Simple Poverty Scorecard® Poverty-Assessment Tool Peru

Mark Schreiner

May 8, 2008

Un índice más actualizado que éste en Castellano está en `SimplePovertyScorecard.com`.
A more-current scorecard than this one is in English at `SimplePovertyScorecard.com`.

Abstract

The Simple Poverty Scorecard-brand poverty-assessment tool uses ten low-cost indicators from Peru's 2003 National Household Survey to estimate the likelihood that a household has income 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 Peru to measure poverty rates, to track changes in poverty rates over time, and to segment clients for targeted services.

Acknowledgements

This paper was funded by the Consultative Group to Assist the Poorest under the CGAP/Ford Social Indicators Project. Data are from Peru's *Instituto Nacional de Estadística e Información*. Thanks go to Malika Anand, Nigel Bigger, Dean Caire, Fielding Chen, Manuel Cortés-Fontcuberta, Frank DeGiovanni, Christiaan Grootaert, Micol Guarneri, Janina León Castillo, Brian McConnell, John Pezzullo, Sergio Olivieri, Tony Sheldon, Don Sillers, Frances Sinha, Jeff Toohig, and Manfred Zeller. "Simple Poverty Scorecard" is a Registered Trademark of Microfinance Risk Management, L.L.C. for its brand of poverty-assessment tools.

Simple Poverty Scorecard® Poverty-Assessment Tool

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

Indicator	Value	Points	Score
1. How many household members are aged 17 or younger?	A. Five or more	0	
	B. Four	6	
	C. Three	13	
	D. Two	23	
	E. One	29	
	F. None	39	
2. What fuel does the household use to cook?	A. Other	0	
	B. Gas, electricity, or does not cook	8	
3. Does the household have a cellular or land-line telephone?	A. No	0	
	B. Yes	13	
4. What is the main construction material for the floors?	A. Dirt	0	
	B. Other	8	
5. If the household farms, how is the majority of its agricultural land watered?	A. Rain, does not farm, none, or no data	0	
	B. Irrigated	3	
6. Does the household own an iron?	A. No	0	
	B. Yes	3	
7. Does any household member work a job that pays monthly?	A. No	0	
	B. Yes	11	
8. Does the household own a blender?	A. No	0	
	B. Yes	6	
9. Where does the household's water come from?	A. Other	0	
	B. Public network in the home or its building	4	
10. Does the household own a color TV?	A. No	0	
	B. Yes	5	

SimplePovertyScorecard.com **Score:**

Simple Poverty Scorecard® Poverty-Assessment Tool Peru

1. Introduction

The Simple Poverty Scorecard poverty-assessment tool is a low-cost way for pro-poor programs in Peru to monitor groups' poverty rates at a point in time, track changes in groups' poverty rates between two points in time, and target services to individuals.

The direct approach to poverty measurement via expenditure surveys is difficult, lengthy, and costly, asking about a long list of consumption items ("How many carrots did you eat last week? If you bought carrots, what price did you pay? If you grew carrots yourself, what price would they have sold for? Now then, how many cabbages did you eat last week? . . .").

In contrast, the indirect approach via the scorecard is simple, quick, and inexpensive. It uses 10 verifiable indicators (such as "Does the household own an iron?" or "What fuel does the household use to cook?") to get a score that is highly correlated with poverty status as measured by the exhaustive expenditure survey.

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

ownership cut-offs or housing indices). Results from these approaches are not comparable across organizations nor across countries, they may be costly, and their accuracy is unknown.

If an organization wants to know what share of its participants are below a poverty line (say, $1/day for the Millenium Development Goals, or what share are among the poorest half below the national poverty line as required of USAID microenterprise grantees, see U.S. Congress, 2004), or if it wants to measure movement across a poverty line (for example, movement across $1/day to report to the Microcredit Summit Campaign), then it needs an expenditure-based, objective tool with known accuracy. While most organizations lack the resources to field expenditure surveys—and even governments cannot survey large shares of all households—many organizations can implement an inexpensive scorecard that can serve for monitoring, management, and targeting.

The statistical approach here aims to be understood by non-specialists. After all, if managers are to adopt the scorecard on their own and apply it to inform their decisions, they must first trust that it works. Transparency and simplicity build trust. Getting "buy-in" matters; proxy means tests and regressions on the "determinants of poverty" have been around for three decades, but they are rarely used to inform decisions, not because they do not work, but because they are presented (when they are presented at all) as tables of regression coefficients incomprehensible to lay people (with cryptic indicator names such as "`HHSIZE_2`", negative values, decimal places, and

standard errors). Thanks to the predictive-modeling phenomenon known as the "flat max", the scorecard is almost as accurate as complex tools.

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 sample-size formula. Although these techniques are simple and/or standard, they have rarely or never been applied to proxy means tests.

The scorecard (Figure 1) is based on Peru's 2003 *Encuesta Nacional de Hogares* (National Household Survey, ENAHO). Indicators are selected to be:

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

All points in the scorecard are non-negative integers, and total scores range from 0 (most likely below a poverty line) to 100 (least likely below a poverty line). Non-specialist field workers can collect data and tally scores on paper in less than 5 minutes.

The scorecard can be used to estimate three basic quantities. First, it can estimate an individual's "poverty likelihood", that is, the probability that the individual has expenditure below a given poverty line.

Second, the scorecard can estimate a group's poverty rate at a point in time. (The "poverty rate" is also known as the "poverty prevalence", "head-count index", or "share below the poverty line"). This is simply the average poverty likelihood among individuals in the group.

Third, the scorecard can estimate changes in the poverty rate for a group between two points in time. This estimate is simply the change in the average poverty likelihood of individuals in the group over time.

The scorecard can also be used for targeting. To help managers choose a targeting cut-off, this paper reports the share of Peru's population who are at or below a given score cut-off and who are also below a given poverty line.

This paper presents a single scorecard (Figure 1) based on Peru's national poverty line and the 2003 ENAHO. Scores from this single scorecard are calibrated to poverty likelihoods for six poverty lines:

- Peru's national poverty line
- Peru's national "food" poverty line
- The "extreme" poverty line used by USAID for microenterprise reporting that divides those below the national poverty line into two equal-sized groups
- $1/day
- $2/day
- $3/day

The accuracy of the scorecard based on the 2003 ENAHO is tested out-of-sample against bootstrapped data from the 2002 and 2004 ENAHO. While all three scoring estimators are unbiased in-sample (that is, they match the true value on average in repeated samples from the same population from which the 2003 ENAHO was drawn), they are biased out-of-sample.

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

status in new future samples will be the same as in the sample used to build the scorecard.[1] Of course, this assumption—ubiquitous and inevitable in predictive modelling—does not hold completely.

Still, for all estimates and poverty lines available for Peru after 2003, bias is smaller than 1 percentage point. Furthermore, for sample sizes of $n = 8{,}192$, these estimators are usually precise to about +/–1 percentage point with 90-percent confidence. (For samples of $n = 512$, they are precise to about +/–4 percentage points.) In fact, for a given sample size, scoring estimates of poverty rates in Peru are sometimes more precise than direct survey estimates.[2]

Section 2 below describes data and poverty lines. Section 3 compares the new scorecard here with six existing tools for Peru. Sections 4 and 5 describe scorecard construction and offer practical guidelines for use. Sections 6 and 7 detail the estimation of individuals' poverty likelihoods and of groups' poverty rates at a point in time. Section 8 discusses estimating changes in poverty rates. Section 9 covers targeting. The final section is a summary.

[1] Bias may also result from changes in data collection, from imperfect adjustment of poverty lines across time or geographic regions, or from sampling variation across expenditure surveys.

[2] In general, accuracy results vary by country, scorecard, and poverty line.

2. Data and poverty lines

The scorecard is based on the 2003 ENAHO. "Panel" households are those scheduled for interviews in more than one year between 2001 and 2004, and they are reserved for testing the accuracy of estimates of changes in poverty rates. Half of the remaining 13,110 "non-panel" households (weighted by household size) are assigned to a "construction" sample used to select indicators and points, while the other half form a "calibration" sample used to associate scores with poverty likelihoods (Figure 2).

Non-panel households in the 2002 and 2004 ENAHO are used in out-of-sample accuracy tests for estimates of individuals' poverty likelihoods, groups' poverty rates at a point in time, and changes in groups' poverty rates between two points in time. Panel households interviewed in 2002/3, 2003/4, or 2002/4 are used in out-of-sample accuracy tests for estimates of changes in groups' poverty rates between two points in time.

Peru has two official poverty lines. The "food" line is based on caloric needs of individual household members, per age and sex. For 2003, the average food line was 3.77 Nuevos Soles/person/day (Figure 3), for a poverty rate of about 24.2 percent (Figure 2).

This paper focuses on the "national" poverty line, which adjusts the food line downwards for economies of size in the household (for example, because of a shared kitchen) and upwards to match the total food plus non-food expenditure observed for households who just meet their caloric needs (Instituto Nacional de Estadística e

Información, 2006). For 2003, the average national poverty line for all of Peru is 6.61 NS/person/day (Figure 3), and the poverty rate is 53.3 percent (Figure 2).

The scorecard is also calibrated to USAID's "extreme" poverty line, defined so as to divide people below the national line into two equal groups. This was implemented by multiplying all household-specific national poverty lines by 0.603 in 2002 and 0.642 in 2003. (National poverty lines are not available for 2004.) On average in 2003, this "extreme" line is 4.33 NS/person/day, for a poverty rate of 27.7 percent.

Finally, the scorecard is calibrated to the international purchase-power parity benchmarks of $1/day, $2/day, and $3/day. Using Sillers (2006) and adjusting for regional cost-of-living by multiplying the country-wide line by the ratio of a given region's average national line to Peru's overall average national line, the 2003 average $1/day line is NS1.98, the $2/day line is NS3.96, and the $3/day line is NS5.95. The corresponding poverty rates for 2003 are 3.85, 26.35, and 48.05 percent. In all but 1 case in Figure 2, the poverty rate for panel households at a point in time is lower—sometimes much lower—than for non-panel households.

In Peru's ENAHO data, poverty rates fell sharply from 2002 to 2004. For non-panel households and $1/day, the poverty rate fell two-thirds in 2002–04, from 6.6 percent to 2.4 percent (Figure 2). The fall for $2/day was 30.2 percent to 20.0 percent (one-third), and for $3/day, it was 51.8 percent to 41.3 percent (one-fifth). Panel households generally had similar decreases.

These three-year decreases in poverty are astounding, and they pose a stringent test for the scorecard. On the one hand, if the decreases are real, then the relationships between indicators and poverty may very well have changed over time, which would decrease scoring's accuracy. On the other hand, if the data are incorrect, then that could only decrease scoring's accuracy. Either way, if scoring still turns out to be accurate, it would be evidence of its robustness and could calm concerns about the need to update scorecards frequently.

Peru has not released household-specific national and food lines for 2004 and after, so the analysis of 2004 here uses only the \$1/day, \$2/day, and \$3/day lines. As discussed below, the 2002 data seem different than the 2003 and 2004 data. Because survey problems can be detected and resolved with time, later rounds should generally be of higher quality. Furthermore, it is more likely that any idiosyncratic problems would affect only one round (2002) than two rounds (2003 and 2004). Furthermore, the scorecard here will be applied in 2008 and beyond, and this future—all else constant—is more likely to resemble 2003 and 2004 than to resemble 2002.

For all these reasons, the analysis here focuses on 2003 and 2004. Results for 2002 are also presented because there are no national, food, and USAID "extreme" poverty lines for 2004 and beyond.

3. Review of poverty-assessment tools for Peru

There are at least six existing poverty-assessment tools for Peru; why one more? First, estimates from the scorecard here are tested out-of-sample, and bias, precision, and sample-size formula are reported. Second, the new scorecard is based on the largest sample, and that sample is nationally representative. Finally, the accuracy of the new scorecard compares well with that of the other tools.

3.1 Grosh and Baker

Grosh and Baker (1995) built the first poverty-assessment tool for Peru. They use data from the 1990 Living Standards Measurement Survey of 1,500 households in Lima (Glewwe and Hall, 1991). The poverty line is set at the 30^{th} percentile of expenditure. Stepwise regression with ordinary least-squares is used to select five simple, verifiable indicators: household size, level of education, and ownership of a telephone, television, and/or car. As is traditional for proxy means tests, the focus is targeting, not estimating poverty rates.

Accuracy is measured as successful "hits" (*coverage* when someone truly below a poverty line is predicted to have per capita expenditure below the line, or *exclusion* when someone truly above a line is predicted to be above) versus unsuccessful "misses" (*undercoverage* when someone truly below a line is predicted to be above, or *leakage* when someone truly above a line is predicted to be below). Grosh and Baker also look at who is mistargeted, and by how far.

Accuracy is overstated to some unknown extent because it is tested "in-sample", that is, using the same data that was used to construct the tool. Bias, precision, and sample-size formula are not reported. Grosh and Baker is a seminal paper in the field, and it is the first to document several key properties of poverty-assessment tools:

- Simple statistical techniques can be almost as accurate as complex ones
- Focusing the tool on poorer segments (supposing those segments can be identified in the first place) can improve accuracy
- Accuracy is robust to households' misrepresentation or to enumerators' errors
- There are rapidly diminishing returns to additional indicators
- Fine-tuning for regional differences has low returns
- "Among all targeting mechanisms, proxy means tests [poverty-assessment tools] produce the best incidence outcomes" (p. 1).

3.2 Meyer, Nagarajan, and Dunn

Meyer, Nagarajan, and Dunn ("MND", 2000) highlight simplicity. The data are from a special-purpose 1997 survey of 700 households in metro Lima (Dunn and Arbuckle, 2001). The poverty line is the then-country-wide national line. Ordinary least-squares is used to estimate per capita expenditure, which is then compared to poverty status from the survey. Three indicators are tested, both individually and jointly: household income (obtained via recall), household size, and a housing index (stories, walls, and roof). Like Grosh and Baker, MND test accuracy in-sample with "hit-or-miss" tables without reporting bias, precision, or sample-size formula.

3.3 Copestake *et al.*

Copestake *et al.* (2005) focus on monitoring poverty accurately and inexpensively. The poverty-assessment tool is constructed from a special-purpose 2001

survey of 1,375 households, some of whom were clients of two microlenders. Accuracy is tested on a 2002 repeat survey of 937 of the original households (Fanning, 2004). This "out-of-sample" test is better than an "in-sample" test because it mimicks how the tool is actually used. Accuracy out-of-sample is about 17 percent less than in-sample.

Copestake *et al.* define poverty in terms of income (NS5.16/person/day in 1997), adjusted for caloric guidelines per age and sex. The tool is constructed using backward stepwise ordinary least-squares, augmented with analyst judgement to ensure that indicators are quantitative and verifiable and that they make sense to users. The 10 indicators include: household size; number of members who are students, self-employed, and/or unemployed; type of floor; cooking fuel; and ownership of refrigerators, VCRs, or cars. As in Grosh and Baker and MND, accuracy is tested with hit-and-miss tables, without bias, precision or sample-size formula.

3.4 Zeller, Alcaraz V., and Johannsen

Zeller, Alcaraz V., and Johannsen ("ZAJ", 2005) discuss more than 20 poverty-assessment tools for Peru, some of them including high-cost indicators such as "Share of food expenditures from total household expenditures" (which, if it could be measured, would eliminate the need for a poverty-assessment tool), "Total value of household assets", and "Average daily per-capita clothing expenditures". ZAJ's Model 9 is the most relevant here, as it uses only indicators available from typical household expenditure surveys (although it still includes some high-cost indicators).

ZAJ conduct a nationally representative expenditure survey of 800 households. They derive a national poverty line by finding, for each of Peru's seven regions in 2004, the income percentile that reproduces regional poverty rates based on expenditure from Peru's 2000 *Encuesta Nacional de Hogares Sobre Medición de Niveles de Vida* and that also matches the national poverty rate in Webb and Fernández (2003). They then use the USAID "extreme" poverty line that defines the poorest half of those under this line.

ZAJ test a wide range of statistical techniques, some estimating expenditure which is then compared to poverty status from the survey, and some estimating poverty likelihood which is then compared to an arbitrary cut-off of 50 percent. Their preferred tool estimates expenditure with quantile regression. They focus on estimating poverty rates at a point in time, and they select indicators using stepwise. Among the Peru poverty-assessment tools reviewed here, ZAJ is the largest (24 indicators) and the most complex (using continuous indicators, averages, squares, medians, and logarithms). ZAJ do not report their tools' points.

Like the others, ZAJ measure accuracy in terms of coverage, exclusion, undercoverage, and leakage. They use in-sample tests and do not report bias, precision, or sample-size formula. They also introduce the Balanced Poverty Assessment Criteria, a measure later adopted as the preferred yardstick for poverty-assessment-tool accuracy by USAID (IRIS Center, 2005a). BPAC considers accuracy both in terms of the poverty rate at a point in time and accuracy in terms of targeting (the formula is discussed later). A higher BPAC means more accuracy; for ZAJ, BPAC is 72.1.

3.5 Johannsen

Johannsen (2006) differs from ZAJ in three ways. First, it classifies a household as "below poverty line" if the percentile of estimated expenditure is below the poverty rate in Peru. Second, it uses the nationally representative 2000 Living Standards Measurement Survey. Third, it follows Schreiner (2006a) in the use of bootstrapped out-of-sample tests to estimate bias and precision. Sample-size formula are not reported. For the 24-indicator tool and the USAID "extreme" line, Johannsen's in-sample BPAC is 65.4. Out-of-sample BPAC is 59.8, a reduction of 8.5 percent.

3.6 IRIS Center

IRIS Center (2007a) is like ZAJ, except that it omits high-cost indicators. Its in-sample BPAC for the USAID "extreme" line is 68.8.

3.7 The scorecard

How is the scorecard here different? In terms of data, it uses the largest sample, and like ZAJ, Johannsen, and IRIS, its data are nationally representative.

In terms of testing, the only other out-of-sample tests are Copestake *et al.* and Johannsen. No other poverty-assessment tool reports sample-size formula, and no one except Johannsen reports bias or precision. The analysis here is the only one to look at estimates for individual poverty likelihoods and at estimates of changes in groups' poverty rates over time (for independent samples and for panels), and it is the only one to measure accuracy for a range of possible targeting cut-offs.

In terms of simplicity, the new scorecard here has 10 indicators (more than NMD and Grosh and Baker, the same as Copestake *et al.*, and fewer than ZAJ, Johannsen, and IRIS), and all indicators are categorical (like Grosh and Baker). Furthermore, the new scorecard has the simplest indicators, the most straightforward derivation, and the simplest weighting scheme.

Finally, the new scorecard is probably about as accurate as ZAJ, Johanssen, and IRIS, the only ones using a similar poverty line. When the new scorecard based on 2003 data is applied out-of-sample to the USAID "extreme" poverty line in the 2002 ENAHO, BPAC is 62.7 (Figure 24). Using non-ENAHO data, ZAJ's in-sample BPAC is 72.1, IRIS's in-sample BPAC is 68.8, and Johannsen's out-of-sample BPAC 59.8. If going from in-sample to out-of-sample causes BPAC in ZAJ and IRIS to fall 8.5 percent (as in Johannsen) to 66.0 and 63.0, or if going from in-sample to out-of-sample causes BPAC to fall 17 percent (as in Copestake *et al.* for non-BPAC accuracy measures) to 59.8 and 57.1, then the new scorecard's accuracy compares well with that of the others.[3]

[3] BPAC tends to increase with the poverty rate. The poverty rate in ZAJ (2005) for the 2004 USAID "extreme" line is 27.9 percent, whereas here it is 28.5 percent for 2002 and 27.7 percent for 2003. This slightly favors the scorecard here. BPAC also depends on the number of score ranges, increasing here to 65.0 with 101 ranges (instead of 20).

4. Scorecard construction

About 150 potential indicators are initially prepared in the areas of:

- Family composition (such as female headship and number of children)
- Housing (such as type of construction and number of rooms)
- Education (such as highest grade completed and school attendance by children)
- Ownership of durable goods (such as land, televisions, and automobiles)

Each indicator is first screened with the entropy-based "uncertainty coefficient" (Goodman and Kruskal, 1979) that measures how well the indicator predicts poverty on its own. Figure 4 lists the top 40, ranked by uncertainty coefficients. Responses for each indicator are ordered starting with those most strongly associated with poverty.

Many indicators in Figure 4 are similar to each other in terms of their association with poverty. For example, most houses with a dirt floor do not have walls of brick or cement blocks, nor do they have a roof of reinforced concrete or wood. If a scorecard includes flooring, then data on the roof and walls do not contribute much. For this reason, many indicators strongly associated with poverty are not in the scorecard, as they add little over and above other indicators that are included.

The scorecard also aims to measure *changes* in poverty through time. Thus, some powerful indicators (such as the highest grade completed by a household member) that are relatively insensitive to changes in poverty are omitted in favor of less-powerful indicators (such as ownership of an iron or a blender) that are more sensitive.

Some indicators are not selected because they are awkward to answer or difficult to verify (such as whether meat was eaten in the past week).

The scorecard itself is built using Logit regression on the construction sample from the 2003 ENAHO (Figure 2). Indicator selection uses both judgment and statistics (forward step-wise based on "c", see below). The first step is to build one scorecard for each candidate indicator, using Logit to derive points. Each scorecard's accuracy is taken as "c", a measure of ability to rank by poverty status (SAS Institute Inc., 2004).

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

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

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

This algorithm is the Logit analogue to the R^2-based stepwise in ZAJ and IRIS Center (2005a and 2005b). Like R^2 in a least-squares regression on expenditure, "c" in a

Logit regression on poverty status is a good measure of global accuracy. The procedure here differs from naïve stepwise in that along with statistical criteria, judgment and non-statistical criteria are also used to select indicators. The use of non-statistical criteria can improve robustness out-of-sample and, more important, helps ensure that indicators are simple and make sense to users.

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

5. 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, 2005b). 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 learn to use it properly (Schreiner, 2002). After all, most reasonable scorecards predict tolerably well, thanks to the empirical phenomenon known as the "flat max" (Hand, 2006; Baesens *et al.*, 2003; Lovie and Lovie, 1986; Kolesar and Showers, 1985; Stillwell, Hutton, and Edwards, 1983; Dawes, 1979; Wainer, 1976; Myers and Forgy, 1963). The bottleneck is less technical and more human, not statistics but organizational change management. Accuracy is easier to achieve than adoption.

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

To this end, the scorecard here fits on one page (Figure 1). The construction process, indicators, and points are simple and transparent. "Extra" work is minimized; non-specialists can compute scores by hand in the field because the scorecard has:

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

A field worker using the paper scorecard would:

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

Of course, field workers must be trained. Quality results depend on quality inputs. If organizations or field workers gather their own data and have an incentive to exaggerate poverty rates (for example, if they are rewarded for having higher poverty rates), then it is wise to implement on-going quality control via data review and random audits (Matul and Kline, 2003).[4] IRIS Center (2007b) and Toohig (2007) are useful nuts-and-bolts guides for budgeting, training field workers and supervisors, logistics, sampling, interviewing, piloting, recording data, and quality control.

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

- Who will do the scoring
- How scores will be recorded

[4] If an organization does not want field workers to know the points associated with indicators, then it is a simple matter to remove the points from the paper scorecard and apply them later in a spreadsheet or database at the central office.

- 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 to participants in the field can be:

- Employees of the organization
- Third-party contractors

Scores 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
- In portable electronic devices in the field and downloaded to a database

The subjects to be scored can be:

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

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

The scorecard's frequency of application can be:

- At in-take 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 interval)
- Each time a field worker visits a participant at home

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

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

An example set of choices were made by BRAC and ASA, two microlenders in Bangladesh (each with 7 million participants) who are applying the Simple Poverty Scorecard tool for Bangladesh (Schreiner, 2006a). Their design is that loan officers in a random sample of branches score all participants each time they visit a homestead as part of their standard due diligence prior to loan disbursement (about once a year). Scores are recorded on paper in the field before being sent to a central office to be entered into a database. ASA's and BRAC's sampling plans cover 50,000–100,000 participants each.

6. Estimates of individual poverty likelihoods

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

To get absolute units, scores must be converted to *poverty likelihoods*, that is, probabilities of being below a poverty line. This is done via simple look-up tables. For the national line, scores of 0–4 have a poverty likelihood of 99.1 percent (Figure 5), and scores of 45–49 have a poverty likelihood of 41.1 percent.

The poverty likelihood associated with a score varies by poverty line. For example, scores of 45–49 are associated with a poverty likelihood of 41.1 percent for the national line but 4.7 percent for the food line.[5]

[5] Starting with Figure 5, most figures have six versions, one for each poverty line. To keep them straight, they are grouped by poverty line. Single tables that pertain to all poverty lines are placed with the tables for the national line.

6.1 Calibrating scores with poverty likelihoods

A given score is associated ("calibrated") with a poverty likelihood non-parametrically by defining the poverty likelihood as the share of people in 100 bootstrapped samples[6] from the 2003 ENAHO calibration sample who have the score and who below a given poverty line. Bootstrapping entails:

- Score each household in the 2003 ENAHO calibration sample
- Draw a new sample *with replacement* from the calibration sample
- For each score and bootstrap sample, compute the share of people who have the score and who also have expenditure below a given poverty line
- Repeat the previous two steps 100 times
- For each score, define the poverty likelihood as the average across the 100 bootstrap samples of the shares of people below a given poverty line

Figure 6 illustrates this for a single hypothetical bootstrap sample. For the example for the national line, there are 5,923 people with a score of 0–4, of whom 5,870 are below the poverty line. The estimated poverty likelihood associated with a score of 0–4 is then 99.1 percent, because 5,870 ÷ 5,923 = 99.1 percent.

As another illustration, with the national line and a score of 45–49, there are 5,439 people in the calibration sample, of whom 2,235 are below the line (Figure 6). Thus, the estimated poverty likelihood for a score of 45–49 is 2,235 ÷ 5,439 = 41.1 percent.

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

[6] Efron and Tibshirani (1993).

Figure 7 shows, for all scores, the likelihood that expenditure falls in a range demarcated by two adjacent poverty lines. For example, the daily expenditure of someone with a score of 35–39 falls in the following ranges with probability:

- 2.0 percent below $1/day
- 14.6 percent between $1/day and the national food line
- 2.4 percent between the national food line and $2/day
- 3.7 percent between $2/day and the USAID "extreme" line
- 28.6 percent between the USAID "extreme" line and $3/day
- 7.2 percent between $3/day and the national line
- 41.6 percent above the national line

The calibration process produces poverty likelihoods that are objective (that is, derived from data and expenditure-based poverty lines) even though the scorecard is constructed partly based on judgment. The poverty likelihoods would be objective even if indicators and/or points were selected without any data at all. In fact, objective scorecards of proven accuracy are often based only on judgment (Fuller, 2006; Caire, 2004; Schreiner *et al.*, 2004). Of course, the scorecard here was constructed with both data and judgement. 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 Peru'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 people 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. This non-parametric calibration can also improve accuracy, especially with large calibration samples.

6.2 Accuracy of estimates of poverty likelihoods

As long as the relationship between indicators and poverty does not change, this calibration process produces unbiased estimates of poverty likelihoods. *Unbiased* means that in repeated samples from the same population, the average estimate matches the true poverty likelihood. The scorecard also produces unbiased estimates of poverty rates at a point in time and of changes in poverty rates between two points in time.[7]

Of course, the relationship between indicators and poverty changes over time, so any scorecard applied out-of-sample (as all are in practice) will generally be biased. Still, estimators that are unbiased in-sample should have less bias out-of-sample.

[7] This follows because these estimates of groups' poverty rates are linear functions of the unbiased estimates of individuals' poverty likelihoods. IRIS also produces unbiased estimates, but none of the other five Peru poverty-assessment tools do.

How accurate are estimates of individual poverty likelihoods? To measure this, the 2003 scorecard is applied to 1,000 bootstrap samples of size $n = 16{,}384$ for non-panel households from 2002 and 2004. The bootstrap process entails:

- Score each non-panel household in the 2002 and 2004 ENAHO
- Draw a new sample *with replacement* from non-panel households in 2002 (or 2004)
- For each score, compute the true poverty likelihood in the bootstrap sample, that is, the share of people with a score and with expenditure below a given poverty line
- For each score, record the difference between the estimated poverty likelihood in Figure 5 and this true poverty likelihood in the bootstrap sample
- Repeat the previous three steps 1,000 times
- For each score, define *bias* as the average difference between estimated and true poverty likelihoods across the 1,000 bootstrap samples
- For each score, define *precision* as the average two-sided interval containing the central 900, 950, or 990 differences between estimated and true poverty likelihoods

For each of the 20 score ranges, Figure 8 shows bias (average difference between estimated and true poverty likelihoods) and precision (average confidence intervals around the estimate).

For the example of the $3/day poverty line (*not* the national line), the average poverty likelihood across bootstrap samples for scores of 0–4 for 2002 was biased upwards by 1.0 percentage points (Figure 8). For a score of 5–9, the estimate is biased downwards by 4.4 percentage points, and for a score of 10–14, the estimate is biased upwards by 1.0 percentage points.

The 90-percent confidence interval for the estimated bias for scores of 0–4 is +/– 0.9 percentage points (Figure 8).[8] This means that in 900 of 1,000 bootstraps, bias is between 0.1 and 1.9 percentage points (because 1.0 – 0.9 = 0.1, and 1.0 + 0.9 = 1.9). In

[8] Confidence intervals are a standard, widely understood measure of precision.

950 of 1,000 bootstraps (95 percent), bias is 1.0 +/–1.1 percentage points, and in 990 of 1,000 bootstraps (99 percent), bias is 1.0 +/– 1.3 percentage points.

The estimated poverty likelihood for almost every score for every poverty line and year are biased, sometimes by a lot (Figure 8).[9] This is because the relationship between indicators and poverty changes over time and because calibration was based on a finite sample. For targeting, however, what matters is less the bias in all score ranges and more the bias in score ranges just above and below the targeting cut-off. This fact greatly reduces the effects of bias on targeting. Section 9 below looks at targeting accuracy in detail.

Of course, if estimates of groups' poverty rates are to be usefully accurate, then errors for individuals must largely cancel out. As discussed later, this is generally what happens, especially for the 2003 scorecard applied to 2004.

Figure 9 (for $1/day, $2/day, and $3/day) shows that bias for the 2003 scorecard is generally greater (and confidence intervals are wider) when applied to 2002 than when applied to 2004. This suggests that the 2002 data is of lower quality.

There are three approaches to mitigating bias. First, poverty likelihoods in Figure 5 could be adjusted to compensate for the biases in Figure 8. For the example of the $3/day line and the 2003 scorecard applied to 2004, Figure 5 shows a poverty likelihood for a score of 0–4 of 99.1, but Figure 8 shows that this is too high by 2.9

[9] Figure 9 summarizes Figure 10 and 11.

percentage points. Changing the poverty likelihood associated with scores of 0–4 to 99.1 – 2.9 = 96.2 would give an unbiased estimate for 2004.

Of course, this approach is not helpful for 2004; if the true poverty likelihoods were known, then the scorecard estimates would not be needed. In practice, what is helpful is a way to use the 2004 data to reduce bias in post-2004 applications. If the 2004 data represent future reality better than data from 2003 (a safe assumption), then this adjustment should indeed reduce—but not eliminate—out-of-sample bias.[10]

A second approach to mitigating bias is to increase the fineness of the points (for example, by allowing points to range from 0 to 200 instead of 0 to 100), to increase the number of ranges into which scores are grouped (for example, from 20 to 40), and/or to increase the number of response categories for given indicators. Of course, all of these approaches add complexity. An experiment with 101 score ranges reduced bias by an average of 20 percent, so grouping scores and rounding points are sources of bias, but they are not the main sources of bias.

By construction, the scorecard here is unbiased in-sample for 2003. But it may still be *overfit* out-of-sample. That is, it may fit the 2003 data so closely that it captures not only some timeless patterns but also some random patterns that, due to sampling variation, show up only in the 2003 data. Or it may be overfit in that it is not robust to changes over time in the relationship between indicators and poverty.

[10] Even better would be to construct the scorecard from 2004 data. This is not done here because the national, food, and USAID poverty lines are not available in 2004.

Overfitting can be mitigated by simplifying the scorecard and not relying completely on the construction data but rather also considering experience, judgment, and theory. Of course, the scorecard here does this. Bootstrapping can also mitigate overfitting by reducing (but not eliminating) dependence on a single sampling instance. The poverty likelihoods here are calibrated using bootstrapping, and scorecard weights in the 10-indicator logit hardly changed when bootstrapped. Combining scorecards can also mitigate overfitting, but that would increase complexity too much in this context.

The third approach to mitigating bias is to do nothing. After all, most errors in individual likelihoods cancel out in the estimates of poverty rates for the 2003 scorecard applied to 2004 (see following sections). Furthermore, up to 80 percent of bias may come from external sources such as changes in the relationship between indicators and poverty, sampling variation, changes in poverty lines, and inconsistencies 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) or by reducing overfitting (which likely has limited returns, given the scorecard's current parsimony).

7. Estimates of group poverty rates 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 individuals in the group.

To illustrate, suppose a program samples three participants on Jan. 1, 2007 and that they have scores of 20, 30, and 40, corresponding to poverty likelihoods of 85.0, 74.8, and 60.0 percent (national line, Figure 5). The group's estimated poverty rate is the participants' average poverty likelihood of $(85.0 + 74.8 + 60.0) \div 3 = 73.3$ percent.[11]

7.1 Accuracy of estimated poverty rates at a point in time

How accurate is this estimate? For a range of sample sizes, Figure 13 reports bias (average difference between estimated and true poverty rates) and precision (average confidence intervals around the estimated bias) for the 2003 scorecard applied to 1,000 bootstrap samples of non-panel households in 2002. The scorecard is biased downward by about 2.4 percentage points ; it estimates a poverty rate of 54.7 percent, but the true value for 2002 is 57.1 percent (Figure 2). For all poverty lines, bias for 2002 is negative and never smaller than 1.0 percentage points (Figure 12).[12]

The presence of some bias is not a surprise; poverty rates in Peru, according to ENAHO, fall steeply between 2002 and 2003; for the national line, the decrease is 3.8 percentage points. The scorecard assumes that the relationship between indicators and

[11] The group's poverty rate is not the poverty likelihood associated with the average score. Here, the average score is $(20 + 30 + 40) \div 3 = 30$, so the poverty likelihood associated with the average score is 74.6 percent. This is not the 73.1 percent found as the average of the three poverty likelihoods associated with each of the three scores.
[12] Figure 12 summarizes Figures 13 and 14.

poverty is the same in 2002 and 2003, but the steep fall in poverty suggests that this relationship may have changed; if so, the scorecard could very well miss it.

But 2004 tells a different story. Poverty again falls steeply—per ENAHO—but bias when applying the 2003 scorecard to 2004 is always +/–0.3 percentage points or less (Figure 12).[13] This could be due to sampling variation in the ENAHO data, due to the relationship between indicators and poverty changing in 2002/3 but then not changing in 2003/4, or due to unintended survey differences. The fact that estimated changes in groups' poverty rates (see below) are more accurate for 2003/4 than for 2002/3 or 2002/4 suggests that the 2002 data are cut from a different cloth.

In terms of precision, the 90-percent confidence interval for the estimated poverty rate at a point in time (2002) with $n = 16{,}384$ is +/–0.7 percentage points (Figure 13). This means that in 900 of 1,000 bootstrap samples, the estimate is between 54.7 – 0.7 = 54.0 percent and 54.7 + 0.7 = 55.4 percent. For $n = 1{,}024$, the 90-percent interval is +/–2.7 percentage points. The 2003 scorecard is more precise (that is, the confidence intervals are smaller) when applied to 2004 than to 2002 (Figure 12).

[13] The only poverty lines available for both 2003 and 2004 are $1, $2, and $3/day.

7.2 Sample-size formula for estimates of poverty rates at a point in time

How many participants should an organization sample if it wants to estimate their poverty rate at a point in time for a desired confidence interval and confidence level? This practical question has yet to be addressed, for Peru or elsewhere.[14]

With direct measurement, the poverty rate can be estimated as the number of people observed to be below the poverty line, divided by the number of all observed people. The formula for sample size n is then (Cochran, 1977):

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

where

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

c is the confidence interval as a proportion (for example, 0.02 for an interval of +/–2 percentage points), and

$\hat{p}$ is the expected (before measurement) proportion of people below the poverty line.

[14] IRIS Center (2007b and 2007c) says that $n = 300$ is sufficient to meet the USAID microenterprise reporting requirements. 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 about +/– 2.2 percentage points. In fact, USAID has not specified confidence levels or confidence 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.

Scorecards, however, do not measure poverty directly, so this formula is not applicable. To derive a similar sample-size formula for the Peru scorecard, consider the national poverty line and the 2003 scorecard applied to 2002. Figure 2 shows that the expected (before measurement) poverty rate $\hat{p}$ is 0.533 (that is, the poverty rate in 2003). In turn, a sample size n of 16,384 and a 90-percent confidence level correspond to a confidence interval of +/–0.68 percentage points (Figure 13).[15] Plugging these into the direct-measurement sample-size formula (1) above gives not $n = 16{,}384$ but rather $n = \left(\frac{1.64}{0.0068}\right)^2 \cdot 0.533 \cdot (1 - 0.533) = 14{,}479$. The ratio of the sample size for scoring (derived empirically) to the sample size for direct measurement (derived from theory) is 16,384 ÷ 14,479 = 1.13.

Applying the same method to $n = 8{,}192$ gives $n = \left(\frac{1.64}{0.0097}\right)^2 \cdot 0.533 \cdot (1 - 0.533) =$ 7,116. This time, the ratio of the sample size using scoring to the sample size using direct measurement is 8,192 ÷ 7,116 = 1.15. This ratio of 1.15 for $n = 8{,}192$ is close to the ratio of 1.13 for $n = 16{,}384$. Indeed, applying this same procedure for all $n \geq 256$ in Figure 13 gives ratios that average to 1.09. This can be used to define a sample-size formula for the 2003 Peru scorecard applied to 2002:

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

where $\alpha = 1.09$ and z, c, and $\hat{p}$ are defined as in (1) above.

[15] Due to rounding, Figure 13 displays 0.7, not 0.68.

To illustrate, if $c = 0.053$ (confidence interval of +/– 5.3 percentage points) and $z = 1.64$ (90-percent confidence), then (2) gives $n = 1.09 \cdot \left(\frac{1.64}{0.053} \right)^2 \cdot 0.533 \cdot (1 - 0.533) =$ 260, which is close to the sample size of 256 for these parameters in Figure 13.

When the sample-size factor α is less than 1.0, it means that the scorecard is more precise than direct measurement. α is less than 1.0 for three of nine estimates of groups' poverty rates at a point in time (Figure 12), all of them for the 2003 scorecard when applied to 2004, again suggesting that the 2002 data are different (and probably of lower-quality).

Of course, the sample-size formula here are specific to Peru, its poverty lines, its poverty rates, and this scorecard. The derivation method, however, is valid for any poverty-assessment tool following the approach in this paper.

In practice, an organization would select a poverty line (say, Peru's national poverty line), select a desired confidence level (say, 90 percent, or $z = 1.64$), select a desired confidence interval (say, +/– 2 percentage points, or $c = 0.02$), make an assumption about $\hat{p}$ (perhaps based on a previous measurement or national figures), assume that the scorecard works out-of-sample,[16] and compute the required sample size. In this illustration, $n = 1.09 \cdot \left(\frac{1.64}{0.02} \right)^2 0.533 \cdot (1 - 0.533) = 1{,}824$.

[16] This paper reports accuracy for the scorecard applied to 2002 and 2004, but it cannot test accuracy for later years. Still, performance after 2004 will most likely resemble performance in 2004, with some deterioriation as time passes.

If the scorecard has already been applied to a sample n, then $\hat{p}$ is the scorecard's estimated poverty rate and the confidence interval c is $+/-z \cdot \sqrt{\frac{\alpha \cdot \hat{p} \cdot (1-\hat{p})}{n}}$.

7.3 Accuracy for scoring versus surveys

How does accuracy for indirect measurement via scoring compare with direct measurement via surveys? On the one hand, direct measurement is more accurate because it is unbiased. On the other hand, the scorecard is sometimes more accurate because it is sometimes more precise.

A common way to compare a biased-but-more-precise estimator (indirect scoring) with an unbiased-but-less-precise estimator (direct survey) is mean-squared error, defined as the average (in repeated samples) of the *squared* differences between the estimate and the true value. For estimates of groups' poverty rates at a point in time, direct measurement has lower mean-squared error than scoring in 7 of 9 cases (Figure 12). The two cases where the scorecard does as well or better are for the 2003 scorecard applied to 2004.

Because the scorecard is constructed from data from direct measurement, it is not surprising that scoring is usually less accurate. Indeed, what is remarkable is that scoring can sometimes do better.

8. Estimates of changes in group poverty rates over time

The change in a group's poverty rate between two points in time is estimated as the change in the average poverty likelihood of the individuals in the group.

8.1 Warning: Change is not impact

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

8.2 Calculating estimated changes in poverty rates over time

Consider the illustration begun in the previous section. On Jan. 1, 2007, a program samples three participants who score 20, 30, and 40 and so have poverty likelihoods of 85.0, 74.8, and 60.0 percent (national line, Figure 5). The group's baseline

estimated poverty rate is the participants' average poverty likelihood of (85.0 + 74.8 + 60.0) ÷ 3 = 73.3 percent.

After baseline, two sampling approaches are possible:

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

By way of illustration, suppose that on Jan. 1, 2008, the program samples three additional people who are in the same cohort as the three people originally sampled (or scores the same three original people) and finds that their scores are 26, 35, and 45 (poverty likelihoods of 78.0, 58.4, and 41.1 percent, national line, Figure 5). Their average poverty likelihood at follow-up is now (78.0 + 58.4 + 41.1) ÷ 3 = 59.2 percent, an improvement of 73.3 – 59.2 = 14.1 percentage points.

This suggests that about 14 of 100 participants crossed the poverty line in 2007.[17] Among those who started below the line, about one in five (14.1 ÷ 73.3 = 19.2 percent) ended up above the line.[18]

8.3 Accuracy for estimated change for two independent samples

Figures 15–18 report bias and precision for the 2003 scorecard applied to 1,000 bootstrap samples of non-panel households from Peru's ENAHO in year-pairs 2002/3, 2003/4, and 2002/4. In each bootstrap, one sample is drawn from each year. Tests involving 2003 are partly in-sample, so their accuracy is overstated to some degree.

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

8.3.1 Bias

In terms of bias, year-pairs involving 2002 are much less accurate, with estimates being too high, in most cases by more than 2.0 percentage points (Figure 15). For example, change for the national line is estimated at –1.1 percentage points, but its true value is –3.8 percentage points (Figure 2), a bias of 2.7 percentage points.

For 2003/4, bias for the $1, $2, and $3/day lines is always less than 0.3 percentage points (Figure 15). This once again suggests that data from 2003 and 2004 are better than 2002.

8.3.2 Precision

In terms of precision, the 90-percent confidence intervals for $n = 16{,}384$ are always +/–1.0 percentage points or less, with estimated changes between 2003/4 being slightly more precise (Figure 15).

The 90-percent intervals for $n = 1{,}024$ in Figures 16–18 are about 4 times the interval for $n = 16{,}384$. This is a general property of all the estimators presented here; quadrupling sample size cuts confidence intervals in half.[19]

8.3.3 Sample-size formula

Under direct measurement, the sample-size formula for estimates of changes in poverty rates in two equal-sized independent samples is:

$$n = 2 \cdot \left(\frac{z}{c} \right)^2 \cdot \hat{p} \cdot (1 - \hat{p}) \,, \tag{3}$$

[19] Equivalently, multiplying sample size by 16 divides the confidence interval by 4.

where z, c, and $\hat{p}$ are defined as in (1). Before measurement, $\hat{p}$ is assumed equal at both baseline and follow-up. n is the sample size at both baseline and follow-up.[20]

The method developed in the previous section can be used again to derive a sample-size formula for indirect measurement via scorecards:

$$n = \alpha \cdot 2 \cdot \left(\frac{z}{c}\right)^2 \cdot \hat{p} \cdot (1 - \hat{p}) \,. \tag{4}$$

As before, α is the average across sample sizes ≥ 256 of the ratio between the empirical sample size required under scoring for a given precision and the theoretical sample size required under direct measurement. In Figure 15, α is less than 1.0 in 2 of 12 cases, so scoring is again usually less precise than direct measurement.

To illustrate how to use (4) 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 \$3/day, and the results for 2003/4 are used to project out-of-sample (so $\alpha = 0.96$ from Figure 15, and $\hat{p} = 0.413$ from Figure 2). Then baseline sample size is $n = 0.96 \cdot 2 \cdot \left(\frac{1.64}{0.02}\right)^2 \cdot 0.413 \cdot (1 - 0.413) = 3{,}130$, and follow-up sample size is also 3,130.

8.3.4 Accuracy for scoring versus direct measurement

As before, scoring is always less accurate than direct measurement in terms of bias but sometimes more accurate in terms of precision. If the bias/precision trade-off is

[20] This means that, for a given precision, estimating the change in a poverty rate between two points in time requires 4 times as many measurements (not twice as many) as does estimating a poverty rate at a point in time.

evaluated in terms of mean-squared error, then direct measurement is usually more accurate, as 10 of 12 ratios of MSE in Figure 15 exceed 1.0.

8.4 Accuracy for estimated change for one sample, scored twice

Figures 19–22 report bias and precision for the 2003 scorecard applied to panel households in 2002/3, 2003/4, and 2002/4. In each of 1,000 bootstraps, one sample is drawn from panel households in the earlier year, and households in this single sample are then scored in both years. The tests are out-of-sample because panel households are excluded from scorecard construction.

8.4.1 Bias

As usual, estimated change for panel samples for year-pairs involving 2002 are more biased (Figure 19), although the bias is smaller than for point-in-time estimates and is quite small for 2002/4. For example, change for the national line for 2002/3 is estimated at –1.7 percentage points, but its true value is –2.9 percentage points (Figure 2), a bias of 1.2 percentage points.

For 2003/4, bias is always less than 0.3 percentage points (Figure 19). In general, bias in estimated changes is usually smaller with panel data than with independent samples (Figures 15 and 19).

8.4.2 Precision

Panel-data estimates also have better precision than independent-sample estimates. All 90-percent confidence intervals in Figure 19 are less than +/– 1.0 percentage points. Estimated changes for 2003/4 are the most precise.

8.4.3 Sample-size formula

Under direct measurement, the sample-size formula for estimates of changes in poverty rates in a single sample scored twice is:[21]

$$n = \left(\frac{z}{c}\right)^2 \cdot \left[\hat{p}_{12} \cdot (1 - \hat{p}_{12}) + \hat{p}_{21} \cdot (1 - \hat{p}_{21}) + 2 \cdot \hat{p}_{12} \cdot \hat{p}_{21}\right], \quad (5)$$

where z and c are defined as in (1), $\hat{p}_{12}$ is the expected (before measurement) share of all sampled cases that move from below the poverty line to above it, and $\hat{p}_{21}$ is the expected share of all sampled cases that move from above the line to below it.

How can a user set $\hat{p}_{12}$ and $\hat{p}_{21}$? Before measurement, a reasonable assumption is that the change in poverty rate is zero. Then $\hat{p}_{12} = \hat{p}_{21}$ and (5) becomes:

$$n = 2 \cdot \left(\frac{z}{c}\right)^2 \hat{p}_*, \quad (6)$$

where $\hat{p}_* = \hat{p}_{12} = \hat{p}_{21}$.

Still, $\hat{p}_*$ could take any value between 0 and 1, so (6) cannot determine sample size. The estimate of $\hat{p}_*$ must be based on data available before baseline measurement.

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

For the three poverty lines in the year-pair 2003/4, it turns out that the observed relationship between $\hat{p}_*$ and the variance of the baseline poverty rate $p_{baseline} \cdot (1 - p_{baseline})$ is close to $\hat{p}_* = 0.0085 + 0.206 \cdot [p_{baseline} \cdot (1 - p_{baseline})]$.[22] Of course, $p_{baseline}$ is not known before the measurement, but it is reasonable to use as its expected value the observed poverty rate from the previous year. Given this and a poverty line, a sample-size formula for a single sample directly measured twice for Peru after 2002 is:

$$n = 2 \cdot \left(\frac{z}{c}\right)^2 \cdot \{0.0085 + 0.206 \cdot [p_{2002} \cdot (1 - p_{2002})]\}. \quad (7)$$

As usual, (7) is modified with α to get the scorecard sample-size formula:

$$n = \alpha \cdot 2 \cdot \left(\frac{z}{c}\right)^2 \cdot \{0.0085 + 0.206 \cdot [p_{2002} \cdot (1 - p_{2002})]\}. \quad (8)$$

In Figure 19, α is less than 1.0 in 9 of 12 year-pairs, and α is generally loan for the higher-quality year-pair 2003/4. Thus, the scorecard is again usually less precise than direct measurement. Furthermore, the tests are in-sample in that p_{2002} is unknown before baseline in 2002, and the 2003/4 data are used to derive the relationship between $\hat{p}_*$ and the variance of pre-baseline poverty.

To illustrate the use of (8), 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 \$3/day, and the panel will be scored in 2003/4. The before-baseline poverty rate

[22] The relationship between $\hat{p}_*$ and $p_{baseline} \cdot (1 - p_{baseline})$ is also about linear for poverty lines in 2002/3 and 2002/4 as $\hat{p}_* = 0.008 + 0.425 \cdot [p_{baseline} \cdot (1 - p_{baseline})]$. Because the 2003 and 2004 data are better and more recent, 2002/3 and 2002/4 are not used here.

is 45.0 percent (p_{2002}=0.450, Figure 2), and α = 1.14 (Figure 19). Then baseline sample size is $n = 1.14 \cdot 2 \cdot \left(\frac{1.64}{0.02}\right)^2 \cdot \{0.0085 + 0.206 \cdot [0.450 \cdot (1 - 0.450)]\} = 912$. Of course, n = 912 for the follow-up sample as well.

Estimating change is more precise with panel data than with two independent samples. For example, the illustrations show that scoring a panel of 912 gives the same precision (and usually less bias) as scoring two independent samples of 3,130 each.

8.4.4 Accuracy for the scorecard versus direct measurement

In terms of mean-squared error (Figure 19), the scorecard is as accurate or better than direct measurement in half the cases. This is remarkable.

9. Targeting

When a program uses the scorecard for targeting, people 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. People with higher scores are *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* (expenditure below a poverty line). Poverty status is a fact that depends on whether expenditure is below a poverty line as directly measured by a survey. In contrast, targeting status is a program's policy choice that depends on a cut-off and an indirect estimate from a scorecard.

Targeting is successful when people truly below a poverty line are targeted (*inclusion*) and people truly above a poverty line are not targeted (*exclusion*). Of course, no scorecard is perfect, and targeting is unsuccessful when people truly below a poverty line are not targeted (*undercoverage*) or people truly above a poverty line are targeted (*leakage*). Figure 23 illustrates 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).

Figures 24 and 25 show the percentage of people by targeting outcome for the 2003 scorecard applied to 2002 and 2004. Given an example cut-off of 15–19, outcomes for the national poverty line applied to all people in 2002 are:

- Inclusion: 21.7 percent are below the line and correctly targeted
- Undercoverage: 35.3 percent are below the line and mistakenly not targeted
- Leakage: 1.4 percent are above the line and mistakenly targeted
- Exclusion: 41.5 percent are above the line and correctly not targeted

Increasing the cut-off to 20–24 improves inclusion and undercoverage but worsens leakage and exclusion:

- Inclusion: 28.3 percent are below the line and correctly targeted
- Undercoverage: 28.8 percent are below the line and mistakenly not targeted
- Leakage: 2.3 percent are above the line and mistakenly targeted
- Exclusion: 40.7 percent are above the line and correctly not targeted

Which cut-off is preferred depends on total net benefit. Suppose each targeting outcome has a per-person benefit or cost. Then total net benefit for a given cut-off is:

Benefit per person correctly included	x	People correctly included	+
Cost per person mistakenly not covered	x	People mistakenly not covered	+
Cost per person mistakenly leaked	x	People mistakenly leaked	+
Benefit per person correctly excluded	x	People correctly excluded.	

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

- Assign benefits and costs to possible outcomes, based on its values and mission
- Tally total net benefits for each cut-off using Figure 24 or 25 for a 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, 2005a).[23] With this, total net benefit is the number of people correctly included or excluded:

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

Figures 24 and 25 show "Total Accuracy" for all cut-offs for the 2003 Peru scorecard applied to 2002 and 2004. For the national line in 2002, total net benefit is greatest (81.2) for a cut-off of 45–49; that cut-off would correctly classify about four in five Peruvians.

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

Beyond "Total Accuracy", IRIS (2005a) proposes a new yardstick called the "Balanced Poverty Accuracy Criterion". BPAC considers two goals:[24]

- Inclusion
- Unbiasedness of the estimated poverty rate

[23] Grootaert and Braithwaite (1998) and ZAJ use this criterion with poverty-assessment tools.

[24] A criterion must consider at least two outcomes among inclusion, undercoverage, leakage, and exclusion. If not, it would imply targeting everyone or no one.

For scorecards that estimate expenditure rather than poverty likelihood, the second goal is optimized by minimizing the absolute value of the difference between undercoverage and leakage. After normalizing by the number of people below the poverty line, the BPAC formula is:

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

BPAC is mostly relevant for poverty-assessment tools—like the other six reviewed here—that estimate expenditure rather than poverty likelihood. This is because when estimating expenditure, the bias of a group's estimated poverty rate is the difference between undercoverage and leakage.

BPAC is less relevant, however, for scorecards (like the one here) that estimate poverty likelihoods. In this case, a group's estimated poverty rate is the average of its members' poverty likelihoods, and this is independent of undercoverage and leakage (which in any case depend on a program-selected cut-off). Nevertheless, the new scorecard's out-of-sample BPAC—as noted earlier—compares well with that of others.

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. Figures 26 and 27 show, for the 2003 Peru scorecard applied to 2002 and 2004, the expected poverty rate among people who score at or below all possible cut-offs. For the example of the national poverty line in 2002,

targeting people who score 45–49 or less would lead to a poverty rate of 80.3 percent and would mean targeting 63.1 percent of all Peruvians.[25]

[25] If some people are not potential participants, then Figures 30 and 31 are valid only if selection into potential participation—whether by the program or the potential participant—is unrelated with poverty status in any way not captured by the scorecard.

10. Conclusion

This paper presents the scorecard. Pro-poor programs in Peru can use it to estimate the likelihood that an individual has expenditure below a given poverty line, to estimate a group's poverty rate at a point in time, and to estimate changes in a group's poverty rate between two points in time. The scorecard can also be used for targeting.

The scorecard is inexpensive to implement 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 so as to speed up their participants' progress out of poverty.

The scorecard is built with data from Peru's 2003 ENAHO, tested out-of-sample on the 2002 and 2004 ENAHO, and calibrated to six poverty lines (national, food, USAID "extreme", $1/day, $2/day, and $3/day).

Bias, precision, and sample-size formula are reported for out-of-sample estimates of individuals' poverty likelihoods, groups' poverty rates at a point in time, and changes in groups' poverty rates over time. Accuracy of estimated changes in poverty rates are better with a single sample scored twice than with two independent samples. Of course, scorecard estimates of change are not the same as estimates of program impact.

When the 2003 scorecard is applied out-of-sample in 2004, bias is less than one percentage point. For $n = 8{,}192$ and 90-percent confidence, precision is usually better than +/–1 percentage point, and for $n = 512$, precision is usually better than +/– 4 percentage points. Compared with direct measurement, scoring is usually—but not

always—less precise. When using mean-squared error to consider both bias and precision, scoring usually—but not always—is less accurate than direct measurement. The accuracy of the new scorecard compares well with that of other recent poverty-assessment tools for Peru.

For targeting, programs can use the results reported here to select a cut-off that fits their values and mission.

Although the statistical technique is innovative, and although technical accuracy is important, the design of the scorecard here focuses on ease-of-use. After all, a perfectly accurate scorecard is worthless if programs feel so daunted by its complexity or its cost that they do not even try to use it. For this reason, the scorecard is kept simple, using 10 indicators that are inexpensive to collect and that are straightforward to observe and verify. Indicator weights are all zeros or positive integers, and scores range from 0 (most likely to be below a poverty line) to 100 (least likely to be 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 can compute scores in the field.

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

References

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

Baesens, B.; Van Gestel, T.; Viaene, S.; Stepanova, M.; Suykens, J.; and J. Vanthienen. (2003) "Benchmarking State-of-the-Art Classification Algorithms for Credit Scoring", *Journal of the Operational Research Society*, Vol. 54, pp. 627–635.

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

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, P.; Fanning, J.-P.; McKay, A.; and K. Wright-Revolledo. (2005) "Monitoring the Diversity of the Poverty Outreach and Impact of Microfinance: A Comparison of Methods Using Data from Peru", *Development Policy Review*, Vol. 24, No. 6, pp. 703–724.

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

Dunn, Elizabeth; and Gordon J. Arbuckle, Jr. (2001) "The Impacts of Microcredit: A Case Study from Peru (Data Set)", `microlinks.org/file_download.php/AIMS_Peru.htm?URL_ID=8009&filename=11208244961AIMS_Peru.htm&filetype=text%2Fhtml&filesize=15748&name=AIMS_Peru.htm&location=user-S/`.

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

Fanning, John-Paul. (2004) "Investigation into a Poverty Correlates Model for PROMUC".

Fuller, Rob. (2006) "Measuring Poverty of Microfinance Clients in Haiti", `microfinance.com/English/Papers/Scoring_Poverty_Haiti_Fuller.pdf`.

Glewwe, Paul; and Gillette Hall. (1991) "The Social Costs of Avoiding Structural Adjustment: Inequality and Poverty in Lima, Peru, from 1985–1986 to 1990", World Bank LSMS Working Paper No. 86.

Goodman, L.A. and Kruskal, W.H. (1979) *Measures of Association for Cross Classification*.

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

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

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

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

Instituto Nacional de Estadística e Informacion (2006) *Condiciones de vida en el Perú: Evolución 1997–2004*, www.inei.gob.pe/biblioineipub/bancopub/Est/Lib0640/Libro.pdf.

IRIS Center. (2007a) "Poverty Assessment Survey: Peru", povertytools.org/USAID_Tools/usaid_pat_per_4-2007.xls.

_____. (2007b) "Manual for the Implementation of USAID Poverty Assessment Tools", povertytools.org/USAID_Tools/USAID_Manual_4-2007.pdf. See also povertytools.org/Tool_Implementation/Tool_Implementation.html.

_____. (2007c) "Introduction to Sampling for the Implementation of PATs", povertytools.org/USAID_Tools/Training/Sampling/Introduction_Sampling.ppt.

_____. (2005a) "Notes on Assessment and Improvement of Tool Accuracy", povertytools.org/documents/Assessing%20and%20Improving%20Accuracy.pdf.

_____. (2005b) "Accuracy Results for 12 Poverty Assessment Tool Countries", povertytools.org/documents/Accuracy%20Results%20for%2012%20Countries.pdf.

Johannsen, Julia. (2006) "Operational Poverty Targeting in Peru—Proxy Means Testing with Non-Income Indicators", IPC Working Paper No. 30, http://www.undp-povertycentre.org/pub/IPCWorkingPaper30.pdf.

Johnson, Glenn. (2007) "Lesson 3: Two-Way Tables—Dependent Samples", www.stat.psu.edu/~jglenn/stat504/03_2way/53_2way_compare.htm.

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

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.

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

Meyer, Richard L.; Nagarajan, Geetha; and Elizabeth Dunn. (2000) "Measuring Depth of Outreach: Tools for Microfinance", *Bangladesh Development Studies*, Vol. 27, No. 2–3, pp. 171–197.

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

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.

SAS Institute Inc. (2004) "The LOGISTIC Procedure: Rank Correlation of Observed Responses and Predicted Probabilities", in *SAS/STAT User's Guide, Version 9*.

Schreiner, Mark. (2006a) "Simple Poverty Scorecard Poverty-Assessment Tool: Bangladesh", SimplePovertyScorecard.com/BGD_2000_ENG.pdf.

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

_____. (2005a) "La Herramienta del Indice de Calificación de Pobreza™: México", SimplePovertyScorecard.com/MEX_2002_SPA.pdf.

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

_____. (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.

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.

Toohig, Jeff. (2007) "Progress Out of Poverty Index: Training Guide", jtoohig@gfusa.org.

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.

Webb, Richard and Graciela Fernández B. (2003) *Anuario Estadístico—Perú en Números 2003*.

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

_____; Alcaraz V., Gabriela; and Julia Johannsen. (2005) "Developing and Testing Poverty-Assessment Tools: Results from Accuracy Tests in Peru", povertytools.org/documents/Peru%20Accuracy%20Report.pdf.

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

			% with expenditure below a poverty line					
				National	USAID	International		
Sub-sample	**Year of Survey**	**Households**	**National**	**food**	**'Extreme'**	**$1/day**	**$2/day**	**$3/day**
Constructing scorecards								
Selecting indicators and weights	'03	6,642	53.3	23.0	26.7	3.8	25.4	48.4
Associating scores with likelihoods	'03	6,468	53.3	23.3	26.7	3.9	25.3	47.7
Measuring Accuracy								
Non-panel	'02	13,088	57.1	27.1	28.5	6.6	30.2	51.8
	'04	12,955	-	-	-	2.4	20.0	41.3
Panel for '02 and '03	'02	4,051	52.5	21.0	22.2	4.8	23.3	45.5
	'03	4,051	49.6	16.7	20.9	2.6	15.7	44.0
Panel for '03 and '04	'03	3,562	50.4	17.9	21.8	3.2	20.5	45.0
	'04	3,562	-	-	-	2.2	16.3	37.3
Panel for '02 and '04	'02	3,256	54.5	22.5	23.6	5.2	24.7	47.5
	'04	3,256	-	-	-	2.2	16.2	37.8
Change in overall poverty rate (percentage points)								
Non-panel	'02 to '03	13,088 and 13,110	-3.8	-4.0	-1.9	-2.7	-4.8	-3.8
	'03 to '04	13,110 and 12,955	-	-	-	-1.4	-5.3	-6.7
	'02 to '04	13,088 and 12,955	-	-	-	-4.2	-10.1	-10.5
Panel	'02 to '03	4,051	-2.9	-4.3	-1.3	-2.2	-7.5	-1.5
	'03 to '04	3,562	-	-	-	-0.9	-4.2	-7.7
	'02 to '04	3,256	-	-	-	-3.0	-8.5	-9.7

Source: ENAHO, 2002-4. Construction sub-samples omit panel households.

Figure 3: Average poverty lines and poverty rates by region and year

		Poverty line (Nuevos soles/person/day) and poverty rate (%)																	
		National			National food			USAID 'Extreme'			International $1/day			International $2/day			International $3/day		
Area	Line or rate	'02	'03	'04	'02	'03	'04	'02	'03	'04	'02	'03	'04	'02	'03	'04	'02	'03	'04
Coast, north	Line	6.33	6.38	—	3.31	3.27	—	3.82	4.18	—	1.91	1.96	2.03	3.82	3.92	4.06	5.74	5.87	6.09
	Rate	61.2	44.1	—	22.9	9.1	—	29.0	17.5	—	6.7	2.0	1.3	33.3	16.8	18.2	58.7	38.3	40.2
Coast, central	Line	6.69	6.67	—	3.67	3.59	—	4.03	4.37	—	2.01	2.06	2.14	4.03	4.12	4.28	6.04	6.19	6.42
	Rate	43.3	34.7	—	8.8	4.3	—	11.2	9.1	—	0.0	0.2	0.0	11.1	7.6	8.2	38.1	29.7	30.9
Coast, south	Line	6.34	6.48	—	3.34	3.41	—	3.83	4.24	—	1.91	1.96	2.03	3.83	3.92	4.06	5.74	5.87	6.09
	Rate	41.2	26.2	—	3.6	2.0	—	4.6	5.9	—	0.0	0.3	0.0	3.6	5.4	4.0	21.1	20.4	23.7
Sierra, north	Line	5.76	5.61	—	3.65	3.44	—	3.47	3.67	—	1.74	1.78	1.85	3.48	3.56	3.69	5.21	5.34	5.53
	Rate	79.7	79.5	—	58.3	49.1	—	53.7	50.7	—	18.1	8.2	4.9	53.4	52.1	34.2	73.8	77.5	59.5
Sierra, central	Line	6.04	6.34	—	3.69	3.91	—	3.64	4.15	—	1.82	1.87	1.94	3.65	3.73	3.87	5.47	5.60	5.81
	Rate	65.8	71.1	—	36.1	44.4	—	35.9	45.5	—	8.3	9.7	1.9	35.4	40.9	26.5	58.8	64.2	51.4
Sierra, south	Line	6.17	6.35	—	3.62	3.77	—	3.72	4.16	—	1.86	1.91	1.98	3.73	3.82	3.96	5.59	5.72	5.94
	Rate	62.3	61.2	—	36.2	30.8	—	36.9	32.9	—	10.0	4.7	7.1	36.5	32.4	24.9	56.4	57.2	47.2
Jungle	Line	6.04	6.25	—	3.84	4.02	—	3.64	4.10	—	1.82	1.87	1.94	3.65	3.73	3.87	5.47	5.60	5.81
	Rate	66.2	63.5	—	34.9	30.3	—	31.1	30.1	—	4.5	2.8	3.1	35.3	29.1	21.6	60.5	56.7	49.1
Metro Lima	Line	8.79	8.89	—	4.08	4.11	—	5.30	5.83	—	2.65	2.71	2.81	5.30	5.42	5.62	7.94	8.13	8.44
	Rate	35.9	30.5	—	1.7	2.5	—	4.4	8.7	—	0.5	0.3	0.6	4.4	7.9	4.1	24.1	25.7	20.0
All Peru:	Line	6.89	6.94	—	3.74	3.80	—	4.16	4.55	—	2.08	2.11	2.18	4.16	4.21	4.35	6.24	6.32	6.53
	Rate	54.5	53.3	—	22.5	23.1	—	23.6	26.6	—	5.2	3.9	2.2	24.7	25.4	16.1	47.5	48.0	37.6

Data for a given year include both panel (whether or not interviewed in all years) and non-panel households.

Figure 4: Poverty indicators by uncertainty coefficient

Uncertainty coefficient	Indicator (Answers ordered starting with those most strongly indicative of poverty)
181	What fuel does the household use to cook? (Other; Gas, electricity, or does not cook)
176	How many household members are aged 17 or younger? (5 or more; 4; 3; 2; 1; None)
171	What is the highest grade completed by a family member? (None or no data, basic education, attended or finished grade school; Attended secondary; Finished secondary; Attended or graduated non-university post-secondary; Attended or graduated university or graduate school)
166	Does the household have a color TV? (No; Yes)
166	Does the household have a cellular telephone or a land-line telephone? (No; Yes)
161	Does the household have an iron? (No; Yes)
159	How does the household dispose of its sewage? (Other; Public sewer network inside the living quarters or outside the living quarters but inside the building)
155	Does the household have a refrigerator/freezer? (No; Yes)
151	What is the main construction material of the floors? (Dirt; Other)
148	Does the household have a blender? (No; Yes)
146	Does the household have a land-line telephone? (No; Yes)
143	Does the household have a gas stove? (No; Yes)
127	What is the main construction material of the exterior walls? (Other; Bricks or cement blocks)
119	If the household farms, how are the majority of its agricultural land watered? (Rain, none, or no data; Does not farm; Irrigated)
117	What does the household use for light? (Candle, other, or none; Kerosene lamp or paraffin lamp or petroleum or gas lamp; Electricity or generator)
115	Does the household have a black-and-white TV? (No; Yes)
114	What is the main construction material of the roof? (Other; Reinforced concrete or wood)
113	How many household members work in agriculture as their principal activity? (Some; None)
113	How many household members receive income from agriculture? (Some; None)
113	How many household members work for others as employees? (None; 1; 2 or more)
111	What is the highest grade passed by the female head/spouse? (Other; Graduated from secondary, or attended or graduated non-university post-secondary, attended or graduated college or graduate school)

Source: 2003 *Encuesta de Hogares* by Peru's Instituto Nacional de Estadística e Información, nacional poverty line.

Figure 4 (continued): Poverty indicators by uncertainty coefficient

Uncertainty coefficient	Indicator (Answers ordered starting with those most strongly associated with poverty)
96	Does the household have a stereo (No; Yes)
88	Where does the household get water from? (Other; Public network inside the living quarters or outside the living quarters but inside the building)
86	What is the highest grade passed by the male head/spouse? (None or no data, basic education, attended or finished grade school, or attended secondary; Graduated secondary or attended non-university post-secondary; Graduated non-university post-secondary, attended or graduated college, or graduate school)
83	Can the female head/spouse read and write (No; Yes)
77	Does the household have a VCR? (No; Yes)
77	Not counting bathrooms, kitchens, hallways, and garages, how many rooms does the house have? (1 or 2; 3 or 4; 5 or more)
68	Has any household member attended a private school? (No; Yes)
65	Does the household have a computer? (No; Yes)
65	How many children ages 6 to 11 attend school? (Not all; All; No children these ages)
62	Does any household member work a job that pays monthly? (No; Yes)
60	Does the household have a washing machine for clothes? (No; Yes)
59	How many children ages 6 to 14 attend school? (Not all; All; No children these ages)
59	Does the household have a cellular telephone? (No; Yes)
53	Does the household have a car, pick-up, or truck? (No; Yes)
51	Does the household have a microwave (No; Yes)
51	How many children ages 6 to 17 attend school? (Not all; All; No children these ages)
43	How many girls ages 6 to 17 attend school? (Not all; All; No girls these ages)
27	How old is the female head/spouse? (44 years or less; 45 to 51; 52 or older)
20	Does the household have a sewing machine? (No; Yes)

Source: 2003 *Encuesta de Hogares* by Peru's Instituto Nacional de Estadística e Información, nacional poverty line.

National Poverty Line Tables

(and tables pertaining to all six poverty lines)

Figure 5 (National poverty line): Estimated poverty likelihoods associated with scores

If an individual's score is . . .	. . . then the likelihood (%) of being below the poverty line is:
0–4	99.1
5–9	93.5
10–14	91.6
15–19	89.6
20–24	85.0
25–29	78.0
30–34	74.8
35–39	58.4
40–44	60.0
45–49	41.1
50–54	34.9
55–59	34.3
60–64	18.3
65–69	16.2
70–74	9.0
75–79	5.2
80–84	5.0
85–89	1.6
90–94	0.0
95–100	2.0

Surveyed cases weighted to represent Peru's population.
Based on the 2003 ENAHO.

Figure 6 (National poverty line): Illustration of derivation of estimated poverty likelihoods associated with scores for a single bootstrap sample

Score	People below poverty line		All people at score		Poverty likelihood (estimated, %)
0–4	5,870	÷	5,923	=	99.1
5–9	5,205	÷	5,565	=	93.5
10–14	6,030	÷	6,585	=	91.6
15–19	3,980	÷	4,443	=	89.6
20–24	5,935	÷	6,984	=	85.0
25–29	6,323	÷	8,107	=	78.0
30–34	4,533	÷	6,064	=	74.8
35–39	3,495	÷	5,985	=	58.4
40–44	3,731	÷	6,221	=	60.0
45–49	2,235	÷	5,439	=	41.1
50–54	2,014	÷	5,775	=	34.9
55–59	1,714	÷	4,994	=	34.3
60–64	875	÷	4,788	=	18.3
65–69	825	÷	5,077	=	16.2
70–74	331	÷	3,683	=	9.0
75–79	162	÷	3,091	=	5.2
80–84	195	÷	3,888	=	5.0
85–89	76	÷	4,807	=	1.6
90–94	0	÷	380	=	0.0
95–100	45	÷	2,200	=	2.0

Number of all people normalized to sum to 100,000.
Based on the 2003 ENAHO.

Figure 7 (All poverty lines): Distribution of poverty likelihoods across ranges demarcated by poverty lines, 2003 scorecard

	Probabilidad de tener gastos en un rango definido por líneas de pobreza por persona por día						
	<$1	≥$1 y <Nacional alimenticia	≥Nacional alimenticia y <$2	≥$2 y <USAID 'extrema'	≥USAID 'extrema' y <$3	≥$3 y <Nacional	≥Nacional
Puntaje	<NS2.11	≥NS2.11 y <NS3.8	≥NS3.8 y <NS4.21	≥NS4.21 y <NS4.55	≥NS4.55 y <NS6.32	≥NS6.32 y <NS6.94	≥NS6.94
0–4	21.4	62.5	4.2	0.0	11.0	0.0	0.9
5–9	14.7	51.8	0.7	0.0	25.8	0.6	6.5
10–14	12.6	43.6	3.5	0.0	31.8	0.2	8.4
15–19	4.3	39.2	0.0	10.8	33.7	1.6	10.4
20–24	4.9	37.7	4.3	0.9	33.3	3.9	15.0
25–29	3.5	27.0	3.4	0.9	36.3	6.9	22.0
30–34	0.3	26.1	3.0	4.6	32.4	8.4	25.2
35–39	2.0	14.6	2.4	3.7	28.6	7.2	41.6
40–44	0.6	8.9	3.9	2.8	32.4	11.4	40.0
45–49	0.2	4.5	2.3	2.1	23.8	8.1	58.9
50–54	0.2	2.5	1.0	1.7	18.9	10.5	65.1
55–59	0.0	1.0	2.7	0.7	15.8	14.1	65.7
60–64	0.0	0.0	3.5	0.2	7.4	7.1	81.7
65–69	0.0	0.3	0.0	0.0	8.1	7.9	83.8
70–74	0.0	0.0	0.0	0.0	2.8	6.2	91.0
75–79	0.0	0.0	0.0	0.0	4.0	1.2	94.8
80–84	0.0	0.0	0.0	0.0	3.1	1.9	95.0
85–89	0.0	0.0	0.0	0.0	0.6	1.0	98.4
90–94	0.0	0.0	0.0	0.0	0.0	0.0	100.0
95–100	0.0	0.0	0.0	0.0	2.0	0.0	98.0

Todas probabilidades en unidades procentuales.

Figure 8 (National poverty line): Bootstrapped differences between estimated and true poverty likelihoods for individuals in a large sample (n=16,384), with confidence intervals, 2003 scorecard applied to non-panel households in 2002 and 2004

	2003 scorecard applied to 2002 non-panel, difference between estimate and true value				2003 scorecard applied to 2004 non-panel, difference between estimate and true value			
		Confidence interval (+/- percentage points)				Confidence interval (+/- percentage points)		
Score	Bias	90-percent	95-percent	99-percent	Bias	90-percent	95-percent	99-percent
0–4	1.7	1.2	1.4	1.8	—	—	—	—
5–9	-3.8	2.5	2.6	2.7	—	—	—	—
10–14	1.4	2.4	3.0	4.0	—	—	—	—
15–19	-1.6	2.5	3.0	3.9	—	—	—	—
20–24	-3.7	2.8	3.0	3.2	—	—	—	—
25–29	-5.7	3.8	4.0	4.4	—	—	—	—
30–34	1.8	3.0	3.7	4.5	—	—	—	—
35–39	-6.9	5.0	5.2	5.8	—	—	—	—
40–44	-4.1	3.7	4.0	5.0	—	—	—	—
45–49	-9.5	6.6	7.0	7.7	—	—	—	—
50–54	-5.2	4.6	4.9	6.2	—	—	—	—
55–59	6.2	3.5	4.3	5.6	—	—	—	—
60–64	-2.9	3.6	4.4	5.6	—	—	—	—
65–69	-2.2	3.2	3.8	4.9	—	—	—	—
70–74	-1.5	3.6	4.2	5.6	—	—	—	—
75–79	-6.2	5.2	5.7	6.6	—	—	—	—
80–84	-1.4	2.7	3.2	4.3	—	—	—	—
85–89	-2.2	2.3	2.6	3.2	—	—	—	—
90–94	-6.6	8.3	9.8	12.8	—	—	—	—
95–100	0.9	1.5	1.7	2.1	—	—	—	—

Based on 2003 scorecard applied to non-panel households from 2002 and 2004.

Figure 9 (All poverty lines): Bias and precision for bootstrapped estimates of individuals' poverty likelihoods from the 2003 scorecard applied to non-panel households in 2002 and 2004

	Poverty line					
Year scorecard applied	National	National food	USAID 'Extreme'	$1/day	$2/day	$3/day
Bias						
2002	-2.5	-2.4	-0.8	-2.2	-3.2	-2.4
2004	—	—	—	0.4	0.2	0.6
Precision						
2002	0.8	0.5	0.5	0.3	0.5	0.8
2004	—	—	—	0.2	0.4	0.5

Precision is measured as 90-percent confidence intervals in units of +/- percentage points.
Scorecard is based on 2003 ENAHO. Scorecard is applied to non-panel households in 2002 and 2004.
Bias and precision estimated from 1,000 bootstraps of size n=16,384.

Figure 10 (National poverty line): Bias and precision for bootstrapped estimates of individuals' poverty likelihoods, by sample size, 2003 scorecard applied to non-panel households in 2002

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	-1.7	49.8	60.4	77.6
4	-1.9	37.8	46.2	59.3
8	-2.5	27.1	34.3	48.3
16	-2.5	20.6	25.2	38.4
32	-2.5	15.3	19.4	25.8
64	-2.6	11.6	14.9	20.8
128	-2.7	8.8	10.5	14.0
256	-2.6	6.4	7.7	10.1
512	-2.5	4.2	5.1	7.0
1,024	-2.5	3.2	4.0	5.6
2,048	-2.5	2.2	2.8	4.0
4,096	-2.5	1.7	2.0	2.5
8,192	-2.5	1.2	1.4	2.0
16,384	-2.5	0.8	1.0	1.3

Figure 11 (National poverty line): Bias and precision for bootstrapped estimates of individuals' poverty likelihoods, by sample size, 2003 scorecard applied to non-panel households in 2004

This poverty line does not exist for 2004.

Figure 12 (All poverty lines): Bias, precision, sample-size α, and mean-squared error for bootstrapped estimates of groups' poverty rates at a point in time from the 2003 scorecard applied to non-panel households in 2002 and 2004

	Poverty line					
Year scorecard applied	National	National food	USAID 'Extreme'	$1/day	$2/day	$3/day
Bias						
2002	-2.4	-3.2	-1.0	-2.7	-4.1	-2.7
2004	—	—	—	0.3	-0.1	0.3
Precision						
2002	0.7	0.6	0.7	0.4	0.6	0.7
2004	—	—	—	0.2	0.5	0.6
α for sample size						
2002	1.09	1.28	1.27	2.50	1.27	1.04
2004	—	—	—	0.77	0.84	0.88
MSE of scorecard as a ratio of MSE of direct measurement						
2002	25	61	5	37	96	31
2004	—	—	—	3	0.3	0.1

Precision is measured as 90-percent confidence intervals in units of +/- percentage points.
Scorecard is based on 2003 ENAHO. Scorecard is applied to non-panel households in 2002 and 2004.
Bias and precision estimated from 1,000 bootstraps of size n=16,384.
α is estimated from 1,000 bootstrap samples of n=256, 512, 1,024, 2,048, 4,096, 8,192, and 16,384.
MSE is mean-squared error. Ratios greater than 1.0 indicate that direct measurement has lower MSE.

Figure 13 (National poverty line): Bias and precision for bootstrapped estimates of groups' poverty rates at a point in time, by sample size, 2003 scorecard applied to non-panel households in 2002

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	-1.7	49.8	60.4	77.6
4	-1.8	37.3	46.1	57.7
8	-2.5	26.2	34.0	47.8
16	-2.7	19.9	23.8	34.1
32	-2.5	13.6	16.6	24.2
64	-2.4	9.9	12.1	16.6
128	-2.4	7.2	8.4	11.2
256	-2.4	5.3	6.3	8.1
512	-2.4	3.7	4.5	5.9
1,024	-2.4	2.7	3.1	4.1
2,048	-2.4	1.8	2.2	2.9
4,096	-2.4	1.4	1.6	2.0
8,192	-2.4	0.97	1.1	1.5
16,384	-2.4	0.7	0.8	1.1

Figure 14 (National poverty line): Bias and precision for bootstrapped estimates of groups' poverty rates at a point in time, by sample size, 2003 scorecard applied to non-panel households in 2004

This poverty line does not exist for 2004.

Figure 15 (All poverty lines): Bias, precision, sample-size α, and mean-squared error for bootstrapped estimates of changes in groups' poverty rates between two points in time from the 2003 scorecard applied to non-panel households in 2002, 2003, and 2004

	Poverty line					
Year scorecard applied	National	National food	USAID 'Extreme'	$1/day	$2/day	$3/day
Bias						
2002 to 2003	2.7	3.5	1.2	2.7	4.2	2.7
2003 to 2004	—	—	—	0.2	-0.3	0.3
2002 to 2004	—	—	—	2.9	4.0	3.0
Precision						
2002 to 2003	1.0	0.8	0.9	0.5	0.9	0.9
2003 to 2004	—	—	—	0.4	0.8	0.9
2002 to 2004	—	—	—	0.5	0.8	0.9
α for sample size						
2002 to 2003	1.09	1.21	1.32	2.10	1.27	1.03
2003 to 2004	—	—	—	1.28	1.01	0.96
2002 to 2004	—	—	—	1.68	1.06	0.97
MSE of scorecard as a ratio of MSE of direct measurement						
2002 to 2003	38	77	11	119	103	38
2003 to 2004	—	—	—	2	1.0	0.7
2002 to 2004	—	—	—	135	93	46

Precision is measured as 90-percent confidence intervals in units of +/- percentage points.
Scorecard is based on 2003 ENAHO. Scorecard is applied to non-panel households in '02, '03, and '04.
Bias and precision estimated from 1,000 bootstraps of size n=16,384.
α is estimated from 1,000 bootstrap samples of n=256, 512, 1,024, 2,048, 4,096, 8,192, and 16,384.
MSE is mean-squared error. Ratios greater than 1.0 indicate that direct measurement has lower MSE.

Figure 16 (National poverty line): Bias and precision for bootstrapped estimates of changes in groups' poverty rates between two points in time, by sample size, 2003 scorecard applied to non-panel households in 2002 and 2003

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	2.8	74.5	91.9	108.6
4	1.4	53.2	64.4	84.7
8	2.5	39.6	47.4	62.2
16	2.5	28.8	35.2	47.1
32	2.6	20.1	25.2	33.8
64	2.8	15.0	18.3	24.8
128	2.8	10.3	12.5	15.8
256	2.7	7.5	9.3	11.5
512	2.8	5.3	6.3	8.4
1,024	2.8	3.8	4.7	5.9
2,048	2.7	2.7	3.1	3.9
4,096	2.7	1.9	2.2	2.9
8,192	2.7	1.3	1.6	2.0
16,384	2.7	1.0	1.1	1.5

Figure 17 (National poverty line): Bias and precision for bootstrapped estimates of changes in groups' poverty rates between two points in time, by sample size, 2003 scorecard applied to non-panel households in 2003 and 2004

This poverty line does not exist for 2004.

Figure 18 (National poverty line): Bias and precision for bootstrapped estimates of changes in groups' poverty rates between two points in time, by sample size, 2003 scorecard applied to non-panel households in 2002 and 2004

This poverty line does not exist for 2004.

Figure 19 (All poverty lines): Bias, precision, sample-size α, and mean-squared error for bootstrapped estimates of changes in groups' poverty rates between two points in time from the 2003 scorecard applied to panel households in 2002, 2003, and 2004

	Poverty line					
Year scorecard applied	National	National food	USAID 'Extreme'	$1/day	$2/day	$3/day
Bias						
2002 to 2003	1.2	3.1	0.2	1.8	2.4	0.0
2003 to 2004	—	—	—	0.3	0.2	-0.1
2002 to 2004	—	—	—	0.2	-0.3	0.3
Precision						
2002 to 2003	0.9	0.6	0.7	0.3	0.7	0.8
2003 to 2004	—	—	—	0.2	0.4	0.5
2002 to 2004	—	—	—	0.4	0.8	0.9
α for sample size						
2002 to 2003	3.70	2.10	2.87	1.62	2.59	3.12
2003 to 2004	—	—	—	0.49	0.98	1.14
2002 to 2004	—	—	—	0.64	1.01	1.15
MSE of scorecard as a ratio of MSE of direct measurement						
2002 to 2003	15.0	27	7.1	29	19.2	8
2003 to 2004	—	—	—	0.9	0.2	0.3
2002 to 2004	—	—	—	1.0	0.3	1

Precision is measured as 90-percent confidence intervals in units of +/- percentage points.
Scorecard is based on 2003 ENAHO. Scorecard is applied to panel households in '02, '03, and '04.
Bias and precision estimated from 1,000 bootstraps of size n=16,384.
α is estimated from 1,000 bootstrap samples of n=256, 512, 1,024, 2,048, 4,096, 8,192, and 16,384.
MSE is mean-squared error. Ratios greater than 1.0 indicate that direct measurement has lower MSE.

Figure 20 (National poverty line): Bias and precision for bootstrapped estimates of changes in groups' poverty rates between two points in time, by sample size, 2003 scorecard applied to panel households in 2002 and 2003

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	0.4	72.3	88.4	106.0
4	1.2	54.4	63.8	80.7
8	1.9	38.0	46.6	64.4
16	1.3	25.8	33.7	46.1
32	1.2	18.0	22.2	28.4
64	1.0	13.1	15.4	22.1
128	1.0	9.0	11.0	14.8
256	1.2	6.9	7.9	10.0
512	1.3	4.9	5.8	7.3
1,024	1.2	3.3	4.0	5.0
2,048	1.2	2.4	2.8	3.6
4,096	1.2	1.7	2.0	2.7
8,192	1.2	1.2	1.4	1.9
16,384	1.2	0.9	1.0	1.3

Figure 21 (National poverty line): Bias and precision for bootstrapped estimates of changes in groups' poverty rates between two points in time, by sample size, 2003 scorecard applied to panel households in 2003 and 2004

This poverty line does not exist for 2004.

Figure 22 (National poverty line): Bias and precision for bootstrapped estimates of changes in groups' poverty rates between two points in time, by sample size, 2003 scorecard applied to panel households in 2002 and 2004

This poverty line does not exist for 2004.

Figure 23 (All poverty lines): Possible types of outcomes from targeting by poverty score

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

Figure 24 (National poverty line): People by targeting classification and score, along with "Total Accuracy" and BPAC, 2003 scorecard applied to non-panel households in 2002

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	6.2	50.8	0.2	42.8	49.0	-77.9
5–9	10.8	46.3	0.3	42.7	53.4	-61.7
10–14	16.6	40.4	0.9	42.0	58.7	-40.0
15–19	21.7	35.3	1.4	41.5	63.2	-21.4
20–24	28.3	28.8	2.3	40.7	68.9	3.0
25–29	35.0	22.1	3.6	39.4	74.4	29.0
30–34	39.8	17.3	5.4	37.6	77.4	48.9
35–39	43.6	13.4	7.4	35.6	79.2	65.8
40–44	48.0	9.1	9.8	33.1	81.1	82.8
45–49	50.7	6.4	12.4	30.5	81.2	78.2
50–54	52.6	4.5	15.3	27.6	80.2	73.2
55–59	54.0	3.1	19.0	24.0	78.0	66.8
60–64	55.0	2.0	22.8	20.1	75.1	60.0
65–69	55.9	1.2	26.6	16.3	72.2	53.3
70–74	56.3	0.8	30.1	12.8	69.1	47.2
75–79	56.6	0.5	32.5	10.5	67.1	43.1
80–84	56.8	0.2	36.0	6.9	63.7	36.8
85–89	57.0	0.1	40.1	2.9	59.9	29.8
90–94	57.0	0.0	40.4	2.5	59.5	29.1
95–100	57.1	0.0	42.9	0.0	57.1	24.7

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

Figure 25 (National poverty line): People by targeting classification and score, along with "Total Accuracy" and BPAC, 2003 scorecard applied to non-panel households in 2004

This poverty line does not exist for 2004.

Figure 26 (National poverty line): People below the poverty line and all people, at a given score or at or below a given score cut-off, 2003 scorecard applied to non-panel households in 2002

	People below poverty line (%)		All people (%)	
Score	At score	At or below score	At score	At or below score
0–4	97.4	97.4	6.4	6.4
5–9	97.3	97.4	4.7	11.1
10–14	90.1	94.7	6.5	17.6
15–19	91.3	93.9	5.6	23.1
20–24	88.7	92.6	7.4	30.5
25–29	83.6	90.7	8.1	38.6
30–34	73.0	88.1	6.6	45.2
35–39	65.2	85.5	5.8	51.0
40–44	64.2	83.0	6.8	57.8
45–49	50.6	80.3	5.3	63.1
50–54	40.0	77.4	4.8	67.9
55–59	28.0	74.0	5.1	73.0
60–64	21.0	70.7	4.9	77.9
65–69	18.5	67.7	4.7	82.5
70–74	10.5	65.1	3.9	86.4
75–79	11.4	63.5	2.7	89.1
80–84	6.4	61.2	3.8	92.9
85–89	3.8	58.7	4.2	97.1
90–94	6.7	58.5	0.4	97.5
95–100	1.1	57.1	2.5	100.0

Figure 27 (National poverty line): People below the poverty line and all people, at a given score or at or below a given score cut-off, 2003 scorecard applied to non-panel households in 2004

This poverty line does not exist for 2004.

National Food Poverty Line Tables

Figure 5 (National food line): Estimated poverty likelihoods associated with scores

If an individual's score is . . .	. . . then the likelihood (%) of being below the poverty line is:
0–4	83.9
5–9	66.4
10–14	56.2
15–19	43.5
20–24	42.6
25–29	30.5
30–34	26.4
35–39	16.5
40–44	9.5
45–49	4.7
50–54	2.7
55–59	1.0
60–64	0.0
65–69	0.3
70–74	0.0
75–79	0.0
80–84	0.0
85–89	0.0
90–94	0.0
95–100	0.0

Surveyed cases weighted to represent Peru's population.
Based on the 2003 ENAHO.

Figure 6 (National food line): Derivation of estimated poverty likelihoods associated with scores

Score	People below poverty line		All people at score		Poverty likelihood (estimated, %)
0–4	4,971	÷	5,923	=	83.9
5–9	3,697	÷	5,565	=	66.4
10–14	3,699	÷	6,585	=	56.2
15–19	1,933	÷	4,443	=	43.5
20–24	2,974	÷	6,984	=	42.6
25–29	2,472	÷	8,107	=	30.5
30–34	1,598	÷	6,064	=	26.4
35–39	989	÷	5,985	=	16.5
40–44	590	÷	6,221	=	9.5
45–49	258	÷	5,439	=	4.7
50–54	156	÷	5,775	=	2.7
55–59	51	÷	4,994	=	1.0
60–64	0	÷	4,788	=	0.0
65–69	16	÷	5,077	=	0.3
70–74	0	÷	3,683	=	0.0
75–79	0	÷	3,091	=	0.0
80–84	0	÷	3,888	=	0.0
85–89	0	÷	4,807	=	0.0
90–94	0	÷	380	=	0.0
95–100	0	÷	2,200	=	0.0

Surveyed cases weighted to represent Peru's population.
Based on the 2003 ENAHO.

Figure 8 (National food line): Bootstrapped differences between estimated and true poverty likelihoods for individuals in a large sample (n=16,384), with confidence intervals, 2003 scorecard applied to non-panel households in 2002 and 2004

	2003 scorecard applied to 2002 non-panel, difference between estimate and true value				2003 scorecard applied to 2004 non-panel, difference between estimate and true value			
		Confidence interval (+/- percentage points)				Confidence interval (+/- percentage points)		
Score	Bias	90-percent	95-percent	99-percent	Bias	90-percent	95-percent	99-percent
0–4	0.6	2.7	3.2	3.7	—	—	—	—
5–9	-7.2	5.4	5.8	6.4	—	—	—	—
10–14	-6.6	5.1	5.4	6.3	—	—	—	—
15–19	-12.2	8.1	8.5	9.4	—	—	—	—
20–24	-5.4	4.3	4.6	5.2	—	—	—	—
25–29	-10.2	6.5	6.9	7.4	—	—	—	—
30–34	0.8	2.9	3.3	4.5	—	—	—	—
35–39	-2.2	2.2	2.5	3.3	—	—	—	—
40–44	-4.6	3.3	3.5	4.0	—	—	—	—
45–49	-1.4	1.7	2.1	2.8	—	—	—	—
50–54	-1.0	1.2	1.5	1.9	—	—	—	—
55–59	-0.2	0.7	0.9	1.2	—	—	—	—
60–64	-0.5	0.5	0.5	0.6	—	—	—	—
65–69	0.3	0.0	0.1	0.1	—	—	—	—
70–74	0.0	0.0	0.0	0.0	—	—	—	—
75–79	-0.4	0.6	0.7	0.9	—	—	—	—
80–84	0.0	0.0	0.0	0.0	—	—	—	—
85–89	0.0	0.0	0.0	0.0	—	—	—	—
90–94	0.0	0.0	0.0	0.0	—	—	—	—
95–100	0.0	0.0	0.0	0.0	—	—	—	—

Based on 2003 scorecard applied to non-panel households from 2002 and 2004.

Figure 10 (National food line): Bias and precision for bootstrapped estimates of individuals' poverty likelihoods, by sample size, 2003 scorecard applied to non-panel households in 2002

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	-5.3	47.4	56.3	70.8
4	-4.2	38.0	44.5	59.3
8	-3.8	27.2	32.2	43.7
16	-3.3	18.5	22.5	30.2
32	-2.7	13.9	16.8	22.0
64	-2.5	9.9	12.2	17.9
128	-2.6	7.0	8.4	11.6
256	-2.5	4.6	5.5	7.4
512	-2.5	3.0	3.6	5.3
1,024	-2.4	2.0	2.4	3.4
2,048	-2.4	1.4	1.8	2.3
4,096	-2.4	1.0	1.1	1.6
8,192	-2.4	0.6	0.8	1.1
16,384	-2.4	0.5	0.6	0.7

Figure 11 (National food line): Bias and precision for bootstrapped estimates of individuals' poverty likelihoods, by sample size, 2003 scorecard applied to non-panel households in 2004

This poverty line does not exist for 2004.

Figure 13 (National food line): Bias and precision for bootstrapped estimates of groups' poverty rates at a point in time, by sample size, 2003 scorecard applied to non-panel households in 2002

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	-5.3	47.4	56.3	70.8
4	-4.3	37.9	43.9	58.5
8	-4.0	26.4	31.8	43.0
16	-3.7	18.3	21.7	29.7
32	-3.3	13.0	16.0	19.5
64	-3.3	9.7	11.7	14.2
128	-3.4	6.8	7.8	10.7
256	-3.3	4.9	5.8	7.4
512	-3.3	3.5	4.4	5.7
1,024	-3.2	2.5	3.0	4.0
2,048	-3.2	1.7	2.1	2.9
4,096	-3.2	1.2	1.4	2.0
8,192	-3.2	0.8	1.0	1.4
16,384	-3.2	0.6	0.7	0.9

Figure 14 (National food line): Bias and precision for bootstrapped estimates of groups' poverty rates at a point in time, by sample size, 2003 scorecard applied to non-panel households in 2004

This poverty line does not exist for 2004.

Figure 16 (National food line): Bias and precision for bootstrapped estimates of changes in groups' poverty rates between two points in time, by sample size, 2003 scorecard applied to non-panel households in 2002 and 2003

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	4.8	72.5	84.5	101.4
4	4.0	52.0	62.5	81.9
8	4.7	37.9	43.9	60.1
16	3.9	26.3	32.7	43.4
32	3.5	18.6	21.9	29.9
64	3.4	13.1	15.7	20.5
128	3.6	9.3	10.9	13.9
256	3.6	6.8	8.1	11.1
512	3.5	4.8	5.7	7.4
1,024	3.5	3.4	4.0	5.2
2,048	3.5	2.4	2.8	3.6
4,096	3.5	1.7	2.0	2.6
8,192	3.5	1.1	1.4	1.7
16,384	3.5	0.8	1.0	1.2

Figure 17 (National food line): Bias and precision for bootstrapped estimates of changes in groups' poverty rates between two points in time, by sample size, 2003 scorecard applied to non-panel households in 2003 and 2004

This poverty line does not exist for 2004.

Figure 18 (National food line): Bias and precision for bootstrapped estimates of changes in groups' poverty rates between two points in time, by sample size, 2003 scorecard applied to non-panel households in 2002 and 2004

This poverty line does not exist for 2004.

Figure 20 (National food line): Bias and precision for bootstrapped estimates of changes in groups' poverty rates between two points in time, by sample size, 2003 scorecard applied to panel households in 2002 and 2003

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	2.8	55.9	73.3	98.4
4	3.4	38.1	47.8	72.1
8	3.3	26.6	34.2	49.7
16	3.4	17.5	23.6	32.9
32	3.1	13.0	15.8	21.9
64	3.3	9.2	11.5	14.6
128	3.3	6.4	7.6	11.1
256	3.2	4.5	5.6	8.0
512	3.1	3.3	4.1	5.3
1,024	3.0	2.4	3.0	4.0
2,048	3.1	1.7	2.0	2.7
4,096	3.1	1.2	1.3	1.9
8,192	3.1	0.8	1.0	1.3
16,384	3.1	0.6	0.7	0.9

Figure 21 (National food line): Bias and precision for bootstrapped estimates of changes in groups' poverty rates between two points in time, by sample size, 2003 scorecard applied to panel households in 2003 and 2004

This poverty line does not exist for 2004.

Figure 22 (National food line): Bias and precision for bootstrapped estimates of changes in groups' poverty rates between two points in time, by sample size, 2003 scorecard applied to panel households in 2002 and 2004

This poverty line does not exist for 2004.

Figure 24 (National food line): People by targeting classification and score, along with "Total Accuracy" and BPAC, 2003 scorecard applied to non-panel households in 2002

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	5.3	21.8	1.1	71.8	77.2	-56.8
5–9	8.8	18.3	2.3	70.6	79.4	-26.9
10–14	12.9	14.3	4.7	68.2	81.0	12.3
15–19	15.9	11.2	7.2	65.7	81.6	44.2
20–24	19.5	7.6	11.0	61.9	81.3	59.3
25–29	22.8	4.3	15.8	57.1	79.9	41.7
30–34	24.4	2.7	20.7	52.2	76.6	23.6
35–39	25.5	1.6	25.5	47.4	73.0	6.1
40–44	26.5	0.6	31.3	41.6	68.1	-15.4
45–49	26.8	0.3	36.3	36.6	63.5	-33.8
50–54	27.0	0.1	40.9	32.0	59.0	-50.9
55–59	27.1	0.0	45.9	27.0	54.1	-69.3
60–64	27.1	0.0	50.8	22.1	49.2	-87.3
65–69	27.1	0.0	55.4	17.5	44.6	-104.5
70–74	27.1	0.0	59.3	13.6	40.7	-118.9
75–79	27.1	0.0	62.0	10.9	38.0	-128.7
80–84	27.1	0.0	65.8	7.1	34.2	-142.7
85–89	27.1	0.0	69.9	2.9	30.1	-158.1
90–94	27.1	0.0	70.4	2.5	29.6	-159.6
95–100	27.1	0.0	72.9	0.0	27.1	-168.9

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

Figure 25 (National food line): People by targeting classification and score, along with "Total Accuracy" and BPAC, 2003 scorecard applied to non-panel households in 2004

This poverty line does not exist for 2004.

Figure 26 (National food line): People below the poverty line and all people, at a given score or at or below a given score cut-off, 2003 scorecard applied to non-panel households in 2002

	People below poverty line (%)		All people (%)	
Score	At score	At or below score	At score	At or below score
0–4	82.3	82.3	6.4	6.4
5–9	74.1	78.9	4.7	11.1
10–14	62.2	72.7	6.5	17.6
15–19	58.7	69.3	5.6	23.1
20–24	51.8	65.1	7.4	30.5
25–29	38.7	59.6	8.1	38.6
30–34	28.2	55.0	6.6	45.2
35–39	21.7	51.2	5.8	51.0
40–44	18.0	47.3	6.8	57.8
45–49	10.7	44.2	5.3	63.1
50–54	6.1	41.5	4.8	67.9
55–59	2.5	38.8	5.1	73.0
60–64	2.1	36.5	4.9	77.9
65–69	1.4	34.5	4.7	82.5
70–74	0.5	33.0	3.9	86.4
75–79	0.7	32.0	2.7	89.1
80–84	0.0	30.7	3.8	92.9
85–89	0.0	29.4	4.2	97.1
90–94	0.0	29.3	0.4	97.5
95–100	0.0	28.5	2.5	100.0

Figure 27 (National food line): People below the poverty line and all people, at a given score or at or below a given score cut-off, 2003 scorecard applied to non-panel households in 2004

This poverty line does not exist for 2004.

USAID "Extreme" Poverty Line Tables

Figure 5 (USAID "extreme" line): Estimated poverty likelihoods associated with scores

If an individual's score is . . .	. . . then the likelihood (%) of being below the poverty line is:
0–4	86.8
5–9	65.7
10–14	57.1
15–19	54.3
20–24	47.8
25–29	34.7
30–34	34.0
35–39	22.6
40–44	16.2
45–49	9.1
50–54	5.4
55–59	4.5
60–64	3.8
65–69	0.2
70–74	0.0
75–79	0.0
80–84	0.0
85–89	0.0
90–94	0.0
95–100	0.0

Surveyed cases weighted to represent Peru's population.
Based on the 2003 ENAHO.

Figure 6 (USAID "extreme" line): Derivation of estimated poverty likelihoods associated with scores

Score	People below poverty line		All people at score		Poverty likelihood (estimated, %)
0–4	5,142	÷	5,923	=	86.8
5–9	3,657	÷	5,565	=	65.7
10–14	3,758	÷	6,585	=	57.1
15–19	2,413	÷	4,443	=	54.3
20–24	3,339	÷	6,984	=	47.8
25–29	2,816	÷	8,107	=	34.7
30–34	2,059	÷	6,064	=	34.0
35–39	1,351	÷	5,985	=	22.6
40–44	1,007	÷	6,221	=	16.2
45–49	497	÷	5,439	=	9.1
50–54	314	÷	5,775	=	5.4
55–59	223	÷	4,994	=	4.5
60–64	180	÷	4,788	=	3.8
65–69	10	÷	5,077	=	0.2
70–74	0	÷	3,683	=	0.0
75–79	0	÷	3,091	=	0.0
80–84	0	÷	3,888	=	0.0
85–89	0	÷	4,807	=	0.0
90–94	0	÷	380	=	0.0
95–100	0	÷	2,200	=	0.0

Surveyed cases weighted to represent Peru's population.
Based on the 2003 ENAHO.

Figure 8 (USAID "extreme" line): Bootstrapped differences between estimated and true poverty likelihoods for individuals in a large sample (n=16,384), with confidence intervals, 2003 scorecard applied to non-panel households in 2002 and 2004

	2003 scorecard applied to 2002 non-panel, difference between estimate and true value				2003 scorecard applied to 2004 non-panel, difference between estimate and true value			
		Confidence interval (+/- percentage points)				Confidence interval (+/- percentage points)		
Score	Bias	90-percent	95-percent	99-percent	Bias	90-percent	95-percent	99-percent
0–4	4.5	2.8	3.3	4.2	—	—	—	—
5–9	-8.5	6.0	6.3	6.8	—	—	—	—
10–14	-5.2	4.3	4.5	5.1	—	—	—	—
15–19	-4.5	4.2	4.8	6.4	—	—	—	—
20–24	-4.0	3.7	4.0	5.4	—	—	—	—
25–29	-3.9	3.3	3.6	4.4	—	—	—	—
30–34	5.7	3.2	3.7	4.9	—	—	—	—
35–39	0.9	2.5	3.0	4.0	—	—	—	—
40–44	-1.7	2.8	3.3	4.2	—	—	—	—
45–49	-1.5	2.7	3.2	4.1	—	—	—	—
50–54	-0.7	1.7	2.1	2.8	—	—	—	—
55–59	2.0	1.1	1.4	1.8	—	—	—	—
60–64	1.6	1.7	2.0	2.6	—	—	—	—
65–69	-1.2	1.1	1.3	1.7	—	—	—	—
70–74	-0.5	0.6	0.6	0.8	—	—	—	—
75–79	-0.8	0.8	0.9	1.2	—	—	—	—
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	—	—	—	—

Based on 2003 scorecard applied to non-panel households from 2002 and 2004.

Figure 10 (USAID "extreme" line): Bias and precision for bootstrapped estimates of individuals' poverty likelihoods, by sample size, 2003 scorecard applied to non-panel households in 2002

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	-1.4	48.3	56.0	70.3
4	-1.1	38.1	45.2	60.2
8	-1.5	27.3	32.6	43.9
16	-1.1	19.5	23.9	30.1
32	-0.8	14.5	17.5	23.8
64	-1.1	10.3	12.3	18.3
128	-1.2	7.4	9.0	14.0
256	-1.1	4.8	5.8	8.5
512	-1.0	3.1	3.7	4.9
1,024	-0.9	2.2	2.6	3.4
2,048	-0.9	1.5	1.7	2.3
4,096	-0.8	1.0	1.2	1.5
8,192	-0.8	0.7	0.8	1.1
16,384	-0.8	0.5	0.6	0.8

Figure 11 (USAID "extreme" line): Bias and precision for bootstrapped estimates of individuals' poverty likelihoods, by sample size, 2003 scorecard applied to non-panel households in 2004

This poverty line does not exist for 2004.

Figure 13 (USAID "extreme" line): Bias and precision for bootstrapped estimates of groups' poverty rates at a point in time, by sample size, 2003 scorecard applied to non-panel households in 2002

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	-1.4	48.3	56.0	70.3
4	-1.1	37.8	44.6	60.2
8	-1.5	26.7	31.4	43.8
16	-1.3	19.1	22.4	29.5
32	-1.0	13.5	16.6	22.0
64	-1.2	9.9	11.8	15.3
128	-1.2	7.2	8.3	11.6
256	-1.2	5.2	6.1	8.0
512	-1.1	3.5	4.2	5.4
1,024	-1.0	2.6	3.1	3.9
2,048	-1.0	1.8	2.2	2.9
4,096	-1.0	1.3	1.5	1.9
8,192	-1.0	0.9	1.0	1.4
16,384	-1.0	0.7	0.8	1.0

Figure 14 (USAID "extreme" line): Bias and precision for bootstrapped estimates of groups' poverty rates at a point in time, by sample size, 2003 scorecard applied to non-panel households in 2004

This poverty line does not exist for 2004.

Figure 16 (USAID "extreme" line): Bias and precision for bootstrapped estimates of changes in groups' poverty rates between two points in time, by sample size, 2003 scorecard applied to non-panel households in 2002 and 2003

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	2.0	74.7	86.8	100.0
4	1.1	55.0	64.6	85.5
8	2.4	38.3	48.7	62.0
16	1.4	28.3	33.4	42.0
32	1.1	19.3	23.3	30.4
64	1.3	14.2	17.0	21.2
128	1.4	10.3	11.9	16.0
256	1.4	7.6	9.0	12.1
512	1.2	5.2	6.2	8.1
1,024	1.2	3.7	4.5	5.7
2,048	1.2	2.5	3.1	3.9
4,096	1.2	1.9	2.3	2.9
8,192	1.2	1.3	1.5	1.9
16,384	1.2	0.9	1.1	1.4

Figure 17 (USAID "extreme" line): Bias and precision for bootstrapped estimates of changes in groups' poverty rates between two points in time, by sample size, 2003 scorecard applied to non-panel households in 2003 and 2004

This poverty line does not exist for 2004.

Figure 18 (USAID "extreme" line): Bias and precision for bootstrapped estimates of changes in groups' poverty rates between two points in time, by sample size, 2003 scorecard applied to non-panel households in 2002 and 2004

This poverty line does not exist for 2004.

Figure 20 (USAID "extreme" line): Bias and precision for bootstrapped estimates of changes in groups' poverty rates between two points in time, by sample size, 2003 scorecard applied to panel households in 2002 and 2003

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	-0.4	63.9	82.2	103.0
4	0.5	42.7	56.9	72.9
8	0.8	29.8	37.4	54.6
16	0.4	21.5	27.6	36.3
32	0.3	15.3	18.9	25.2
64	0.6	10.8	13.0	18.0
128	0.5	7.5	9.6	12.4
256	0.4	5.6	6.8	9.0
512	0.2	4.0	4.8	6.4
1,024	0.2	2.8	3.4	4.4
2,048	0.2	2.0	2.3	3.1
4,096	0.2	1.4	1.7	2.1
8,192	0.2	1.0	1.1	1.5
16,384	0.2	0.7	0.8	1.1

Figure 21 (USAID "extreme" line): Bias and precision for bootstrapped estimates of changes in groups' poverty rates between two points in time, by sample size, 2003 scorecard applied to panel households in 2003 and 2004

This poverty line does not exist for 2004.

Figure 22 (USAID "extreme" line): Bias and precision for bootstrapped estimates of changes in groups' poverty rates between two points in time, by sample size, 2003 scorecard applied to panel households in 2002 and 2004

This poverty line does not exist for 2004.

Figure 24 (USAID "extreme" line): People by targeting classification and score, along with "Total Accuracy" and BPAC, 2003 scorecard applied to non-panel households in 2002

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	5.3	23.3	1.1	70.4	75.6	-59.2
5–9	8.7	19.8	2.3	69.1	77.9	-30.6
10–14	12.8	15.7	4.8	66.7	79.5	6.4
15–19	16.0	12.5	7.1	64.4	80.4	37.4
20–24	19.9	8.7	10.6	60.8	80.7	62.7
25–29	23.0	5.5	15.6	55.9	78.9	45.3
30–34	24.8	3.7	20.3	51.2	76.0	28.7
35–39	26.1	2.4	24.9	46.6	72.7	12.7
40–44	27.3	1.2	30.5	41.0	68.3	-6.8
45–49	27.9	0.6	35.2	36.3	64.2	-23.4
50–54	28.2	0.3	39.7	31.8	60.0	-39.2
55–59	28.3	0.2	44.6	26.8	55.2	-56.5
60–64	28.4	0.1	49.4	22.0	50.5	-73.3
65–69	28.5	0.0	54.0	17.4	45.9	-89.5
70–74	28.5	0.0	57.9	13.6	42.1	-103.1
75–79	28.5	0.0	60.6	10.9	39.4	-112.3
80–84	28.5	0.0	64.4	7.1	35.6	-125.7
85–89	28.5	0.0	68.5	2.9	31.5	-140.3
90–94	28.5	0.0	69.0	2.5	31.0	-141.8
95–100	28.5	0.0	71.5	0.0	28.5	-150.6

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

Figure 25 (USAID "extreme" line): People by targeting classification and score, along with "Total Accuracy" and BPAC, 2003 scorecard applied to non-panel households in 2004

This poverty line does not exist for 2004.

Figure 26 (USAID "extreme" line): People below the poverty line and all people, at a given score or at or below a given score cut-off, 2003 scorecard applied to non-panel households in 2002

	People below poverty line (%)		All people (%)	
Score	At score	At or below score	At score	At or below score
0–4	82.3	82.3	6.4	6.4
5–9	74.1	78.9	4.7	11.1
10–14	62.2	72.7	6.5	17.6
15–19	58.7	69.3	5.6	23.1
20–24	51.8	65.1	7.4	30.5
25–29	38.7	59.6	8.1	38.6
30–34	28.2	55.0	6.6	45.2
35–39	21.7	51.2	5.8	51.0
40–44	18.0	47.3	6.8	57.8
45–49	10.7	44.2	5.3	63.1
50–54	6.1	41.5	4.8	67.9
55–59	2.5	38.8	5.1	73.0
60–64	2.1	36.5	4.9	77.9
65–69	1.4	34.5	4.7	82.5
70–74	0.5	33.0	3.9	86.4
75–79	0.7	32.0	2.7	89.1
80–84	0.0	30.7	3.8	92.9
85–89	0.0	29.4	4.2	97.1
90–94	0.0	29.3	0.4	97.5
95–100	0.0	28.5	2.5	100.0

Figure 27 (USAID "extreme" line): People below the poverty line and all people, at a given score or at or below a given score cut-off, 2003 scorecard applied to non-panel households in 2004

This poverty line does not exist for 2004.

$1/Day Poverty Line Tables

Figure 5 ($1/day line): Estimated poverty likelihoods associated with scores

If an individual's score is . . .	. . . then the likelihood (%) of being below the poverty line is:
0–4	88.1
5–9	67.1
10–14	59.6
15–19	42.8
20–24	46.9
25–29	33.9
30–34	29.4
35–39	18.9
40–44	13.3
45–49	7.1
50–54	3.7
55–59	3.8
60–64	3.5
65–69	0.1
70–74	0.0
75–79	0.0
80–84	0.0
85–89	0.0
90–94	0.0
95–100	0.0

Surveyed cases weighted to represent Peru's population.

Based on the 2003 ENAHO.

Figure 6 ($1/day line): Derivation of estimated poverty likelihoods associated with scores

Score	People below poverty line		All people at score		Poverty likelihood (estimated, %)
0–4	5,221	÷	5,923	=	88.1
5–9	3,735	÷	5,565	=	67.1
10–14	3,926	÷	6,585	=	59.6
15–19	1,903	÷	4,443	=	42.8
20–24	3,275	÷	6,984	=	46.9
25–29	2,747	÷	8,107	=	33.9
30–34	1,781	÷	6,064	=	29.4
35–39	1,132	÷	5,985	=	18.9
40–44	830	÷	6,221	=	13.3
45–49	384	÷	5,439	=	7.1
50–54	215	÷	5,775	=	3.7
55–59	187	÷	4,994	=	3.8
60–64	169	÷	4,788	=	3.5
65–69	6	÷	5,077	=	0.1
70–74	0	÷	3,683	=	0.0
75–79	0	÷	3,091	=	0.0
80–84	0	÷	3,888	=	0.0
85–89	0	÷	4,807	=	0.0
90–94	0	÷	380	=	0.0
95–100	0	÷	2,200	=	0.0

Surveyed cases weighted to represent Peru's population.
Based on the 2003 ENAHO.

Figure 8 ($1/day line): Bootstrapped differences between estimated and true poverty likelihoods for individuals in a large sample (n=16,384), with confidence intervals, 2003 scorecard applied to non-panel households in 2002 and 2004

	2003 scorecard applied to 2002 non-panel, difference between estimate and true value				2003 scorecard applied to 2004 non-panel, difference between estimate and true value			
		Confidence interval (+/- percentage points)				Confidence interval (+/- percentage points)		
Score	Bias	90-percent	95-percent	99-percent	Bias	90-percent	95-percent	99-percent
0–4	1.3	2.4	2.8	3.6	11.0	10.8	11.7	13.7
5–9	-13.0	8.1	8.4	8.8	-1.3	5.1	6.4	10.2
10–14	-9.4	6.4	6.7	7.5	0.8	5.3	6.0	7.8
15–19	-17.5	10.6	11.0	11.8	-3.6	5.2	6.9	9.2
20–24	-8.9	6.1	6.5	7.2	2.5	6.0	6.9	8.4
25–29	-7.9	5.3	5.7	6.1	-3.4	3.1	3.3	4.1
30–34	-0.3	3.3	4.0	4.9	2.2	5.2	6.1	7.8
35–39	-4.7	3.7	4.0	4.4	-5.0	3.6	3.8	4.3
40–44	-4.6	3.7	3.9	4.3	-1.5	2.1	2.7	4.2
45–49	-2.3	2.6	3.3	3.9	0.2	2.7	3.3	4.2
50–54	-1.5	1.6	2.0	2.6	-0.4	1.5	1.7	2.3
55–59	1.9	1.1	1.3	1.7	1.7	2.5	2.7	3.2
60–64	1.6	1.7	2.0	2.7	2.4	2.4	2.6	2.9
65–69	-0.7	0.9	1.0	1.4	0.1	0.1	0.1	0.1
70–74	-0.3	0.4	0.5	0.7	-0.1	0.1	0.1	0.1
75–79	-0.8	0.8	0.9	1.2	0.0	0.0	0.0	0.0
80–84	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
85–89	0.0	0.0	0.0	0.0	-0.4	0.5	0.6	0.7
90–94	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
95–100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Based on 2003 scorecard applied to non-panel households from 2002 and 2004.

Figure 10 ($1/day line): Bias and precision for bootstrapped estimates of individuals' poverty likelihoods, by sample size, 2003 scorecard applied to non-panel households in 2002

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	-5.7	45.8	56.9	71.3
4	-4.5	37.6	44.3	59.9
8	-4.8	26.8	33.0	45.0
16	-4.0	19.4	23.4	31.7
32	-3.5	14.5	17.5	23.8
64	-3.7	10.3	11.9	16.8
128	-3.7	7.5	9.2	12.7
256	-3.6	4.7	5.6	8.0
512	-3.4	3.2	3.7	5.0
1,024	-3.3	2.1	2.6	3.6
2,048	-3.3	1.4	1.8	2.3
4,096	-3.2	1.0	1.2	1.5
8,192	-3.2	0.7	0.8	1.0
16,384	-3.2	0.5	0.6	0.7

Figure 11 ($1/day line): Bias and precision for bootstrapped estimates of individuals' poverty likelihoods, by sample size, 2003 scorecard applied to non-panel households in 2004

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	1.1	47.2	57.9	71.4
4	0.3	34.3	41.5	53.4
8	-0.2	24.4	29.3	40.5
16	0.4	17.5	20.9	28.6
32	0.1	12.4	15.3	19.2
64	0.2	8.9	10.6	14.6
128	0.2	6.1	7.4	9.3
256	0.2	3.7	4.5	6.1
512	0.2	2.4	2.9	3.8
1,024	0.2	1.7	2.0	2.6
2,048	0.2	1.2	1.4	2.0
4,096	0.2	0.8	1.0	1.3
8,192	0.2	0.6	0.7	0.9
16,384	0.2	0.4	0.5	0.6

Figure 13 ($1/day line): Bias and precision for bootstrapped estimates of groups' poverty rates at a point in time, by sample size, 2003 scorecard applied to non-panel households in 2002

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	-5.7	45.8	56.9	71.3
4	-4.5	37.7	44.3	59.5
8	-4.9	25.8	32.1	43.8
16	-4.3	19.2	23.1	30.5
32	-3.9	13.0	16.3	21.7
64	-4.1	10.0	11.8	15.0
128	-4.3	7.3	8.5	10.8
256	-4.3	5.1	5.8	7.6
512	-4.2	3.7	4.3	5.5
1,024	-4.1	2.5	3.0	4.0
2,048	-4.1	1.7	2.2	2.8
4,096	-4.1	1.2	1.5	2.0
8,192	-4.1	0.9	1.0	1.4
16,384	-4.1	0.6	0.7	1.0

Figure 14 ($1/day line): Bias and precision for bootstrapped estimates of groups' poverty rates at a point in time, by sample size, 2003 scorecard applied to non-panel households in 2004

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	1.1	47.2	57.9	71.4
4	0.3	34.1	41.4	54.8
8	-0.2	24.1	28.7	37.4
16	0.2	16.5	20.0	27.9
32	-0.2	11.4	13.8	19.0
64	-0.2	8.3	9.9	13.4
128	-0.2	5.9	7.1	9.3
256	-0.2	4.0	4.8	6.3
512	-0.1	2.7	3.2	4.2
1,024	-0.1	2.1	2.5	3.3
2,048	-0.1	1.4	1.7	2.3
4,096	-0.1	1.0	1.2	1.6
8,192	-0.1	0.8	0.9	1.1
16,384	-0.1	0.5	0.6	0.8

Figure 16 ($1/day line): Bias and precision for bootstrapped estimates of changes in groups' poverty rates between two points in time, by sample size, 2003 scorecard applied to non-panel households in 2002 and 2003

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	5.0	73.9	84.9	101.5
4	3.7	53.3	66.7	83.1
8	5.3	37.9	45.7	61.7
16	4.2	27.2	32.3	44.2
32	4.0	20.2	23.7	30.2
64	4.3	14.4	16.8	21.5
128	4.6	10.0	12.3	15.7
256	4.5	7.4	8.8	11.4
512	4.4	4.9	5.9	8.3
1,024	4.3	3.7	4.3	5.5
2,048	4.2	2.5	3.1	4.1
4,096	4.2	1.8	2.1	2.7
8,192	4.2	1.3	1.5	1.8
16,384	4.2	0.9	1.0	1.3

Figure 17 ($1/day line): Bias and precision for bootstrapped estimates of changes in groups' poverty rates between two points in time, by sample size, 2003 scorecard applied to non-panel households in 2003 and 2004

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	1.9	69.6	81.2	102.2
4	1.1	50.1	59.2	78.4
8	-0.6	34.7	43.4	58.9
16	0.3	25.7	31.6	41.0
32	-0.3	18.1	22.3	30.0
64	-0.4	12.7	15.3	19.7
128	-0.5	9.1	10.8	14.1
256	-0.4	6.5	8.0	10.5
512	-0.3	4.6	5.4	6.7
1,024	-0.3	3.2	3.8	5.1
2,048	-0.3	2.2	2.6	3.5
4,096	-0.3	1.6	1.9	2.6
8,192	-0.3	1.1	1.3	1.8
16,384	-0.3	0.8	0.9	1.2

Figure 18 ($1/day line): Bias and precision for bootstrapped estimates of changes in groups' poverty rates between two points in time, by sample size, 2003 scorecard applied to non-panel households in 2002 and 2004

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	6.8	70.0	80.5	102.6
4	4.8	49.7	61.5	80.2
8	4.7	36.1	43.5	54.4
16	4.5	24.8	30.3	41.6
32	3.7	17.9	21.2	28.9
64	3.9	12.9	15.1	19.8
128	4.1	9.4	11.0	13.5
256	4.1	6.4	7.7	10.5
512	4.1	4.5	5.1	7.0
1,024	4.0	3.3	4.0	5.4
2,048	4.0	2.3	2.7	3.8
4,096	4.0	1.6	1.9	2.4
8,192	4.0	1.2	1.4	1.7
16,384	4.0	0.8	1.0	1.3

Figure 20 ($1/day line): Bias and precision for bootstrapped estimates of changes in groups' poverty rates between two points in time, by sample size, 2003 scorecard applied to panel households in 2002 and 2003

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	2.6	58.2	77.2	102.3
4	3.1	41.5	53.3	80.5
8	2.7	29.8	39.3	54.4
16	2.3	21.0	27.1	36.8
32	2.4	14.8	18.0	26.6
64	2.8	10.5	12.3	18.4
128	2.7	7.3	8.7	11.8
256	2.6	5.1	6.2	9.1
512	2.4	3.9	4.8	6.5
1,024	2.4	2.7	3.3	4.5
2,048	2.4	1.9	2.1	2.9
4,096	2.4	1.3	1.6	2.0
8,192	2.4	0.9	1.1	1.4
16,384	2.4	0.7	0.8	1.1

Figure 21 ($1/day line): Bias and precision for bootstrapped estimates of changes in groups' poverty rates between two points in time, by sample size, 2003 scorecard applied to panel households in 2003 and 2004

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	0.0	49.5	64.9	87.9
4	-0.4	30.1	39.1	61.1
8	-0.4	20.3	26.4	36.3
16	-0.1	13.7	17.1	22.0
32	0.2	9.4	11.3	15.6
64	0.1	6.8	8.3	10.9
128	0.2	4.8	5.7	7.6
256	0.2	3.2	4.0	5.2
512	0.2	2.3	2.8	3.6
1,024	0.2	1.6	2.0	2.5
2,048	0.2	1.2	1.4	1.9
4,096	0.2	0.8	0.9	1.3
8,192	0.2	0.6	0.7	0.9
16,384	0.2	0.4	0.5	0.6

Figure 22 ($1/day line): Bias and precision for bootstrapped estimates of changes in groups' poverty rates between two points in time, by sample size, 2003 scorecard applied to panel households in 2002 and 2004

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	-0.3	0.8	0.9	1.2
4	-0.3	0.8	0.9	1.2
8	-0.3	0.8	0.9	1.2
16	-0.3	0.8	0.9	1.2
32	-0.3	0.8	0.9	1.2
64	-0.3	0.8	0.9	1.2
128	-0.3	0.8	0.9	1.2
256	-0.3	0.8	0.9	1.2
512	-0.3	0.8	0.9	1.2
1,024	-0.3	0.8	0.9	1.2
2,048	-0.3	0.8	0.9	1.2
4,096	-0.3	0.8	0.9	1.2
8,192	-0.3	0.8	0.9	1.2
16,384	-0.3	0.8	0.9	1.2

Figure 24 ($1/day line): People by targeting classification and score, along with "Total Accuracy" and BPAC, 2003 scorecard applied to non-panel households in 2002

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	2.2	4.4	4.2	89.2	91.3	29.3
5–9	3.4	3.2	7.7	85.7	89.1	-16.4
10–14	4.5	2.1	13.0	80.3	84.9	-97.7
15–19	5.2	1.4	17.9	75.5	80.7	-171.2
20–24	5.8	0.8	24.7	68.7	74.5	-274.4
25–29	6.2	0.4	32.3	61.1	67.3	-389.9
30–34	6.3	0.3	38.8	54.6	60.9	-488.1
35–39	6.5	0.1	44.5	48.9	55.3	-574.7
40–44	6.6	0.0	51.2	42.2	48.8	-675.7
45–49	6.6	0.0	56.5	36.9	43.5	-755.7
50–54	6.6	0.0	61.3	32.1	38.7	-828.5
55–59	6.6	0.0	66.3	27.0	33.7	-905.1
60–64	6.6	0.0	71.2	22.1	28.8	-979.3
65–69	6.6	0.0	75.9	17.5	24.1	-1,050.0
70–74	6.6	0.0	79.8	13.6	20.2	-1,109.2
75–79	6.6	0.0	82.5	10.9	17.5	-1,149.4
80–84	6.6	0.0	86.3	7.1	13.7	-1,207.1
85–89	6.6	0.0	90.5	2.9	9.5	-1,270.2
90–94	6.6	0.0	90.9	2.5	9.1	-1,276.6
95–100	6.6	0.0	93.4	0.0	6.6	-1,314.8

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

Figure 25 ($1/day line): People by targeting classification and score, along with "Total Accuracy" and BPAC, 2003 scorecard applied to non-panel households in 2004

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
0–4	0.4	2.0	2.8	94.7	95.1
5–9	0.7	1.7	5.4	92.1	92.9
10–14	1.3	1.2	10.0	87.5	88.8
15–19	1.4	1.1	12.7	84.9	86.3
20–24	1.6	0.8	18.3	79.2	80.9
25–29	2.0	0.5	25.5	72.1	74.1
30–34	2.0	0.4	31.0	66.6	68.6
35–39	2.4	0.1	40.5	57.0	59.4
40–44	2.4	0.0	47.7	49.9	52.3
45–49	2.4	0.0	53.8	43.7	46.2
50–54	2.4	0.0	59.8	37.7	40.2
55–59	2.4	0.0	65.7	31.9	34.3
60–64	2.4	0.0	70.9	26.7	29.1
65–69	2.4	0.0	76.4	21.1	23.6
70–74	2.4	0.0	80.9	16.7	19.1
75–79	2.4	0.0	84.5	13.0	15.5
80–84	2.4	0.0	88.7	8.9	11.3
85–89	2.4	0.0	94.4	3.2	5.6
90–94	2.4	0.0	94.9	2.7	5.1
95–100	2.4	0.0	97.6	0.0	2.4

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

Figure 26 ($1/day line): People below the poverty line and all people, at a given score or at or below a given score cut-off, 2003 scorecard applied to non-panel households in 2002

	People below poverty line (%)		All people (%)	
Score	At score	At or below score	At score	At or below score
0–4	33.8	33.8	6.4	6.4
5–9	26.1	30.5	4.7	11.1
10–14	17.7	25.8	6.5	17.6
15–19	12.6	22.6	5.6	23.1
20–24	7.7	19.0	7.4	30.5
25–29	5.5	16.2	8.1	38.6
30–34	1.5	14.0	6.6	45.2
35–39	2.2	12.7	5.8	51.0
40–44	1.7	11.4	6.8	57.8
45–49	0.3	10.5	5.3	63.1
50–54	0.0	9.7	4.8	67.9
55–59	0.0	9.0	5.1	73.0
60–64	0.0	8.5	4.9	77.9
65–69	0.0	8.0	4.7	82.5
70–74	0.0	7.6	3.9	86.4
75–79	0.0	7.4	2.7	89.1
80–84	0.0	7.1	3.8	92.9
85–89	0.0	6.8	4.2	97.1
90–94	0.0	6.8	0.4	97.5
95–100	0.0	6.6	2.5	100.0

Figure 27 ($1/day line): People below the poverty line and all people, at a given score or at or below a given score cut-off, 2003 scorecard applied to non-panel households in 2004

	People below poverty line (%)		All people (%)	
Score	At score	At or below score	At score	At or below score
0–4	12.5	12.5	3.2	3.2
5–9	11.5	12.0	2.9	6.2
10–14	10.1	11.2	5.1	11.3
15–19	3.6	9.7	2.7	14.0
20–24	4.8	8.2	6.0	20.0
25–29	4.6	7.3	7.5	27.5
30–34	1.0	6.2	5.6	33.0
35–39	3.1	5.5	9.8	42.9
40–44	0.6	4.8	7.2	50.1
45–49	0.5	4.3	6.1	56.2
50–54	0.2	3.9	6.0	62.3
55–59	0.0	3.6	5.8	68.1
60–64	0.0	3.3	5.2	73.3
65–69	0.0	3.1	5.5	78.9
70–74	0.0	2.9	4.5	83.3
75–79	0.0	2.8	3.6	87.0
80–84	0.0	2.7	4.1	91.1
85–89	0.0	2.5	5.7	96.8
90–94	0.0	2.5	0.5	97.3
95–100	0.0	2.4	2.7	100.0

$2/Day Poverty Line Tables

Figure 5 ($2/day line): Estimated poverty likelihoods associated with scores

If an individual's score is . . .	. . . then the likelihood (%) of being below the poverty line is:
0–4	88.1
5–9	67.1
10–14	59.6
15–19	42.8
20–24	46.9
25–29	33.9
30–34	29.4
35–39	18.9
40–44	13.3
45–49	7.1
50–54	3.7
55–59	3.8
60–64	3.5
65–69	0.1
70–74	0.0
75–79	0.0
80–84	0.0
85–89	0.0
90–94	0.0
95–100	0.0

Surveyed cases weighted to represent Peru's population.
Based on the 2003 ENAHO.

Figure 6 ($2/day line): Derivation of estimated poverty likelihoods associated with scores

Score	People below poverty line		All people at score		Poverty likelihood (estimated, %)
0–4	5,221	÷	5,923	=	88.1
5–9	3,735	÷	5,565	=	67.1
10–14	3,926	÷	6,585	=	59.6
15–19	1,903	÷	4,443	=	42.8
20–24	3,275	÷	6,984	=	46.9
25–29	2,747	÷	8,107	=	33.9
30–34	1,781	÷	6,064	=	29.4
35–39	1,132	÷	5,985	=	18.9
40–44	830	÷	6,221	=	13.3
45–49	384	÷	5,439	=	7.1
50–54	215	÷	5,775	=	3.7
55–59	187	÷	4,994	=	3.8
60–64	169	÷	4,788	=	3.5
65–69	6	÷	5,077	=	0.1
70–74	0	÷	3,683	=	0.0
75–79	0	÷	3,091	=	0.0
80–84	0	÷	3,888	=	0.0
85–89	0	÷	4,807	=	0.0
90–94	0	÷	380	=	0.0
95–100	0	÷	2,200	=	0.0

Surveyed cases weighted to represent Peru's population.
Based on the 2003 ENAHO.

Figure 8 ($2/day line): Bootstrapped differences between estimated and true poverty likelihoods for individuals in a large sample (n=16,384), with confidence intervals, 2003 scorecard applied to non-panel households in 2002 and 2004

	2003 scorecard applied to 2002 non-panel, difference between estimate and true value				2003 scorecard applied to 2004 non-panel, difference between estimate and true value			
		Confidence interval (+/- percentage points)				Confidence interval (+/- percentage points)		
Score	Bias	90-percent	95-percent	99-percent	Bias	90-percent	95-percent	99-percent
0–4	1.3	2.4	2.8	3.6	11.0	10.8	11.7	13.7
5–9	-13.0	8.1	8.4	8.8	-1.3	5.1	6.4	10.2
10–14	-9.4	6.4	6.7	7.5	0.8	5.3	6.0	7.8
15–19	-17.5	10.6	11.0	11.8	-3.6	5.2	6.9	9.2
20–24	-8.9	6.1	6.5	7.2	2.5	6.0	6.9	8.4
25–29	-7.9	5.3	5.7	6.1	-3.4	3.1	3.3	4.1
30–34	-0.3	3.3	4.0	4.9	2.2	5.2	6.1	7.8
35–39	-4.7	3.7	4.0	4.4	-5.0	3.6	3.8	4.3
40–44	-4.6	3.7	3.9	4.3	-1.5	2.1	2.7	4.2
45–49	-2.3	2.6	3.3	3.9	0.2	2.7	3.3	4.2
50–54	-1.5	1.6	2.0	2.6	-0.4	1.5	1.7	2.3
55–59	1.9	1.1	1.3	1.7	1.7	2.5	2.7	3.2
60–64	1.6	1.7	2.0	2.7	2.4	2.4	2.6	2.9
65–69	-0.7	0.9	1.0	1.4	0.1	0.1	0.1	0.1
70–74	-0.3	0.4	0.5	0.7	-0.1	0.1	0.1	0.1
75–79	-0.8	0.8	0.9	1.2	0.0	0.0	0.0	0.0
80–84	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
85–89	0.0	0.0	0.0	0.0	-0.4	0.5	0.6	0.7
90–94	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
95–100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Based on 2003 scorecard applied to non-panel households from 2002 and 2004.

Figure 10 ($2/day line): Bias and precision for bootstrapped estimates of individuals' poverty likelihoods, by sample size, 2003 scorecard applied to non-panel households in 2002

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	-5.7	45.8	56.9	71.3
4	-4.5	37.6	44.3	59.9
8	-4.8	26.8	33.0	45.0
16	-4.0	19.4	23.4	31.7
32	-3.5	14.5	17.5	23.8
64	-3.7	10.3	11.9	16.8
128	-3.7	7.5	9.2	12.7
256	-3.6	4.7	5.6	8.0
512	-3.4	3.2	3.7	5.0
1,024	-3.3	2.1	2.6	3.6
2,048	-3.3	1.4	1.8	2.3
4,096	-3.2	1.0	1.2	1.5
8,192	-3.2	0.7	0.8	1.0
16,384	-3.2	0.5	0.6	0.7

Figure 11 ($2/day line): Bias and precision for bootstrapped estimates of individuals' poverty likelihoods, by sample size, 2003 scorecard applied to non-panel households in 2004

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	1.1	47.2	57.9	71.4
4	0.3	34.3	41.5	53.4
8	-0.2	24.4	29.3	40.5
16	0.4	17.5	20.9	28.6
32	0.1	12.4	15.3	19.2
64	0.2	8.9	10.6	14.6
128	0.2	6.1	7.4	9.3
256	0.2	3.7	4.5	6.1
512	0.2	2.4	2.9	3.8
1,024	0.2	1.7	2.0	2.6
2,048	0.2	1.2	1.4	2.0
4,096	0.2	0.8	1.0	1.3
8,192	0.2	0.6	0.7	0.9
16,384	0.2	0.4	0.5	0.6

Figure 13 ($2/day line): Bias and precision for bootstrapped estimates of groups' poverty rates at a point in time, by sample size, 2003 scorecard applied to non-panel households in 2002

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	-5.7	45.8	56.9	71.3
4	-4.5	37.7	44.3	59.5
8	-4.9	25.8	32.1	43.8
16	-4.3	19.2	23.1	30.5
32	-3.9	13.0	16.3	21.7
64	-4.1	10.0	11.8	15.0
128	-4.3	7.3	8.5	10.8
256	-4.3	5.1	5.8	7.6
512	-4.2	3.7	4.3	5.5
1,024	-4.1	2.5	3.0	4.0
2,048	-4.1	1.7	2.2	2.8
4,096	-4.1	1.2	1.5	2.0
8,192	-4.1	0.9	1.0	1.4
16,384	-4.1	0.6	0.7	1.0

Figure 14 ($2/day line): Bias and precision for bootstrapped estimates of groups' poverty rates at a point in time, by sample size, 2003 scorecard applied to non-panel households in 2004

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	1.1	47.2	57.9	71.4
4	0.3	34.1	41.4	54.8
8	-0.2	24.1	28.7	37.4
16	0.2	16.5	20.0	27.9
32	-0.2	11.4	13.8	19.0
64	-0.2	8.3	9.9	13.4
128	-0.2	5.9	7.1	9.3
256	-0.2	4.0	4.8	6.3
512	-0.1	2.7	3.2	4.2
1,024	-0.1	2.1	2.5	3.3
2,048	-0.1	1.4	1.7	2.3
4,096	-0.1	1.0	1.2	1.6
8,192	-0.1	0.8	0.9	1.1
16,384	-0.1	0.5	0.6	0.8

Figure 16 ($2/day line): Bias and precision for bootstrapped estimates of changes in groups' poverty rates between two points in time, by sample size, 2003 scorecard applied to non-panel households in 2002 and 2003

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	5.0	73.9	84.9	101.5
4	3.7	53.3	66.7	83.1
8	5.3	37.9	45.7	61.7
16	4.2	27.2	32.3	44.2
32	4.0	20.2	23.7	30.2
64	4.3	14.4	16.8	21.5
128	4.6	10.0	12.3	15.7
256	4.5	7.4	8.8	11.4
512	4.4	4.9	5.9	8.3
1,024	4.3	3.7	4.3	5.5
2,048	4.2	2.5	3.1	4.1
4,096	4.2	1.8	2.1	2.7
8,192	4.2	1.3	1.5	1.8
16,384	4.2	0.9	1.0	1.3

Figure 17 ($2/day line): Bias and precision for bootstrapped estimates of changes in groups' poverty rates between two points in time, by sample size, 2003 scorecard applied to non-panel households in 2003 and 2004

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	1.9	69.6	81.2	102.2
4	1.1	50.1	59.2	78.4
8	-0.6	34.7	43.4	58.9
16	0.3	25.7	31.6	41.0
32	-0.3	18.1	22.3	30.0
64	-0.4	12.7	15.3	19.7
128	-0.5	9.1	10.8	14.1
256	-0.4	6.5	8.0	10.5
512	-0.3	4.6	5.4	6.7
1,024	-0.3	3.2	3.8	5.1
2,048	-0.3	2.2	2.6	3.5
4,096	-0.3	1.6	1.9	2.6
8,192	-0.3	1.1	1.3	1.8
16,384	-0.3	0.8	0.9	1.2

Figure 18 ($2/day line): Bias and precision for bootstrapped estimates of changes in groups' poverty rates between two points in time, by sample size, 2003 scorecard applied to non-panel households in 2002 and 2004

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	6.8	70.0	80.5	102.6
4	4.8	49.7	61.5	80.2
8	4.7	36.1	43.5	54.4
16	4.5	24.8	30.3	41.6
32	3.7	17.9	21.2	28.9
64	3.9	12.9	15.1	19.8
128	4.1	9.4	11.0	13.5
256	4.1	6.4	7.7	10.5
512	4.1	4.5	5.1	7.0
1,024	4.0	3.3	4.0	5.4
2,048	4.0	2.3	2.7	3.8
4,096	4.0	1.6	1.9	2.4
8,192	4.0	1.2	1.4	1.7
16,384	4.0	0.8	1.0	1.3

Figure 20 ($2/day line): Bias and precision for bootstrapped estimates of changes in groups' poverty rates between two points in time, by sample size, 2003 scorecard applied to panel households in 2002 and 2003

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	2.6	58.2	77.2	102.3
4	3.1	41.5	53.3	80.5
8	2.7	29.8	39.3	54.4
16	2.3	21.0	27.1	36.8
32	2.4	14.8	18.0	26.6
64	2.8	10.5	12.3	18.4
128	2.7	7.3	8.7	11.8
256	2.6	5.1	6.2	9.1
512	2.4	3.9	4.8	6.5
1,024	2.4	2.7	3.3	4.5
2,048	2.4	1.9	2.1	2.9
4,096	2.4	1.3	1.6	2.0
8,192	2.4	0.9	1.1	1.4
16,384	2.4	0.7	0.8	1.1

Figure 21 ($2/day line): Bias and precision for bootstrapped estimates of changes in groups' poverty rates between two points in time, by sample size, 2003 scorecard applied to panel households in 2003 and 2004

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	0.0	49.5	64.9	87.9
4	-0.4	30.1	39.1	61.1
8	-0.4	20.3	26.4	36.3
16	-0.1	13.7	17.1	22.0
32	0.2	9.4	11.3	15.6
64	0.1	6.8	8.3	10.9
128	0.2	4.8	5.7	7.6
256	0.2	3.2	4.0	5.2
512	0.2	2.3	2.8	3.6
1,024	0.2	1.6	2.0	2.5
2,048	0.2	1.2	1.4	1.9
4,096	0.2	0.8	0.9	1.3
8,192	0.2	0.6	0.7	0.9
16,384	0.2	0.4	0.5	0.6

Figure 22 ($2/day line): Bias and precision for bootstrapped estimates of changes in groups' poverty rates between two points in time, by sample size, 2003 scorecard applied to panel households in 2002 and 2004

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	-0.3	0.8	0.9	1.2
4	-0.3	0.8	0.9	1.2
8	-0.3	0.8	0.9	1.2
16	-0.3	0.8	0.9	1.2
32	-0.3	0.8	0.9	1.2
64	-0.3	0.8	0.9	1.2
128	-0.3	0.8	0.9	1.2
256	-0.3	0.8	0.9	1.2
512	-0.3	0.8	0.9	1.2
1,024	-0.3	0.8	0.9	1.2
2,048	-0.3	0.8	0.9	1.2
4,096	-0.3	0.8	0.9	1.2
8,192	-0.3	0.8	0.9	1.2
16,384	-0.3	0.8	0.9	1.2

Figure 24 ($2/day line): People by targeting classification and score, along with "Total Accuracy" and BPAC, 2003 scorecard applied to non-panel households in 2002

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	5.5	24.6	0.8	69.0	74.5	-60.5
5–9	9.3	20.9	1.8	68.0	77.3	-32.6
10–14	13.8	16.4	3.8	66.0	79.8	3.9
15–19	17.1	13.0	6.0	63.8	81.0	33.4
20–24	21.3	8.9	9.3	60.6	81.8	69.3
25–29	24.6	5.6	14.0	55.9	80.5	53.8
30–34	26.6	3.6	18.6	51.2	77.8	38.4
35–39	28.0	2.2	23.1	46.8	74.7	23.6
40–44	29.2	1.0	28.6	41.2	70.4	5.2
45–49	29.7	0.5	33.4	36.4	66.1	-10.7
50–54	29.9	0.3	38.0	31.8	61.8	-25.8
55–59	30.0	0.2	42.9	26.9	56.9	-42.2
60–64	30.1	0.1	47.7	22.1	52.2	-58.1
65–69	30.2	0.0	52.4	17.4	47.6	-73.5
70–74	30.2	0.0	56.3	13.6	43.7	-86.4
75–79	30.2	0.0	58.9	10.9	41.1	-95.1
80–84	30.2	0.0	62.7	7.1	37.3	-107.7
85–89	30.2	0.0	66.9	2.9	33.1	-121.5
90–94	30.2	0.0	67.3	2.5	32.7	-122.9
95–100	30.2	0.0	69.8	0.0	30.2	-131.3

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

Figure 25 ($2/day line): People by targeting classification and score, along with "Total Accuracy" and BPAC, 2003 scorecard applied to non-panel households in 2004

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	2.5	17.5	0.7	79.2	81.7	-71.4
5–9	4.5	15.5	1.7	78.3	82.8	-46.7
10–14	7.5	12.5	3.8	76.2	83.7	-6.0
15–19	8.8	11.3	5.2	74.7	83.5	13.8
20–24	11.4	8.6	8.5	71.4	82.9	56.7
25–29	14.2	5.8	13.2	66.7	80.9	33.9
30–34	15.7	4.3	17.3	62.7	78.4	13.7
35–39	18.1	1.9	24.8	55.2	73.3	-23.7
40–44	19.2	0.9	30.9	49.0	68.2	-54.3
45–49	19.6	0.5	36.6	43.3	62.9	-82.8
50–54	19.8	0.2	42.4	37.5	57.4	-111.7
55–59	20.0	0.1	48.1	31.8	51.8	-140.2
60–64	20.0	0.0	53.3	26.6	46.7	-166.0
65–69	20.0	0.0	58.9	21.1	41.1	-193.7
70–74	20.0	0.0	63.3	16.6	36.7	-215.9
75–79	20.0	0.0	66.9	13.0	33.0	-234.0
80–84	20.0	0.0	71.1	8.9	28.9	-254.7
85–89	20.0	0.0	76.8	3.2	23.2	-283.0
90–94	20.0	0.0	77.3	2.7	22.7	-285.7
95–100	20.0	0.0	80.0	0.0	20.0	-298.9

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

Figure 26 ($2/day line): People below the poverty line and all people, at a given score or at or below a given score cut-off, 2003 scorecard applied to non-panel households in 2002

	People below poverty line (%)		All people (%)	
Score	At score	At or below score	At score	At or below score
0–4	86.9	86.9	6.4	6.4
5–9	80.1	84.0	4.7	11.1
10–14	69.0	78.5	6.5	17.6
15–19	60.2	74.1	5.6	23.1
20–24	55.8	69.7	7.4	30.5
25–29	41.8	63.8	8.1	38.6
30–34	29.6	58.8	6.6	45.2
35–39	23.6	54.8	5.8	51.0
40–44	18.0	50.5	6.8	57.8
45–49	9.4	47.0	5.3	63.1
50–54	5.3	44.1	4.8	67.9
55–59	1.9	41.2	5.1	73.0
60–64	1.9	38.7	4.9	77.9
65–69	0.8	36.5	4.7	82.5
70–74	0.3	34.9	3.9	86.4
75–79	0.7	33.9	2.7	89.1
80–84	0.0	32.5	3.8	92.9
85–89	0.0	31.1	4.2	97.1
90–94	0.0	31.0	0.4	97.5
95–100	0.0	30.2	2.5	100.0

Figure 27 ($2/day line): People below the poverty line and all people, at a given score or at or below a given score cut-off, 2003 scorecard applied to non-panel households in 2004

	People below poverty line (%)		All people (%)	
Score	At score	At or below score	At score	At or below score
0–4	77.1	77.1	3.2	3.2
5–9	68.5	73.0	2.9	6.2
10–14	58.8	66.6	5.1	11.3
15–19	46.5	62.7	2.7	14.0
20–24	44.5	57.3	6.0	20.0
25–29	37.3	51.8	7.5	27.5
30–34	27.2	47.6	5.6	33.0
35–39	24.0	42.2	9.8	42.9
40–44	14.8	38.3	7.2	50.1
45–49	6.9	34.8	6.1	56.2
50–54	4.2	31.9	6.0	62.3
55–59	2.0	29.3	5.8	68.1
60–64	1.2	27.3	5.2	73.3
65–69	0.0	25.4	5.5	78.9
70–74	0.1	24.0	4.5	83.3
75–79	0.0	23.0	3.6	87.0
80–84	0.0	22.0	4.1	91.1
85–89	0.4	20.7	5.7	96.8
90–94	0.0	20.6	0.5	97.3
95–100	0.0	20.0	2.7	100.0

$3/Day Poverty Line Tables

Figure 5 ($3/day line): Estimated poverty likelihoods associated with scores

If an individual's score is . . .	. . . then the likelihood (%) of being below the poverty line is:
0–4	99.1
5–9	92.9
10–14	91.4
15–19	88.0
20–24	81.1
25–29	71.1
30–34	66.4
35–39	51.2
40–44	48.6
45–49	33.0
50–54	24.4
55–59	20.2
60–64	11.2
65–69	8.4
70–74	2.8
75–79	4.0
80–84	3.1
85–89	0.6
90–94	0.0
95–100	2.0

Surveyed cases weighted to represent Peru's population.
Based on the 2003 ENAHO.

Figure 6 ($3/day line): Derivation of estimated poverty likelihoods associated with scores

Score	People below poverty line		All people at score		Poverty likelihood (estimated, %)
0–4	5,870	÷	5,923	=	99.1
5–9	5,170	÷	5,565	=	92.9
10–14	6,019	÷	6,585	=	91.4
15–19	3,911	÷	4,443	=	88.0
20–24	5,663	÷	6,984	=	81.1
25–29	5,761	÷	8,107	=	71.1
30–34	4,025	÷	6,064	=	66.4
35–39	3,064	÷	5,985	=	51.2
40–44	3,020	÷	6,221	=	48.6
45–49	1,792	÷	5,439	=	33.0
50–54	1,406	÷	5,775	=	24.4
55–59	1,010	÷	4,994	=	20.2
60–64	536	÷	4,788	=	11.2
65–69	425	÷	5,077	=	8.4
70–74	102	÷	3,683	=	2.8
75–79	124	÷	3,091	=	4.0
80–84	122	÷	3,888	=	3.1
85–89	28	÷	4,807	=	0.6
90–94	0	÷	380	=	0.0
95–100	45	÷	2,200	=	2.0

Surveyed cases weighted to represent Peru's population.
Based on the 2003 ENAHO.

Figure 8 ($3/day line): Bootstrapped differences between estimated and true poverty likelihoods for individuals in a large sample (n=16,384), with confidence intervals, 2003 scorecard applied to non-panel households in 2002 and 2004

	2003 scorecard applied to 2002 non-panel, difference between estimate and true value				2003 scorecard applied to 2004 non-panel, difference between estimate and true value			
		Confidence interval (+/- percentage points)				Confidence interval (+/- percentage points)		
Score	Bias	90-percent	95-percent	99-percent	Bias	90-percent	95-percent	99-percent
0–4	1.0	0.9	1.1	1.3	2.9	3.8	4.3	5.4
5–9	-4.4	2.8	2.9	3.1	0.3	3.4	4.4	5.8
10–14	1.0	2.4	2.8	3.7	1.0	3.6	4.2	5.5
15–19	-0.3	2.8	3.4	4.4	0.3	4.5	5.5	6.8
20–24	-4.4	3.2	3.4	3.7	1.7	4.7	5.7	7.5
25–29	-9.2	5.7	5.9	6.4	-2.4	2.7	3.5	5.0
30–34	-0.6	3.2	3.9	4.8	0.5	4.8	5.6	8.0
35–39	-5.4	4.2	4.5	5.1	-5.5	3.9	4.1	4.3
40–44	-4.4	4.0	4.3	5.8	2.9	5.9	6.8	8.3
45–49	-5.7	4.8	5.2	6.1	2.4	5.9	6.7	8.8
50–54	-5.1	4.4	4.7	6.1	-0.9	4.2	5.1	7.1
55–59	3.0	3.2	3.9	5.2	5.1	6.4	7.1	8.6
60–64	-0.4	3.0	3.6	4.8	2.2	4.3	4.9	6.3
65–69	-4.3	3.8	4.0	4.5	1.1	3.9	4.6	5.6
70–74	-3.2	3.3	3.7	5.2	-1.3	1.8	2.2	3.0
75–79	-2.1	3.3	3.8	5.1	1.5	2.9	3.2	3.9
80–84	-0.5	2.1	2.5	3.4	0.8	2.5	2.8	3.7
85–89	-0.8	0.9	1.1	1.6	-2.2	1.8	1.9	2.3
90–94	-6.6	8.3	9.8	12.8	0.0	0.0	0.0	0.0
95–100	0.9	1.5	1.7	2.1	2.0	1.0	1.0	1.0

Based on 2003 scorecard applied to non-panel households from 2002 and 2004.

Figure 10 ($3/day line): Bias and precision for bootstrapped estimates of individuals' poverty likelihoods, by sample size, 2003 scorecard applied to non-panel households in 2002

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	-2.0	46.7	59.8	76.1
4	-2.1	37.8	46.8	60.9
8	-2.5	26.8	33.6	50.4
16	-2.6	20.8	24.9	37.6
32	-2.3	15.5	19.5	27.2
64	-2.3	11.6	14.1	19.5
128	-2.6	8.5	10.4	14.6
256	-2.5	6.0	7.3	10.0
512	-2.5	4.0	4.9	7.3
1,024	-2.4	2.9	3.7	5.3
2,048	-2.4	2.2	2.7	3.7
4,096	-2.4	1.6	1.8	2.5
8,192	-2.4	1.1	1.3	1.9
16,384	-2.4	0.8	1.0	1.3

Figure 11 ($3/day line): Bias and precision for bootstrapped estimates of individuals' poverty likelihoods, by sample size, 2003 scorecard applied to non-panel households in 2004

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	1.5	50.1	59.7	78.5
4	1.1	38.6	47.1	56.7
8	0.2	28.2	34.7	45.5
16	0.8	20.7	24.8	32.8
32	0.5	14.0	17.0	23.7
64	0.9	10.0	12.1	16.7
128	0.5	6.8	7.8	11.4
256	0.6	4.3	5.2	7.0
512	0.6	2.8	3.4	4.4
1,024	0.6	2.0	2.4	3.1
2,048	0.6	1.3	1.5	2.0
4,096	0.6	0.9	1.1	1.3
8,192	0.6	0.6	0.8	1.1
16,384	0.6	0.5	0.6	0.7

Figure 13 ($3/day line): Bias and precision for bootstrapped estimates of groups' poverty rates at a point in time, by sample size, 2003 scorecard applied to non-panel households in 2002

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	-2.0	46.7	59.8	76.1
4	-2.0	37.3	46.6	60.4
8	-2.5	26.9	33.8	49.6
16	-2.8	19.7	24.0	34.5
32	-2.6	14.9	17.8	24.8
64	-2.6	10.7	12.5	16.3
128	-2.7	7.1	8.4	11.5
256	-2.7	5.4	6.3	7.9
512	-2.7	3.8	4.5	5.8
1,024	-2.7	2.6	3.0	4.0
2,048	-2.7	1.8	2.1	2.8
4,096	-2.7	1.3	1.5	2.0
8,192	-2.7	0.9	1.1	1.4
16,384	-2.7	0.7	0.8	1.1

Figure 14 ($3/day line): Bias and precision for bootstrapped estimates of groups' poverty rates at a point in time, by sample size, 2003 scorecard applied to non-panel households in 2004

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	1.5	50.1	59.7	78.5
4	1.1	38.9	46.3	56.5
8	0.2	27.1	33.2	45.5
16	0.7	19.3	23.5	30.1
32	0.4	13.6	16.3	21.9
64	0.7	9.6	11.2	14.5
128	0.3	7.1	8.0	11.0
256	0.3	4.9	5.8	7.4
512	0.3	3.4	4.1	5.6
1,024	0.3	2.5	3.0	3.9
2,048	0.3	1.6	2.1	2.7
4,096	0.3	1.2	1.4	1.8
8,192	0.3	0.8	1.0	1.4
16,384	0.3	0.6	0.8	0.9

Figure 16 ($3/day line): Bias and precision for bootstrapped estimates of changes in groups' poverty rates between two points in time, by sample size, 2003 scorecard applied to non-panel households in 2002 and 2003

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	3.0	70.7	87.8	105.4
4	1.7	53.3	64.9	85.1
8	2.4	38.7	46.0	62.3
16	2.2	28.0	33.9	44.1
32	2.3	19.8	25.4	36.3
64	2.5	15.2	18.6	25.1
128	2.8	10.2	12.8	16.4
256	2.8	7.6	9.0	11.5
512	2.8	5.3	6.6	8.2
1,024	2.8	3.7	4.4	5.8
2,048	2.8	2.6	2.9	3.8
4,096	2.8	1.8	2.1	2.9
8,192	2.7	1.3	1.5	1.9
16,384	2.7	0.9	1.1	1.4

Figure 17 ($3/day line): Bias and precision for bootstrapped estimates of changes in groups' poverty rates between two points in time, by sample size, 2003 scorecard applied to non-panel households in 2003 and 2004

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	0.4	71.7	86.2	107.1
4	1.4	53.8	67.0	86.3
8	0.3	40.8	49.4	66.2
16	1.2	29.1	35.8	49.4
32	0.6	20.0	24.1	31.5
64	0.7	13.8	17.2	22.3
128	0.2	10.1	12.0	14.8
256	0.1	7.0	8.2	10.8
512	0.2	5.1	6.0	7.6
1,024	0.2	3.5	4.4	5.7
2,048	0.2	2.5	3.0	4.0
4,096	0.2	1.8	2.1	2.8
8,192	0.3	1.2	1.4	1.9
16,384	0.3	0.9	1.1	1.4

Figure 18 ($3/day line): Bias and precision for bootstrapped estimates of changes in groups' poverty rates between two points in time, by sample size, 2003 scorecard applied to non-panel households in 2002 and 2004

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	3.4	70.2	85.3	106.3
4	3.1	54.8	64.7	77.6
8	2.7	39.8	48.0	63.8
16	3.5	27.4	32.6	42.8
32	2.9	19.1	24.4	33.0
64	3.2	13.7	16.7	22.0
128	3.0	10.1	11.9	15.5
256	3.0	7.4	8.7	12.1
512	3.0	5.1	6.0	7.6
1,024	3.0	3.5	4.2	5.6
2,048	3.0	2.5	2.9	3.8
4,096	3.0	1.7	2.1	2.8
8,192	3.0	1.2	1.5	2.0
16,384	3.0	0.9	1.1	1.5

Figure 20 ($3/day line): Bias and precision for bootstrapped estimates of changes in groups' poverty rates between two points in time, by sample size, 2003 scorecard applied to panel households in 2002 and 2003

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	0.8	65.0	83.9	106.7
4	1.7	49.6	59.3	80.8
8	1.6	34.2	43.4	58.1
16	0.3	23.7	29.3	38.2
32	0.2	16.5	20.5	27.7
64	0.0	12.1	14.6	19.5
128	0.0	8.3	10.1	13.2
256	0.1	6.0	7.2	9.8
512	0.1	4.4	5.2	6.8
1,024	-0.1	3.1	3.7	4.8
2,048	-0.1	2.2	2.7	3.5
4,096	0.0	1.6	1.9	2.5
8,192	-0.1	1.1	1.4	1.8
16,384	0.0	0.8	0.9	1.3

Figure 21 ($3/day line): Bias and precision for bootstrapped estimates of changes in groups' poverty rates between two points in time, by sample size, 2003 scorecard applied to panel households in 2003 and 2004

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	-1.0	48.9	64.7	88.4
4	-0.6	30.2	39.6	64.3
8	-0.8	21.5	27.2	40.7
16	-0.3	14.5	17.8	28.5
32	-0.2	10.8	12.8	18.6
64	-0.1	7.4	9.5	12.7
128	-0.1	5.4	6.4	8.1
256	0.0	4.0	4.7	6.2
512	0.0	2.7	3.3	4.1
1,024	0.0	1.9	2.3	3.0
2,048	0.0	1.4	1.6	2.2
4,096	-0.1	0.9	1.1	1.5
8,192	-0.1	0.7	0.8	1.0
16,384	-0.1	0.5	0.5	0.7

Figure 22 ($3/day line): Bias and precision for bootstrapped estimates of changes in groups' poverty rates between two points in time, by sample size, 2003 scorecard applied to panel households in 2002 and 2004

	Difference between estimate and true value			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	0.3	0.9	1.1	1.4
4	0.3	0.9	1.1	1.4
8	0.3	0.9	1.1	1.4
16	0.3	0.9	1.1	1.4
32	0.3	0.9	1.1	1.4
64	0.3	0.9	1.1	1.4
128	0.3	0.9	1.1	1.4
256	0.3	0.9	1.1	1.4
512	0.3	0.9	1.1	1.4
1,024	0.3	0.9	1.1	1.4
2,048	0.3	0.9	1.1	1.4
4,096	0.3	0.9	1.1	1.4
8,192	0.3	0.9	1.1	1.4
16,384	0.3	0.9	1.1	1.4

Figure 24 ($3/day line): People by targeting classification and score, along with "Total Accuracy" and BPAC, 2003 scorecard applied to non-panel households in 2002

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	6.3	45.6	0.1	48.0	54.3	-75.6
5–9	10.8	41.0	0.2	47.9	58.7	-57.8
10–14	16.7	35.1	0.9	47.3	64.0	-33.9
15–19	21.6	30.2	1.5	46.6	68.2	-13.7
20–24	27.9	23.9	2.6	45.6	73.5	12.7
25–29	34.4	17.5	4.2	44.0	78.4	40.7
30–34	38.8	13.0	6.4	41.8	80.6	61.9
35–39	42.1	9.7	8.9	39.3	81.4	79.6
40–44	45.7	6.1	12.1	36.1	81.8	76.7
45–49	47.8	4.1	15.3	32.8	80.6	70.4
50–54	49.2	2.7	18.7	29.4	78.6	63.9
55–59	50.0	1.8	22.9	25.2	75.3	55.8
60–64	50.6	1.2	27.2	20.9	71.5	47.5
65–69	51.2	0.6	31.3	16.8	68.0	39.6
70–74	51.4	0.4	35.0	13.2	64.6	32.5
75–79	51.6	0.3	37.5	10.7	62.3	27.7
80–84	51.7	0.1	41.2	7.0	58.7	20.6
85–89	51.8	0.1	45.3	2.9	54.7	12.7
90–94	51.8	0.0	45.7	2.5	54.3	11.9
95–100	51.8	0.0	48.2	0.0	51.8	7.1

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

Figure 25 ($3/day line): People by targeting classification and score, along with "Total Accuracy" and BPAC, 2003 scorecard applied to non-panel households in 2004

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	3.1	38.2	0.1	58.6	61.7	-84.6
5–9	5.8	35.5	0.3	58.3	64.2	-70.9
10–14	10.5	30.8	0.8	57.8	68.3	-47.3
15–19	12.8	28.5	1.2	57.5	70.4	-35.0
20–24	17.6	23.7	2.4	56.3	73.9	-9.1
25–29	23.1	18.2	4.4	54.3	77.4	22.4
30–34	26.8	14.6	6.3	52.4	79.1	44.7
35–39	32.3	9.0	10.5	48.1	80.5	74.5
40–44	35.6	5.7	14.5	44.2	79.8	65.0
45–49	37.5	3.8	18.7	40.0	77.5	54.7
50–54	39.0	2.3	23.2	35.5	74.5	43.8
55–59	39.9	1.4	28.2	30.5	70.4	31.8
60–64	40.4	0.9	32.9	25.7	66.1	20.2
65–69	40.8	0.5	38.1	20.6	61.4	7.8
70–74	41.0	0.3	42.4	16.3	57.3	-2.5
75–79	41.1	0.3	45.9	12.8	53.8	-11.1
80–84	41.2	0.2	49.9	8.7	49.9	-20.9
85–89	41.3	0.0	55.5	3.2	44.5	-34.3
90–94	41.3	0.0	56.0	2.7	44.0	-35.6
95–100	41.3	0.0	58.7	0.0	41.3	-42.0

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

Figure 26 ($3/day line): People below the poverty line and all people, at a given score or at or below a given score cut-off, 2003 scorecard applied to non-panel households in 2002

	People below poverty line (%)		All people (%)	
Score	At score	At or below score	At score	At or below score
0–4	98.1	98.1	6.4	6.4
5–9	97.3	97.7	4.7	11.1
10–14	90.4	95.0	6.5	17.6
15–19	88.4	93.4	5.6	23.1
20–24	85.4	91.5	7.4	30.5
25–29	80.3	89.1	8.1	38.6
30–34	67.0	85.9	6.6	45.2
35–39	56.6	82.5	5.8	51.0
40–44	53.1	79.1	6.8	57.8
45–49	38.8	75.7	5.3	63.1
50–54	29.3	72.4	4.8	67.9
55–59	17.3	68.6	5.1	73.0
60–64	11.5	65.0	4.9	77.9
65–69	12.7	62.0	4.7	82.5
70–74	6.0	59.5	3.9	86.4
75–79	6.2	57.9	2.7	89.1
80–84	3.6	55.7	3.8	92.9
85–89	1.4	53.4	4.2	97.1
90–94	6.7	53.2	0.4	97.5
95–100	1.1	51.8	2.5	100.0

Figure 27 ($3/day line): People below the poverty line and all people, at a given score or at or below a given score cut-off, 2003 scorecard applied to non-panel households in 2004

	People below poverty line (%)		All people (%)	
Score	At score	At or below score	At score	At or below score
0–4	96.1	96.1	3.2	3.2
5–9	92.6	94.5	2.9	6.2
10–14	90.4	92.6	5.1	11.3
15–19	87.7	91.7	2.7	14.0
20–24	79.4	88.0	6.0	20.0
25–29	73.5	84.0	7.5	27.5
30–34	65.8	81.0	5.6	33.0
35–39	56.8	75.4	9.8	42.9
40–44	45.6	71.1	7.2	50.1
45–49	30.6	66.7	6.1	56.2
50–54	25.3	62.7	6.0	62.3
55–59	15.1	58.6	5.8	68.1
60–64	9.0	55.1	5.2	73.3
65–69	7.2	51.7	5.5	78.9
70–74	4.1	49.2	4.5	83.3
75–79	2.5	47.2	3.6	87.0
80–84	2.3	45.2	4.1	91.1
85–89	2.8	42.7	5.7	96.8
90–94	0.0	42.4	0.5	97.3
95–100	0.0	41.3	2.7	100.0