for use by its microenterprise partners in the West Bank²¹ for reporting on their participants’ poverty rates. Given this mandate, IRIS considers only the USAID “extreme” poverty line, using the 2007 PECS to estimate consumption with a two-step quantile regression (IRIS Center, 2005; Koenker and Hallock, 2001). IRIS’ 21 indicators are:²²

- Household demographics:
 - Number of members (regardless of age)
 - Number of members 15-years-old or younger
 - Number of members between 16 and 64 years of age, inclusive
 - Age of head
 - Marital status of head
 - Whether any member is a refugee

²¹ USAID has no microenterprise partners in the Gaza Strip.

²² IRIS does not report the actual tool, so this list is based on its questionnaire.

- Characteristics of the residence:
 - Tenancy status
 - Number of rooms
 - Type of exterior walls
 - Source of energy for heating the residence
 - Source of energy for baking
 - Source of energy for heating water
 - Source of water
 - Type of sewage connection
- Asset ownership:
 - Solar water heater
 - Bookcase
 - Television
 - VCR
 - Vacuum cleaner
 - Washing machine
 - Private car

Like the indicators in the new scorecard here, all of IRIS' indicators are simple to collect and verify; indeed, six indicators appear in both tools.

IRIS does not report its tools' points;²³ total points can be computed only with free IRIS-provided software which reports not total points for individual households but rather only an estimate of a group's poverty rate, as this is all the USAID mandate requires. This precludes use for targeting.

²³ IRIS does not reveal points so as to reduce the opportunity for manipulation. Points, indicator definitions, and measures of precision are available from IRIS on request.

IRIS’ preferred measure of accuracy is the “Balanced Poverty Accuracy Criterion” (IRIS Center, 2005), and USAID uses BPAC as its criterion for certifying poverty-assessment tools. BPAC depends on inclusion and on the difference between the estimated poverty rate and its true value (equivalent to the difference between undercoverage and leakage). The BPAC formula is:

$$(\text{Inclusion} - |\text{Undercoverage} - \text{Leakage}|) \times [100 \div (\text{Inclusion} + \text{Undercoverage})].$$

A higher BPAC is better. Both IRIS and this paper use the 2007 PECS, so accuracy comparisons would seem straightforward. Unfortunately, this is not the case, for five reasons.

First, the two poverty-assessment tools appear to use different versions of the 2007 PECS and/or different poverty lines. Different versions of the data are known to exist (UNRWA, 2009). Also, this paper uses a national poverty line of NIS12.92 per person per day (leading to a USAID “extreme” line of NIS8.91 and a corresponding person-level poverty rate in the West Bank of 7.9 percent). In contrast, IRIS uses a national line of JOD124 per adult-equivalent per day (leading to a USAID “extreme” line of JOD94 and a person-level poverty rate in the West Bank of 12.9 percent). It is also possible that the two tools use different definitions of consumption.

Second, IRIS—due to limited sample size—tests its tool *in-sample*, that is, using the same data that is used to construct the tool in the first place. In contrast, the scorecard here—in spite of using an even more data-hungry algorithm—is tested *out-of-sample*, that is, using data that is not used to construct the scorecard. In-sample testing

overstates accuracy; for example, Johanssen (2006, for BPAC) and Copestake *et al.* (2005, for a variety of measures) find that accuracy measures for poverty-assessment tools can deteriorate 8 to 17 percent going from in-sample to out-of-sample. Out-of-sample is more relevant because, in practice, poverty-assessment tools are applied to data on households that are not used to construct the tools.

Third, the reason in-sample testing overstates accuracy is *overfitting*, that is, when a poverty-assessment tool fits the construction data so closely that it captures not only some real patterns but also some false patterns that, due to sampling variation, show up only in the construction data but not in general in other samples. Overfitting is a greater risk for IRIS' West Bank tool because it estimates a more complex model. In particular, IRIS uses only data for the West Bank (not all of Palestine, as in this paper), and it includes 15 indicators (plus six “controls”) for each of its two steps.

Fourth, the IRIS tool applies only to the West Bank, whereas the scorecard here applies to the West Bank and the Gaza Strip. In general and with all else constant, accuracy is better the more homogeneous is the area to which a poverty-assessment tool is applied (Tarozzi and Deaton, 2007), and poverty rates are much higher in the Gaza Strip than the West Bank, implying little homogeneity.

Fifth, IRIS' tool is fine-tuned to the USAID “extreme” poverty line; IRIS does not report accuracy for other lines. The scorecard here is constructed based on the national poverty line and then calibrated to seven lines, one of which is the USAID line.

IRIS reports an in-sample BPAC of 51.9. When the all-Palestine scorecard here is applied out-of-sample to the West Bank for the USAID “extreme” line and a cut-off of 24 or below, BPAC is 37.0. It is difficult to know how much of this gap in BPACs would remain if all the differences listed above could be removed.

The main distinction between the new scorecard here and IRIS’ tool is transparency and ease-of-use: IRIS requires more indicators (21 versus 10), it cannot be used for targeting, it only applies to the West Bank, and it does not report tool points, indicators, or precision.

10. Conclusion

This paper presents scorecard. It can be used by pro-poor programs in Palestine (West Bank and Gaza Strip) to estimate the likelihood that a household has consumption below a given poverty line, to estimate the poverty rate of a group of households at a point in time, and to estimate changes in the poverty rate of a group of households between two points in time. The scorecard can also be used for targeting.

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

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

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

When the scorecard is applied to the validation sample with $n = 16,384$, the absolute difference between estimates and true poverty rates at a point in time is 3.0 percentage points or less and averages—across the seven poverty lines—1.1 percentage points. With 90-percent confidence, the precision of these differences is ± 0.5 percentage points or less. The scorecard is more precise than direct measurement.

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 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 straightforward to verify. Points are all zeros or positive integers, and scores range from 0 (most likely below a poverty line) to 100 (least likely below a poverty line). Scores are related to poverty likelihoods via simple look-up tables, and targeting cut-offs are likewise simple to apply. The design attempts to facilitate adoption by helping managers understand and trust scoring and by allowing non-specialists to generate scores quickly in the field.

In sum, the scorecard is a practical, objective way for pro-poor programs in Palestine (West Bank and Gaza Strip) to monitor poverty rates, track changes in poverty rates over time, and target services, provided that the scorecard is applied in a time period similar to that of 2007, the period when the data used to construct the scorecard was collected. The same approach can be applied to any country with similar data from a national income, consumption, or expenditure survey.

References

- Adams, Niall M.; and David J. Hand. (2000) “Improving the Practice of Classifier Performance Assessment”, *Neural Computation*, Vol. 12, pp. 305–311.
- Astrup, Claus; and Sébastien Dessus. (2005) “Targeting the Poor beyond Gaza or the West Bank: The Geography of Poverty in the Palestinian Territories”, *Région et Développement*, No. 21, pp. 173–197.
- 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.
- Benson, Todd. (2002) “Correlates of Poverty Measures in Malawi: An Application of Poverty Mapping at the Sub-District Scale”.
- Bigman, David; and P.V. Srinivasan. (2002) “Geographic Targeting of Poverty Alleviation Programs: Methodology and Applications to Rural India”, *Journal of Policy Modeling*, Vol. 24, pp. 237–255.
- Caire, Dean. (2004) “Building Credit Scorecards for Small Business Lending in Developing Markets”, microfinance.com/English/Papers/Scoring_SMEs_Hybrid.pdf, retrieved 10 April 2010.
- Chen, Shiyuan; and Mark Schreiner. (2009) “Simple Poverty Scorecard Poverty-Assessment Tool: Vietnam”, SimplePovertyScorecard.com/VNM_2006_ENG.pdf, retrieved 7 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*.
- Copestake, James G.; Dawson, Peter.; Fanning, John-Paul; McKay, Andrew; and Katie Wright-Revollo. (2005) “Monitoring the Diversity of the Poverty Outreach and Impact of Microfinance: A Comparison of Methods Using Data from Peru”, *Development Policy Review*, Vol. 23, No. 6, pp. 703–723.
- Daley-Harris, Sam. (2009) *State of the Microcredit Summit Campaign Report 2009*, microcreditsummit.org/state_of_the_campaign_report/, retrieved 10 April 2010.

- Dawes, Robyn M. (1979) “The Robust Beauty of Improper Linear Models in Decision Making”, *American Psychologist*, Vol. 34, No. 7, pp. 571–582.
- Deaton, Angus; and Salman Zaidi. (2002) “Guidelines for Constructing Consumption Aggregates for Welfare Analysis”, World Bank LSMS Working Paper No. 135, go.worldbank.org/8YRCR9ERJ0, retrieved 10 April 2010.
- Demombynes, Gabriel; Elbers, Chris; Lanjouw, Jenny; Lanjouw, Peter; Mistiaen, Johan; and Berk Özler. (2002) “Producing an Improved Geographic Profile of Poverty: Methodology and Evidence from Three Developing Countries”, World Institute for Development Economics Research Discussion Paper No. 2002/39, go.worldbank.org/UMQCZ1BW00, retrieved 10 April 2010.
- Efron, Bradley; and Robert J. Tibshirani. (1993) *An Introduction to the Bootstrap*.
- Elbers, Chris; Lanjouw, Jean O.; and Peter Lanjouw. (2003) “Micro-Level Estimation of Poverty and Inequality”, *Econometrica*, Vol. 71, No. 1, pp. 355–364.
- ; Fujii, Tomoki; Lanjouw, Peter; Özler, Berk; and Wesley Yin. (2007) “Poverty Alleviation through Geographic Targeting: How Much Does Disaggregation Help?”, *Journal of Development Economics*, Vol. 83, pp. 198–213.
- Falkenstein, Eric. (2008) “DefProbTM: A Corporate Probability of Default Model”, papers.ssrn.com/sol3/papers.cfm?abstract_id=1103404, retrieved 10 April 2010.
- 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 10 April 2010.
- Goodman, Leo A.; and Kruskal, William H. (1979) *Measures of Association for Cross Classification*.
- Grootaert, Christiaan; and Jeanine Braithwaite. (1998) “Poverty Correlates and Indicator-Based Targeting in Eastern Europe and the Former Soviet Union”, World Bank Policy Research Working Paper No. 1942, go.worldbank.org/VPMWVLU8E0, retrieved 10 April 2010.

- Grosh, Margaret; and Judy L. Baker. (1995) “Proxy Means Tests for Targeting Social Programs: Simulations and Speculation”, LSMS Working Paper No. 118, go.worldbank.org/W90WN57PD0, retrieved 10 April 2010.
- Hand, David J. (2006) “Classifier Technology and the Illusion of Progress”, *Statistical Science*, Vol. 22, No. 1, pp. 1–15.
- Hentschel, Jesko; Olsen Lanjouw, Jean; Lanjouw, Peter; and Javier Poggi. (2000) “Combining Census and Survey Data to Trace the Spatial Dimensions of Poverty: A Case Study of Ecuador”, *World Bank Economic Review*, Vol. 14, No. 1, pp. 147–165.
- Hoadley, Bruce; and Robert M. Oliver. (1998) “Business Measures of Scorecard Benefit”, *IMA Journal of Mathematics Applied in Business and Industry*, Vol. 9, pp. 55–64.
- IRIS Center. (2009) “Client Assessment Survey—West Bank”, povertytools.org/USAID_documents/Tools/Current_Tools/USAID_PAT_WB_10_2009.xls, retrieved 10 April 2010.
- (2007a) “Manual for the Implementation of USAID Poverty Assessment Tools”, povertytools.org/training_documents/Manuals/USAID_PAT_Manual_Eng.pdf, retrieved 10 April 2010.
- (2007b) “Introduction to Sampling for the Implementation of PATs”, povertytools.org/training_documents/Sampling/Introduction_Sampling.pdf, retrieved 10 April 2010.
- (2005) “Notes on Assessment and Improvement of Tool Accuracy”, povertytools.org/other_documents/AssessingImproving_Accuracy.pdf, retrieved 10 April 2010.
- Johannsen, Julia. (2006) “Operational Poverty Targeting in Peru—Proxy Means Testing with Non-Income Indicators”, IPC Working Paper No. 30, www.undp-povertycentre.org/pub/IPCWorkingPaper30.pdf, accessed 10 April 2010.
- Johnson, Glenn. (2007) “Lesson 3: Two-Way Tables—Dependent Samples”, www.stat.psu.edu/online/development/stat504/03_2way/53_2way_compare.htm, retrieved 10 April 2010.
- Koenker, Roger; and Kevin F. Hallock. (2001) “Quantile Regression”, *Journal of Economic Perspectives*, Vol. 15, No. 4, pp. 143–156

- 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”, ciep.itam.mx/~martinel/lies4.pdf, retrieved 10 April 2010.
- Matul, Michal; and Sean Kline. (2003) “Scoring Change: Prizma’s Approach to Assessing Poverty”, Microfinance Centre for Central and Eastern Europe and the New Independent States Spotlight Note No. 4, www.mfc.org.pl/doc/Research/ImpAct/SN/MFC_SN04_eng.pdf, retrieved 10 April 2010.
- McNemar, Quinn. (1947) “Note on the Sampling Error of the Difference between Correlated Proportions or Percentages”, *Psychometrika*, Vol. 17, pp. 153–157.
- Minot, Nicholas. (2000) “Generating Disaggregated Poverty Maps: An Application to Vietnam”, *World Development*, Vol. 28, No. 2, pp. 319–331.
- 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.
- Palestinian Central Bureau of Statistics. (2008) “Press Release: Poverty and Living Conditions in the Palestinian Territory, 2007”, files.tiggroups.org/68187/Poverty%20and%20Living%20Conditions%20in%20the%20Palestinian%20Territory,%202007.doc, retrieved 10 April 2010.
- (2006) “The Palestinian Expenditure and Consumption Survey 2005: Levels of Living in the Palestinian Territory, Final Report (January 2005 – January 2006)”, www.pcbs.gov.ps/Portals/_pcbs/living/report_english.pdf, retrieved 10 April 2010.

- 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/statug_logistic_sect035.htm, retrieved 10 April 2010.
- Schreiner, Mark. (2013) “Simple Poverty Scorecard Poverty-Assessment Tool: Peru”, SimplePovertyScorecard.com/PER_2010_ENG.pdf, retrieved 7 July 2016.
- (2009a) “Simple Poverty Scorecard Poverty-Assessment Tool: Peru”, SimplePovertyScorecard.com/PER_2007_ENG.pdf, retrieved 7 July 2016.
- (2009b) “Simple Poverty Scorecard Poverty-Assessment Tool: Philippines”, SimplePovertyScorecard.com/PHL_2002_ENG.pdf, retrieved 7 July 2016.
- (2009c) “Simple Poverty Scorecard Poverty-Assessment Tool: Pakistan”, SimplePovertyScorecard.com/PAK_2005_ENG.pdf, retrieved 7 July 2016.
- (2009d) “Simple Poverty Scorecard Poverty-Assessment Tool: Bolivia”, SimplePovertyScorecard.com/BOL_2007_ENG.pdf, retrieved 7 July 2016.
- (2009e) “Simple Poverty Scorecard Poverty-Assessment Tool: Mexico”, SimplePovertyScorecard.com/MEX_2008_ENG.pdf, retrieved 7 July 2016.
- (2008a) “Simple Poverty Scorecard Poverty-Assessment Tool: Peru”, SimplePovertyScorecard.com/PER_2003_ENG.pdf, retrieved 7 July 2016.
- (2008b) “Simple Poverty Scorecard Poverty-Assessment Tool: India”, SimplePovertyScorecard.com/IND_2005_ENG.pdf, retrieved 7 July 2016.
- (2008c) “Simple Poverty Scorecard Poverty-Assessment Tool: Ecuador”, SimplePovertyScorecard.com/ECU_2005_ENG.pdf, retrieved 7 July 2016.
- (2006a) “La Herramienta del Índice de Calificación de la PobrezaTM: México”, SimplePovertyScorecard.com/MEX_2002_SPA.pdf, retrieved 7 July 2016.
- (2006b) “Is One Simple Poverty Scorecard Poverty-Assessment Tool Enough for India?” microfinance.com/English/Papers/Scoring_Poverty_India_Segments.pdf, retrieved 10 April 2010.
- (2005) “IRIS Questions on the Simple Poverty Scorecard Poverty-Assessment Tool”, microfinance.com/English/Papers/Scoring_Poverty_Response_to_IRIS.pdf, retrieved 10 April 2010.

- (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 10 April 2010.
- Sillers, Don. (2006) “National and International Poverty Lines: An Overview”, pdf.usaid.gov/pdf_docs/Pnadh069.pdf, retrieved 13 May 2016.
- 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. (2007) “Using Census and Survey Data to Estimate Poverty and Inequality for Small Areas”, princeton.edu/~deaton/downloads/20080301SmallAreas_FINAL.pdf, retrieved 10 April 2010.
- Toohig, Jeff. (2008) “PPI Pilot Training Guide”, progressoutofpoverty.org/toolkit, retrieved 10 April 2010.
- United Nations Relief and Works Agency for Palestine Refugees in the Near East. (2009) “Poverty in the Occupied Palestinian Territory, Briefing Paper”, unrwa.org/userfiles/20100118142147.pdf, retrieved 10 April 2010.
- (2008) “Prolonged Crisis in the Occupied Palestinian Territory: Socio-Economic Developments in 2007”, unrwa.org/userfiles/20100118155922.pdf, retrieved 10 April 2010.
- United States Congress. (2004) “Microenterprise Results and Accountability Act of 2004 (HR 3818 RDS)”, November 20, smith4nj.com/laws/108-484.pdf, retrieved 13 May 2016.
- 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-benchmark.pdf,[sic] retrieved 10 April 2010.

----- (2004) “Deep Palestinian Poverty in the Midst of Economic Crisis”,
go.worldbank.org/KZNKZZBYX0, retrieved 10 April 2010.

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

Figure 2: Poverty lines and poverty rates, for all of Palestine, by West Bank/Gaza Strip, by scoring sub-sample, and by household-/person-level

Grouping	Line/Rate	Households	% with consumption below a poverty line						
			National				USAID	Intl. 2005 PPP	
			Natl.	Subsistence	150%	200%	'Extreme'	\$3.75/day	\$5.00/day
All Palestine	Poverty line	1,231	12.92	10.33	19.39	25.85	8.91	9.23	12.31
	Household rate	1,231	30.0	19.7	53.9	68.9	13.5	14.8	27.7
	Person rate	1,231	37.4	25.9	62.3	76.4	18.6	20.2	34.9
West Bank	Household rate	835	18.1	10.0	41.4	58.3	5.6	6.2	16.2
	Person rate	835	22.5	13.1	48.9	66.3	7.9	8.7	20.3
Gaza Strip	Household rate	396	53.0	38.6	77.9	89.5	28.9	31.5	50.0
	Person rate	396	62.7	47.5	85.0	93.7	36.7	39.7	59.8
Scoring sub-sample									
Construction									
Selecting indicators and points	Household rate	397	29.7	20.0	53.6	69.5	13.2	14.4	27.3
Calibration									
Associating scores with likelihoods	Household rate	397	30.0	19.0	53.9	69.8	13.4	14.2	28.2
Validation									
Measuring accuracy	Household rate	437	30.1	20.1	54.1	67.6	14.0	15.8	27.6
Change in poverty rate (percentage points)									
From construction/calibration to validation			-0.2	-0.7	-0.4	+2.0	-0.7	-1.6	+0.2

Source: 2007 PECS. Poverty lines do not vary by region and are in units of NIS per person per day.

Figure 3: Poverty indicators by uncertainty coefficient

<u>Uncertainty coefficient</u>	<u>Indicator (Answers ordered starting with those most strongly indicative of poverty)</u>
1566	Does the household have a vacuum cleaner? (No; Yes)
1329	Does the household have a computer? (No; Yes)
1321	How many household members are 17-years-old or younger? (Four or more; Three; Two; One; None)
1242	How many household members are 16-years-old or younger? (Four or more; Three; Two; One; None)
1235	How many household members are 15-years-old or younger? (Four or more; Three; Two; One; None)
1194	How many household members are 18-years-old or younger? (Four or more; Three; Two; One; None)
1186	How many household members are paid employees? (Six or more; Five; Four; Three; Two; None or one)
1177	How many household members are 14-years-old or younger? (Four or more; Three; Two; One; None)
1134	What is the main source of energy for baking? (Wood, coal, olive cakes, electricity, or other; Gas or none)
1129	How many members does the household have? (Ten or more; Nine; Eight; Five, six, or seven; Four; One, two, or three)
1061	How many household members are 13-years-old or younger? (Four or more; Three; Two; One; None)
1010	Does the household have a land-line and/or cellular telephone? (None; Only cellular; Land-line (regardless of cellular))
939	How many household members are 12-years-old or younger? (Four or more; Three; Two; One; None)
900	What is the main building material used in the exterior walls of the residence? (Cement cob, mud, or other; Cleaned stone, stone and cement, old stone, or concrete)
865	What is the occupation of the male head/spouse? (Elementary occupation; Skilled agriculture and fishery worker; Skilled worker and service, shop, or market worker; Does not work; Craft and related trade worker, or pant and machine operator and assembler; No male head/spouse; Legislator, senior official, or manager, professional, technician or associate professional, clerk)
796	How many household members are 11years-old or younger? (Four or more; Three; Two; One; None)
765	Do all children ages 6 to 18 attend school? (No; Yes; No children in this age range)
761	Do all children ages 6 to 12 attend school? (No; Yes; No children in this age range)
758	Do all children ages 6 to 17 attend school? (No; Yes; No children in this age range)

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

<u>Uncertainty coefficient</u>	<u>Indicator (Answers ordered starting with those most strongly associated with poverty)</u>
749	How many household members work as a professional, technician or associate professional, clerk, legislator, senior official, or manager (None; One; Two or more)
705	Do all children ages 6 to 13 attend school? (No; Yes; No children in this age range)
652	Does the household have a private car? (No; Yes)
652	Do all children ages 6 to 16 attend school? (No; Yes; No children in this age range)
632	What is the main source of energy for heating water? (Kerosene, wood, coal, or other; Gas; Sun; Electricity)
599	What is the employment status of the male head/spouse? (Unemployed, but has worked in the past; Unable to work; Employed for 15 hours or more; Employed for 1 to 14 hours, unemployed and has never worked, full-time student, housewife, does not work and is not seeking a job, or other; Absent from work; No male head/spouse)
595	Do all children ages 6 to 14 attend school? (No; Yes; No children in this age range)
592	Does the household have a T.V. and a VCR and/or DVD? (None; Only T.V.; T.V. and VCR and/or DVD)
586	Does the household have a bookcase? (No; Yes)
583	Does the household have a satellite dish? (No; Yes)
567	Do all children ages 6 to 15 attend school? (No; Yes; No children in this age range)
566	Do all children ages 6 to 11 attend school? (No; Yes; No children in this age range)
554	What is the highest educational level that the female head/spouse has completed? (Can read and write; Elementary school; Secondary school; Preparatory school; None or illiterate; Associate diploma, bachelor degree, high diploma, master degree, doctorate, or no female head/spouse)
540	Does the household have a VCR and/or DVD? (No; Yes)
533	How many household members work in elementary occupations or in agriculture, fishing, or forestry? (Two or more; One; None)
448	Do any household members work in elementary occupations? (Yes; No)

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

<u>Uncertainty coefficient</u>	<u>Indicator (Answers ordered starting with those most strongly associated with poverty)</u>
401	What is the highest educational level that the female head/spouse has completed? (Elementary school; Can read and write; Preparatory school; Secondary school; None or illiterate; No male head/spouse; Associate diploma, bachelor degree, high diploma, master degree, or doctorate)
401	What is the highest educational level that a household member has completed? (Elementary school; Preparatory school; Secondary school; Can read and write; Illiterate; Associate diploma; Bachelor degree, high diploma, master degree, doctorate)
401	What is the main source of energy for air conditioning? (None, or other; Electricity)
399	Is the male head/spouse employed? (No; Yes; No male head/spouse)
392	How many household members are 6-years-old or younger? (Three or more; Two; One; None)
384	In what sector does the male head/spouse work? (Outside establishments; non-profit, UNRWA, or other international organization; National private; No data; National government; Does not work; No male head/spouse; Foreign private or foreign government;)
360	Is the male head/spouse unemployed, unable to work, or not working? (Yes; No; No male head/spouse)
337	How many household members work for a national government, foreign government, non-profit organization, UNRWA, or an international organization? (None; One; Two or more)
331	How old is the male head/spouse? (35 to 44; 29 to 34; 45 to 64; No male head/spouse; 65 or older; 28 or younger)
298	In what sector does the female head/spouse work? (Outside establishments; No data; National private or foreign private; National government, foreign government, non-profit, UNRWA or other international organization, does not work, or no female head/spouse)
298	What is the occupation of the female head/spouse? (Skilled worker and service, shop, or market worker, skilled agriculture and fishery worker, craft and related trade worker, plant and machine operator and assembler, elementary occupation, or no female head/spouse; Does not work or no data; Legislator, senior official, or manager, professional, technician or associate professional, or clerk)

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

<u>Uncertainty coefficient</u>	<u>Indicator (Answers ordered starting with those most strongly associated with poverty)</u>
282	What is the employment type of the male head/spouse? (Self-employed; Paid employee; Unpaid employee; Employer; No male head/spouse)
273	What is the employment type of the female head/spouse? (Employer or self-employed; Paid employee; Unpaid employee or no female head/spouse)
272	Does the household have a solar water heater? (No; Yes)
255	Are any household members registered refugees? (Yes; No)
253	How old is the female head/spouse? (28 to 49; 50 or older; No female head/spouse; 27 or younger)
252	Do any household members work in the sector of agriculture, fishing, or forestry? (Yes; No)
248	How many household members are unemployed, unable to work, or not working? (Two or more; One; None)
247	If the household has any agricultural land (be it owned, rented, or free), how many male household members worked agricultural land during the last agricultural season? (Has agricultural land, and three or more males worked in agriculture; Does not have any agricultural land; Has agricultural land, and one male worked in agriculture; Has agricultural land, and two males worked in agriculture; Has agricultural land, but no males worked in agriculture)
230	What is the marital status of the female head/spouse? (Married; Other)
227	What is the main source of energy for heating the residence? (None; Other)
223	Does the household have a refrigerator? (No; Yes)
209	How many household members worked agricultural land during the last agricultural season? (Three or more; None; Two; One)
193	Where does the male/head spouse work? (At home, in same locality, or does not work; Outside same locality, or no male head/spouse)
192	Where is the place of work of the female/head spouse? (At home; In the same locality; No data; Other)
183	Does the household have a clothes dryer? (No; Yes)
177	Are any household members refugees (be they registered or unregistered)? (Yes; No)

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

<u>Uncertainty coefficient</u>	<u>Indicator (Answers ordered starting with those most strongly associated with poverty)</u>
174	Is the female head/spouse unemployed, unable to work, or not working? (No; Yes, or no female head/spouse)
167	What type of residence does the household have? (House, separate room, tent, marginal shelter, or other; Villa or apartment)
165	What is the employment type of the female head/spouse? (Employer or self-employed; Paid employee; Unpaid employee, or no female head/spouse)
164	How many household members are employed? (None; One; Two or more)
162	Does the toilet/water closet used by the household have piped water? (No, or no toilet; Yes)
150	What is the structure of household headship? (Both male and female heads/spouses; Female head/spouse only, or male head/spouse only)
150	What is the marital status of the male head/spouse? (Married; Other)
146	Does the bathroom used by the household have piped water? (No, or no bathroom; Yes)
137	What type of electrical network is the residence connected to? (Private, or no electricity; Public)
136	If the household has any agricultural land (be it owned, rented, or free), how many male household members worked agricultural land during the last agricultural season? (Has agricultural land, and three or more males worked in agriculture; Does not have any agricultural land; Has agricultural land, and one male worked in agriculture; Has agricultural land, and two males worked in agriculture; Has agricultural land, but no males worked in agriculture;)
133	If the household has any agricultural land (be it owned, rented, or free), how many female household members worked agricultural land during the last agricultural season? (Has agricultural land, and two or more females worked in agriculture; Does not have any agricultural land; Has agricultural land, and one female worked in agriculture; Has agricultural land, but no females worked in agriculture)
132	Does the household have a T.V.? (No; Yes)

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

<u>Uncertainty coefficient</u>	<u>Indicator (Answers ordered starting with those most strongly associated with poverty)</u>
130	What is the main means of garbage disposal? (Thrown randomly, thrown in open garbage area/open container, or other; Collected by sanitation worker; Thrown away in nearby garbage container; Burned)
129	Do any household members work in Israel, in the settlements, or abroad? (No; Yes)
104	What the the tenancy status of the household in its residence? (Owned; Other)
104	How many rooms does the residence have? (One; More than one)
101	If the household has any agricultural land (owned, rented, or free), does it have any cattle, sheep, goats, poultry, horses, mules, or beehives? (No agricultural land; Has land, but no animals; Has land and animals)
98	What is the household's main source of income? (Wages and salaries from private sector, international aid, social aid/affairs, or remittances; Farming, family business, property income, national insurance (Jerusalem), wages and salaries from the public sector, Israeli sector, or international organizations, or other)
94	Does the household have any agricultural land (be it owned, rented, or free)? (No; Yes)
90	Does the male head/spouse work in agriculture, fishing, or forestry? (Yes; No; No male head/spouse)
76	Do any household members work in agriculture, fishing, or forestry? (Yes; No)
73	Can the male head/spouse read and write? (Yes; No; No male head/spouse)
73	Does the household have any cattle, sheep, goats, poultry, horses, mules, or beehives? (Yes; No)
69	How many bedrooms does the residence have? (Four or more; Three; Two; One; None)
64	Does the kitchen used by the household have piped water? (No, no kitchen, Yes)
63	What is the main source of energy for cooking? (Kerosene, coal/firewood, or other; Gas or electricity)
61	Does the female head/spouse work in agriculture, fishing, or forestry? (Yes; No, or no female head/spouse)
55	How many household members are self-employed (with or without other employees)? (Two or more; One; None)

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

<u>Uncertainty coefficient</u>	<u>Indicator (Answers ordered starting with those most strongly associated with poverty)</u>
52	How many male household members worked agricultural land during the last agricultural season? (Three or more; None; One; Two)
46	Can the female head/spouse read and write? (Yes; No or no female head/spouse)
45	Does the household have a washing machine? (No; Yes)
22	How many female household members worked agricultural land during the last agricultural season? (Three or more; None; One; Two)
22	Can any household member read and write? (No; Yes)
19	What type of water network is the residence connected to? (Private or no piped water; Public)
12	Does the household have a dish washer? (No; Yes)
3	Does the household have a cooking stove? (No; Yes)
0.8	Are any household members unregistered refugees? (Yes; No)
0.3	What type of sewage system is the residence connected to? (Cesspit or none; Public)
0.2	Is the female head/spouse employed? (Yes; No or no female head/spouse)

Source: 2007 PECS and the national poverty line.

National Poverty Line

(and tables pertaining to all poverty lines)

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

If a household's score is then the likelihood (%) of being below the poverty line is:
0-4	100.0
5-9	100.0
10-14	100.0
15-19	81.7
20-24	82.2
25-29	63.1
30-34	43.1
35-39	41.6
40-44	30.2
45-49	8.3
50-54	14.4
55-59	6.0
60-64	4.8
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 5 (National line): Derivation of estimated poverty likelihoods associated with scores

Score	Households below poverty line		All households at score		Poverty likelihood (estimated, %)
0-4	446	÷	446	=	100.0
5-9	91	÷	91	=	100.0
10-14	806	÷	806	=	100.0
15-19	3,191	÷	3,905	=	81.7
20-24	5,094	÷	6,196	=	82.2
25-29	4,960	÷	7,860	=	63.1
30-34	5,002	÷	11,615	=	43.1
35-39	5,417	÷	13,022	=	41.6
40-44	3,349	÷	11,108	=	30.2
45-49	613	÷	7,388	=	8.3
50-54	1,250	÷	8,667	=	14.4
55-59	391	÷	6,499	=	6.0
60-64	326	÷	6,761	=	4.8
65-69	0	÷	8,132	=	0.0
70-74	0	÷	3,913	=	0.0
75-79	0	÷	1,860	=	0.0
80-84	0	÷	1,059	=	0.0
85-89	0	÷	671	=	0.0
90-94	0	÷	0	=	0.0
95-100	0	÷	0	=	0.0

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

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

Score	Likelihood of having consumption in range demarcated by poverty lines per day per capita							
	<'Extreme'	=>'Extreme'	=>\$3.75/day	=>Subsistence	=>\$5.00/day	=>Natl.	=>150% Natl.	=>200% Natl.
		and	and	and	and	and	and	
	<\$3.75/day	<Subsistence	<\$5.00/day	<Natl.	<150% Natl.	<200% Natl.		
	=>NIS8.91	=>NIS9.23	=>NIS10.33	=>NIS12.31	=>NIS12.92	=>NIS19.39		
	<NIS8.91	and	and	and	and	and	and	=>NIS25.85
		<NIS9.23	<NIS10.33	<NIS12.31	<NIS12.92	<NIS19.39	<NIS25.85	
0-4	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5-9	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
10-14	73.7	0.0	13.2	13.2	0.0	0.0	0.0	0.0
15-19	50.6	0.0	6.7	18.2	6.3	13.4	4.9	0.0
20-24	65.5	0.0	5.1	8.7	3.0	9.3	4.1	4.4
25-29	23.5	0.4	13.5	22.2	3.6	33.7	0.0	3.2
30-34	22.7	0.0	5.9	11.1	3.4	38.9	8.3	9.7
35-39	5.1	5.2	10.3	18.2	2.8	31.3	15.4	11.7
40-44	8.3	0.0	2.6	16.2	3.1	31.6	20.2	18.0
45-49	0.0	0.0	1.8	6.5	0.0	30.1	35.1	26.5
50-54	5.2	0.0	0.0	6.4	2.8	17.5	25.5	42.6
55-59	0.0	0.0	6.0	0.0	0.0	23.4	28.8	41.8
60-64	0.0	0.0	4.8	0.0	0.0	22.0	19.6	53.6
65-69	0.0	0.0	0.0	0.0	0.0	20.1	13.6	66.3
70-74	0.0	0.0	0.0	0.0	0.0	12.6	7.6	79.8
75-79	0.0	0.0	0.0	0.0	0.0	0.0	21.4	78.6
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 (National line): Bootstrapped differences between estimated and true household poverty likelihoods with confidence intervals in a large sample ($n = 16,384$), scorecard applied to the validation sample

Score	Difference between estimate and true value			
	Diff.	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	+20.7	5.8	6.8	8.7
15-19	-6.8	4.4	4.6	4.9
20-24	-6.6	4.0	4.3	4.6
25-29	+12.1	2.4	2.8	3.6
30-34	-6.0	4.0	4.2	4.4
35-39	+10.9	1.7	2.0	2.6
40-44	+13.7	1.6	1.8	2.4
45-49	-28.0	15.2	15.4	15.9
50-54	-1.3	1.7	1.9	2.5
55-59	-7.0	4.4	4.6	4.8
60-64	+4.8	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 (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							
	<u>National</u>					<u>USAID</u>	<u>Intl. 2005 PPP</u>	
	Natl.	Subsistence	150%	200%	'Extreme'	\$3.75/day	\$5.00/day	
<u>Estimate minus true value</u>								
Scorecard applied to validation sample	+0.4	-1.3	+0.5	+3.0	-0.1	-1.4	+0.8	
<u>Precision of difference</u>								
Scorecard applied to validation sample	0.5	0.5	0.5	0.5	0.4	0.4	0.5	
<u>α factor for standard errors</u>								
Scorecard applied to validation sample	0.88	0.88	0.85	0.86	0.90	0.88	0.89	
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 $n = 256, 512, 1,024, 2,048, 4,096, 8,192, \text{ and } 16,384$.								

Figure 9 (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	+1.5	66.5	77.4	87.5
4	-0.2	34.0	40.5	51.7
8	+0.1	24.4	28.5	36.4
16	+0.8	17.4	20.9	26.1
32	+0.3	11.7	14.2	19.4
64	+0.4	8.5	10.3	14.2
128	+0.4	6.0	7.2	9.8
256	+0.4	4.0	4.8	5.9
512	+0.3	2.9	3.3	4.5
1,024	+0.3	2.2	2.5	3.3
2,048	+0.4	1.5	1.8	2.5
4,096	+0.4	1.0	1.2	1.7
8,192	+0.4	0.7	0.9	1.1
16,384	+0.4	0.5	0.6	0.9

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

		<u>Targeting segment</u>	
		<u>Targeted</u>	<u>Non-targeted</u>
<u>True poverty status</u>	<u>Below poverty line</u>	<u>Inclusion</u> Under poverty line Correctly Targeted	<u>Undercoverage</u> Under poverty line Mistakenly Non-targeted
	<u>Above poverty line</u>	<u>Leakage</u> Above poverty line Mistakenly Targeted	<u>Exclusion</u> Above poverty line Correctly Non-targeted

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

Score	<u>Inclusion:</u>	<u>Undercoverage:</u>	<u>Leakage:</u>	<u>Exclusion:</u>	<u>Total Accuracy</u>	<u>BPAC</u>
	< poverty line correctly targeted	< poverty line mistakenly non-targeted	=> poverty line mistakenly targeted	=> poverty line non-targeted	Inclusion + Exclusion	See text
0–4	0.4	29.6	0.0	69.9	70.4	–97.0
5–9	0.5	29.5	0.0	69.9	70.5	–96.4
10–14	1.2	28.9	0.2	69.8	70.9	–91.7
15–19	4.6	25.5	0.7	69.3	73.9	–67.3
20–24	10.0	20.1	1.5	68.5	78.4	–28.8
25–29	13.9	16.2	5.4	64.5	78.4	+10.3
30–34	19.9	10.2	11.0	58.9	78.8	+63.3
35–39	23.6	6.5	20.4	49.6	73.1	+32.2
40–44	25.3	4.8	29.8	40.2	65.5	+1.0
45–49	27.9	2.2	34.6	35.4	63.2	–15.0
50–54	29.3	0.8	41.8	28.1	57.4	–39.2
55–59	30.1	0.0	47.5	22.4	52.5	–58.1
60–64	30.1	0.0	54.3	15.6	45.7	–80.6
65–69	30.1	0.0	62.4	7.5	37.6	–107.7
70–74	30.1	0.0	66.3	3.6	33.7	–120.7
75–79	30.1	0.0	68.2	1.7	31.8	–126.9
80–84	30.1	0.0	69.3	0.7	30.7	–130.4
85–89	30.1	0.0	69.9	0.0	30.1	–132.6
90–94	30.1	0.0	69.9	0.0	30.1	–132.6
95–100	30.1	0.0	69.9	0.0	30.1	–132.6

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

Figure 12 (National line): Households below the poverty line and all households at a given score or at or below a given score cut-off, scorecard applied to 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	0.4	100.0	1.5	Only poor targeted
5-9	0.5	100.0	1.8	Only poor targeted
10-14	1.3	86.3	3.9	6.3:1
15-19	5.2	87.5	15.3	7.0:1
20-24	11.4	87.0	33.1	6.7:1
25-29	19.3	71.8	46.1	2.5:1
30-34	30.9	64.3	66.1	1.8:1
35-39	43.9	53.6	78.4	1.2:1
40-44	55.0	45.9	84.1	0.9:1
45-49	62.4	44.6	92.7	0.8:1
50-54	71.1	41.2	97.4	0.7:1
55-59	77.6	38.7	100.0	0.6:1
60-64	84.4	35.6	100.0	0.6:1
65-69	92.5	32.5	100.0	0.5:1
70-74	96.4	31.2	100.0	0.5:1
75-79	98.3	30.6	100.0	0.4:1
80-84	99.3	30.3	100.0	0.4:1
85-89	100.0	30.1	100.0	0.4:1
90-94	100.0	30.1	100.0	0.4:1
95-100	100.0	30.1	100.0	0.4:1

Subsistence Poverty Line

Figure 4 (Subsistence 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	86.8
15-19	57.2
20-24	70.6
25-29	37.4
30-34	28.5
35-39	20.6
40-44	10.9
45-49	1.8
50-54	5.2
55-59	6.0
60-64	4.8
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 (Subsistence line): Bootstrapped differences between estimated and true household poverty likelihoods with confidence intervals in a large sample ($n = 16,384$), scorecard applied to the validation sample

Score	Difference between estimate and true value			
	Diff.	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	+7.6	5.8	6.8	8.7
15-19	-23.5	13.1	13.3	13.7
20-24	-5.1	3.6	3.9	4.2
25-29	-4.6	3.5	3.7	4.1
30-34	+1.6	1.7	2.0	2.5
35-39	+4.0	1.5	1.8	2.5
40-44	+7.7	0.7	0.8	1.1
45-49	-19.9	10.9	11.1	11.5
50-54	-0.6	1.1	1.3	1.8
55-59	-2.3	1.9	2.0	2.3
60-64	+4.8	0.0	0.0	0.0
65-69	+0.0	0.0	0.0	0.0
70-74	+0.0	0.0	0.0	0.0
75-79	+0.0	0.0	0.0	0.0
80-84	+0.0	0.0	0.0	0.0
85-89	+0.0	0.0	0.0	0.0
90-94	+0.0	0.0	0.0	0.0
95-100	+0.0	0.0	0.0	0.0

Figure 9 (Subsistence 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	-0.4	56.4	65.8	84.4
4	-1.1	27.4	33.1	46.1
8	-1.0	20.8	24.1	30.6
16	-0.9	14.5	16.8	23.1
32	-1.3	10.7	12.4	16.4
64	-1.4	7.4	9.0	11.8
128	-1.4	5.3	6.4	8.0
256	-1.3	3.7	4.5	5.8
512	-1.4	2.6	3.0	4.1
1,024	-1.4	1.8	2.2	3.0
2,048	-1.3	1.3	1.5	2.1
4,096	-1.3	0.9	1.1	1.5
8,192	-1.3	0.6	0.8	1.0
16,384	-1.3	0.5	0.5	0.7

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

Score	Inclusion:	Undercoverage:	Leakage:	Exclusion:	Total Accuracy	BPAC
	< poverty line correctly targeted	< poverty line mistakenly non-targeted	=> poverty line mistakenly targeted	=> poverty line correctly non-targeted	Inclusion + Exclusion	See text
0-4	0.4	19.7	0.0	79.9	80.3	-95.6
5-9	0.5	19.6	0.0	79.9	80.4	-94.7
10-14	1.2	19.0	0.2	79.7	80.8	-87.6
15-19	4.3	15.8	0.9	78.9	83.2	-52.5
20-24	8.9	11.2	2.5	77.3	86.3	+1.1
25-29	12.0	8.1	7.3	72.6	84.6	+55.4
30-34	15.5	4.7	15.4	64.4	79.9	+23.4
35-39	17.2	2.9	26.7	53.2	70.4	-32.6
40-44	17.6	2.5	37.5	42.4	60.0	-85.9
45-49	19.1	1.0	43.3	36.6	55.7	-114.9
50-54	19.6	0.5	51.5	28.4	48.0	-155.5
55-59	20.1	0.0	57.5	22.4	42.5	-185.3
60-64	20.1	0.0	64.2	15.6	35.8	-218.8
65-69	20.1	0.0	72.4	7.5	27.6	-259.2
70-74	20.1	0.0	76.3	3.6	23.7	-278.6
75-79	20.1	0.0	78.1	1.7	21.9	-287.8
80-84	20.1	0.0	79.2	0.7	20.8	-293.1
85-89	20.1	0.0	79.9	0.0	20.1	-296.4
90-94	20.1	0.0	79.9	0.0	20.1	-296.4
95-100	20.1	0.0	79.9	0.0	20.1	-296.4

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

Figure 12 (Subsistence line): Households below the poverty line and all households at a given score or at or below a given score cut-off, scorecard applied to 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	0.4	100.0	2.2	Only poor targeted
5-9	0.5	100.0	2.7	Only poor targeted
10-14	1.3	86.3	5.8	6.3:1
15-19	5.2	82.3	21.4	4.7:1
20-24	11.4	78.0	44.3	3.5:1
25-29	19.3	62.2	59.6	1.6:1
30-34	30.9	50.1	76.9	1.0:1
35-39	43.9	39.2	85.6	0.6:1
40-44	55.0	32.0	87.4	0.5:1
45-49	62.4	30.7	95.0	0.4:1
50-54	71.1	27.6	97.4	0.4:1
55-59	77.6	26.0	100.0	0.4:1
60-64	84.4	23.9	100.0	0.3:1
65-69	92.5	21.8	100.0	0.3:1
70-74	96.4	20.9	100.0	0.3:1
75-79	98.3	20.5	100.0	0.3:1
80-84	99.3	20.3	100.0	0.3:1
85-89	100.0	20.1	100.0	0.3:1
90-94	100.0	20.1	100.0	0.3:1
95-100	100.0	20.1	100.0	0.3:1

150% of the National Poverty Line

Figure 4 (150% of the 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	95.1
20-24	91.6
25-29	96.8
30-34	82.0
35-39	72.9
40-44	61.8
45-49	38.4
50-54	31.9
55-59	29.4
60-64	26.9
65-69	20.1
70-74	12.6
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 national line): Bootstrapped differences between estimated and true household poverty likelihoods with confidence intervals in a large sample ($n = 16,384$), scorecard applied to the validation sample

Score	Difference between estimate and true value			
	Diff.	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	+20.7	5.8	6.8	8.7
15-19	-4.9	2.5	2.5	2.5
20-24	-5.1	2.9	3.0	3.1
25-29	+15.3	1.7	2.1	2.6
30-34	+7.7	1.7	1.9	2.7
35-39	-7.6	4.5	4.6	4.9
40-44	+5.8	2.1	2.4	3.1
45-49	-28.2	15.2	15.4	15.9
50-54	-1.9	2.1	2.7	3.7
55-59	-13.0	7.8	8.0	8.6
60-64	+26.9	0.0	0.0	0.0
65-69	+0.4	2.1	2.4	3.2
70-74	+12.6	0.0	0.0	0.0
75-79	+0.0	0.0	0.0	0.0
80-84	+0.0	0.0	0.0	0.0
85-89	+0.0	0.0	0.0	0.0
90-94	+0.0	0.0	0.0	0.0
95-100	+0.0	0.0	0.0	0.0

Figure 9 (150% of the 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	+1.2	70.5	76.3	88.4
4	+0.0	35.2	40.6	49.9
8	+0.4	25.3	29.9	36.6
16	+0.7	17.1	20.9	27.5
32	+0.4	12.5	14.7	18.7
64	+0.5	9.0	10.2	13.8
128	+0.5	6.2	7.7	10.5
256	+0.5	4.2	5.2	7.4
512	+0.4	3.1	3.7	5.0
1,024	+0.4	2.2	2.5	3.2
2,048	+0.5	1.5	1.8	2.5
4,096	+0.5	1.1	1.3	1.8
8,192	+0.5	0.8	0.9	1.2
16,384	+0.5	0.5	0.6	0.9

Figure 11 (150% of the 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	0.4	53.7	0.0	45.9	46.3	-98.4
5-9	0.5	53.6	0.0	45.9	46.4	-98.0
10-14	1.2	52.9	0.2	45.7	46.9	-95.4
15-19	5.1	49.0	0.2	45.7	50.8	-80.9
20-24	11.0	43.1	0.5	45.4	56.4	-58.6
25-29	17.3	36.8	2.0	43.9	61.2	-32.3
30-34	25.8	28.3	5.1	40.8	66.6	+4.9
35-39	35.9	18.2	8.0	37.9	73.8	+47.6
40-44	42.3	11.8	12.7	33.2	75.5	+76.5
45-49	47.2	6.9	15.3	30.6	77.8	+71.8
50-54	50.1	4.0	21.0	24.9	75.0	+61.2
55-59	52.7	1.4	24.9	21.0	73.8	+54.0
60-64	52.7	1.4	31.6	14.3	67.0	+41.5
65-69	54.1	0.0	38.4	7.5	61.6	+29.0
70-74	54.1	0.0	42.3	3.6	57.7	+21.8
75-79	54.1	0.0	44.2	1.7	55.8	+18.4
80-84	54.1	0.0	45.2	0.7	54.8	+16.4
85-89	54.1	0.0	45.9	0.0	54.1	+15.2
90-94	54.1	0.0	45.9	0.0	54.1	+15.2
95-100	54.1	0.0	45.9	0.0	54.1	+15.2

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

Figure 12 (150% of the national line): Households below the poverty line and all households at a given score or at or below a given score cut-off, scorecard applied to 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	0.4	100.0	0.8	Only poor targeted
5-9	0.5	100.0	1.0	Only poor targeted
10-14	1.3	86.3	2.1	6.3:1
15-19	5.2	96.5	9.4	27.4:1
20-24	11.4	95.9	20.3	23.5:1
25-29	19.3	89.6	32.0	8.7:1
30-34	30.9	83.5	47.7	5.1:1
35-39	43.9	81.8	66.4	4.5:1
40-44	55.0	76.9	78.2	3.3:1
45-49	62.4	75.6	87.2	3.1:1
50-54	71.1	70.5	92.6	2.4:1
55-59	77.6	67.9	97.5	2.1:1
60-64	84.4	62.5	97.5	1.7:1
65-69	92.5	58.5	100.0	1.4:1
70-74	96.4	56.1	100.0	1.3:1
75-79	98.3	55.1	100.0	1.2:1
80-84	99.3	54.5	100.0	1.2:1
85-89	100.0	54.1	100.0	1.2:1
90-94	100.0	54.1	100.0	1.2:1
95-100	100.0	54.1	100.0	1.2:1

200% of the National Poverty Line

Figure 4 (200% of the 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	95.6
25-29	96.8
30-34	90.3
35-39	88.3
40-44	82.0
45-49	73.5
50-54	57.4
55-59	58.2
60-64	46.4
65-69	33.7
70-74	20.2
75-79	21.4
80-84	0.0
85-89	0.0
90-94	0.0
95-100	0.0

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

Score	Difference between estimate and true value			
	Diff.	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	+20.7	5.8	6.8	8.7
15-19	+0.0	0.0	0.0	0.0
20-24	-4.4	2.2	2.2	2.2
25-29	+6.3	1.2	1.5	1.9
30-34	-0.8	1.1	1.3	1.7
35-39	+1.7	1.2	1.4	1.8
40-44	+15.1	2.0	2.2	3.0
45-49	-8.2	5.0	5.2	5.6
50-54	-5.3	3.8	4.0	4.4
55-59	+5.9	2.6	3.0	4.1
60-64	+11.1	2.5	3.0	3.9
65-69	-1.6	2.2	2.6	3.2
70-74	+12.9	1.6	1.9	2.4
75-79	+21.4	0.0	0.0	0.0
80-84	+0.0	0.0	0.0	0.0
85-89	+0.0	0.0	0.0	0.0
90-94	+0.0	0.0	0.0	0.0
95-100	+0.0	0.0	0.0	0.0

Figure 9 (200% of the 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	+2.2	67.8	77.3	81.6
4	+2.1	31.2	38.4	49.2
8	+2.6	22.8	27.6	35.9
16	+3.0	16.4	19.2	24.8
32	+2.9	11.3	13.4	18.2
64	+3.0	8.2	10.1	13.2
128	+3.1	5.9	7.0	9.4
256	+3.0	4.1	4.9	6.7
512	+3.0	2.9	3.4	4.8
1,024	+2.9	2.0	2.4	3.4
2,048	+3.0	1.5	1.8	2.4
4,096	+3.0	1.1	1.2	1.6
8,192	+3.0	0.7	0.9	1.1
16,384	+3.0	0.5	0.6	0.8

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

Score	Inclusion:	Undercoverage:	Leakage:	Exclusion:	Total Accuracy	BPAC
	< poverty line correctly targeted	< poverty line mistakenly non-targeted	=> poverty line mistakenly targeted	=> poverty line correctly non-targeted	Inclusion + Exclusion	See text
0-4	0.4	67.2	0.0	32.4	32.8	-98.7
5-9	0.5	67.1	0.0	32.4	32.9	-98.4
10-14	1.2	66.5	0.2	32.2	33.3	-96.3
15-19	5.1	62.6	0.2	32.2	37.2	-84.8
20-24	11.3	56.4	0.2	32.2	43.4	-66.4
25-29	18.3	49.3	1.0	31.4	49.7	-44.4
30-34	28.9	38.7	2.0	30.4	59.3	-11.5
35-39	40.0	27.6	3.9	28.4	68.5	+24.1
40-44	47.8	19.9	7.3	25.1	72.9	+52.0
45-49	53.6	14.0	8.8	23.6	77.2	+71.6
50-54	59.0	8.6	12.1	20.3	79.3	+82.1
55-59	62.3	5.4	15.3	17.0	79.3	+77.3
60-64	64.5	3.1	19.9	12.5	77.0	+70.7
65-69	67.3	0.4	25.2	7.1	74.4	+62.7
70-74	67.6	0.0	28.8	3.6	71.2	+57.5
75-79	67.6	0.0	30.6	1.7	69.4	+54.7
80-84	67.6	0.0	31.7	0.7	68.3	+53.2
85-89	67.6	0.0	32.4	0.0	67.6	+52.2
90-94	67.6	0.0	32.4	0.0	67.6	+52.2
95-100	67.6	0.0	32.4	0.0	67.6	+52.2

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

Figure 12 (200% of the national line): Households below the poverty line and all households at a given score or at or below a given score cut-off, scorecard applied to 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	0.4	100.0	0.7	Only poor targeted
5-9	0.5	100.0	0.8	Only poor targeted
10-14	1.3	86.3	1.7	6.3:1
15-19	5.2	96.5	7.5	27.4:1
20-24	11.4	98.4	16.6	61.0:1
25-29	19.3	94.9	27.1	18.7:1
30-34	30.9	93.6	42.8	14.5:1
35-39	43.9	91.1	59.2	10.2:1
40-44	55.0	86.8	70.6	6.6:1
45-49	62.4	85.9	79.3	6.1:1
50-54	71.1	83.0	87.2	4.9:1
55-59	77.6	80.3	92.1	4.1:1
60-64	84.4	76.5	95.4	3.2:1
65-69	92.5	72.7	99.5	2.7:1
70-74	96.4	70.2	100.0	2.4:1
75-79	98.3	68.8	100.0	2.2:1
80-84	99.3	68.1	100.0	2.1:1
85-89	100.0	67.6	100.0	2.1:1
90-94	100.0	67.6	100.0	2.1:1
95-100	100.0	67.6	100.0	2.1:1

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	0.0
10–14	73.7
15–19	50.6
20–24	65.5
25–29	23.5
30–34	22.7
35–39	5.1
40–44	8.3
45–49	0.0
50–54	5.2
55–59	0.0
60–64	0.0
65–69	0.0
70–74	0.0
75–79	0.0
80–84	0.0
85–89	0.0
90–94	0.0
95–100	0.0

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

Score	Difference between estimate and true value			
	Diff.	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	-5.6	5.8	6.8	8.7
15-19	-21.8	12.3	12.6	13.1
20-24	+15.3	2.6	3.2	3.9
25-29	+6.0	1.8	2.0	2.7
30-34	+4.7	1.5	1.7	2.1
35-39	-3.8	2.4	2.5	2.7
40-44	+8.1	0.1	0.1	0.2
45-49	-13.4	7.5	7.7	8.0
50-54	-0.6	1.1	1.3	1.8
55-59	-8.3	4.9	5.1	5.3
60-64	+0.0	0.0	0.0	0.0
65-69	+0.0	0.0	0.0	0.0
70-74	+0.0	0.0	0.0	0.0
75-79	+0.0	0.0	0.0	0.0
80-84	+0.0	0.0	0.0	0.0
85-89	+0.0	0.0	0.0	0.0
90-94	+0.0	0.0	0.0	0.0
95-100	+0.0	0.0	0.0	0.0

Figure 9 (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	+0.1	50.4	80.2	82.7
4	-0.4	26.5	31.6	44.9
8	-0.2	18.4	22.5	29.8
16	+0.2	12.0	14.9	19.4
32	-0.1	8.6	10.7	15.3
64	-0.0	6.5	7.7	10.6
128	-0.1	4.5	5.5	7.7
256	-0.1	3.3	3.7	5.1
512	-0.2	2.2	2.7	3.5
1,024	-0.1	1.6	1.9	2.4
2,048	-0.1	1.1	1.4	1.8
4,096	-0.1	0.8	1.0	1.3
8,192	-0.1	0.6	0.7	0.9
16,384	-0.1	0.4	0.5	0.6

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:	Undercoverage:	Leakage:	Exclusion:	Total Accuracy	BPAC
	< poverty line correctly targeted	< poverty line mistakenly non-targeted	=> poverty line mistakenly targeted	=> poverty line correctly non-targeted	Inclusion + Exclusion	See text
0-4	0.4	13.5	0.0	86.0	86.5	-93.6
5-9	0.4	13.5	0.1	85.9	86.4	-93.0
10-14	1.1	12.9	0.3	85.8	86.8	-82.7
15-19	3.9	10.0	1.3	84.7	88.7	-34.2
20-24	7.0	7.0	4.5	81.5	88.5	+31.7
25-29	8.4	5.6	10.9	75.1	83.5	+22.0
30-34	10.8	3.1	20.1	66.0	76.8	-43.7
35-39	11.9	2.1	32.0	54.0	65.9	-129.3
40-44	12.0	2.0	43.1	43.0	54.9	-208.2
45-49	13.0	1.0	49.5	36.6	49.5	-254.0
50-54	13.5	0.5	57.6	28.4	41.8	-312.6
55-59	14.0	0.0	63.6	22.4	36.4	-355.4
60-64	14.0	0.0	70.4	15.6	29.6	-403.8
65-69	14.0	0.0	78.5	7.5	21.5	-462.0
70-74	14.0	0.0	82.4	3.6	17.6	-490.0
75-79	14.0	0.0	84.3	1.7	15.7	-503.3
80-84	14.0	0.0	85.4	0.7	14.6	-510.9
85-89	14.0	0.0	86.0	0.0	14.0	-515.7
90-94	14.0	0.0	86.0	0.0	14.0	-515.7
95-100	14.0	0.0	86.0	0.0	14.0	-515.7

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

Figure 12 (USAID “extreme” line): Households below the poverty line and all households at a given score or at or below a given score cut-off, scorecard applied to 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	0.4	100.0	3.2	Only poor targeted
5–9	0.5	83.0	3.2	4.9:1
10–14	1.3	79.5	7.6	3.9:1
15–19	5.2	75.1	28.2	3.0:1
20–24	11.4	60.9	49.8	1.6:1
25–29	19.3	43.5	60.1	0.8:1
30–34	30.9	35.1	77.6	0.5:1
35–39	43.9	27.1	85.2	0.4:1
40–44	55.0	21.8	85.8	0.3:1
45–49	62.4	20.8	92.8	0.3:1
50–54	71.1	18.9	96.3	0.2:1
55–59	77.6	18.0	100.0	0.2:1
60–64	84.4	16.6	100.0	0.2:1
65–69	92.5	15.1	100.0	0.2:1
70–74	96.4	14.5	100.0	0.2:1
75–79	98.3	14.2	100.0	0.2:1
80–84	99.3	14.1	100.0	0.2:1
85–89	100.0	14.0	100.0	0.2:1
90–94	100.0	14.0	100.0	0.2:1
95–100	100.0	14.0	100.0	0.2:1

\$3.75/day 2005 PPP Poverty Line

Figure 4 (\$3.75/day 2005 PPP line): Estimated poverty likelihoods associated with scores

If a household's score is then the likelihood (%) of being below the poverty line is:
0-4	100.0
5-9	100.0
10-14	73.7
15-19	50.6
20-24	65.5
25-29	23.9
30-34	22.7
35-39	10.3
40-44	8.3
45-49	0.0
50-54	5.2
55-59	0.0
60-64	0.0
65-69	0.0
70-74	0.0
75-79	0.0
80-84	0.0
85-89	0.0
90-94	0.0
95-100	0.0

Figure 7 (\$3.75/day 2005 PPP line): Bootstrapped differences between estimated and true household poverty likelihoods with confidence intervals in a large sample ($n = 16,384$), scorecard applied to the validation sample

Score	Difference between estimate and true value			
	Diff.	Confidence interval (+/- percentage points)		
		90-percent	95-percent	99-percent
0-4	+0.0	0.0	0.0	0.0
5-9	+100.0	0.0	0.0	0.0
10-14	-5.6	5.8	6.8	8.7
15-19	-21.8	12.3	12.6	13.1
20-24	+4.9	2.6	3.1	3.8
25-29	-4.5	3.4	3.6	4.1
30-34	-0.1	1.6	1.9	2.4
35-39	+1.5	1.0	1.2	1.5
40-44	+8.1	0.1	0.1	0.2
45-49	-13.4	7.5	7.7	8.0
50-54	-0.6	1.1	1.3	1.8
55-59	-8.3	4.9	5.1	5.3
60-64	+0.0	0.0	0.0	0.0
65-69	+0.0	0.0	0.0	0.0
70-74	+0.0	0.0	0.0	0.0
75-79	+0.0	0.0	0.0	0.0
80-84	+0.0	0.0	0.0	0.0
85-89	+0.0	0.0	0.0	0.0
90-94	+0.0	0.0	0.0	0.0
95-100	+0.0	0.0	0.0	0.0

Figure 9 (\$3.75/day 2005 PPP line): Differences and precision of differences for bootstrapped estimates of poverty rates for groups of households at a point in time, by sample size, scorecard applied to the validation sample

Sample Size n	Difference between estimate and true value			
	Diff.	Confidence interval (+/- percentage points)		
		90-percent	95-percent	99-percent
1	-0.9	50.6	77.6	82.7
4	-1.6	26.5	31.3	42.9
8	-1.5	18.6	22.4	28.2
16	-1.2	12.4	15.4	19.7
32	-1.4	9.3	10.7	14.1
64	-1.4	6.8	7.9	10.6
128	-1.5	4.8	5.7	7.9
256	-1.4	3.3	4.2	5.5
512	-1.5	2.3	2.8	3.9
1,024	-1.4	1.6	1.9	2.6
2,048	-1.4	1.2	1.4	1.9
4,096	-1.4	0.8	1.0	1.3
8,192	-1.4	0.6	0.7	0.9
16,384	-1.4	0.4	0.5	0.6

Figure 11 (\$3.75/day 2005 PPP line): Households by targeting classification and score, along with “Total Accuracy” and BPAC, scorecard applied to the validation sample

Score	Inclusion:	Undercoverage:	Leakage:	Exclusion:	Total Accuracy	BPAC
	< poverty line correctly targeted	< poverty line mistakenly non-targeted	=> poverty line mistakenly targeted	=> poverty line correctly non-targeted	Inclusion + Exclusion	See text
0-4	0.4	15.4	0.0	84.2	84.6	-94.4
5-9	0.4	15.4	0.1	84.1	84.5	-93.8
10-14	1.1	14.8	0.3	83.9	85.0	-84.8
15-19	3.9	11.9	1.3	82.9	86.8	-41.9
20-24	7.6	8.2	3.8	80.4	88.0	+20.5
25-29	9.7	6.1	9.6	74.6	84.3	+39.4
30-34	12.7	3.1	18.2	66.0	78.6	-15.1
35-39	13.8	2.1	30.2	54.0	67.8	-90.7
40-44	13.8	2.0	41.2	43.0	56.8	-160.4
45-49	14.8	1.0	47.6	36.6	51.4	-200.9
50-54	15.3	0.5	55.8	28.4	43.7	-252.5
55-59	15.8	0.0	61.8	22.4	38.2	-290.4
60-64	15.8	0.0	68.5	15.6	31.5	-333.1
65-69	15.8	0.0	76.7	7.5	23.3	-384.5
70-74	15.8	0.0	80.6	3.6	19.4	-409.2
75-79	15.8	0.0	82.4	1.7	17.6	-420.9
80-84	15.8	0.0	83.5	0.7	16.5	-427.6
85-89	15.8	0.0	84.2	0.0	15.8	-431.9
90-94	15.8	0.0	84.2	0.0	15.8	-431.9
95-100	15.8	0.0	84.2	0.0	15.8	-431.9

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

Figure 12 (\$3.75/day 2005 PPP line): Households below the poverty line and all households at a given score or at or below a given score cut-off, scorecard applied to 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	0.4	100.0	2.8	Only poor targeted
5-9	0.5	83.0	2.8	4.9:1
10-14	1.3	79.5	6.7	3.9:1
15-19	5.2	75.1	24.9	3.0:1
20-24	11.4	66.6	48.2	2.0:1
25-29	19.3	50.3	61.4	1.0:1
30-34	30.9	41.1	80.2	0.7:1
35-39	43.9	31.3	86.9	0.5:1
40-44	55.0	25.1	87.4	0.3:1
45-49	62.4	23.7	93.7	0.3:1
50-54	71.1	21.5	96.7	0.3:1
55-59	77.6	20.4	100.0	0.3:1
60-64	84.4	18.8	100.0	0.2:1
65-69	92.5	17.1	100.0	0.2:1
70-74	96.4	16.4	100.0	0.2:1
75-79	98.3	16.1	100.0	0.2:1
80-84	99.3	15.9	100.0	0.2:1
85-89	100.0	15.8	100.0	0.2:1
90-94	100.0	15.8	100.0	0.2:1
95-100	100.0	15.8	100.0	0.2:1

\$5.00/day 2005 PPP Poverty Line

Figure 4 (\$5.00/day 2005 PPP line): Estimated poverty likelihoods associated with scores

If a household's score is then the likelihood (%) of being below the poverty line is:
0-4	100.0
5-9	100.0
10-14	100.0
15-19	75.4
20-24	79.3
25-29	59.5
30-34	39.6
35-39	38.8
40-44	27.1
45-49	8.3
50-54	11.6
55-59	6.0
60-64	4.8
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 (\$5.00/day 2005 PPP line): Bootstrapped differences between estimated and true household poverty likelihoods with confidence intervals in a large sample ($n = 16,384$), scorecard applied to the validation sample

Score	Difference between estimate and true value			
	Diff.	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	+20.7	5.8	6.8	8.7
15-19	-13.1	7.6	7.8	8.0
20-24	-0.3	2.1	2.5	3.4
25-29	+11.3	2.4	2.8	3.5
30-34	-6.8	4.4	4.6	4.8
35-39	+11.6	1.7	1.9	2.7
40-44	+12.6	1.5	1.8	2.3
45-49	-20.5	11.3	11.6	11.9
50-54	-2.1	1.8	2.0	2.4
55-59	-7.0	4.4	4.6	4.8
60-64	+4.8	0.0	0.0	0.0
65-69	+0.0	0.0	0.0	0.0
70-74	+0.0	0.0	0.0	0.0
75-79	+0.0	0.0	0.0	0.0
80-84	+0.0	0.0	0.0	0.0
85-89	+0.0	0.0	0.0	0.0
90-94	+0.0	0.0	0.0	0.0
95-100	+0.0	0.0	0.0	0.0

Figure 9 (\$5.00/day 2005 PPP line): Differences and precision of differences for bootstrapped estimates of poverty rates for groups of households at a point in time, by sample size, scorecard applied to the validation sample

Sample Size n	Difference between estimate and true value			
	Diff.	Confidence interval (+/- percentage points)		
		90-percent	95-percent	99-percent
1	+2.0	60.4	75.6	86.1
4	+0.3	33.8	40.9	50.6
8	+0.8	24.2	27.4	36.0
16	+1.4	17.0	19.2	24.2
32	+1.0	11.9	14.0	18.9
64	+1.0	8.4	10.2	13.2
128	+0.9	6.1	7.3	9.5
256	+0.9	4.3	4.8	6.1
512	+0.8	2.8	3.3	4.3
1,024	+0.8	2.0	2.4	3.1
2,048	+0.8	1.5	1.7	2.3
4,096	+0.8	1.0	1.3	1.7
8,192	+0.8	0.7	0.9	1.1
16,384	+0.8	0.5	0.6	0.8

Figure 11 (\$5.00/day 2005 PPP line): Households by targeting classification and score, along with “Total Accuracy” and BPAC, scorecard applied to the validation sample

Score	Inclusion: < poverty line correctly targeted	Undercoverage: < poverty line mistakenly non-targeted	Leakage: => poverty line mistakenly targeted	Exclusion: => poverty line correctly non-targeted	Total Accuracy Inclusion + Exclusion	BPAC See text
0-4	0.4	27.1	0.0	72.4	72.9	-96.8
5-9	0.5	27.0	0.0	72.4	73.0	-96.1
10-14	1.2	26.4	0.2	72.2	73.4	-90.9
15-19	4.6	23.0	0.7	71.8	76.4	-64.3
20-24	9.4	18.1	2.0	70.4	79.9	-24.3
25-29	13.1	14.5	6.2	66.2	79.3	+17.6
30-34	18.8	8.7	12.1	60.4	79.2	+56.2
35-39	22.0	5.5	21.9	50.5	72.6	+20.6
40-44	23.5	4.0	31.5	40.9	64.5	-14.3
45-49	25.6	2.0	36.8	35.6	61.2	-33.6
50-54	26.8	0.8	44.3	28.1	54.9	-60.7
55-59	27.6	0.0	50.0	22.4	50.0	-81.4
60-64	27.6	0.0	56.8	15.6	43.2	-105.9
65-69	27.6	0.0	64.9	7.5	35.1	-135.4
70-74	27.6	0.0	68.8	3.6	31.2	-149.6
75-79	27.6	0.0	70.7	1.7	29.3	-156.3
80-84	27.6	0.0	71.8	0.7	28.2	-160.2
85-89	27.6	0.0	72.4	0.0	27.6	-162.6
90-94	27.6	0.0	72.4	0.0	27.6	-162.6
95-100	27.6	0.0	72.4	0.0	27.6	-162.6

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

Figure 12 (\$5.00/day 2005 PPP line): Households below the poverty line and all households at a given score or at or below a given score cut-off, scorecard applied to 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	0.4	100.0	1.6	Only poor targeted
5-9	0.5	100.0	1.9	Only poor targeted
10-14	1.3	86.3	4.2	6.3:1
15-19	5.2	87.5	16.7	7.0:1
20-24	11.4	82.5	34.2	4.7:1
25-29	19.3	67.9	47.6	2.1:1
30-34	30.9	61.0	68.3	1.6:1
35-39	43.9	50.2	79.9	1.0:1
40-44	55.0	42.8	85.4	0.7:1
45-49	62.4	41.0	92.8	0.7:1
50-54	71.1	37.7	97.1	0.6:1
55-59	77.6	35.5	100.0	0.6:1
60-64	84.4	32.7	100.0	0.5:1
65-69	92.5	29.8	100.0	0.4:1
70-74	96.4	28.6	100.0	0.4:1
75-79	98.3	28.1	100.0	0.4:1
80-84	99.3	27.8	100.0	0.4:1
85-89	100.0	27.6	100.0	0.4:1
90-94	100.0	27.6	100.0	0.4:1
95-100	100.0	27.6	100.0	0.4:1