Simple Poverty Scorecard®
Russia

Mark Schreiner and Dean Caire

17 March 2010

Этот документ доступен на русском языке на `SimplePovertyScorecard.com`.
This document is available at `SimplePovertyScorecard.com`.

Abstract

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

Acknowledgements

This paper was funded by Grameen Foundation (GF) with a grant from Ford Foundation. Data are from Russia's Federal State Statistical Institute. Thanks go to Nigel Biggar, Irina Denisova, Victor Sulla, David Tarr, and Jeff Toohig. The Simple Poverty Scorecard® is the same as what GF calls the Progress out of Poverty Index®. The PPI® is a performance-management tool that GF promotes so that institutions are able to achieve their social objectives more effectively.

Authors

Mark Schreiner directs Microfinance Risk Management, L.L.C. He is also a Senior Scholar at the Center for Social Development, Washington University in Saint Louis. Dean Caire is a scoring consultant, `dean_caire@hotmail.com`.

Simple Poverty Scorecard®

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

Indicator	Response	Points	Score
1. How many members does the household have?	A. Five or more	0	
	B. Four	5	
	C. Three	11	
	D. Two	22	
	E. One	35	
2. In their main line of work, are any household members administrators or heads/leaders of government, agencies, organizations, or state businesses, highly skilled specialists/professionals, skilled technicians, or white-collar employees?	A. No	0	
	B. Yes	5	
3. What is the total area of the residence in meters squared?	A. 24 or less	0	
	B. 25 to 39	3	
	C. 40 to 59	4	
	D. 60 to 99	6	
	E. 100 or more	8	
4. What is the source of hot water for the residence?	A. None	0	
	B. Individual water heater	4	
	C. Centralized	6	
5. How many color televisions does the household own?	A. None	0	
	B. One	3	
	C. Two or more	7	
6. How many VCRs and DVDs does the household own?	A. None	0	
	B. One	3	
	C. Two or more	7	
7. Does the household own a microwave?	A. No	0	
	B. Yes	2	
8. Does the household own a personal computer?	A. No	0	
	B. Yes	4	
9. How many land-line and cellular telephones does the household own?	A. None	0	
	B. One or more land-lines, and no cellular	6	
	C. No land-lines, and one cellular	10	
	D. One or more land-lines, and one cellular	13	
	E. No land-lines, and two or more cellular	15	
	F. One or more land-lines, and two cellular	16	
	G. One or more land-lines, and three or more cellular	21	
10. Does the household own an automobile?	A. No	0	
	B. Yes	5	

 Score:

Simple Poverty Scorecard®
Russia

1. Introduction

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

The direct approach to poverty measurement via surveys is difficult and costly. In the case of Russia's Household Budget Survey, each participating household keeps a daily record for a month of 300 categories of items bought or consumed as well as a log of non-food purchases for two more months (Ovtcharova and Tesliuc, 2006).

In contrast, the indirect approach via the scorecard is simple, quick, and inexpensive. It uses ten verifiable indicators (such as "What is the source of hot water for the residence?" and "How many color televisions does the household own?") to get a score that is highly correlated with poverty status as measured by consumption from the exhaustive consumption survey.

The Simple Poverty Scorecard® differs from "proxy means tests" (Coady, Grosh, and Hoddinott, 2002) in that it is tailored to the capabilities and purposes not of national governments but rather of local, pro-poor organizations. The feasible poverty-measurement options for local organizations are typically subjective and relative (such as participatory wealth ranking by skilled field workers) or blunt (such as rules based on land-ownership or housing quality). These approaches may be costly, their results are not comparable across organizations or across countries, and their accuracy and precision are unknown.

The scorecard here can be used by organizations who want to know what share of their participants are below a poverty line, perhaps because they want to relate their poverty status to the Millennium Development Goals' $1.25/day poverty line at 2005 purchase-power parity (PPP). It can also be used by USAID microenterprise partners who want to report how many of their participants are among the poorest half of people below the national poverty line. Or it can be used by organizations who want to measure movement across a poverty line (for example, Daley-Harris, 2009). The Simple Poverty Scorecard® is an consumption-based, objective tool with known accuracy that can serve for monitoring, management, and/or targeting. While consumption surveys are difficult and costly even for governments, a simple, inexpensive poverty-assessment tool can be feasible for many local, pro-poor organizations.

The statistical approach here aims to be understood by non-specialists. After all, if managers are to adopt the scorecard on their own and apply it to inform their decisions, they must first trust that it works. Transparency and simplicity build trust. Getting "buy-in" matters; proxy means tests and regressions on the "determinants of poverty" have been around for three decades, but they are rarely used to inform decisions by local pro-poor organizations. This is not because these tools do not work, but because they are presented (when they are presented at all) as tables of regression coefficients incomprehensible to non-specialists (with indicator names such as "`LGHHSZ_2`", negative points, and points with many decimal places). Thanks to the predictive-modeling phenomenon known as the "flat maximum", simple, transparent scorecards are about as accurate as complex, opaque ones.

The technical approach here is innovative in how it associates scores with poverty likelihoods, in the extent of its accuracy tests, and in how it derives formulas for standard errors. Although the accuracy tests are simple and commonplace in statistical practice and in the for-profit field of credit-risk scoring, they have rarely been applied to poverty-assessment tools.

The scorecard is based on data from Russia's Household Budget Survey (HBS) for the fourth quarter of 2007, conducted by the Federal State Statistical Institute. 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 zeroes or positive integers, and total scores range from 0 (most likely below a poverty line) to 100 (least likely below a poverty line). Non-specialists can collect data and tally scores on paper in the field in five to ten minutes.

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

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

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

The scorecard can also be used for targeting services to poorer households.[1] To help managers choose a targeting cut-off, this paper reports several measures of targeting accuracy for a range of possible cut-offs.

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

The scorecard is constructed and calibrated using a sub-sample from the 2007 HBS. Its accuracy is then validated on a different sub-sample from the 2007 HBS. While all three scoring estimators are unbiased when applied to the population from which they were derived (that is, they match the true value on average in repeated

[1] World Bank (2005) recommends that Russia use scoring to improve the targeting of social programs, and this is also the central recommendation of World Bank (2009).

samples from the same population from which the scorecard is built), they are—like all predictive models—biased to some extent when applied to a different population.[2]

Thus, while the indirect scoring approach is less costly than the direct survey approach, it is also biased. (The direct survey approach is unbiased by definition.) There is bias because scoring must assume that the relationships between indicators and poverty will be the same in the future as they are in the data used to build the scorecard. It must also assume that these relationships will be the same in all sub-groups as in the population as a whole. Of course, these assumptions—ubiquitous and inevitable in predictive modeling—hold only partly.

When applied to the validation sample for Russia with the national poverty line and n = 16,384, the average difference between scorecard estimates of groups' poverty rates and true rates at a point in time is +2.2 percentage points. Across all three lines, the average absolute difference is 1.3 percentage points.

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

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

For $n = 16,384$, the 90-percent confidence intervals for these estimates are +/–0.1 percentage points. For $n = 1,024$, these intervals are +/–0.5 percentage points.

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

2. Data and poverty lines

This section discusses the data used to construct and validate the Simple Poverty Scorecard®. It also documents the poverty lines to which scores are calibrated.

2.1 Data

The scorecard is based on data from the 53,042 households in Russia's Household Budget Survey (HBS) for the fourth quarter of 2007, conducted by the Federal State Statistical Institute. This is the most recent national consumption survey available for Russia.[3] Households are randomly divided into three sub-samples (Figure 2):

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

2.2 Poverty rates and poverty lines

2.2.1 Rates

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

[3] According to Lokshin (2008, p. 11), "The HBS is the only official, nationally and regionally representative source of data about living conditions and poverty in Russia."

Beyond this general definition, there two special cases, *household-level poverty rates* and *person-level poverty rates*. With household-level rates, each household is counted as if it had only one person, regardless of true household size, so all households are counted equally. With person-level rates (the "head-count index"), each household is weighted by the number of people in it, so larger households have greater weight.

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

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

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

The Simple Poverty Scorecard® is constructed using Russia's 2007 HBS and household-level lines, scores are calibrated to household-level poverty likelihoods, and accuracy is measured for household-level rates. This use of household-level rates reflects the belief that they are the most relevant for most pro-poor organizations.

Organizations can estimate person-level poverty rates by taking a household-size-weighted average of household-level poverty likelihoods. It is also possible to use person-level weights to construct, calibrate, and validate a scorecard, but it is not done here.

2.2.2 Poverty lines

Figure 2 reports poverty lines and household-level poverty rates based on Russia's 2007 HBS. Figure 3 reports the same information—plus person-level poverty rates—by region.

Russia's national line (also known as the "minimum subsistence level") is based on the cost—derived from HBS data—of a nutritionist-defined food basket that varies by region (World Bank, 2005). It also includes the cost of a non-food bundle of goods and services, again normatively defined by regional government agencies and then adjusted at the federal level. The national line ignores economies of scale in household size (that is, the fact that a two-person household requires less than twice the consumption of a one-person household). Overall, the national line diverges from international standards,[4] but it is all that is available for the 2007 HBS. The average

[4] See also Gibson and Poduzov (2003).

national line is RUB132 per person per day, giving a household-level poverty rate of 9.7 percent and a person-level poverty rate of 12.2 percent (Figures 2 and 3).[5]

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

- National
- USAID "extreme"
- $6.25/day 2005 PPP

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

The $6.25/day 2005 PPP line is derived from:

- 2005 PPP exchange rate for "individual consumption expenditure by households" (World Bank, 2008): RUB13.39 per $1.00
- Price deflators for Russia overall: 121.97 for 2007 on average, and 100 for 2005 on average[6]

[5] The officially reported rates are somewhat different because they are based on income rather than consumption and because they are adjusted to reconcile HBS figures with national accounts (World Bank, 2009 and 2005).
[6] Derived from `http://www.gks.ru/bgd/regl/b09_12/IssWWW.exe/stg/d02/25-01.htm`, retrieved 19 February 2010.

Using the formula in Sillers (2006), the \$6.25/day 2005 PPP line for Russia as a whole in 2007 is:

$$(\text{2005 PPP exchange rate}) \cdot \$6.25 \cdot \frac{\text{CPI}_{\text{Ave. 2007}}}{\text{CPI}_{\text{Ave. 2005}}} =$$

$$\left(\frac{\text{RUB13.39}}{\$1.00}\right) \cdot \$6.25 \cdot \frac{121.97}{100} = \text{RUB102.}$$

This line is not adjusted for regional differences in cost-of-living.

3. Scorecard construction

For the Russia scorecard, about 80 potential indicators are initially prepared in the areas of:

- Family composition (such as household size)
- Education (such as school attendance by children)
- Employment (such as the number of household members with white-collar jobs)
- Housing (such as the source of hot water for the residence)
- Ownership of durable goods (such as televisions and microwaves)

Figure 4 lists all the candidate indicators, ranked by the entropy-based "uncertainty coefficient" that is a measure of how well an indicator predicts poverty on its own (Goodman and Kruskal, 1979). For a given indicator, responses are ordered starting with those associated with higher poverty likelihoods.

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

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

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

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

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

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

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

4. Practical guidelines for scorecard use

The main challenge of scorecard design is not to squeeze out the last drops of accuracy but rather to improve the chances that scoring is actually used (Schreiner, 2005). When scoring projects fail, the reason is not usually technical inaccuracy but rather the failure of an organization to decide to do what is needed to integrate scoring in its processes and to learn to use it properly (Schreiner, 2002). After all, most reasonable scorecards predict tolerably well, thanks to the empirical phenomenon known as the "flat maximum" (Falkenstein, 2008; Hand, 2006; Baesens *et al.*, 2003; Lovie and Lovie, 1986; Kolesar and Showers, 1985; Stillwell, Barron, and Edwards, 1983; Dawes, 1979; Wainer, 1976; Myers and Forgy, 1963). The bottleneck is less technical and more human, not statistics but organizational-change management. Accuracy is easier to achieve than adoption.

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

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

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

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

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

4.1 Quality control

Of course, field workers must be trained. High-quality outputs require high-quality inputs. If organizations or field workers gather their own data and if they believe that they have an incentive to exaggerate poverty rates (for example, if funders reward them for higher poverty rates), then it is wise to do on-going quality control via data review and random audits (Matul and Kline, 2003).[7] IRIS Center (2007a) and Toohig (2008) are useful nuts-and-bolts guides for planning, budgeting, training field

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

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

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

For a Mexican social program that uses self-reported indicators in the first stage of targeting with a poverty-assessment tool, Martinelli and Parker (2007) find that "underreporting [of asset ownership] is widespread but not overwhelming, except for a few goods . . . [and] overreporting is common for a few goods, which implies that self-reporting may lead to the exclusion of deserving households" (pp. 24–25). Still, as done in the second stage of the Mexican program, field agents using the scorecard can verify responses with a home visit and correct any false reports.

4.2 Implementation and sampling

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

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

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

- Employees of the organization
- Third-party contractors

Responses, scores, and poverty likelihoods can be recorded:

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

The subjects to be scored can be:

- All participants
- A representative sample of all participants
- All participants in a representative sample of branches
- A representative sample of all participants in a representative sample of branches
- A representative sample of a sub-group that is relevant for a particular question

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

Frequency of application can be:

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

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

- Different sets of participants, with each set representative of a given group
- A single set of participants

An example bundle of implementation and design choices is provided by BRAC and ASA, two microlenders in Bangladesh (each with more than 7 million participants) who are applying the Simple Poverty Scorecard® (Schreiner, 2013). Their design is that loan officers in a random sample of branches score all their clients each time they visit a homestead (about once a year) as part of their standard due diligence prior to loan disbursement. Responses in the field are recorded on paper before being sent to a central office to be entered into a spreadsheet database. The sampling plans of ASA and BRAC cover 50,000–100,000 participants each, which is far more than would be required to inform most decisions at a typical pro-poor organization.

5. Estimates of household poverty likelihoods

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

To get absolute units, scores must be converted to *poverty likelihoods*, that is, probabilities of being below a poverty line. This is done via simple look-up tables. For the example of the national line with the 2007 HBS, scores of 30–34 correspond to a poverty likelihood of 35.1 percent, and scores of 35–39 correspond to a poverty likelihood of 23.5 percent (Figure 5).

The poverty likelihood associated with a score varies by poverty line. For example, scores of 30–34 are associated with a poverty likelihood of 35.1 percent for the national line but 16.2 percent for the USAID "extreme" line.[8]

5.1 Calibrating scores with poverty likelihoods

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

[8] Starting with Figure 5, some figures have one version for each of the three poverty lines. Single tables that pertain to all lines are placed with tables for the national line.

For the example of the national line (Figure 6), there are 4,187 (normalized) households in the calibration sub-sample with a score of 30–34, of whom 1,468 (normalized) are below the poverty line. The estimated poverty likelihood associated with a score of 30–34 is then 35.1 percent, as 1,468 ÷ 4,187 = 0.351.

As another illustration, consider the national line and a score of 35–39. Now there are 7,480 (normalized) households in the calibration sample, of whom 1,760 (normalized) are below the line (Figure 6). Thus, the poverty likelihood for this score is 1,760 ÷ 7,480 = 0.235, or 23.5 percent.

The same method is used to calibrate scores with estimated poverty likelihoods for all three poverty lines.

Figure 7 shows, for all scores, the likelihood that consumption falls in a range demarcated by two adjacent poverty lines. For example, the daily consumption of someone with a score of 30–34 falls in the following ranges with probability:

- 16.2 percent less than the USAID "extreme" line
- 18.9 percent between the USAID "extreme" line and the national line
- 65.0 percent more than the national line

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

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

5.2 Accuracy of estimates of households' poverty likelihoods

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

The scorecard also produces unbiased estimates of poverty rates at a point in time, as well as unbiased estimates of changes in poverty rates between two points in time.[9]

But the relationships between indicators and poverty do change with time, and they also change across sub-groups in Russia's population. Thus, the scorecard will generally be biased when applied after the end date of fieldwork for the 2007 HBS (as it must be applied in practice) or when applied with non-nationally representative groups (as it probably will be applied by local, pro-poor organizations).

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

- Score each household in the validation sample
- Draw a new bootstrap sample *with replacement* from the validation sample
- For each score, compute the true poverty likelihood in the bootstrap sample, that is, the share of households with the score who have consumption below a poverty line
- For each score, record the difference between the estimated poverty likelihood (Figure 5) and the true poverty likelihood in the bootstrap sample
- Repeat the previous three steps 1,000 times
- For each score, report the average difference between estimated and true poverty likelihoods across the 1,000 bootstrap samples
- For each score, report the two-sided interval containing the central 900, 950, or 990 differences between estimated and true poverty likelihoods

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

For each score range and for $n = 16{,}384$, Figure 8 shows the average difference between estimated and true poverty likelihoods as well as confidence intervals for the differences.

For the national line in the validation sample, the average poverty likelihood across bootstrap samples for scores of 30–34 is too high by 8.1 percentage points. For scores of 35–39, the estimate is too high by 16.3 percentage points.[10]

The 90-percent confidence interval for the differences for scores of 30–34 is +/–3.7 percentage points (Figure 8). This means that in 900 of 1,000 bootstraps, the difference between the estimate and the true value is between +4.4 and +11.8 percentage points (because +8.1 – 3.7 = +4.4, and +8.1 + 3.7 = +11.8). In 950 of 1,000 bootstraps (95 percent), the difference is +8.1 +/–4.4 percentage points, and in 990 of 1,000 bootstraps (99 percent), the difference is +8.1 +/–5.9 percentage points.

For many scores, Figure 8 shows differences—sometimes large—between estimated poverty likelihoods and true values. The differences are not all zero because the validation sub-sample is a single sample that—thanks to sampling variation—differs in distribution from the construction/calibration sub-samples and from Russia's population. Also, some score ranges have few households in them, increasing the importance of sampling variation.

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

For targeting, what matters is less the differences across all score ranges and more the differences in score ranges just above and just below the targeting cut-off. This mitigates the effects of bias and sampling variation on targeting (Friedman, 1997). Section 8 below looks at targeting accuracy in detail.

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

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

Overfitting can be mitigated by simplifying the scorecard and by not relying only on data but rather also considering experience, judgment, and theory. Of course, the scorecard here does this. Bootstrapping scorecard construction—which is not done here—can also mitigate overfitting by reducing (but not eliminating) dependence on a single sampling instance. Combining scorecards can also help, at the cost of complexity. Simplifying the scorecard can also reduce overfitting (at the cost of decreased precision), although the scorecard here has limited scope for additional simplification. A good option is simply to update the scorecard as soon as new data becomes available.

In any case, errors in individual households' poverty likelihoods generally balance out in the estimates of groups' poverty rates (see the next section). Furthermore, much of the differences between scorecard estimates and true values may come from non-scorecard sources. These factors can be addressed only by improving data quantity and quality, which is beyond the scope of the scorecard.

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

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

To illustrate, suppose a program samples three households on 1 January 2010 and that they have scores of 20, 30, and 40, corresponding to poverty likelihoods of 59.2, 35.1, and 15.0 percent (national line, Figure 5). The group's estimated poverty rate is the households' average poverty likelihood of (59.2 + 35.1 + 15.0) ÷ 3 = 36.4 percent.[11]

6.1 Accuracy of estimated poverty rates at a point in time

How accurate is this estimate? For a range of sample sizes, Figure 10 reports average differences between estimated and true poverty rates as well as precision (confidence intervals for the differences) for the scorecard applied to 1,000 bootstrap samples from the validation sample.

Summarizing Figure 10 across poverty lines and years for n = 16,384, Figure 9 shows that the absolute differences between estimated poverty rates and true rates for the scorecard applied to the validation sample are 2.2 percentage points or less. The

[11] The group's poverty rate is *not* the poverty likelihood associated with the average score. Here, the average score is (20 + 30 + 40) ÷ 3 = 30, and the poverty likelihood associated with the average score is 35.1 percent. This is not the 36.4 percent found as the average of the three poverty likelihoods associated with each of the three scores.

average absolute difference across the three poverty lines for the validation sample is 1.3 percentage points.

In terms of precision, the 90-percent confidence interval for a group's estimated poverty rate at a point in time with n = 16,384 is +/–0.1 percentage points (Figure 9). This means that in 900 of 1,000 bootstraps of this size, the absolute difference between the estimate and the average estimate is 0.1 percentage points.

In the specific case of the national line and the validation sample, 90 percent of all samples of n = 16,384 produce estimates that differ from the true value in the range of +2.2 – 0.1 = +2.1 to +2.2 + 0.1 = +2.3 percentage points. This is because +2.2 is the average difference and +/–0.1 is its 90-percent confidence interval. The average difference is +2.2 because the average scorecard estimate is too high by 2.2 percentage points; the scorecard tends to estimate a poverty rate of 11.6 percent for the validation sample, but the true value is 9.4 percent (Figure 2).

Part of these differences is due to sampling variation in the division of the 2007 HBS into three sub-samples. Of course, estimates of poverty rates at a point in time from now on will be most accurate for periods that resemble the fourth quarter of 2007, the period of fieldwork for the 2007 HBS.

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

How precise are the point-in-time estimates? Because they are averages, the estimates have a Normal distribution and can be characterized by their average difference vis-à-vis true values, along with the standard error of the average difference.

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

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

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

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

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

n is the sample size.

For example, with a sample n = 16,384, 90-percent confidence (z = 1.64), and a poverty rate p of 9.4 percent (the true rate in the validation sample for the national line in Figure 2), the confidence interval c is

$$+/- z \cdot \sqrt{\frac{p \cdot (1-p)}{n}} = +/- 1.64 \cdot \sqrt{\frac{0.094 \cdot (1-0.094)}{16{,}384}} = +/-0.374 \text{ percentage points.}$$

The scorecard, however, does not measure poverty directly, so this formula is not applicable. To derive a formula for the Russia scorecard, consider Figure 10, which reports empirical confidence intervals c for the differences for the scorecard applied to 1,000 bootstrap samples of various sample sizes from the validation sample. For n = 16,384, the national line, and the validation sub-sample, the 90-percent confidence interval is +/–0.120 percentage points.[12] Thus, the ratio of confidence intervals for the scorecard versus direct measurement is 0.120 ÷ 0.374 = 0.32.

Now consider the same case, but with n = 8,192. The confidence interval under direct measurement is $+/- 1.64 \cdot \sqrt{\frac{0.094 \cdot (1-0.094)}{8{,}192}} = +/-0.529$ percentage points. The empirical confidence interval with the Russia scorecard for the national line (Figure 10) is +/–0.160 percentage points. Thus for n = 8,192, the ratio for the scorecard to direct measurement is 0.160 ÷ 0.529 = 0.30.

This ratio of 0.30 for n = 8,192 is close to the ratio of 0.32 for n = 16,384. Indeed, across all sample sizes of 256 or more in Figure 10, the average ratio turns out to be 0.31, implying that confidence intervals for indirect estimates of poverty rates via

[12] Due to rounding, Figure 10 displays 0.1, not 0.120.

the Russia scorecard and this poverty line are about one-third as large as for direct estimates. This 0.31 appears in Figure 9 as the "α factor" because if $\alpha = 0.31$, then the formula relating confidence intervals c and standard errors σ for the Russia scorecard is $c = +/- z \cdot \alpha \cdot \sigma$. The standard error σ for point-in-time estimates of poverty rates via scoring is $\alpha \cdot \sqrt{\frac{p \cdot (1-p)}{n}}$.

In general, α could be more or less than 1.00. When α is less than 1.00, it means that the scorecard is more precise than direct measurement. This occurs for all three poverty lines for the validation sample in Figure 9.

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

To illustrate how to use this, suppose $c = 0.00955$ and $z = 1.64$ (90-percent confidence), and $\hat{p} = 0.0985$ (the average poverty rate for the national line in the construction and calibration sub-samples, Figure 2). Then the formula gives

[13] IRIS Center (2007a and 2007b) says that a sample size of $n = 300$ is sufficient for reporting estimated poverty rates to USAID. If a poverty-assessment tool is as precise as direct measurement, if the expected (before measurement) poverty rate is 50 percent, and if the confidence level is 90 percent, then $n = 300$ implies a confidence interval of +/–2.2 percentage points. In fact, USAID has not specified confidence levels nor 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.

$$n = \left(\frac{0.31 \cdot 1.64}{0.00955}\right)^2 \cdot 0.0985 \cdot (1 - 0.0985) = 252$$, which is almost the same as the sample size of 256 observed for these parameters in Figure 10.

Of course, the α factors in Figure 9 are specific to Russia, its poverty lines, its poverty rates, and this scorecard. The method for deriving the formulas, however, is valid for any poverty-assessment tool following the approach in this paper.

In practice after the end of the HBS field work in the fourth quarter of 2007, an organization would select a poverty line (say, the national line), select a desired confidence level (say, 90 percent, or $z = 1.64$), select a desired confidence interval (say, +/–2.0 percentage points, or $c = 0.02$), make an assumption about $\hat{p}$ (perhaps based on a previous measurement such as the 9.7 percent overall average for the national line in Figure 2), look up α (here, 0.31, Figure 9), assume that the scorecard will work the same in the future and/or for non-nationally representative sub-groups,[14] and then compute the required sample size. In this illustration,

$$n = \left(\frac{0.31 \cdot 1.64}{0.02}\right)^2 \cdot 0.097 \cdot (1 - 0.097) = 57.$$

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

7. Estimates of changes in group poverty rates over time

The change in a group's poverty rate between two points in time is estimated as the change in the average poverty likelihood of the households in the group. With data for 2007 only, this paper cannot estimate changes over time, nor can it present sample-size formula. Nevertheless, the relevant concepts are presented here because, in practice, pro-poor organizations can apply the scorecard to measure change over time.

7.1 Warning: Change is not impact

Scoring can estimate change. Of course, change could be for the better or for the worse, and scoring does not indicate what caused change. This point is often forgotten, confused, or ignored, so it bears repeating: the scorecard simply estimates change, and it does not, in and of itself, indicate the reason for the change. In particular, estimating the impact of program participation on poverty status requires knowing what would have happened to participants if they had not been participants. Knowing this requires either strong assumptions or a control group that resembles participants in all ways except participation. To belabor the point, the scorecard can help estimate program impact only if there is some way to know what would have happened in the absence of the program. And that information must come from somewhere beyond the scorecard.

7.2 Calculating estimated changes in poverty rates over time

Consider the illustration begun in the previous section. On 1 January 2010, a program samples three households who score 20, 30, and 40 and so have poverty likelihoods of 59.2, 35.1, and 15.0 percent (national line, Figure 5). The group's baseline estimated poverty rate is the households' average poverty likelihood of (59.2 + 35.1 + 15.0) ÷ 3 = 36.4 percent.

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

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

By way of illustration, suppose that a year later on 1 January 2011, the program samples three additional households who are in the same cohort as the three households originally sampled (or suppose that the program scores the same three original households a second time) and finds that their scores are now 25, 35, and 45 (poverty likelihoods of 41.5, 23.5, and 8.8 percent, national line, Figure 5). Their average poverty likelihood at follow-up is (41.5 + 23.5 + 8.8) ÷ 3 = 24.6 percent, an improvement of 36.4 – 24.6 = 11.8 percentage points.[15]

This suggests that about one of nine participants crossed the poverty line in 2010. (This is a net figure; some people start above the line and end below it, and vice versa.) Among those who started below the line, about one in three (11.8 ÷ 36.4 = 32.4

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

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

7.3 Accuracy for estimated change in two independent samples

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

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

z, c, and p are defined as before, n is the sample size at both baseline and follow-up,[16] and α is the average (across a range of bootstrapped sample sizes) of the ratio of the observed confidence intervals from the scorecard and the theoretical confidence intervals from the textbook formula for direct measurement for two equal-sized independent samples.

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

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

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

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

To illustrate the use of the formula above to determine sample size for estimating changes in poverty rates across two independent samples, suppose the desired confidence level is 90 percent ($z = 1.64$), the desired confidence interval is 2 percentage points ($c = 0.02$), the poverty line is the national line, $\alpha = 1.11$, and $\hat{p} = 0.097$ (from Figure 2). Then the baseline sample size is $n = 2 \cdot \left(\frac{1.11 \cdot 1.64}{0.02}\right)^2 \cdot 0.097 \cdot (1 - 0.097) =$ 1,452, and the follow-up sample size is also 1,452.

7.4 Accuracy for estimated change for one sample, scored twice

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

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

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

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

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

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

$\hat{p}_*$ could be anything between 0 and 0.5, so more information is needed before applying this formula. Suppose that the observed relationship between $\hat{p}_*$, the number of years y between baseline and follow-up, and $p_{\text{baseline}} \cdot (1 - p_{\text{baseline}})$ is—as in Peru (Schreiner, 2009a)—close to:

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

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

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

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

To illustrate the use of this formula, suppose the desired confidence level is 90 percent ($z = 1.64$), the desired confidence interval is 2.0 percentage points ($c = 0.02$), the poverty line is the national line, and the sample will be scored first in 2010 and then again in 2013 ($y = 3$). The before-baseline poverty rate is 9.7 percent ($p_{2007} = 0.097$, Figure 2), and suppose $\alpha = 1.30$. Then the baseline sample size is

$n = 2 \cdot \left(\frac{1.3 \cdot 1.64}{0.02}\right)^2 \cdot \{-0.02 + 0.016 \cdot 3 + 0.47 \cdot [0.097 \cdot (1 - 0.097)]\} = 1{,}572$. The same group of 1,572 households is scored at follow-up as well.

8. Targeting

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

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

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

Figure 11 depicts these four possible targeting outcomes. Targeting accuracy varies by cut-off; a higher cut-off has better inclusion (but greater leakage), while a lower cut-off has better exclusion (but higher undercoverage).

A program should weigh these trade-offs when setting a cut-off. A formal way to do this is to assign net benefits—based on a program's values and mission—to each of the four possible targeting outcomes and then to choose the cut-off that maximizes total net benefits (Adams and Hand, 2000; Hoadley and Oliver, 1998).

Figure 12 shows the distribution of households by targeting outcome. For an example cut-off of 34 or less and the scorecard applied to the validation sample, outcomes for the national line are:

- Inclusion: 4.0 percent are below the line and correctly targeted
- Undercoverage: 5.4 percent are below the line and mistakenly not targeted
- Leakage: 5.1 percent are above the line and mistakenly targeted
- Exclusion: 85.5 percent are above the line and correctly not targeted

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

- Inclusion: 5.5 percent are below the line and correctly targeted
- Undercoverage: 3.9 percent are below the line and mistakenly not targeted
- Leakage: 11.1 percent are above the line and mistakenly targeted
- Exclusion: 79.5 percent are above the line and correctly not targeted

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

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

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

- Assign benefits and costs to possible outcomes, based on its values and mission
- Tally total net benefits for each cut-off using Figure 12 for a given poverty line
- Select the cut-off with the highest total net benefit

The most difficult step is assigning benefits and costs to targeting outcomes. Any program that uses targeting—with or without scoring—should thoughtfully consider how it values successful inclusion and exclusion versus errors of undercoverage and leakage. It is healthy to go through a process of thinking explicitly and intentionally about how possible targeting outcomes are valued.

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

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

Figure 12 shows "Total Accuracy" for all cut-offs for Russia's scorecard. For the national line in the validation sample, total net benefit is greatest (91.2) for a cut-off of 24 or less, with about 10 in 11 households in Russia correctly classified.

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

[18] Figure 12 also reports "BPAC", the Balanced Poverty Accuracy Criteria adopted by USAID as its criterion for certifying poverty-assessment tools. IRIS Center (2005) says that BPAC considers accuracy both in terms of the estimated poverty rate and in terms of targeting inclusion. After normalizing by the number of people below a poverty line, BPAC = (Inclusion – |Undercoverage – Leakage|) x [100 ÷ (Inclusion+Undercoverage)].

As an alternative to assigning benefits and costs to targeting outcomes and then choosing a cut-off to maximize total net benefits, a program could set a cut-off to achieve a desired poverty rate among targeted households. The third column of Figure 13 ("% targeted who are poor") shows the expected poverty rate among Russia households who score at or below a given cut-off. For the example of the national line and the validation sample, targeting households who score 34 or less would target 9.1 percent of all households (second column) and produce a poverty rate among those targeted of 44.1 percent (third column).

Figure 13 also reports two other measures of targeting accuracy. The first is a version of inclusion ("% of poor who are targeted"). For the example of the national line and the validation sample with a cut-off of 34 or less, 42.9 percent of all poor households are covered.

The final targeting measure in Figure 13 is the number of successfully targeted poor households for each non-poor household mistakenly targeted (right-most column). For the national line, the validation sample, and a cut-off of 34 or less, covering 0.8 poor households means leaking to one non-poor household.

9. Context of poverty-assessment tools for Russia

This section discusses three existing poverty-assessment tools for Russia in terms of their goals, methods, poverty lines, indicators, cost, accuracy, and precision. The advantages of the new scorecard here are its use of the latest nationally representative data, its focus on feasibility for local, pro-poor organizations, its testing of accuracy and precision out-of-sample, and its reporting of formulas for standard errors.

9.1 Struyk and Kolodeznikova

Struyk and Kolodeznikova (SK, 1999) test whether a poverty-assessment tool can improve the targeting of Russia's Housing Allowance Program (HAP). They construct three tools with least-squares regression on the logarithm of household income, using data gathered by the Urban Institute that are representative of housing units in Moscow (1996, n = 2,239), Vladimir (1995, n = 626), and Gorodetz (1995, n = 376). At the time, HAP targeted based on self-reported income—a non-verifiable indicator subject to severe understatement—and the World Bank was recommending poverty-assessment tools as an alternative way to target.

The "poverty line" in SK corresponds to an expected amount of income devoted to housing. This comes from a HAP formula that depends on the cost of a housing unit determined to be appropriate for a given household's size, self-reported household income, and the share of income that HAP expects families to devote to housing. If the expected cost of housing exceeds the amount of income expected to be designated for housing, then the family qualifies for a subsidy to close the gap. In 1996, 7 percent of Russian households received the subsidy.

The tools in SK use 10 indicators, all verifiable and inexpensive to collect:

- Characteristics of up to three individual income earners:
  - — Sex
  - — Age (and age squared)
  - — Marital status
  - — Sector of work
  - — Type of occupation
  - — Education
- Number of household members with income from a given source:
  - — Pensions
  - — Other social benefits
  - — Stipends
  - — Alimony

SK deliberately exclude indicators of asset ownership. They say that for HAP, the targeting indicator should be sensitive to short-term swings in income. They argue that asset ownership is unlikely to respond quickly to changes in income and may be weakly linked with current income due to the legacy of broad distribution of asset ownership in the Soviet era combined with uneven household fortunes with income in transition. Likewise, SK say that they want to predict income—rather than consumption—because of its greater short-term volatility.

SK test their tools *in-sample*, that is, they check accuracy with the same data that is used to construct the tools in the first place. In-sample tests overstate accuracy. In contrast, this paper reports only *out-of-sample* tests with data that is not used to construct the scorecard. Houssou *et al.* (2007), Johanssen (2006), and Copestake *et al.* (2005) find that accuracy for poverty-assessment tools can deteriorate 8 to 17 percent going from in-sample to out-of-sample. Out-of-sample is also more relevant because, in practice, poverty-assessment tools are applied to data on households that are not used to construct the tool.

Even though in-sample tests are too sanguine, SK find what they call "substantial errors of undercoverage" on a "quite large scale" that is "disturbing", "startling", and "sobering". They find that up to about one-third of households that should qualify—given their survey measure of income—would not qualify based on their poverty-assessment tool. Their central conclusion is (pp. 1885–6):

> Very great caution should be used in proposing [poverty-assessment tools] of the type analyzed here in the countries of the Soviet bloc *for any needs-tested program.* [italics original] . . . The welfare losses from undercoverage associated with proxy tests and absent in the income-testing approach should simply be unacceptable to policymakers.

If correct, this critique would cripple the scorecard. Of course, SK are entitled to their judgment. But do their arguments in support of their judgment hold up?

First, it seems a stretch to conclude—based on one test for one program in one country—that the scorecard should not be used for targeting any social program anywhere in the former Soviet bloc.

Second, SK omit many powerful indicators; a better tool would do better.

Third, SK investigate scoring because income is severely underreported, yet they take the measure of income from their survey as being correct. More likely, it is less underreported than simple self-reports, but still underreported to some unknown extent.

While SK compare income in their Moscow survey to that in other rounds of the same survey and find it to be similar, this does not address the issue of underreporting.

Fourth, while SK are entitled to argue from their value-based standard that undercoverage should be avoided at all costs, this standard is not realistic. After all, nothing is worth doing at all costs. Furthermore, it is at least possible that savings from reduced leakage in HAP could increase subsidies to included participants enough to compensate for undercoverage. In any case, if SK took their own standard of "no undercoverage" seriously, then they would have to recommend universal coverage.

Fifth, SK implicitly assume that the HAP does not consider the fact that income is systematically underreported with it sets its rules for housing standards, expected housing costs, and expected income devoted to housing. If HAP adjusts its rules for underreporting, then there may be undercoverage even with income-based targeting.

Sixth, SK implicitly assume that HAP's income-based system has no undercoverage. That seems unlikely; in any case, undercoverage under the *status quo* is unknown. While SK measure scoring's accuracy (given their assumptions), they have no measure of HAP's current accuracy.[19] Thus, SK's critique of scoring depends on the assumption that the current system has less undercoverage than scoring.

[19] Braithwaite and Ivanova (1999, p. 10) state: "Preliminary evidence from the housing allowance subsidy program in Ukraine and Russia, which are based on official income (wages plus transfer income), suggest that this official income test has a very high error of [undercoverage] (those who are actually poor are not receiving the benefit)."

9.2 Grootaert and Braithwaite

Like SK, Grootaert and Braithwaite ("GB", 1998) seek test whether poverty-assessment tools can improve the targeting of Russia's social transfers. To this end, they use the Household Expenditure and Income for Transition Economies Data Set[20] to construct all-Russia poverty-assessment tools using forward stepwise least-squares regression on the logarithm of household consumption per adult equivalent. The poverty line is defined as two-thirds of average household consumption per adult equivalent, giving a person-level poverty rate of 39.4 percent.

GB build four all-Russia tools, two with five indicators and two with ten. All the indicators are verifiable and inexpensive to collect:

- Demographics: Number of elderly household members
- Presence of sources of income:
    - Employment
    - Transfers
- Education of head
- Employment:
    - Presence of a household enterprise
    - Presence of an inactive head
- Asset ownership:
    - Color television
    - Refrigerator
    - Sewing machine
    - Car
    - Residence
- Location: Urban/other-urban/rural

[20] GB do not report the name of the original survey in Russia nor its year.

Using in-sample tests and a five-indicator tool, GB report person-level inclusion of 22.2 percent and exclusion of 46.2 percent.[21] They also stumble onto the flat maximum, finding that 10 indicators target only slightly better than five.

GB also test a two-step procedure. In the first step, a tool is applied, and households who rank above the median are not targeted and set aside. In the second step, a second tool is applied to the remaining households, and those with estimated consumption below the poverty line are targeted. This approach is common in credit-risk scoring (Hand and Vinciotti, 2002; Shapire, 2001; Myers and Forgy, 1963). IRIS Center (2005) has since adopted the two-step approach as its preferred method for poverty-assessment tools for most countries, although the approach's greater complexity generally offers only small improvements in targeting accuracy.

According to GB, "Overall, the results are impressive . . . This is very respectable performance, and suggests that such an approach is worth considering for real-life application" (pp. 86, 88). Unfortunately, such application would be difficult, as GB do not report tool points.

[21] Targeting accuracy in GB cannot be compared to the new scorecard here because GB report person-level figures and because the poverty rate in GB (39.4 percent) is far from the highest poverty rate here.

Is the scorecard in Russia "impressive" (as GB say) or "sobering" (as SK say)? SK base their judgment on a comparison with the ideal of perfect targeting. In contrast, GB base their judgment on a comparison of actual targeting, which they note is usually regressive and has more beneficiaries who are poor than non-poor (pp. 92–93).

9.3 Christiaensen *et al.*

Christiaensen *et al.* (2008) use the poverty-mapping approach (Elbers, Lanjouw, and Lanjouow, 2003) to construct a poverty-assessment tool for Russia based on the 1994 Russia Longitudinal Monitoring Survey (RLMS). Their goal is to check the stability of the relationships between indicators and poverty over time, a prerequisite for using poverty-assessment tools to track changes in poverty rates (that is, the third use of the scorecard discussed in this paper).[22] To this end, Christiaensen *et al.* apply their tools from the 1994 RLMS to the 2003 RLMS, comparing estimates to true values out-of-sample.[23]

Using consumption, Christiaensen *et al.*'s estimate for the 2004 person-level poverty rate is 3.5 percentage points too low. Using income, their estimate is 4.3 percentage points too low. Given the upheaval in Russia and the wide swings in poverty

---

[22] Although Christiaensen *et al.* (p. 5) say that their paper is the "first contribution" in this regard, it is preceded by Schreiner (2008a and 2008b).

[23] They also apply the tool to the 1998 RLMS, collected during that year's crisis. Not surprisingly, the relationships between indicators and poverty status in 1998 did not resemble those in 1994 or 2003, so estimated poverty rates were not very accurate.

in this nine-year period, this accuracy is remarkable. It cannot, however, be compared with the scorecard here, as change cannot be estimated with only the 2007 HBS.[24]

While Christiaensen *et al.* report the standard error of the 2004 estimates, they do not report sample sizes, standard errors of estimated changes, nor standard-error formula. Thus, precision cannot be compared with that of the poverty-scoring approach in the other countries that have estimates of change.

Christiaensen *et al.* report whether their 2004 estimate is statistically different (with 95-percent confidence) from the true value, finding that it usually is not different. They do not report tool points or indicators, although they indicate that they are available on request.

[24] Rosstat's web site omits the 2006 HBS file on the residence and asset ownership. The 2005 HBS data is complete, but we lack national poverty lines for Q4 2005.

10. Conclusion

Pro-poor programs in Russia can use the Simple Poverty Scorecard® to segment clients for differentiated treatment as well as to estimate:

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

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

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

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

When the scorecard is applied to the validation sample with $n = 16{,}384$, the absolute difference between estimates and true poverty rates at a point in time is 2.2 percentage points or less and averages (across the three poverty lines) 1.3 percentage points. With 90-percent confidence, the precision of these differences for all lines is +/– 0.1 percentage points.

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

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

In sum, the Simple Poverty Scorecard® is a practical, objective way for pro-poor programs in Russia to monitor poverty rates, track changes in poverty rates over time, and target services, provided that it is applied during a period similar to that of the fourth quarter of 2007, the period when the data used to construct the scorecard was collected. The same approach can be applied to any country with similar data from a national income or consumption survey.

References

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

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

Braithwaite, Jeanine; and Anna Ivanova. (1999) "Proxy Means Tests for Russia, 1994–98", Annex 4 in World Bank, *Russia: Targeting and the Longer-Term Poor, Volume II: Annexes*, Report No. 19377-RU, `www-wds.worldbank.org/servlet/WDSContentServer/WDSP/IB/1999/09/10/000094946_99072102535886/Rendered/PDF/multi_page.pdf`, retrieved 14 February 2010.

Caire, Dean. (2004) "Building Credit Scorecards for Small Business Lending in Developing Markets", `microfinance.com/English/Papers/Scoring_SMEs_Hybrid.pdf`, retrieved 15 February 2010.

Chen, Shiyuan; and Mark Schreiner. (2009) "Simple Poverty Scorecard®: Vietnam", `SimplePovertyScorecard.com/VNM_2006_ENG.pdf`, retrieved 9 July 2016.

Christiaensen, Luc; Lanjouw, Peter; Luoto, Jill; and David Stifel. (2008) "The Reliability of Small-Area Estimation Prediction Methods to Track Poverty", `ww2.lafayette.edu/~stifeld/papers/christiaensen_et_al_predict_poverty_2008.pdf`, retrieved 15 February 2010.

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

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

Copestake, James G.; Dawson, Peter; Fanning, John-Paul; McKay, Andrew; and Katie Wright-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. 23, No. 6, pp. 703–723.

Daley-Harris, Sam. (2009) *State of the Microcredit Summit Campaign Report 2009*, `microcreditsummit.org/state_of_the_campaign_report/`, retrieved 15 February 2010.

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

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

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

Falkenstein, Eric. (2008) "DefProb™: A Corporate Probability of Default Model", `defprob.com/publications/DefProb.pdf`, retrieved 2 December 2009.

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

Fuller, Rob. (2006) "Measuring the Poverty of Microfinance Clients in Haiti", `microfinance.com/English/Papers/Scoring_Poverty_Haiti_Fuller.pdf`, retrieved 15 February 2010.

Gibson, John; and Alexander A. Poduzov. (2003) "Assessing Welfare Indicators for Poverty Measurement in Russia".

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

Grootaert, Christiaan; and Jeanine Braithwaite. (1998) "Poverty Correlates and Indicator-Based Targeting in Eastern Europe and the Former Soviet Union", World Bank Policy Research Working Paper No. 1942, `go.worldbank.org/VPMWVLU8E0`, retrieved 15 February 2010.

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 15 February 2010.

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

_____; and Veronica Vinciotti. (2003) "Local Versus Global Models for Classification Problems: Fitting Models Where It Matters", *American Statistician*, Vol. 57, No. 2, pp. 124–131.

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.

Houssou, Nazaire; Zeller, Manfred; Alcaraz V., Gabriela; Schware, Stefan; and Julia Johannsen. (2007) “Proxy Means Tests for Targeting the Poorest Households: Applications to Uganda”, ageconsearch.umn.edu/bitstream/7946/1/sp07ho01.pdf, retrieved 15 February 2010.

IRIS Center. (2007a) “Manual for the Implementation of USAID Poverty Assessment Tools”, povertytools.org/training_documents/Manuals/USAID_PAT_Manual_Eng.pdf, retrieved 15 February 2010.

_____. (2007b) “Introduction to Sampling for the Implementation of PATs”, povertytools.org/training_documents/Sampling/Introduction_Sampling.ppt, retrieved 15 February 2010.

_____. (2005) “Notes on Assessment and Improvement of Tool Accuracy”, povertytools.org/other_documents/AssessingImproving_Accuracy.pdf, retrieved 15 February 2010.

Johannsen, Julia. (2006) “Operational Poverty Targeting in Peru—Proxy Means Testing with Non-Income Indicators”, IPC Working Paper No. 30, undp-povertycentre.org/pub/IPCWorkingPaper30.pdf, retrieved 15 February 2010.

Johnson, Glenn. (2007) “Lesson 3: Two-Way Tables—Dependent Samples”, www.stat.psu.edu/online/development/stat504/03_2way/53_2way_compare.htm, retrieved 15 February 2010.

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

Lokshin, Michael. (2008) “Does Poverty Research in Russia Follow the Scientific Method?” World Bank Policy Research Working Paper No. 4528, go.worldbank.org/KUVSZ8PGC0, retrieved 14 February 2010.

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

Martinelli, César; and Susan W. Parker. (2007) “Deception and Misreporting in a Social Program”, ciep.itam.mx/~martinel/lies4.pdf, retrieved 15 February 2010.

Matul, Michal; and Sean Kline. (2003) "Scoring Change: Prizma's Approach to Assessing Poverty", Microfinance Centre for Central and Eastern Europe and the New Independent States Spotlight Note No. 4, www.mfc.org.pl/doc/Research/ImpAct/SN/MFC_SN04_eng.pdf, retrieved 15 February 2010.

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

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

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

Ovtcharova, Lilia; and Emil Daniel Tesliuc. (2006) "Sensitivity of Poverty and Inequality Statistics to Alternative Definitions of Household Welfare: Illustration Using the NOBUS Survey", socpol.ru/news/docs/Sensitivity_Poverty_Russia.pdf, retrieved 14 February 2010.

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

SAS Institute Inc. (2004) "The LOGISTIC Procedure: Rank Correlation of Observed Responses and Predicted Probabilities", in *SAS/STAT User's Guide, Version 9*, support.sas.com/documentation/cdl/en/statug/63033/HTML/default/statug_logistic_sect035.htm, retrieved 15 February 2010.

Schapire, Robert E. (2001) "The Boosting Approach to Machine Learning: An Overview", www.cs.princeton.edu/courses/archive/fall06/cos402/papers/boosting-survey.pdf, retrieved 15 February 2010.

Schreiner, Mark. (2013) "Simple Poverty Scorecard®: Bangladesh", SimplePovertyScorecard.com/BGD_2010_ENG.pdf, retrieved 9 July 2016.

_____. (2009a) "Simple Poverty Scorecard®: Peru", SimplePovertyScorecard.com/PER_2007_ENG.pdf, retrieved 9 July 2016.

_____. (2009b) "Simple Poverty Scorecard®: Philippines", SimplePovertyScorecard.com/PHL_2002_ENG.pdf, retrieved 9 July 2016.

_____. (2009c) "Simple Poverty Scorecard®: Pakistan", SimplePovertyScorecard.com/PAK_2005_ENG.pdf, retrieved 9 July 2016.

_____. (2009d) "Simple Poverty Scorecard®: Bolivia", SimplePovertyScorecard.com/BOL_2007_ENG.pdf, retrieved 9 July 2016.

_____. (2009e) "Simple Poverty Scorecard®: Mexico", SimplePovertyScorecard.com/MEX_2008_ENG.pdf, retrieved 9 July 2016.

_____. (2008a) "Simple Poverty Scorecard®: Peru", SimplePovertyScorecard.com/PER_2003_ENG.pdf, retrieved 9 July 2016.

_____. (2008b) "Simple Poverty Scorecard®: India", SimplePovertyScorecard.com/IND_2005_ENG.pdf, retrieved 9 July 2016.

_____. (2006a) "Índice de Calificación de la Pobreza™: México", SimplePovertyScorecard.com/MEX_2002_SPA.pdf, retrieved 9 July 2016.

_____. (2006b) "Is One Simple Poverty Scorecard® Enough for India?", microfinance.com/English/Papers/Scoring_Poverty_India_Segments.pdf, retrieved 15 February 2010.

_____. (2005) "IRIS Questions on the Simple Poverty Scorecard®", microfinance.com/English/Papers/Scoring_Poverty_Response_to_IRIS.pdf, retrieved 15 February 2010.

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

_____; Matul, Michal; Pawlak, Ewa; and Sean Kline. (2004) "The Simple Poverty Scorecard®: Lessons from a Microlender in Bosnia-Herzegovina", microfinance.com/English/Papers/Scoring_Poverty_in_BiH_Short.pdf, retrieved 15 February 2010.

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

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

Struyk, Raymond; and Anastasia Kolodeznikova. (1999) “Needs-Based Targeting without Knowing Household Incomes: How Would It Work in Russia?” *Urban Studies*, Vol. 36, No. 11, pp. 1875–1889.

Tarozzi, Alesandro; and Angus Deaton. (2007) “Using Census and Survey Data to Estimate Poverty and Inequality for Small Areas”, `princeton.edu/~deaton/downloads/20080301SmallAreas_FINAL.pdf`, retrieved 15 February 2010.

Toohig, Jeff. (2008) “PPI Pilot Training Guide”, `progressoutofpoverty.org/toolkit`, retrieved 15 February 2010.

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

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

World Bank. (2009) *Russian Federation: Addressing the Challenge of Chronic Poverty and Vulnerability to Poverty.*

_____. (2008) “International Comparison Project: Tables of Results”, `siteresources.worldbank.org/ICPINT/Resources/icp-final-tables.pdf`, retrieved 15 February 2010.

_____. (2005) *Russian Federation: Reducing Poverty through Growth and Social Policy Reform, Report No. 28923–RU.*

Zeller, Manfred. (2004) “Review of Poverty Assessment Tools”, `povertytools.org/other_documents/Review%20of%20PAT%20Tools.pdf`, retrieved 1 February 2010.

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

Sub-sample	Number of households	% with expenditure below a poverty line: National	USAID 'Extreme'	$6.25/day 2005 PPP
All Russia	53,042	9.7	4.7	5.0
Construction				
Selecting indicators and weights	17,750	9.9	4.8	5.1
Calibration				
Associating scores with likelihoods	17,666	9.8	4.8	5.1
Validation				
Measuring accuracy	17,626	9.4	4.5	4.8
Change in poverty rate (percentage points)				
From construction/calibration to validation		+0.5	+0.3	+0.3

Source: 2007 HBS

Figure 3: Poverty lines and poverty rates by krai/oblast/republic, at household- and person-levels

	HH	Poverty line (RUB/person/day) and poverty rate (%, household level and person level)								
	in	National			USAID "extreme"			$6.25/day 2005 PPP		
Krai/Oblast/Republic	HBS	Line	Household	Person	Line	Household	Person	Line	Household	Person
All Russia	53,042	132	9.7	12.2	105	4.7	6.3	102	5.0	6.7
Altai Krai	799	117	8.5	10.3	93	4.2	5.2	102	5.7	7.1
Krasnodar Krai	959	126	14.2	17.5	99	6.4	8.8	102	7.1	9.5
Krasnoyarsk Krai	1,466	137	7.2	8.8	114	3.4	4.5	102	2.3	3.1
Primorsky Krai	684	168	13.2	16.0	135	7.0	8.4	102	1.3	1.4
Stavropol Krai	739	116	12.9	16.2	89	6.2	8.2	102	8.1	10.7
Khabarovsk Krai	662	179	8.2	10.5	137	4.1	5.5	102	1.6	2.3
Amur Oblast	529	165	21.1	24.5	129	10.9	13.4	102	3.3	3.8
Arkhangelsk Oblast	896	161	11.0	14.3	133	5.1	7.3	102	1.5	2.4
Astrakhan Oblast	504	114	10.0	14.2	91	5.1	7.4	102	7.4	10.6
Belgorod Oblast	529	109	4.0	4.8	99	2.0	2.4	102	2.8	3.5
Bryansk Oblast	502	112	9.8	10.6	95	4.7	5.4	102	6.4	7.5
Vladimir Oblast	604	124	11.6	14.1	100	6.0	7.1	102	6.4	7.6
Volgograd Oblast	824	122	6.6	7.8	104	3.4	4.2	102	2.7	3.1
Vologda Oblast	559	134	7.2	10.5	119	3.8	5.7	102	2.0	3.0
Voronezh Oblast	744	119	12.6	14.4	91	5.8	7.4	102	8.3	10.2
Nizhny Novgorod Oblast	884	127	11.7	13.8	105	5.8	7.0	102	5.2	6.2
Ivanovo Oblast	456	122	13.5	17.7	101	6.4	8.9	102	7.1	9.7
Irkutsk Oblast	1,030	130	10.5	14.5	96	5.1	7.2	102	5.3	7.4
Ingushetia Republic	290	104	37.5	42.2	72	18.8	21.4	102	37.5	42.2
Kaliningrad Oblast	544	135	8.8	11.6	107	3.8	6.0	102	3.1	4.0
Tver Oblast	734	127	7.5	8.8	101	3.5	4.4	102	3.6	4.5
Kaluga Oblast	508	118	3.6	4.2	103	1.9	2.1	102	1.7	1.9
Kamchatka Krai	854	255	13.2	18.6	187	6.3	9.4	102	0.1	0.1
Kemerovo Oblast	814	117	7.0	9.2	87	2.9	4.6	102	4.5	6.2
Kirov Oblast	678	123	7.0	8.7	100	3.4	4.4	102	3.5	4.6
Kostroma Oblast	534	120	4.2	5.5	107	2.2	3.1	102	1.2	1.6
Samara Oblast	918	141	8.0	9.3	103	3.5	4.6	102	3.4	4.6
Kurgan Oblast	518	115	11.6	14.4	90	5.8	7.5	102	8.9	11.4
Kursk Oblast	549	113	4.8	6.3	95	2.5	3.2	102	3.3	4.2
Saint Petersburg	957	137	1.2	1.8	134	1.1	1.8	102	0.0	0.0
Leningrad Oblast	561	126	3.0	4.3	107	1.7	2.5	102	1.2	1.6
Lipetsk Oblast	499	112	2.4	3.1	103	1.2	1.6	102	1.1	1.5
Magadan Oblast	523	206	13.0	17.3	161	6.1	8.9	102	2.1	3.6
Moscow Oblast	1,407	147	1.1	1.5	119	0.6	0.9	102	0.4	0.5
Moscow	1,063	192	21.9	25.5	131	10.0	12.8	102	5.9	8.5
Murmansk Oblast	492	195	2.5	3.6	156	1.5	2.5	102	0.0	0.0
Novgorod Oblast	503	130	24.6	28.5	104	12.9	14.9	102	10.0	10.7
Novosibirsk Oblast	739	138	13.0	13.7	113	6.5	6.9	102	4.8	5.4
Omsk Oblast	703	129	8.6	12.5	101	4.1	6.4	102	4.4	6.8
Orenburg Oblast	699	113	10.3	13.2	93	5.0	6.7	102	7.2	9.3
Oryol Oblast	554	104	4.2	5.8	95	2.5	3.3	102	3.9	5.5
Penza Oblast	529	113	10.4	12.9	89	5.2	6.5	102	7.1	8.9
Perm Krai	1,179	139	14.8	18.2	104	7.0	9.2	102	6.3	8.4
Pskov Oblast	529	117	8.7	11.5	95	4.2	5.8	102	5.3	7.3
Rostov Oblast	984	123	6.2	8.5	100	3.0	4.3	102	3.2	4.5
Ryazan Oblast	559	121	13.0	14.8	99	6.1	7.5	102	7.9	9.2
Saratov Oblast	824	123	18.6	23.5	92	9.2	12.1	102	11.7	14.9
Sakhalin Oblast	459	217	16.5	20.9	153	8.0	10.5	102	3.1	3.9
Sverdlovsk Oblast	1,049	128	8.1	9.8	95	3.7	5.0	102	4.4	5.7
Smolensk Oblast	502	122	30.4	32.9	89	14.6	17.1	102	20.3	22.6
Tambov Oblast	604	101	10.5	13.9	80	4.5	7.0	102	10.6	14.0
Tomsk Oblast	558	136	6.5	8.6	123	4.6	6.2	102	0.6	0.3
Tula Oblast	629	119	4.3	6.1	103	2.2	3.3	102	1.9	2.9
Tyumen Oblast	1,108	131	3.2	3.7	106	1.6	1.8	102	1.1	1.2
Ulyanovsk Oblast	563	115	13.8	17.2	93	6.6	8.7	102	9.3	11.9
Chelyabinsk Oblast	809	118	4.8	5.7	98	2.3	2.9	102	2.9	3.6
Zabaykalsky Krai	882	131	12.8	17.8	92	6.1	9.6	102	7.9	12.2
Chukotka Autonomous Okrug	425	295	30.2	36.8	197	14.8	18.8	102	3.1	4.8
Yaroslavl Oblast	627	131	12.3	15.2	111	5.8	7.6	102	4.3	5.8
Adygea Republic	519	116	7.6	10.4	97	3.9	5.5	102	4.7	6.7
Bashkortostan Republic	979	111	6.2	8.5	92	3.4	4.7	102	5.1	7.0
Buryat Republic	584	136	19.5	24.1	94	8.9	12.1	102	11.0	14.6
Dagestan Republic	666	104	17.6	23.7	77	8.3	12.0	102	17.2	23.2
Kabardino-Balkar Republic	529	100	4.2	5.9	84	2.1	3.1	102	4.9	6.8
Altai Republic	443	149	27.5	36.0	105	14.8	20.4	102	11.1	14.2
Kalmykia Republic	509	109	27.4	33.1	79	13.5	16.7	102	24.1	29.8
Karelia Republic	479	152	12.4	15.2	125	6.2	8.1	102	3.3	4.1
Komi Republic	624	170	18.7	20.9	127	9.5	10.8	102	6.0	7.1
Mari El Republic	529	108	14.9	19.5	86	7.3	9.8	102	12.6	16.5
Mordovia Republic	494	107	15.0	18.6	78	6.6	9.4	102	13.0	16.4
North Ossetia-Alania Republic	534	100	7.1	9.7	88	3.7	5.0	102	7.6	10.6
Karachay-Cherkess Republic	459	103	7.9	11.1	80	4.2	5.7	102	7.9	11.1
Tatarstan Republic	994	104	12.0	14.8	83	5.9	7.9	102	11.5	14.2
Tyva Republic	514	132	25.2	31.4	93	12.9	16.1	102	15.3	19.0
Udmurt Republic	480	119	6.9	9.4	94	3.3	5.0	102	3.8	5.6
Khakassia Republic	509	122	7.7	10.8	103	3.8	5.5	102	3.5	4.9
Chuvashia Republic	554	110	6.1	8.0	97	3.2	4.2	102	4.4	5.9
Sakha (Yakutia) Republic	504	221	16.2	22.1	160	7.3	11.1	102	2.2	3.4
Jewish Autonomous Oblast	309	154	12.6	16.0	117	5.8	8.0	102	5.1	7.1

Figure 4: Poverty indicators by uncertainty coefficient

Uncertainty coefficient	Indicator (responses are ordered starting with those associated with higher poverty likelihoods)
826	How many land-line and cellular telephones does the household own? (None; One or more land-lines, and no cellular; No land-lines, and one cellular; One or more land-lines, and one cellular; No land-lines, and two or more cellular; One or more land-lines, and two cellular; One or more land-lines, and three or more cellular)
785	Does the household own a land-line telephone and/or a cellular telephone? (None; Only cellular; Only land-line; Both)
559	What is the source of hot water for the residence? (None; Individual water heater; Centralized)
554	How many members does the household have? (Five or more; Four; Three; Two; One)
522	What is the highest level of education that the female head/spouse has completed? (None or less than primary, general primary (primary), or major general (incomplete secondary); Initial vocational training (vocational) without a certificate of secondary (full) general education; Secondary (complete) general, or initial vocational training (vocational) with a certificate of secondary (full) general education; Secondary vocational training (secondary vocational), or incomplete higher professional (incomplete higher); No female head/spouse; Higher professional, or graduate professional)
517	Does the residence have a land-line telephone? (No; Yes)
515	What is the highest level of education that a member of the household has completed? (None or less than primary, general primary (primary), major general (incomplete secondary), secondary (complete) general, initial vocational training (vocational) with a certificate of secondary (full) general education, or initial vocational training (vocational) without a certificate of secondary (full) general education; Secondary vocational training (secondary vocational); Incomplete higher professional (incomplete higher); Higher professional, or graduate professional)
496	Is there a flush toilet in the residence? (No; Yes)

Figure 4 (cont.): Poverty indicators by uncertainty coefficient

Uncertainty coefficient	Indicator (responses are ordered starting with those associated with higher poverty likelihoods)
495	What is the occupation of the female head/spouse in her main job? (Unskilled workers; Skilled workers in agriculture, forestry, hunting, and fishing, military personnel, or does not work; Service workers in housing, communal services, trade, and related activities; Skilled workers of large and small industrial enterprises, crafts, construction, transport, communications, geology, and mining, or operators and drivers and machinery and equipment fitters/assemblers; No female head/spouse; Skilled technicians; White-collar employees; Highly skilled specialists/professionals; Administrators and government leaders, including heads of agencies, organizations, and state businesses)
437	Does the residence have a bathroom with a toilet and/or shower? (No; Yes)
425	Is the residence connected to public sewerage? (No; Yes)
411	How many cellular telephones does the household own? (None; One; Two; Three; Four or more)
407	In their main line of work, are any household members administrators or heads/leaders of government, agencies, organizations, and state businesses, highly skilled specialists/professionals, skilled technicians, or white-collar employees? (No; Yes)
381	What type of heat does the residence have? (Other; Central heat (from public source or source common to a multi-unit building))
378	How many household members are 18-years-old or younger? (Two or more; One; None)
376	Does the household own a personal computer? (No; Yes)
368	What is the occupation of the male head/spouse in his main job? (Does not work, skilled worker in agriculture, forestry, hunting, and fishing, or unskilled worker; Operator, driver, machinery and equipment fitters/assemblers, or military personnel; No male head/spouse; Skilled workers of large and small industrial enterprises, crafts, construction, transport, communications, geology, and mining; Service workers in housing, communal services, trade, and related activities; Skilled technicians, or white-collar employees; Highly skilled specialists/professionals; Government leaders and administrators, including heads of agencies, organizations, and state businesses)

Figure 4 (cont.): Poverty indicators by uncertainty coefficient

Uncertainty coefficient	Indicator (responses are ordered starting with those associated with higher poverty likelihoods)
352	Does the household own a vacuum cleaner? (No; Yes)
350	How many household members are 17-years-old or younger? (Two or more; One; None)
348	Does the household own a microwave? (No; Yes)
339	What is the highest level of education that the male head/spouse has completed? (None or less than primary, general primary (primary), or major general (incomplete secondary); Initial vocational training (vocational) without a certificate of secondary (full) general education; Secondary (complete) general, initial vocational training (vocational) with a certificate of secondary (full) general education, or incomplete higher professional (incomplete higher); Secondary vocational training (secondary vocational); No male head/spouse; Higher professional, or graduate professional)
328	What type of residence does the household have? (Detached house or part of a house with a separate entrance; Communal residence, dorm/hostel, none, or other (yurt, trailer, barge); Private apartment)
320	What is the source of drinking water of the household? (Other; Piped-in to the residence)
320	How many color televisions does the household own? (None; One; Two or more)
312	How many household members are 16-years-old or younger? (Two or more; One; None)
290	How many household members are 15-years-old or younger? (Two or more; One; None)
284	How many televisions (color or black-and-white) does the household own? (None, or one; Two; Three or more)?

Figure 4 (cont.): Poverty indicators by uncertainty coefficient

Uncertainty coefficient	Indicator (responses are ordered starting with those associated with higher poverty likelihoods)
274	What is the main activity of the enterprise where the female head of the household mainly works? (Does not work; Agriculture, hunting, forestry, fishing, or fish farming; Health care and social services; Education; Wholesale and retail trade, repair of motor vehicles, motorcycles, household goods, and personal items, or hotels and restaurants; Mining, manufacturing, transport and communications, provision of other community, social, and personal services, provision of household services or activities of international organizations; No female head/spouse; Production and distribution of electricity, gas, and water, construction, finance, real estate, renting, and services, or public administration, military security, and compulsory social service)
260	How many VCRs and DVDs does the household own? (None; One; Two or more)
259	What is the main activity of the enterprise where the male head of the household mainly works? (Does not work; Agriculture, hunting, forestry, fishing, and fish farming; Construction, finance, education, health care and social services, or provision of other community, social, and personal services; No male head/spouse; Transport and communications; Manufacturing, production and distribution of electricity, gas, and water, wholesale and retail trade, repair of motor vehicles, motorcycles, household goods, and personal items, provision of household services, or activities of international organizations; Mining, hotels and restaurants, real estate, renting, and services, or public administration, military security, and compulsory social service)
250	Do all children ages 7 to 18 attend school? (No; Yes; No children these ages)
224	Does the household own a bicycle, motorcycle, or an automobile? (None; Only bicycle; Motorcycle, but no car (regardless of bicycle); Car (regardless of bicycle or motorcycle))
223	What was the employment status of the male head/spouse in his main line of work? (Does not work, farmer. member of a production cooperative (*artel*), or domestic worker; Employed by an individual or entrepreneur; No male head/spouse; Employed in an enterprise, institution, or organization; Owner or co-owner of a business with employees, or self-employed with no employees)

Figure 4 (cont.): Poverty indicators by uncertainty coefficient

Uncertainty coefficient	Indicator (responses are ordered starting with those associated with higher poverty likelihoods)
222	Do all children ages 7 to 17 attend school? (No; Yes; No children these ages)
221	How many household members are 14-years-old or younger? (One or more; None)
204	How many household members are 13-years-old or younger? (One or more; None)
196	Does the household own an automobile? (No; Yes)
187	Does the household own a television (color or black-and-white) and a VCR or a DVD? (No television (regardless of VCR or DVD), television without VCR or DVD; Television with VCR or DVD)
186	Do all children ages 7 to 16 attend school? (No; Yes; No children these ages)
171	Does the household own a radio, tape recorder, or stereo system? (None; Radio, but no tape recorder nor stereo system; Tape recorder, but no stereo system (regardless of radio); Stereo system (regardless of radio or tape recorder))
171	How many household members are 12-years-old or younger? (One or more; None)
170	Do all children ages 7 to 15 attend school? (No; Yes; No children these ages)
158	Do all children ages 7 to 14 attend school? (No; Yes; No children these ages)
150	Does the household own a VCR? (No; Yes)
148	What was the employment status of the female head/spouse in her main line of work? (Does not work, farmer, member of a production cooperative (*artel*), self-employed with no employees, or domestic worker; Employed by an individual or entrepreneur; No female head/spouse; Employed in an enterprise, institution, or organization, or owner or co-owner of a business with employees)
142	How many household members are 11-years-old or younger? (One or more; None)
142	Do all children ages 7 to 13 attend school? (No; Yes; No children these ages)
140	Does the household own a stereo system? (No; Yes)
137	Does the residence have piped-in gas, LPG, or neither? (LPG; Piped-in gas; Neither)
118	How many refrigerators does the household own? (None; One; Two or more)

Figure 4 (cont.): Poverty indicators by uncertainty coefficient

Uncertainty coefficient	Indicator (responses are ordered starting with those associated with higher poverty likelihoods)
114	Do all children ages 7 to 12 attend school? (No; Yes; No children these ages)
89	Did the female head/spouse work or have gainful employment continuously in the past three months? (No; Yes; No female head/spouse)
87	Does the household own a refrigerator or freezer? (None; Refrigerator, but no freezer; Freezer (regardless of refrigerator))
85	How old is the female head/spouse? (60 or older; 26 to 34; 35 to 44; No female head/spouse; 25 or younger; 45 to 59)
83	How many household members are 5-years-old or younger? (One or more; None)
79	Does the household own an air conditioner? (No; Yes)
78	Do all children ages 7 to 11 attend school? (No; Yes; No children these ages)
76	Are any household members employed in an enterprise, institution, or organization in their main line of work? (No; Yes)
75	Does the household own a DVD? (No; Yes)
68	Did the male head/spouse work or have gainful employment continuously in the past three months? (No; Yes; No male head/spouse)
66	Does the household own a video camera? (No; Yes)
59	Are any household members employed by an individual or entrepreneur in their main line of work? (Yes; No)
57	Does the household own a freezer? (No; Yes)
52	How old is the male head/spouse? (60 or older; 25 to 47; No male head/spouse; 24 or younger; 48 to 59)
47	Does the household own a dishwasher? (No; Yes)
41	Does the household own a sewing machine? (No; Yes)
39	Does the household own a sewing machine or a knitting machine? (No; Yes)

Figure 4 (cont.): Poverty indicators by uncertainty coefficient

Uncertainty coefficient	Indicator (responses are ordered starting with those associated with higher poverty likelihoods)
36	Does the household own a radio? (No; Yes)
35	What is the tenancy status of the household in its residence? (State or municipality, another entity, or no one; Owned by the household (privatized, bought, etc.); Private person or entity)
34	How many household members worked or had gainful employment continuously in the past three months? (None; Two; One; Three or more)
29	How many household members are employed in an enterprise, institution, or organization or employed by an individual or entrepreneur in their main line of work? (None; One; Two; Three or more)
29	Does the household own a clothes washer? (No; Yes)
19	What is the structure of household headship? (Both male and female heads/spouses; Female head/spouse only; Male head/spouse only)
15	How many rooms does the residence have? (Four or more; Three; Two; One)
14	Does the household own a tape recorder? (No; Yes)
13	What is the total area of the residence in meters squared? (24 or less; 25 to 39; 40 to 59; 60 to 99; 100 or more)
12	Are any household members self-employed with no employees? (No; Yes)
9	Does the household own a bicycle? (Yes; No)
6	Does the household own a motorcycle? (Yes; No)
4	Does the household own a knitting machine? (No; Yes)
2	Does the household own a black-and-white television? (Yes; No)

Source: 2007 HBS the national poverty line

National Poverty Line

(and tables pertaining to all three poverty lines)

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

If a household's score is . . .	. . . then the likelihood (%) of being below the poverty line is:
0–4	100.0
5–9	88.8
10–14	68.0
15–19	72.7
20–24	59.2
25–29	41.5
30–34	35.1
35–39	23.5
40–44	15.0
45–49	8.8
50–54	4.5
55–59	2.4
60–64	1.4
65–69	0.5
70–74	0.2
75–79	0.5
80–84	0.0
85–89	0.0
90–94	0.0
95–100	0.0

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

Score	Households below poverty line		All households at score		Poverty likelihood (estimated, %)
0–4	7	÷	7	=	100.0
5–9	93	÷	105	=	88.8
10–14	240	÷	354	=	68.0
15–19	425	÷	585	=	72.7
20–24	713	÷	1,203	=	59.2
25–29	1,125	÷	2,708	=	41.5
30–34	1,468	÷	4,187	=	35.1
35–39	1,760	÷	7,480	=	23.5
40–44	1,668	÷	11,111	=	15.0
45–49	1,050	÷	11,923	=	8.8
50–54	691	÷	15,245	=	4.5
55–59	347	÷	14,597	=	2.4
60–64	194	÷	14,146	=	1.4
65–69	44	÷	9,433	=	0.5
70–74	6	÷	3,777	=	0.2
75–79	14	÷	2,961	=	0.5
80–84	0	÷	168	=	0.0
85–89	0	÷	8	=	0.0
90–94	0	÷	0	=	0.0
95–100	0	÷	0	=	0.0

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

Figure 7 (All poverty lines): Distribution of household poverty likelihoods across ranges demarcated by poverty lines

	Likelihood of expenditure in range demarcated by poverty lines per day per person		
	<USAID	=>USAID and <National	=>National
Score	<RUB102	=>RUB102 and <RUB132	=>RUB132
0–4	100.0	0.0	0.0
5–9	81.9	6.9	11.2
10–14	56.4	11.6	32.0
15–19	42.0	30.7	27.3
20–24	36.3	23.0	40.8
25–29	26.9	14.7	58.5
30–34	16.2	18.9	65.0
35–39	10.2	13.3	76.5
40–44	6.4	8.7	85.0
45–49	3.4	5.4	91.2
50–54	1.6	3.0	95.5
55–59	0.9	1.5	97.6
60–64	0.6	0.8	98.6
65–69	0.2	0.3	99.5
70–74	0.1	0.1	99.8
75–79	0.5	0.0	99.5
80–84	0.0	0.0	100.0
85–89	0.0	0.0	100.0
90–94	0.0	0.0	100.0
95–100	0.0	0.0	100.0

All poverty likelihoods in percentage units.

$6.25/day 2005 PPP line omitted because it is almost the same as the USAID "extreme" line.

Figure 8 (National line): Bootstrapped differences between estimated and true poverty likelihoods for households in a large sample ($n = 16{,}384$) from the validation sample, with confidence intervals, scorecard applied to the validation sample

	Difference between estimate and true value			
		Confidence interval (+/− percentage points)		
Score	Diff.	90-percent	95-percent	99-percent
0–4	+0.0	0.0	0.0	0.0
5–9	–3.2	10.2	11.5	15.6
10–14	–1.8	13.3	15.6	21.3
15–19	+5.2	9.3	11.1	14.6
20–24	+8.3	8.2	9.8	12.0
25–29	+7.9	6.5	7.6	9.7
30–34	+8.1	3.7	4.4	5.9
35–39	+16.3	1.2	1.5	1.8
40–44	+10.9	0.7	0.8	1.0
45–49	+6.0	0.5	0.5	0.7
50–54	+3.5	0.2	0.2	0.3
55–59	+1.2	0.3	0.4	0.5
60–64	+1.3	0.0	0.0	0.0
65–69	+0.4	0.0	0.0	0.0
70–74	+0.1	0.1	0.1	0.1
75–79	+0.5	0.0	0.0	0.0
80–84	+0.0	0.0	0.0	0.0
85–89	+0.0	0.0	0.0	0.0
90–94	+0.0	0.0	0.0	0.0
95–100	+0.0	0.0	0.0	0.0

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

		Poverty line				
				USAID		**$6.25/day**
		National		**'Extreme'**		**2005 PPP**
Estimate minus true value						
Scorecard applied to validation		+2.2		+0.9		+0.9
Precision of difference						
Scorecard applied to validation		0.1		0.1		0.1
α factor						
Scorecard applied to validation		0.31		0.27		0.26

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

Figure 10 (National line): Differences and precision of differences for bootstrapped estimates of poverty rates for groups of households at a point in time, by sample size, scorecard applied to the validation sample

Sample Size n	Difference between estimate and true value			
		Confidence interval (+/− percentage points)		
	Diff.	90-percent	95-percent	99-percent
1	–0.0	52.9	63.3	78.4
4	+2.2	16.6	25.6	39.3
8	+2.7	9.9	13.4	23.7
16	+2.5	5.7	7.9	13.4
32	+2.5	3.5	4.5	7.3
64	+2.3	2.0	2.5	4.2
128	+2.3	1.5	1.8	2.7
256	+2.2	1.0	1.1	1.5
512	+2.2	0.7	0.8	1.1
1,024	+2.2	0.5	0.6	0.8
2,048	+2.2	0.3	0.4	0.5
4,096	+2.2	0.2	0.3	0.4
8,192	+2.2	0.2	0.2	0.3
16,384	+2.2	0.1	0.1	0.2

Figure 11 (All poverty lines): Possible outcomes from targeting by poverty score

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

Figure 12 (National line): Households by targeting classification and score, along with "Total Accuracy" and BPAC, scorecard applied to the validation sample

Score	Inclusion: < poverty line correctly targeted	Undercoverage: < poverty line mistakenly non-targeted	Leakage: => poverty line mistakenly targeted	Exclusion: => poverty line correctly non-targeted	Total Accuracy Inclusion + Exclusion	BPAC See text
0–4	0.0	9.4	0.0	90.6	90.6	–99.8
5–9	0.1	9.3	0.0	90.6	90.7	–97.7
10–14	0.4	9.0	0.1	90.5	90.9	–91.1
15–19	0.7	8.6	0.3	90.3	91.0	–80.9
20–24	1.4	8.0	0.8	89.8	91.2	–60.6
25–29	2.6	6.8	2.4	88.2	90.8	–19.7
30–34	4.0	5.4	5.1	85.5	89.5	+40.3
35–39	5.5	3.9	11.1	79.5	85.1	–18.1
40–44	7.0	2.4	20.7	69.9	76.9	–120.3
45–49	8.0	1.4	31.6	59.0	67.0	–236.6
50–54	8.7	0.7	46.2	44.4	53.2	–391.5
55–59	9.2	0.2	60.3	30.3	39.5	–542.1
60–64	9.3	0.1	74.3	16.3	25.6	–691.1
65–69	9.4	0.0	83.7	6.9	16.3	–791.1
70–74	9.4	0.0	87.5	3.1	12.5	–831.1
75–79	9.4	0.0	90.4	0.2	9.6	–862.6
80–84	9.4	0.0	90.6	0.0	9.4	–864.4
85–89	9.4	0.0	90.6	0.0	9.4	–864.5
90–94	9.4	0.0	90.6	0.0	9.4	–864.5
95–100	9.4	0.0	90.6	0.0	9.4	–864.5

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

Figure 13 (National line): For a given score cut-off, the percentage of all households who are targeted, the percentage of targeted households who are poor, the percentage of poor households who are targeted, and the number of poor households who are successful targeted (inclusion) per non-poor household mistakenly targeted (leakage), scorecard applied to the validation sample

Targeting cut-off	% all households who are targeted	% targeted who are poor	% of poor who are targeted	Poor households targeted per non-poor household targeted
0–4	0.0	100.0	0.1	Only poor targeted
5–9	0.1	90.2	1.1	9.2:1
10–14	0.5	79.3	3.9	3.8:1
15–19	1.1	70.9	7.9	2.4:1
20–24	2.3	64.0	15.4	1.8:1
25–29	5.0	52.1	27.5	1.1:1
30–34	9.1	44.1	42.9	0.8:1
35–39	16.6	33.3	59.0	0.5:1
40–44	27.7	25.4	75.0	0.3:1
45–49	39.7	20.3	85.6	0.3:1
50–54	54.9	15.9	93.0	0.2:1
55–59	69.5	13.2	97.8	0.2:1
60–64	83.7	11.2	99.4	0.1:1
65–69	93.1	10.1	99.8	0.1:1
70–74	96.9	9.7	100.0	0.1:1
75–79	99.8	9.4	100.0	0.1:1
80–84	100.0	9.4	100.0	0.1:1
85–89	100.0	9.4	100.0	0.1:1
90–94	100.0	9.4	100.0	0.1:1
95–100	100.0	9.4	100.0	0.1:1

USAID “Extreme” Line

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

If a household's score is . . .	. . . then the likelihood (%) of being below the poverty line is:
0–4	100.0
5–9	81.9
10–14	56.4
15–19	42.0
20–24	36.3
25–29	26.9
30–34	16.2
35–39	10.2
40–44	6.4
45–49	3.4
50–54	1.6
55–59	0.9
60–64	0.6
65–69	0.2
70–74	0.1
75–79	0.5
80–84	0.0
85–89	0.0
90–94	0.0
95–100	0.0

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

Score	Households below poverty line		All households at score		Poverty likelihood (estimated, %)
0–4	7	÷	7	=	100.0
5–9	86	÷	105	=	81.9
10–14	199	÷	354	=	56.4
15–19	246	÷	585	=	42.0
20–24	436	÷	1,203	=	36.3
25–29	728	÷	2,708	=	26.9
30–34	677	÷	4,187	=	16.2
35–39	763	÷	7,480	=	10.2
40–44	707	÷	11,111	=	6.4
45–49	404	÷	11,923	=	3.4
50–54	236	÷	15,245	=	1.6
55–59	127	÷	14,597	=	0.9
60–64	78	÷	14,146	=	0.6
65–69	17	÷	9,433	=	0.2
70–74	3	÷	3,777	=	0.1
75–79	14	÷	2,961	=	0.5
80–84	0	÷	168	=	0.0
85–89	0	÷	8	=	0.0
90–94	0	÷	0	=	0.0
95–100	0	÷	0	=	0.0

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

Figure 8 (USAID "extreme" line): Bootstrapped differences between estimated and true poverty likelihoods for households in a large sample (n = 16,384) from the validation sample, with confidence intervals, scorecard applied to the validation sample

	Difference between estimate and true value			
		Confidence interval (+/− percentage points)		
Score	**Diff.**	**90-percent**	**95-percent**	**99-percent**
0–4	+23.3	50.0	50.0	50.0
5–9	+9.7	19.1	23.1	30.9
10–14	–9.9	13.2	15.2	22.7
15–19	+12.6	8.8	10.4	13.2
20–24	+3.5	7.4	9.0	12.1
25–29	–0.7	6.4	7.5	9.9
30–34	–1.6	3.2	3.9	5.2
35–39	+6.5	0.9	1.1	1.4
40–44	+4.7	0.4	0.5	0.7
45–49	+2.3	0.3	0.4	0.4
50–54	+1.2	0.1	0.1	0.2
55–59	+0.4	0.2	0.2	0.3
60–64	+0.5	0.0	0.0	0.0
65–69	+0.2	0.0	0.0	0.0
70–74	+0.1	0.0	0.0	0.0
75–79	+0.5	0.0	0.0	0.0
80–84	+0.0	0.0	0.0	0.0
85–89	+0.0	0.0	0.0	0.0
90–94	+0.0	0.0	0.0	0.0
95–100	+0.0	0.0	0.0	0.0

Figure 10 (USAID "extreme" line): Differences and precision of differences for bootstrapped estimates of poverty rates for groups of households at a point in time, by sample size, scorecard applied to the validation sample

Sample Size n	Difference between estimate and true value			
		Confidence interval (+/− percentage points)		
	Diff.	90-percent	95-percent	99-percent
1	–0.5	37.1	56.8	70.2
4	+0.6	11.2	18.1	36.9
8	+1.1	6.0	9.3	17.6
16	+0.9	3.6	5.5	10.1
32	+0.9	2.2	3.0	5.1
64	+0.9	1.3	1.7	2.5
128	+0.9	0.9	1.1	1.7
256	+0.9	0.6	0.7	1.0
512	+0.9	0.4	0.5	0.7
1,024	+0.9	0.3	0.4	0.5
2,048	+0.9	0.2	0.2	0.3
4,096	+0.9	0.1	0.2	0.2
8,192	+0.9	0.1	0.1	0.2
16,384	+0.9	0.1	0.1	0.1

Figure 12 (USAID "extreme" line): Households by targeting classification and score, along with "Total Accuracy" and BPAC, scorecard applied to the validation sample

Score	Inclusion: < poverty line correctly targeted	Undercoverage: < poverty line mistakenly non-targeted	Leakage: => poverty line mistakenly targeted	Exclusion: => poverty line correctly non-targeted	Total Accuracy Inclusion + Exclusion	BPAC See text
0–4	0.0	4.5	0.0	95.5	95.5	–99.7
5–9	0.1	4.4	0.0	95.5	95.5	–95.8
10–14	0.3	4.2	0.2	95.3	95.7	–82.7
15–19	0.5	4.0	0.5	95.0	95.5	–64.6
20–24	1.0	3.5	1.3	94.2	95.2	–28.5
25–29	1.6	2.9	3.3	92.2	93.8	+25.8
30–34	2.4	2.1	6.7	88.8	91.2	–49.0
35–39	3.1	1.4	13.5	82.0	85.1	–199.9
40–44	3.7	0.8	24.1	71.4	75.1	–434.7
45–49	4.1	0.4	35.6	59.9	64.0	–691.1
50–54	4.3	0.2	50.6	44.9	49.1	–1,025.5
55–59	4.4	0.1	65.1	30.4	34.8	–1,346.7
60–64	4.5	0.0	79.2	16.3	20.8	–1,659.7
65–69	4.5	0.0	88.6	6.9	11.4	–1,869.2
70–74	4.5	0.0	92.4	3.1	7.6	–1,953.1
75–79	4.5	0.0	95.3	0.2	4.7	–2,018.9
80–84	4.5	0.0	95.5	0.0	4.5	–2,022.6
85–89	4.5	0.0	95.5	0.0	4.5	–2,022.8
90–94	4.5	0.0	95.5	0.0	4.5	–2,022.8
95–100	4.5	0.0	95.5	0.0	4.5	–2,022.8

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

Figure 13 (USAID "extreme" line): For a given score cut-off, the percentage of all households who are targeted, the percentage of targeted households who are poor, the percentage of poor households who are targeted, and the number of poor households who are successful targeted (inclusion) per non-poor household mistakenly targeted (leakage), scorecard applied to the validation sample

Targeting cut-off	% all households who are targeted	% targeted who are poor	% of poor who are targeted	Poor households targeted per non-poor household targeted
0–4	0.0	65.3	0.1	1.9:1
5–9	0.1	68.2	1.7	2.1:1
10–14	0.5	67.2	7.0	2.0:1
15–19	1.1	51.5	12.0	1.1:1
20–24	2.3	42.7	21.4	0.7:1
25–29	5.0	32.7	36.1	0.5:1
30–34	9.1	26.8	54.4	0.4:1
35–39	16.6	18.9	69.7	0.2:1
40–44	27.7	13.3	81.9	0.2:1
45–49	39.7	10.3	90.6	0.1:1
50–54	54.9	7.8	95.0	0.1:1
55–59	69.5	6.4	98.3	0.1:1
60–64	83.7	5.4	99.7	0.1:1
65–69	93.1	4.8	99.9	0.1:1
70–74	96.9	4.6	100.0	0.0:1
75–79	99.8	4.5	100.0	0.0:1
80–84	100.0	4.5	100.0	0.0:1
85–89	100.0	4.5	100.0	0.0:1
90–94	100.0	4.5	100.0	0.0:1
95–100	100.0	4.5	100.0	0.0:1

$6.25/day 2005 PPP Line

Figure 5 ($6.25/day 2005 PPP line): Estimated poverty likelihoods associated with scores

If a household's score is . . .	. . . then the likelihood (%) of being below the poverty line is:
0–4	100.0
5–9	81.4
10–14	56.0
15–19	49.3
20–24	45.8
25–29	27.2
30–34	19.0
35–39	11.6
40–44	6.8
45–49	3.1
50–54	1.4
55–59	0.6
60–64	0.5
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

Figure 6 ($6.25/day 2005 PPP line): Derivation of estimated poverty likelihoods associated with scores

Score	Households below poverty line		All households at score		Poverty likelihood (estimated, %)
0–4	7	÷	7	=	100.0
5–9	86	÷	105	=	81.4
10–14	198	÷	354	=	56.0
15–19	289	÷	585	=	49.3
20–24	551	÷	1,203	=	45.8
25–29	736	÷	2,708	=	27.2
30–34	797	÷	4,187	=	19.0
35–39	871	÷	7,480	=	11.6
40–44	751	÷	11,111	=	6.8
45–49	366	÷	11,923	=	3.1
50–54	210	÷	15,245	=	1.4
55–59	80	÷	14,597	=	0.6
60–64	64	÷	14,146	=	0.5
65–69	8	÷	9,433	=	0.1
70–74	0	÷	3,777	=	0.0
75–79	0	÷	2,961	=	0.0
80–84	0	÷	168	=	0.0
85–89	0	÷	8	=	0.0
90–94	0	÷	0	=	0.0
95–100	0	÷	0	=	0.0

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

Figure 8 ($6.25/day 2005 PPP line): Bootstrapped differences between estimated and true poverty likelihoods for households in a large sample (n = 16,384) from the validation sample, with confidence intervals, scorecard applied to the validation sample

	Difference between estimate and true value			
		Confidence interval (+/− percentage points)		
Score	**Diff.**	**90-percent**	**95-percent**	**99-percent**
0–4	+0.0	0.0	0.0	0.7
5–9	–9.0	11.1	12.8	17.4
10–14	–10.9	13.1	15.8	20.7
15–19	+16.9	9.3	11.0	13.8
20–24	+10.9	7.8	9.0	11.9
25–29	+14.2	2.9	3.4	4.6
30–34	+0.1	3.3	4.0	5.2
35–39	+7.7	0.9	1.1	1.4
40–44	+4.7	0.5	0.6	0.8
45–49	+2.1	0.3	0.3	0.4
50–54	+1.1	0.1	0.1	0.1
55–59	+0.1	0.2	0.2	0.3
60–64	+0.4	0.0	0.0	0.0
65–69	+0.1	0.0	0.0	0.0
70–74	+0.0	0.0	0.0	0.0
75–79	+0.0	0.0	0.0	0.0
80–84	+0.0	0.0	0.0	0.0
85–89	+0.0	0.0	0.0	0.0
90–94	+0.0	0.0	0.0	0.0
95–100	+0.0	0.0	0.0	0.0

Figure 10 ($6.25/day 2005 PPP line): Differences and precision of differences for bootstrapped estimates of poverty rates for groups of households at a point in time, by sample size, scorecard applied to the validation sample

Sample Size n	Difference between estimate and true value			
		Confidence interval (+/− percentage points)		
	Diff.	90-percent	95-percent	99-percent
1	–0.6	36.6	57.8	74.0
4	+0.7	11.1	17.1	33.3
8	+1.1	6.0	9.0	17.6
16	+0.9	3.6	4.9	9.4
32	+0.9	2.1	2.8	4.8
64	+0.9	1.2	1.5	2.3
128	+0.9	0.9	1.1	1.5
256	+0.9	0.6	0.7	0.9
512	+0.9	0.4	0.5	0.6
1,024	+0.9	0.3	0.4	0.5
2,048	+0.9	0.2	0.2	0.3
4,096	+0.9	0.1	0.2	0.2
8,192	+0.9	0.1	0.1	0.2
16,384	+0.9	0.1	0.1	0.1

Figure 12 ($6.25/day 2005 PPP line): Households by targeting classification and score, along with "Total Accuracy" and BPAC, scorecard applied to the validation sample

Score	Inclusion: < poverty line correctly targeted	Undercoverage: < poverty line mistakenly non-targeted	Leakage: => poverty line mistakenly targeted	Exclusion: => poverty line correctly non-targeted	Total Accuracy Inclusion + Exclusion	BPAC See text
0–4	0.0	4.8	0.0	95.2	95.2	–99.7
5–9	0.1	4.7	0.0	95.2	95.3	–95.7
10–14	0.3	4.5	0.1	95.1	95.4	–83.1
15–19	0.6	4.2	0.5	94.7	95.3	–65.9
20–24	1.1	3.7	1.2	94.0	95.1	–31.0
25–29	1.7	3.1	3.2	92.0	93.7	+33.0
30–34	2.7	2.2	6.5	88.7	91.4	–34.8
35–39	3.4	1.4	13.2	82.0	85.4	–174.2
40–44	4.1	0.8	23.7	71.5	75.6	–392.0
45–49	4.4	0.4	35.3	59.9	64.3	–632.6
50–54	4.6	0.2	50.3	44.9	49.5	–944.9
55–59	4.8	0.0	64.7	30.4	35.2	–1,245.0
60–64	4.8	0.0	78.8	16.3	21.1	–1,538.1
65–69	4.8	0.0	88.3	6.9	11.7	–1,733.9
70–74	4.8	0.0	92.0	3.1	8.0	–1,812.4
75–79	4.8	0.0	95.0	0.2	5.0	–1,873.9
80–84	4.8	0.0	95.2	0.0	4.8	–1,877.4
85–89	4.8	0.0	95.2	0.0	4.8	–1,877.5
90–94	4.8	0.0	95.2	0.0	4.8	–1,877.5
95–100	4.8	0.0	95.2	0.0	4.8	–1,877.5

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

Figure 13 ($6.25/day 2005 PPP line): For a given score cut-off, the percentage of all households who are targeted, the percentage of targeted households who are poor, the percentage of poor households who are targeted, and the number of poor households who are successful targeted (inclusion) per non-poor household mistakenly targeted (leakage), scorecard applied to the validation sample

Targeting cut-off	% all households who are targeted	% targeted who are poor	% of poor who are targeted	Poor households targeted per non-poor household targeted
0–4	0.0	99.0	0.2	104.0:1
5–9	0.1	85.6	2.0	6.0:1
10–14	0.5	74.3	7.2	2.9:1
15–19	1.1	56.2	12.3	1.3:1
20–24	2.3	47.4	22.2	0.9:1
25–29	5.0	35.1	36.1	0.5:1
30–34	9.1	29.1	55.3	0.4:1
35–39	16.6	20.6	71.3	0.3:1
40–44	27.7	14.6	84.3	0.2:1
45–49	39.7	11.1	91.4	0.1:1
50–54	54.9	8.4	95.9	0.1:1
55–59	69.5	6.9	99.0	0.1:1
60–64	83.7	5.7	99.9	0.1:1
65–69	93.1	5.2	100.0	0.1:1
70–74	96.9	5.0	100.0	0.1:1
75–79	99.8	4.8	100.0	0.1:1
80–84	100.0	4.8	100.0	0.1:1
85–89	100.0	4.8	100.0	0.1:1
90–94	100.0	4.8	100.0	0.1:1
95–100	100.0	4.8	100.0	0.1:1