Simple Poverty Scorecard® Poverty-Assessment Tool Tajikistan

Mark Schreiner

27 April 2015

This document is at `SimplePovertyScorecard.com`

Abstract

The Simple Poverty Scorecard-brand poverty-assessment tool uses 10 low-cost indicators from Tajikistan's 2007 Living Standards 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 Tajikistan to measure poverty rates, to track changes in poverty rates over time, and to segment clients for differentiated treatment.

Revision note

This document revises one from 21 October 2015. It replaces the indicator "Does the household own and use a color television and a video player or satellite dish?" with "Does the household own and use a color television?" The First MicroFinance Bank of Tajikistan sponsored the revision because the relationship between poverty and ownership of a video player or satellite dish has changed greatly since 2007.

Acknowledgements

This work was funded by Grameen Foundation (GF). Data are from Tajikistan's GosKomStat. Thanks go to Diane Steele, Sasun Tsirunyan, and Matt Walsh. This scorecard was re-branded by GF as a Progress out of Poverty Index® tool. The PPI® is a performance-management tool that GF promotes to help organizations achieve their social objectives more effectively. "Progress out of Poverty Index" and "PPI" are Registered Trademarks of Innovations for Poverty Action. "Simple Poverty Scorecard" is a Registered Trademark of Microfinance Risk Management, L.L.C. for its brand of poverty-assessment tools.

Simple Poverty Scorecard® Poverty-Assessment Tool

			Name	Identifier
Interview ID:				
Interview date:		Participant:		
Country:	TJK	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. Eight or more	0	
	B. Seven	4	
	C. Six	9	
	D. Five	13	
	E. Four	19	
	F. Three	29	
	G. One or two	37	
2. Do all household members 8- to 17-years-old currently attend (or plan to attend) an educational institution in this academic year?	A. No	0	
	B. Yes	1	
	C. No one 8- to 17-years-old	2	
3. What is the highest educational level that the female head/spouse has achieved (do not include incomplete levels)?	A. None, primary (grades 1–4), or basic (grades 1–8(9))	0	
	B. Secondary general (grades 9–10(11))	3	
	C. No female head/spouse	7	
	D. Secondary special, secondary technical, higher education, or graduate school/*aspirantura*	8	
4. How many rooms does the household occupy (excluding any kitchens, balconies, or hallways)?	A. One	0	
	B. Two	6	
	C. Three	8	
	D. Four	11	
	E. Five or more	14	
5. What is the major construction material of the external walls of the residence?	A. Adobe, mud bricks, wood, logs, tin, or other	0	
	B. Mud, stone, baked bricks, or concrete	7	
6. Does the residence have a separate bath/shower?	A. No	0	
	B. Yes	6	
7. Does the household own and use a gas oven, electric oven, or a microwave oven?	A. No	0	
	B. Yes	5	
8. Does the household own and use an electric iron?	A. No	0	
	B. Yes	3	
9. Does the household own and use a color television?	A. No	0	
	B. Yes	7	
10. Does the household currently own any cattle (beef or dairy), or any sheep or goats?	A. No	0	
	B. Only sheep or goats	5	
	C. Only cattle	6	
	D. Both	11	

Score:

Back-page Worksheet:
Household Members, Age, and School Attendance

Write down the name and identifier of the client and of yourself as the enumerator, as well as the service point that the client uses. Record the date of the interview and the date when the client first participated with the organization.

Then read to the respondent: *Please tell me the first name and the age of each member of your household. A* household *is a group of people who live together in a residence as a single unit who share common housekeeping arrangements and a common budget. To be a* household member*, a person must have been present for at least one of the past 12 months. The head of the household is always a household member, even if he/she has been absent for the past 12 months or more.* Write down the first name and the age of each household member, noting who (if anyone) is the female head/spouse. Then write the total number of members in the scorecard header next to "# HH members:", and circle the response to the first scorecard indicator.

For each household member ages 8 to 17, ask: *Does <name> currently attend (or plan to attend) an educational institution in this academic year?* Then circle the response to the second indicator ("A" if some members are 8 to 17 and at least one does not attend school; "B" if some members are 8 to 17 but all attend school; and "C" if no members are ages 8 to 17).

Please keep in mind the full definition of *household* found in the "Guidelines for the Interpretation of Scorecard Indicators".

First name	Age	If <name> is 8- to 17-years-old, does he/she currently attend (or plan to atttend) an educational institution in this academic year?		
1.		Not 8 to 17	No	Yes
2.		Not 8 to 17	No	Yes
3.		Not 8 to 17	No	Yes
4.		Not 8 to 17	No	Yes
5.		Not 8 to 17	No	Yes
6.		Not 8 to 17	No	Yes
7.		Not 8 to 17	No	Yes
8.		Not 8 to 17	No	Yes
9.		Not 8 to 17	No	Yes
10.		Not 8 to 17	No	Yes
11.		Not 8 to 17	No	Yes
12.		Not 8 to 17	No	Yes
13.		Not 8 to 17	No	Yes
14.		Not 8 to 17	No	Yes
# household members:		# "No":		

Look-up table to convert scores to poverty likelihoods

	Poverty likelihood (%)							
	National					Intl. 2005 PPP		
Score	**Food**	**100%**	**150%**	**200%**	**Median**	**$1.25**	**$2.00**	**$2.50**
0–4	48.2	100.0	100.0	100.0	70.9	18.2	34.0	54.3
5–9	48.2	100.0	100.0	100.0	70.9	18.2	34.0	54.3
10–14	48.2	100.0	100.0	100.0	68.1	13.6	33.4	54.3
15–19	48.2	100.0	100.0	100.0	62.1	3.6	29.2	54.3
20–24	39.0	84.8	99.3	100.0	54.5	3.6	28.4	44.7
25–29	37.8	77.9	96.7	99.5	48.7	2.1	20.0	43.7
30–34	28.3	73.3	95.0	98.9	42.8	2.0	16.4	34.5
35–39	19.2	63.3	91.4	98.1	32.0	0.7	10.3	24.8
40–44	13.3	51.5	85.4	95.7	23.1	0.7	7.5	17.4
45–49	9.7	45.3	81.8	93.0	15.7	0.7	4.4	11.1
50–54	8.2	37.5	72.6	88.0	13.5	0.5	3.6	11.1
55–59	3.9	23.4	56.4	79.6	8.6	0.3	2.5	6.5
60–64	3.6	17.1	47.0	75.5	3.9	0.2	2.5	3.9
65–69	3.3	13.2	39.8	66.6	3.7	0.0	2.4	3.6
70–74	1.3	12.8	29.7	46.4	1.9	0.0	1.3	1.6
75–79	0.7	4.8	23.1	38.1	1.0	0.0	0.7	1.0
80–84	0.0	0.0	22.4	38.1	0.0	0.0	0.0	0.0
85–89	0.0	0.0	0.0	8.7	0.0	0.0	0.0	0.0
90–94	0.0	0.0	0.0	8.7	0.0	0.0	0.0	0.0
95–100	0.0	0.0	0.0	8.7	0.0	0.0	0.0	0.0

Simple Poverty Scorecard® Poverty-Assessment Tool Tajikistan

1. Introduction

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

The direct approach to poverty measurement via consumption surveys is difficult and costly. As a case in point, Tajikistan's 2007 Living Standards Survey (TLSS) runs 86 pages. Enumerators completed interviews at a rate of about two households per day, asking—in addition to hundreds of non- consumption items—about 200 holds per day, asking—in addition to hundreds of non-consumption items. Many holds per day, asking—in addition to hundreds of non-consumption items have a number of sub-items, for example: "Did your household consume onions during the last seven days? What was the total amount of onions consumed? How much of the consumed onions was purchased? How much of the consumed onions was produced by the household? How much of the consumed onions was received as a gift or humanitarian aid? How much of the consumed onions was received as part of salary/business? How much of the consumed onions was taken from stocks? What is your estimate of the total value in

current prices of the consumed onions that was not purchased in the last seven days? Now then, did your household consume any garlic in the last seven days? . . ."

In comparison, the indirect approach via the scorecard is simple, quick, and inexpensive. It uses ten verifiable indicators (such as "Does the household own and use an electric iron?" and "Does the household currently own any cattle (beef or dairy), or any sheep or goats?") to get a score that is highly correlated with poverty status as measured by the exhaustive TLSS survey.

The scorecard differs from "proxy-means tests" (Coady, Grosh, and Hoddinott, 2004) in that it is transparent, it is freely available,[1] and 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 blunt (such as rules based on land-ownership or housing quality) or subjective and relative (such as participatory wealth ranking facilitated by skilled field workers). Estimates from these approaches may be costly, their accuracy is unknown, and they are not comparable across places, organizations, nor time.

The scorecard can be used to measure the share of a program's participants who are below a given poverty line, for example, the international "very poor" benchmark of $1.25/day per person at 2005 purchase-power parity (PPP). USAID microenterprise partners in Tajikistan can use scoring with the median poverty line to report how many

[1] The Simple Poverty Scorecard tool for Tajikistan is not, however, in the public domain. Copyright is held by the sponsor and by Microfinance Risk Management, L.L.C.

of their participants are "very poor" by USAID's definition.[2] Scoring can also be used to measure net movement across a poverty line over time. In all these cases, the scorecard provides a consumption-based, objective tool with known accuracy. While consumption surveys are costly even for governments, some local pro-poor organizations may be able to implement an inexpensive poverty-assessment tool to help with poverty monitoring and (if desired) targeting and segmenting participants for differentiated servcies.

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, then 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 they do not work, but because they are presented (when they are presented at all) as tables of regression coefficients incomprehensible to non-specialists (with cryptic indicator names such as "`LGHHSZ_2`" and with points with negative values and many decimal places). Thanks to the predictive-modeling phenomenon known as the "flat maximum", simple, transparent scoring approaches can be about as accurate as complex, opaque ones (Schreiner, 2012a; Caire and Schreiner, 2012).

[2] USAID defines a household as *very poor* if its daily per-capita consumption is less than the highest of $1.25/day (TJS1.55 on average, Figure 1) or the median line that divides people in households below the national line into two equal-size groups (TJS3.37).

Beyond its simplicity and transparency, the scorecard's technical approach 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 the 2007 TLSS done by Tajikistan's GosKomStat). 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
- Cover a variety of types of indicators
- Applicable in all regions of Tajikistan

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

The scorecard can be used to estimate three basic quantities. First, it can estimate a particular household's *poverty likelihood*, that is, the probability that the household has per-capita 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 estimate is the average poverty likelihood among the households in the group.

Third, the scorecard can estimate changes in the poverty rate for a group of households (or for two independent samples of households, both of which are

representative of the same population) between two points in time. This estimate is the baseline/follow-up change in the average poverty likelihood of the group(s).

The scorecard can also be used to target participants or to segment them for differential treatment. To help managers choose an appropriate targeting/segmenting cut-off for their purposes, this paper reports several measures of targeting accuracy for a range of possible cut-offs.

The scorecard's indicators and points are derived from household consumption data and Tajikistan's national poverty line. Scores from this one scorecard are calibrated to poverty likelihoods for eight poverty lines.

The scorecard is constructed and calibrated using half of the data from the 2007 TLSS. The other half is used to validate the scorecard's accuracy for estimating households' poverty likelihoods, for estimating groups' poverty rates at a point in time, and for targeting/segmentation.

Scoring's estimators of poverty rates and of changes in poverty rates are *unbiased*. That is, they match the true value on average in repeated samples when constructed from (and applied to) a single, unchanging population. Like all predictive models, the scorecard here is constructed from a single sample and so misses the mark to some unknown extent when applied to a different population or when applied after 2007.[3]

[3] Important examples include nationally representative samples at a later point in time or sub-groups that are not nationally representative (Tarozzi and Deaton, 2009).

Thus, while the indirect scoring approach is less costly than the direct survey approach, it is also biased when applied in practice. (The survey approach is unbiased by definition.) There is bias because the scorecard necessarily assumes that the future relationships between indicators and poverty in all possible groups of households will be the same as in the construction data. Of course, this assumption—inevitable in predictive modeling—holds only partly.

On average when applied to the validation sample with 1,000 bootstraps of $n =$ 16,384, the difference between scorecard estimates of groups' poverty rates and the true rates at a point in time for the national poverty line is –0.6 percentage points. The average absolute difference across all eight poverty lines is 0.4 percentage points, and the maximum absolute difference for any poverty line is 0.9 percentage points. These differences are due to sampling variation, not bias; the average difference would be zero if the whole 2007 TLSS were to be repeatedly re-fielded and divided into sub-samples before repeating the entire process of constructing and validating scorecards.

The 90-percent confidence intervals for these estimates are ±0.6 percentage points or less. For $n =$ 1,024, the 90-percent intervals are ±2.4 percentage points or less.

Section 2 below documents data and poverty lines. Sections 3 and 4 describe scorecard construction and offer guidelines for use in practice. Sections 5 and 6 tell how to estimate households' poverty likelihoods and groups' poverty rates at a point in time. Section 7 discusses estimating changes in poverty rates over time, and Section 8 covers targeting/segmentation. Section 9 places the scorecard here in the context of related exercises for Tajikistan. The last section is a summary.

2. Data and poverty lines

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

2.1 Data

The scorecard is based on data from the 4,860 households in the 2007 TLSS. This is Tajikistan's most recent national consumption survey that includes all types of indicators typically used in poverty-assessment tools.[4]

The 2007 TLSS was fielded in two rounds, from 7 September to 13 October 2007, and from 29 October to 10 November. The data provided by the World Bank[5] uses non-food consumption from the first round and food consumption collected in the second round (because the first round partly overlapped with Ramadan). Non- consumption items (including the measure of household size on which measures of per-capita consumption are based) come from the first round. All data for households in Sughd division is from the second round. All consumption figures here are in average prices during the second round.

[4] The 2009 TLSS interviewed 4,490 of the 4,860 households in the 2007 TLSS, but the TLSS 2009 is not used here because it did not ask about characteristics of the residence.
[5] `microdata.worldbank.org/index.php/catalog/72`, retrieved 19 October 2014.

For the purposes of the scorecard, the households in the 2007 TLSS are randomly divided into two sub-samples:

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

2.2 Poverty rates

A *poverty rate* is the share of units in households in which total household consumption (divided by the number of household members) is below a given poverty line. The unit is either the household itself or a person in the household. Each household member is defined to have the same poverty status (or estimated poverty likelihood) as does the household as a whole.

Suppose a program serves two households. The first household is poor (its per-capita consumption is less than a given poverty line), and it has three members, one of whom is a program participant. The second household is non-poor, and it has four members, two of whom are program participants.

Poverty rates are at the level of either households or people. If the program defines its *participants* as households, then the household level is relevant. The estimated household-level poverty rate is the equal-weighted average of poverty statuses (or estimated poverty likelihoods) across participants' households. In the example here, this is $\frac{1 \cdot 1 + 1 \cdot 0}{1 + 1} = \frac{1}{2} = 0.5 = 50$ percent. In the "$1 \cdot 1$" term in the numerator, the first "1" is the first household's weight, and the second "1" is the first household's poverty

status (poor). In the "$1 \cdot 0$" term in the numerator, the "1" is the second household's weight, and the "0" is the second household's poverty status (non-poor). The "$1 + 1$" in the denominator is the sum of the weights of the two households. Each household has a weight of one (1) because the unit of analysis is the household.

Alternatively, a person-level rate is relevant if a program defines all people in households that benefit from its services as *participants*. In the example here, the person-level rate is the household-size-weighted average of poverty statuses for households with participants, or $\frac{3 \cdot 1 + 4 \cdot 0}{3 + 4} = \frac{3}{7} = 0.43 = 43$ percent. In the "$3 \cdot 1$" term in the numerator, the "3" is the first household's weight because it has three members, and the "1" is its poverty status (poor). In the "$4 \cdot 0$" term in the numerator, the "4" is the second household's weight because it has four members, and the zero is its poverty status (non-poor). The "$3 + 4$" in the denominator is the sum of the weights of the two households. A household's weight is its number of members because the unit of analysis is the household member.

As a final example, a program counts as *participants* only those household members with whom it deals with directly. For the example here, this means that some—but not all—household members are counted. The person-level rate is now the participant-weighted average of the poverty statuses of households with participants, or $\frac{1 \cdot 1 + 2 \cdot 0}{1 + 2} = \frac{1}{3} = 0.33 = 33$ percent. The first "1" in the "$1 \cdot 1$" in the numerator is the first household's weight because it has one participant, and the second "1" is its poverty

status (poor). In the "$2 \cdot 0$" term in the numerator, the "2" is the second household's weight because it has two participants, and the zero is its poverty status (non-poor). The "$1 + 2$" in the denominator is the sum of the weights of the two households. Each household's weight is its number of participants because the unit of analysis is the participant.

To sum up, estimated poverty rates are weighted averages of households' poverty statuses (or estimated poverty likelihoods), where the weights are the number of relevant units in the household. When reporting, programs should explain who is counted as a *participant* and why.

Figure 1 reports poverty rates for eight poverty lines at the levels of households and people for Tajikistan as a whole in 2007 and for the construction and validation samples. Person-level poverty rates are included in Figure 1 because these are the rates reported by governments and used in most policy discussions. Household-level poverty rates are also reported because—as shown above—household-level poverty likelihoods can be straightforwardly converted into poverty rates for other units of analysis. This is also why the scorecard is constructed, calibrated, and validated with household weights.

2.3 Poverty lines

According to Steele (2008), the derivation of Tajikistan's national poverty line (sometimes called here "100% of the national line") follows the "cost-of-basic-needs" method of Ravaillon (1998). It begins with a food-poverty line defined as the cost—adjusted for differences in food prices across nine regions—of an 59-item food basket with 2,250 Calories. The reference group for the food basket and its regional prices are households in the 2007 TLSS in the 30^{th} to 50^{th} percentiles of total per-capita consumption. The average food line for Tajikistan as a whole is TJS2.92 per person per day, giving poverty rates of 14.6 percent (households) and 17.1 percent (people, Figure 1).

The national poverty line is then defined as this food line, plus the average non-food consumption observed in the 2007 TLSS for households whose food consumption is between 100 and 120 percent of the food line. This non-food component, like the food component, is adjusted for regional differences in food prices (not non-food prices). This national (food-plus-non-food) poverty line is TJS4.56 per person per day, giving poverty rates of 47.7 percent (households) and 53.5 percent (people, Figure 1). The person-level poverty rates for the food line and the national line here match those in State Committee on Statistics (2009, p. 10) and World Bank (2009, p. 4).

The scorecard is constructed using the national poverty line. Because local, pro-poor programs in Tajikistan may want to use different or various poverty lines, this paper calibrates scores from its single scorecard to poverty likelihoods for eight poverty lines:

- Food
- 100% of national
- 150% of national
- 200% of national
- Median
- $1.25/day 2005 PPP
- $2.00/day 2005 PPP
- $2.50/day 2005 PPP

The median line is defined (for each of the nine regions separately) as the median per-capita consumption of people (not households) who are below 100% of the national line (United States Congress, 2004).

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

- 2005 PPP exchange rate of TJS0.927 per $1.00 (World Bank, 2008)
- Consumer Price Index for Tajikistan:[6]
  - — Average in 2005: 124.5684
  - — Average from 29 October to 10 November 2007: 166.8277
- Average all-Tajikistan national line (Figure 1): TJS4.56
- The value of the national line in each of nine poverty-line regions (Figure 2)

[6] Appendix G of Steele (2008), and `tradingeconomics.com/tajikistan/consumer-price-index-cpi`, retrieved 5 November 2013.

Using the formula from Sillers (2006), the all-Tajikistan $1.25/day 2005 PPP line is:

$$(\text{2005 PPP exchange rate}) \cdot \$1.25 \cdot \left(\frac{\text{CPI}_{\text{29oct to 10nov2007}}}{\text{CPI}_{\text{2005}}} \right) =$$

$$\left(\frac{\text{TJS}0.927}{\$1.00} \right) \cdot \$1.25 \cdot \left(\frac{166.8277}{124.5684} \right) = \text{TJS}1.55.$$

This line applies to Tajikistan on average. In a given poverty-line region, the $1.25/day line is the all-Tajikistan $1.25/day line of TJS1.55, multiplied by the national line in that poverty-line region (for example, TJS5.01 for Dushanbe, Figure 2), and then divided by Tajikistan's average national line of TJS4.56 (giving, in the example of Dushanbe, 1.55 x 5.01 ÷ 4.56 = TJS1.71).

For Tajikistan overall, the person-level poverty rate for the $1.25/day 2005 PPP poverty line is 1.0 percent (Figure 1).[7]

USAID microenterprise partners in Tajikistan who use the scorecard to report poverty rates to USAID should use the median line. This is because USAID defines the "very poor" as those people in households whose per-capita consumption is below the highest of two lines (Figure 1):

- $1.25/day 2005 PPP (TJS1.55)
- Median line (TJS3.37).

[7] The person-level poverty rate for the 2007 TLSS on the World Bank's PovCalNet (`iresearch.worldbank.org/PovcalNet/index.htm`, retrieved 19 October 2014) is 12.2 percent, far from 1.0 percent. The World Bank has been alerted to the likely error, and it bears the burden of explanation. It is not uncommon for PovCalNet to be mistaken, and—unlike this paper—PovCalNet does not report the version of the data that it uses, the $1.25/day line in local currency, nor how (or whether) it adjusts for price differences over time and across regions.

3. Scorecard construction

For Tajikistan, about 140 candidate indicators are initially prepared in the areas of:

- Household composition (such as number of members)
- Education (such as school attendance)
- Housing (such as the number of rooms and the type of exterior walls)
- Ownership of durable assets (such as color televisions or ovens)
- Employment (such as the number of household members who work)
- Agriculture (such as ownership of livestock)

One aim of the scorecard is to measure *changes* in poverty through time. This means that, when selecting indicators and holding other considerations constant, preference is given to more sensitive indicators. For example, the ownership of an oven or a color television 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. The first step is to use Logit to build one scorecard for each candidate indicator. Each scorecard's power to rank households by poverty status is measured as "c" (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,

variety among indicators, applicability across regions, relevance for distinguishing among households at the poorer end of the distribution of consumption, and verifiability.

A series of two-indicator scorecards are then built, each based on the one-indicator scorecard selected from the first round, with a second candidate indicator added. The best two-indicator scorecard is then selected, again using judgment to balance "c" with the non-statistical criteria. These steps are repeated until the scorecard has 10 indicators that work well together.

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

This algorithm is similar to the common R^2-based stepwise least-squares regression. It differs from naïve stepwise in that the selection of indicators considers both statistical[8] and non-statistical criteria. The non-statistical criteria can improve robustness through time and help ensure that indicators are simple, sensible, and acceptable to users.

[8] The statistical criterion for selecting an indicator is not the p value of its coefficient but rather its contribution to the ranking of households by poverty status.

The single scorecard here applies to all of Tajikistan. Tests for Indonesia (World Bank, 2012), Bangladesh (Sharif, 2009), India and Mexico (Schreiner, 2006 and 2005a), 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 accuracy much. In general, segmentation of poverty-assessment tools by sub-national groups or regions may improve the bias and precision of estimates of poverty rates (Tarozzi and Deaton, 2009) at the risk of overfitting (Haslett, 2012).

4. Practical guidelines for scorecard use

The main challenge of scorecard design is not to maximize statistical accuracy but rather to improve the chances that the scorecard is actually used (Schreiner, 2005b). When scoring projects fail, the reason is not usually statistical inaccuracy but rather the failure of an organization to decide to do what is needed to integrate scoring in its processes and to train and convince its employees to use the scorecard properly (Schreiner, 2002). After all, most reasonable scorecards have similar targeting accuracy, thanks to the empirical phenomenon known as the "flat maximum" (Caire and Schreiner, 2012; 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 on their own and use it properly. Of course, accuracy matters, but it must be balanced with 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, Tajikistan's scorecard fits on one page. The construction process, indicators, and points are simple and transparent. Additional work is minimized; non-specialists can compute scores by hand in the field because the scorecard has:

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

The scorecard (and its back-page worksheet) is ready to be photocopied.

A field worker using Tajikistan's paper scorecard would:

- Record the names and identifiers of the participant, of the field worker, and of the relevant organizational service point
- Record the date that the participant first participated with the organization
- Record the date of the scorecard interview
- Complete the back-page worksheet with each household member's name, age, and school attendance
- Record household size in the scorecard header, and record the responses to the scorecard's first and second indicators based on the back-page worksheet
- Read each of the remaining eight questions one-by-one from the scorecard, drawing a circle around the relevant responses and their points, and writing each point value in the far right-hand column
- Add up the points to get a total score
- Implement targeting policy (if any)
- Deliver the paper scorecard to a central office for data entry and filing

Of course, field workers must be trained. The quality of outputs depends on the quality of inputs. If organizations or field workers gather their own data and believe that they have an incentive to exaggerate poverty rates (for example, if funders reward them for higher poverty rates), then it is wise to do on-going quality control via data

review and random audits (Matul and Kline, 2003).[9] IRIS Center (2007a) and Toohig (2008) are useful nuts-and-bolts guides for budgeting, training field workers and supervisors, logistics, sampling, interviewing, piloting, recording data, and controlling quality.

In particular, while collecting scorecard indicators is relatively easier than alternative ways of measuring consumption-based poverty, it is still absolutely difficult. Training and explicit definitions of terms and concepts in the scorecard are essential, and field workers should scrupulously study and follow the "Guidelines for the Interpretation of Indicators" found at the end of this paper as well as the "Back-page worksheet", as they are an integral part of the Simple Poverty Scorecard tool.[10]

For the example of Nigeria, one study (Onwujekwe, Hanson, and Fox-Rushby, 2006) found distressingly low inter-rater and test-retest correlations for indicators as seemingly simple as whether the household owns an automobile. At the same time, Grosh and Baker (1995) suggest that gross underreporting of assets does not affect targeting. For the first stage of targeting in a conditional cash-transfer program in

[9] If a program does not want field workers to know the points associated with responses, then it can use a version of the scorecard that does not display the points and then apply the points and compute scores later at a central office. Schreiner (2012b) argues that hiding points in Colombia (Camacho and Conover, 2011) did little to deter cheating and that, in any case, cheating by the user's central office was more damaging than cheating by field workers and respondents. Even if points are hidden, field workers and respondents can apply common sense to guess which response options are linked with greater poverty.

[10] The guidelines here are the only ones that field workers should receive. All other issues of interpretation should be left to the judgment of field workers and respondents, as this seems to be what Tajikistan's GosKomStat did when it fielded the 2007 TLSS.

Mexico, Martinelli and Parker (2007, pp. 24–25) 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". Still, as is done in Mexico in the second stage of its targeting process, most false self-reports can be corrected (or avoided in the first place) by field workers who make a home visit. This is the recommended procedure for local, pro-poor organizations in Tajikistan.

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

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

In general, the sampling design should follow from the organization's goals for the exercise, the questions to be answered, and the budget. The main goal should be to make sure that the sample is representative of a well-defined population and that the scorecard will inform an issue that matters to the organization.

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

- Employees of the organization
- Third parties

Responses, scores, and poverty likelihoods can be recorded on:

- Paper in the field, and then filed at a central office
- Paper in the field, and then keyed into a database or spreadsheet at a central office
- Portable electronic devices in the field, and then uploaded to a database

Given a population of participants relevant for a particular business question, the participants to be scored can be:

- All relevant participants (a census)
- A representative sample of relevant participants
- All relevant participants in a representative sample of relevant field offices
- A representative sample of relevant participants in a representative sample of relevant field offices

If not determined by other factors, the number of participants to be scored can be derived from sample-size formulas (presented later) to achieve a desired confidence level and a desired confidence interval. To be clear, however, the focus should not be on having a sample size large enough to achieve some arbitrary level of statistical significance but rather to get a representative sample from a well-defined population so that the analysis of the results can meaningfully inform questions that matter to the organization.

Frequency of application can be:

- As a once-off project (precluding measuring change)
- Every two years (or at any other fixed or variable time interval, allowing measuring change)
- Each time a field worker visits a participant at home (allowing measuring change)

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

- With two independent samples of participants from the same population
- With a single set of participants

An example set of choices is illustrated by BRAC and ASA, two microfinance organizations in Bangladesh who each have about 7 million participants and who apply the Simple Poverty Scorecard tool for Bangladesh (Schreiner, 2013a) with a sample of about 25,000. Their design is that all loan officers in a random sample of branches score all participants each time they visit a homestead (about once a year) as part of their standard due diligence prior to loan disbursement. They record responses on paper in the field before sending the forms to a central office to be entered into a database and converted to poverty likelihoods.

5. Estimates of household poverty likelihoods

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

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 100% of the national line, scores of 35–39 have a poverty likelihood of 63.3 percent, and scores of 30–34 have a poverty likelihood of 73.3 percent (Figure 3).

The poverty likelihood associated with a score varies by poverty line. For example, scores of 35–39 are associated with a poverty likelihood of 63.3 percent for the national line but of 0.7 percent for the $1.25/day 2005 PPP line.[11]

5.1 Calibrating scores with poverty likelihoods

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

[11] Starting with Figure 3, many figures have eight versions, one for each of the eight poverty lines. To keep them straight, they are grouped by poverty line. Single tables pertaining to all eight lines are placed with the tables for 100% of the national line.

For the example of the national line (Figure 4), there are 13,988 (normalized) households in the calibration sub-sample with a score of 35–39. Of these, 8,852 (normalized) are below the poverty line. The estimated poverty likelihood associated with a score of 35–39 is then 63.3 percent, because 8,852 ÷ 13,988 = 63.3 percent.

To illustrate with the national line and a score of 30–34, there are 9,213 (normalized) households in the calibration sample, of whom 6,757 (normalized) are below the line (Figure 4). The poverty likelihood for this score range is then 6,757 ÷ 9,213 = 73.3 percent.

The same method is used to calibrate scores with estimated poverty likelihoods for the other seven poverty lines.[12]

Figure 5 shows, for all scores, the likelihood that a given household's per-capita consumption falls in a range demarcated by two adjacent poverty lines.

[12] To ensure that poverty likelihoods never increase as scores increase, likelihoods across series of adjacent scores are sometimes iteratively averaged before grouping scores into ranges. This preserves unbiasedness, and it keeps users from balking when sampling variation in score ranges with few households would otherwise lead to higher scores being linked with higher poverty likelihoods.

For example, the probability that a household with a score of 35–39 falls between two adjacent poverty lines is:

- 0.7 percent below $1.25/day
- 9.6 percent between $1.25/day and $2.00/day
- 9.0 percent between $2.00/day and the food line
- 5.6 percent between the food line and $2.50/day
- 7.2 percent between $2.50/day and the median line
- 31.3 percent between the median line and 100% of the national line
- 28.1 percent between 100% and 150% of the national line
- 6.7 percent between 150% and 200% of the national line
- 1.9 percent above 200% of the national line

Even though the scorecard is constructed partly based on judgment related to non-statistical criteria, the calibration process produces poverty likelihoods that are objective, that is, derived from quantitative poverty lines and from survey data on consumption. The calibrated poverty likelihoods would be objective even if the process of selecting indicators and points did not use any data at all. In fact, objective scorecards of proven accuracy are often constructed using only expert judgment to select indicators and points (Fuller, 2006; Caire, 2004; Schreiner *et al.*, 2004). Of course, the scorecard here is constructed with both data and judgment. The fact that this paper acknowledges that some choices in scorecard construction—as in any statistical analysis—are informed by judgment in no way impugns the objectivity of the poverty likelihoods, as this depends on using data in score calibration, not on using data (and nothing else) in scorecard construction.

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

5.2 Accuracy of estimates of households' poverty likelihoods

As long as the relationships between indicators and poverty do not change over time, and as long as the scorecard is applied to households that are representative of the same population from which the scorecard was originally constructed, then 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 value. The scorecard also produces unbiased estimates of poverty rates at a point in time and unbiased estimates of changes in poverty rates between two points in time.[13]

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

Of course, the relationships between indicators and poverty do change to some unknown extent over time and also across sub-groups in Tajikistan's population. Thus, the scorecard will generally be biased when applied after November 2007 (the last month of fieldwork for the 2007 TLSS) or when applied with sub-groups that are not nationally representative.

How accurate are estimates of households' poverty likelihoods, given the assumption of unchanging relationships between indicators and poverty over time and the assumption of a sample that is representative of Tajikistan as a whole? To find out, the scorecard is applied to 1,000 bootstrap samples of size $n = 16{,}384$ from the validation sample. Bootstrapping means to:

- Score each household in the validation sample
- Draw a bootstrap sample *with replacement* from the validation sample
- For each score, compute the true poverty likelihood in the bootstrap sample, that is, the share of households with the score and with consumption below a poverty line
- For each score, record the difference between the estimated poverty likelihood (Figure 3) 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 intervals containing the central 900, 950, and 990 differences between estimated and true poverty likelihoods

For each score range and for $n = 16{,}384$, Figure 6 shows the average difference between estimated and true poverty likelihoods as well as confidence intervals for the differences.

For the national line, the average poverty likelihood across bootstrap samples for scores of 35–39 in the validation sample is too low by 2.7 percentage points. For scores of 30–34, the estimate is too low by 1.5 percentage points.[14]

The 90-percent confidence interval for the differences for scores of 35–39 is ±2.2 percentage points (national line, Figure 6). This means that in 900 of 1,000 bootstraps, the difference between the estimate and the true value is between –4.9 and –0.5 percentage points (because –2.7 – 2.2 = –4.9, and –2.7 + 2.2 = –0.5). In 950 of 1,000 bootstraps (95 percent), the difference is –2.7 ± 2.4 percentage points, and in 990 of 1,000 bootstraps (99 percent), the difference is –2.7 ± 2.8 percentage points.

A few differences between estimated poverty likelihoods and true values in Figure 6 are large. There are differences because the validation sample is a single sample that—thanks to sampling variation—differs in distribution from the construction/calibration sub-samples and from Tajikistan's population. For targeting, however, what matters is less the difference in all score ranges and more the difference in score ranges just above and below the targeting cut-off. This mitigates the effects of bias and sampling variation on targeting (Friedman, 1997). Section 8 below looks at targeting accuracy in detail.

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

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

Another possible source of differences between estimates and true values is overfitting. The scorecard here is unbiased, but it may still be *overfit* when applied after the end of the TLSS fieldwork in November 2007. That is, it may fit the data from the 2007 TLSS so closely that it captures not only some real patterns but also some random patterns that, due to sampling variation, show up only in the 2007 TLSS but not in the overall population of Tajikistan. Or the scorecard may be overfit in the sense that it is not robust when relationships between indicators and poverty change over time or when the scorecard is applied to non-nationally representative samples.

Overfitting can be mitigated by simplifying the scorecard and by not relying only on data but rather also considering theory, experience, and judgment. Of course, the scorecard here does this. Combining scorecards can also reduce overfitting, at the cost of greater complexity.

Most errors in individual households' likelihoods do balance out in the estimates of groups' poverty rates (see the next section). Furthermore, at least some of the differences will come from non-scorecard sources such as changes in the relationships between indicators and poverty, sampling variation, changes in poverty lines, inconsistencies in data quality across time, and imperfections in cost-of-living adjustments across time and across geographic regions. These factors can be addressed

only by improving data quantity and quality (which is beyond the scope of the scorecard) or by reducing overfitting (which likely has limited returns, given the scorecard's parsimony).

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

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

To illustrate, suppose an organization samples three households on 1 January 2015 and that they have scores of 20, 30, and 40, corresponding to poverty likelihoods of 84.8, 73.3, and 51.5 percent (national line, Figure 3). The group's estimated poverty rate is the households' average poverty likelihood of $(84.8 + 73.3 + 51.5) \div 3 = 69.9$ percent.

Be careful; the group's poverty rate is *not* the poverty likelihood associated with the average score. Here, the average score is 30, which corresponds to a poverty likelihood of 73.3 percent. This differs from the 69.9 percent found as the average of the three individual poverty likelihoods associated with each of the three scores. Unlike poverty likelihoods, scores are ordinal symbols, like letters in the alphabet or colors in the spectrum. Because scores are not cardinal numbers, they cannot be added up or averaged across households. Only three operations are valid for scores: conversion to poverty likelihoods, analysis of distributions (Schreiner, 2012a), or comparison—if desired—with a cut-off for targeting/segmentation. The safest rule to follow is: Always use poverty likelihoods, never scores.

6.1 Accuracy of estimated poverty rates at a point in time

For the Tajikistan scorecard applied to 1,000 bootstraps of $n = 16{,}384$ from the validation sample, the maximum absolute difference between the estimated poverty rate at a point in time and the true rate is 0.9 percentage points (Figure 8, summarizing Figure 7 across all eight poverty lines). The average absolute difference across poverty lines is 0.4 percentage points. At least part of these differences is due to sampling variation in the division of the 2007 TLSS into two sub-samples as well as to degradation of the bootstrap with unequally weighted observations.

When estimating poverty rates at a point in time, the bias reported in Figure 8 should be subtracted from the average poverty likelihood to make the estimate unbiased. For the Tajikistan scorecard and the national line, bias is –0.6 percentage points (Figure 8), so the unbiased estimate in the three-household example above is 69.9 – (–0.6) = 70.5 percent.

In terms of precision, the 90-percent confidence interval for a group's estimated poverty rate at a point in time with $n = 16{,}384$ is ±0.6 percentage points or better (Figure 8). This means that in 900 of 1,000 bootstraps of this size, the estimate (after subtracting off bias) is within 0.6 percentage points of the true value.

For example, suppose that the average poverty likelihood in a sample of $n =$ 16,384 with the Tajikistan scorecard and the national line is 69.9 percent. Then estimates in 90 percent of such samples would be expected to fall in the range of 69.9 – (–0.6) – 0.6 = 69.9 percent to 69.9 – (–0.6) + 0.6 = 71.1 percent, with the most likely

true value being the unbiased estimate in the middle of this range (69.9 – (–0.6) = 70.5 percent). This is because the original (biased) estimate is 69.9 percent, bias is –0.6 percentage points, and the 90-percent confidence interval for the national line and this sample size is ±0.6 percentage points (Figure 8).

6.2 Formula for standard errors for estimates of poverty rates

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

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

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

z is from the Normal distribution and is $\begin{cases} 1.04 \text{ for confidence levels of 70 percent} \\ 1.28 \text{ for confidence levels of 80 percent} \\ 1.64 \text{ for confidence levels of 90 percent} \end{cases}$,

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

$\hat{p}$ is the estimated proportion of households below the poverty line in the sample,

ϕ is the finite population correction factor $\sqrt{\frac{N - n}{N - 1}}$,

N is the population size, and

n is the sample size.

For example, Tajikistan's 2007 TLSS gives a direct-measurement estimate of the household-level poverty rate for the national line of $\hat{p} = 47.7$ percent (Figure 1). If this

estimate came from a sample of $n = 16{,}384$ households from a population N of 1,117,949 (the number of households in Tajikistan in 2007), then the finite population correction ϕ is $\sqrt{\frac{1{,}117{,}949 - 16{,}384}{1{,}117{,}949 - 1}} = 0.9926$, which can be taken as $\phi = 1$. If the desired confidence level is 90-percent ($z = 1.64$), then the confidence interval $\pm c$ is

$$\pm\, z \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} \cdot \sqrt{\frac{N - n}{N - 1}} = \pm 1.64 \cdot \sqrt{\frac{0.477 \cdot (1 - 0.477)}{16{,}384}} \cdot 1 = \pm 0.640 \text{ percentage points.}$$

The scorecard, however, does not measure poverty directly, so this formula is not applicable. To derive a formula for the Tajikistan scorecard, consider Figure 7, which reports empirical confidence intervals $\pm c$ for the differences for the scorecard applied to 1,000 bootstraps of various sizes from the validation sample. For example, with $n =$ 16,384 and the national line, the 90-percent confidence interval is ± 0.595 percentage points.[15]

Thus, the 90-percent confidence interval with $n = 16{,}384$ is ± 0.595 percentage points for the Tajikistan scorecard and ± 0.640 percentage points for direct measurement. The ratio of the two intervals is $0.595 \div 0.640 = 0.93$.

[15] Due to rounding, Figure 7 displays 0.6, not 0.595.

Now consider the same exercise, but with $n = 8{,}192$. The confidence interval under direct measurement and the national line is $\pm 1.64 \cdot \sqrt{\frac{0.477 \cdot (1 - 0.477)}{8{,}192}} \cdot 1 =$ ± 0.905 percentage points. The empirical confidence interval with the Tajikistan scorecard (Figure 7) is ± 0.835 percentage points. Thus for $n = 8{,}192$, the ratio of the two intervals is $0.835 \div 0.905 = 0.92$.

This ratio of 0.92 for $n = 8{,}192$ is close to the ratio of 0.93 for $n = 16{,}384$. Across all sample sizes of 256 or more in Figure 7, the average ratio turns out to be 0.95, implying that confidence intervals for indirect estimates of poverty rates via the Tajikistan scorecard and the national poverty line are—for a given sample size—about 5-percent narrower than confidence intervals for direct estimates via the 2007 TLSS. This 0.95 appears in Figure 8 as the "α factor" because if $\alpha = 0.95$, then the formula for confidence intervals c for the Tajikistan scorecard is $\pm c = \pm z \cdot \alpha \cdot \sigma$. That is, the formula for the standard error σ for point-in-time estimates of poverty rates via scoring is $\alpha \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} \cdot \sqrt{\frac{N - n}{N - 1}}$.

In general, α can be more or less than 1.00. When α is less than 1.00, it means that the scorecard is more precise than direct measurement. This is the cases for three of the eight poverty lines in Figure 8.

The formula relating confidence intervals with standard errors for the scorecard can be rearranged to give a formula for determining sample size before measurement. If $\widetilde{p}$ is the expected poverty rate before measurement, then the formula for sample size n

from a population of size N that is based on the desired confidence level that corresponds to z and the desired confidence interval $\pm c$ is

$n = N \cdot \left(\frac{z^2 \cdot \alpha^2 \cdot \widetilde{p} \cdot (1 - \widetilde{p})}{z^2 \cdot \alpha^2 \cdot \widetilde{p} \cdot (1 - \widetilde{p}) + c^2 \cdot (N - 1)} \right)$. If the population N is "large" relative to the sample size n, then the finite population correction factor ϕ can be taken as one (1), and the formula becomes $n = \left(\frac{\alpha \cdot z}{c} \right)^2 \cdot \widetilde{p} \cdot (1 - \widetilde{p})$.

To illustrate how to use this, suppose the population N is 1,117,949 (the number of households in Tajikistan in 2007), suppose $c = 0.05031$, $z = 1.64$ (90-percent confidence), and the relevant poverty line is the national line so that the most sensible expected poverty rate $\widetilde{p}$ is Tajikistan's overall poverty rate for that line in 2007 (47.7 percent at the household level, Figure 1). The α factor is 0.95 (Figure 8). Then the sample-size formula gives

$n = 1{,}117{,}949 \cdot \left(\frac{1.64^2 \cdot 0.95^2 \cdot 0.477 \cdot (1 - 0.477)}{1.64^2 \cdot 0.95^2 \cdot 0.477 \cdot (1 - 0.477) + 0.05031^2 \cdot (1{,}117{,}949 - 1)} \right) = 239$, which is not far from the sample size of 256 observed for these parameters in Figure 7 for the national line. Taking the finite population correction factor ϕ as one (1) gives the same answer, as $n = \left(\frac{0.95 \cdot 1.64}{0.05031} \right)^2 \cdot 0.477 \cdot (1 - 0.477) = 239$.[16]

[16] Although USAID has not specified confidence levels nor intervals, IRIS Center (2007a and 2007b) says that a sample size of $n = 300$ is sufficient for USAID reporting. USAID microenterprise partners in Tajikistan should report using the median line. Given the α factor of 1.02 for this line (Figure 8), an expected before-measurement household-level poverty rate of 22.8 percent (the all-Tajikistan rate for 2007, Figure 1), and a

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

In practice after the end of fieldwork for the TLSS in November 2007, a program would select a poverty line (say, the national line), note its participants' population size (for example, $N = 10{,}000$ participants), select a desired confidence level (say, 90 percent, or $z = 1.64$), select a desired confidence interval (say, ±2.0 percentage points, or $c = \pm 0.02$), make an assumption about $\widetilde{p}$ (perhaps based on a previous measurement such as the household-level poverty rate for the national line for Tajikistan of 47.7 percent in the 2007 TLSS in Figure 1), look up α (here, 0.95 in Figure 8), assume that the scorecard will still work in the future and for non-nationally representative sub-groups,[17] and then compute the required sample size. In this illustration, $n = 10{,}000 \cdot \left(\frac{1.64^2 \cdot 0.95^2 \cdot 0.477 \cdot (1 - 0.477)}{1.64^2 \cdot 0.95^2 \cdot 0.477 \cdot (1 - 0.477) + 0.02^2 \cdot (10{,}000 - 1)} \right) = 1{,}315.$

---

confidence level of 90 percent, then $n = 300$ implies a confidence interval of $\pm 1.64 \cdot 1.02 \cdot \sqrt{\frac{0.228 \cdot (1 - 0.228)}{300}} = \pm 4.1$ percentage points.

[17] This paper reports accuracy for the scorecard applied to the validation sample, but it does not test accuracy for later years nor for sub-groups. Performance after November 2007 will resemble that in the 2007 TLSS with deterioration over time to the extent that the relationships between indicators and poverty status change and to the extent that the Tajikistan population changes.

7. Estimates of changes in poverty rates over time

The change in a group's poverty rate between two points in time is estimated as the change in the average poverty likelihood of the households in the group. With data only from the 2007 TLSS,[18] this paper cannot test estimates of change over time for Tajikistan, and it can only suggest approximate formulas for standard errors. Nonetheless, the relevant concepts are presented here because, in practice, local pro-poor organizations can apply the scorecard to collect their own data and measure change through time.

7.1 Warning: Change is not impact

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

[18] The 2009 TLSS could be used to measure the accuracy of estimates of changes in poverty rates over time, but it did not collect indicators related to the residence, in particular, the number of rooms and the type of the exterior walls.

way to know—or explicit assumptions about—what would have happened in the absence of participation. And that information must come from beyond the scorecard.

7.2 Estimating changes in poverty rates over time

Consider the illustration begun in the previous section. On 1 January 2015, an organization samples three households who score 20, 30, and 40 and so have poverty likelihoods of 84.8, 73.3, and 51.5 percent (national line, Figure 3). Adjusting for the known bias of –0.6 percentage points (Figure 8), the group's baseline estimated poverty rate is the households' average poverty likelihood of [(84.8 + 73.3 + 51.5) ÷ 3] – (–0.6) = 70.5 percent.

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

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

By way of illustration, suppose that two years later on 1 January 2017, the organization samples three additional households who are in the same population as the three original households (or suppose that the same three original households are scored a second time) and finds that their scores are 25, 35, and 45 (poverty likelihoods of 77.9, 63.3, and 45.3 percent, national line, Figure 3). Adjusting for known bias, the average poverty likelihood at follow-up is [(77.9 + 63.3 + 45.3) ÷ 3] – (–0.6) = 62.8

percent, an improvement of 69.9 – 62.8 = 7.1 percentage points. In the two years between measurements, the annual rate of reduction is 7.1 ÷ 2 = 3.6 percentage points.[19]

Thus, about one in 28 participants in this hypothetical example progress across the poverty line per year in 2015/7.[20] Among those who start a year below the line, about one in 19 (3.6 ÷ 69.9 = 5.2 percent) on net end up above the line.[21]

7.3 Accuracy for estimated change in two independent samples

With only the 2007 TLSS, it is not possible to measure the accuracy of scorecard estimates of changes in groups' poverty rates over time. In practice, of course, local pro-poor organizations in Tajikistan can still use the scorecard to estimate change. The rest of this section suggests approximate formulas for standard errors that may be used until there is additional data.

[19] Of course, such a huge annual reduction in poverty is highly unlikely, but this is just an example to show how the scorecard can be used to estimate change.
[20] This is a net figure; some start above the line and end below it, and vice versa.
[21] The scorecard does not reveal the reasons for this change.

7.4 Precision for estimates of change in two 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 $\pm c$ with the standard error σ of a scorecard's estimate of the change in poverty rates over time:

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

Here, z, c, $\hat{p}$ and N are defined as above, n is the sample size at both baseline and follow-up,[22] and α is the average (across a range of bootstrapped sample sizes) of the ratio of the observed confidence interval from a scorecard and the theoretical confidence interval under direct measurement.

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

$n = 2 \cdot N \cdot \left(\frac{z^2 \cdot \alpha^2 \cdot \tilde{p} \cdot (1 - \tilde{p})}{z^2 \cdot \alpha^2 \cdot \tilde{p} \cdot (1 - \tilde{p}) + c^2 \cdot (N - 1)} \right)$. If ϕ can be taken as one, then the formula becomes $n = 2 \cdot \left(\frac{\alpha \cdot z}{c} \right)^2 \cdot \tilde{p} \cdot (1 - \tilde{p})$.

This α has been measured for 11 countries (Schreiner, 2013a, 2013b, 2012c, 2010, 2009a, 2009b, 2009c, 2009d; Chen and Schreiner, 2009; and Schreiner and Woller, 2010a and 2010b). The simple average of α across countries—after averaging α across poverty

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

lines and survey years within each country—is 1.15. This is as reasonable a figure as any to use for Tajikistan.

To illustrate the use of this formula 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 ($\pm c = \pm 0.02$), the poverty line is the national line, $\alpha = 1.15$, $\hat{p} = 0.477$ (the household-level poverty rate in 2007 for the national line in Figure 1), and the population N is large enough relative to the expected sample size n that the finite population correction ϕ can be taken as one. Then the baseline sample size is $n = 2 \cdot \left(\frac{1.15 \cdot 1.64}{0.02} \right)^2 \cdot 0.477 \cdot (1 - 0.477) \cdot 1 = 4{,}437$, and the follow-up sample size is also 4,437.

7.5 Precision for estimated change for one sample, scored twice

Analogous to previous derivations, the general formula relating the confidence interval $\pm c$ to the standard error σ when using a scorecard to estimate change for a single group of households, all of whom are scored at two points in time, is:[23]

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

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

where z, c, α, N, and n are defined as usual, $\hat{p}_{12}$ is the share of all sampled households that move from below the poverty line to above it, and $\hat{p}_{21}$ is the share of all sampled households that move from above the line to below it.

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

$$n = 2 \cdot \left(\frac{\alpha \cdot z}{c} \right)^2 \cdot \tilde{p}_{*} \cdot \sqrt{\frac{N - n}{n - 1}}.$$

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

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

Given this, a sample-size formula for a group of households to whom the Tajikistan scorecard is applied twice (once after November 2007 and then again later) is

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

In Peru (the only source of a data-based estimate, Schreiner, 2009e), 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 ($\pm c = \pm 0.02$), the poverty line is the national line, the sample will first be scored in 2015 and then again in 2018 ($y = 3$), and the population N is so large relative to the expected sample size n that the finite population correction ϕ can be taken as one. The pre-baseline poverty rate p_{2007} is taken as 47.7 percent (Figure 1), and α is assumed to be 1.30. Then the baseline sample size is

$n = 2 \cdot \left(\frac{1.30 \cdot 1.64}{0.02} \right)^2 \cdot \{-0.02 + 0.016 \cdot 3 + 0.47 \cdot [0.477 \cdot (1 - 0.477)]\} \cdot 1 = 3{,}302$. The same group of 3,302 households is scored at follow-up as well.

8. Targeting and segmentation

When an organization uses the scorecard for targeting and segmentation, 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 is defined by whether consumption is below a poverty line as directly measured by a survey. In contrast, targeting/segmentation status is an organization'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 9 depicts these four possible targeting outcomes. Targeting accuracy varies by the cut-off score; a higher cut-off has better inclusion (but worse leakage), while a lower cut-off has better exclusion (but worse undercoverage).

Programs should weigh these trade-offs when setting a cut-off. A formal way to do this is to assign net benefits—based on a program's values and mission—to each of

the four possible targeting outcomes and then to choose the cut-off that maximizes total net benefits (Adams and Hand, 2000; Hoadley and Oliver, 1998).

Figure 10 shows the distribution of households by targeting outcome for Tajikistan. For an example cut-off of 39 or less, outcomes for the national line in the validation sample are:

- Inclusion: 23.8 percent are below the line and correctly targeted
- Undercoverage: 24.0 percent are below the line and mistakenly not targeted
- Leakage: 9.2 percent are above the line and mistakenly targeted
- Exclusion: 43.0 percent are above the line and correctly not targeted

Increasing the cut-off to 44 or less improves inclusion and undercoverage but worsens leakage and exclusion:

- Inclusion: 23.8 percent are below the line and correctly targeted
- Undercoverage: 24.0 percent are below the line and mistakenly not targeted
- Leakage: 9.2 percent are above the line and mistakenly targeted
- Exclusion: 43.0 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 10 for a given poverty line
- Select the cut-off with the highest total net benefit

The most difficult step is assigning benefits and costs to targeting outcomes. A program that uses targeting—with or without scoring—should thoughtfully consider

how it values successful inclusion 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 correctly included or correctly excluded:

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

Figure 10 shows "Total Accuracy" for all cut-offs for the Tajikistan scorecard. For the national line in the validation sample, total net benefit is greatest (68.4) for a cut-off of 44 or less, with about two in three households in Tajikistan 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 values inclusion more (say, twice as much) than exclusion, it can reflect this by setting the benefit for inclusion to 2 and the benefit for exclusion to 1. Then the chosen cut-off will maximize (2 x Households correctly included) + (1 x Households correctly excluded).[24]

[24] Figure 10 also reports "BPAC", discussed in the next section.

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 11 ("% targeted HHs who are poor") shows, for the Tajikistan scorecard applied to the validation sample, the expected poverty rate among households who score at or below a given cut-off. For the example of the national line, targeting households who score 39 or less would target 33.0 percent of all households (second column) and produce a poverty rate among those targeted of 72.1 percent (third column).

Figure 11 also reports two other measures of targeting accuracy. The first is a version of coverage ("% poor HHs who are targeted"). For the example of the national line with the validation sample and a cut-off of 39 or less, 49.8 percent of all poor households are covered.

The final targeting measure in Figure 11 is the number of successfully targeted poor households for each non-poor household mistakenly targeted (right-most column). For the national line with the validation sample and a cut-off of 39 or less, covering 2.6 poor households means leaking to 1 non-poor household.

9. Context of poverty-assessment tools in Tajikistan

This section discusses two existing poverty-assessment tools for Tajikistan in terms of their goals, methods, definitions of *poverty*, data, indicators, cost, bias, and precision. In general, the advantages of the scorecard are its:

- Use of data from a more recent nationally representative consumption
- Use of a definition of *poverty* that is simple to understand and used by government
- Reporting of bias and precision from out-of-sample tests, including formulas for standard errors
- Feasibility for local, pro-poor programs, due to its simplicity and transparency

9.1 Baschieri and Falkingham

Baschieri and Falkingham (2005) use the "poverty-mapping" approach of Elbers, Lanjouw, and Lanjouw (2003) to estimate poverty rates down to the lowest administrative level (*jamoat*) in Tajikistan.

They use generalized least-squares regression with data from the 2003 TLSS to construct nine poverty-assessment tools (one for each survey stratum) to estimate the logarithm of per-capita consumption, considering only indicators that match items in Tajikistan's 2000 census, along with community-level census averages and geographic indicators. The resulting tools are then applied to the census data to estimate poverty rates for the \$2.15/day 1993 PPP line at the level of Tajikistan's five regions, 58 districts, and 369 *jamoats*. Such estimates would not be possible with only the 2003 TLSS, due to its smaller sample size. Finally, Baschieri and Falkingham make "poverty

maps" that quickly show how estimated poverty rates vary across areas in a way that makes sense to non-specialists.

Poverty mapping in Baschieri and Falkingham has much in common with the scorecard here in that they both:

- Build poverty-assessment tools with nationally representative survey data and then apply them to other data on groups that may not be nationally representative
- Estimate poverty rates for groups
- Provide unbiased estimates when their assumptions hold
- Report standard errors
- Seek to be useful in practice and so aim to be transparent to non-specialists

Strengths of poverty mapping include that it:

- Has formally established theoretical properties
- Can be applied straightforwardly to distributional measures of well-being (such as the poverty gap or the Gini coefficient) that go beyond head-count poverty rates
- Accounts for uncertainty in the estimation of tool points when estimating standard errors
- Requires data on fewer households for construction and calibration
- Includes community-level indicators and geographic indicators, increasing accuracy and precision

Strengths of the scorecard include that it:

- Uses simple, verifiable indicators that are quick and inexpensive to collect
- Is simpler in terms of both construction and application
- Tests accuracy empirically
- Reports bias
- Associates poverty likelihoods with scores non-parametrically
- Surfaces estimates of poverty likelihoods for individual households
- Reduces overfitting by selecting indicators with statistical and non-statistical criteria
- Reports simple formulas for standard errors

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

pro-poor organizations to manage their social performance.[25] On a technical level for Tajikistan, Baschieri and Falkingham estimate consumption directly, whereas the scorecard estimates poverty likelihoods.[26] Also, the nine tools in Baschieri and Falkingham are more likely to be overfit than the single scorecard here.

Baschieri and Falkingham's tools for Tajikistan use an average of about 14 indicators from among the following 62 indicators:

- Indicators from the 2003 TLSS (Baschieri and Falkingham report their internal variable names, but do not provide definitions):
  - `hh_size`
  - `hh_work`
  - `hh_marr`
  - `hh_fem`
  - `sepapart`
  - `dwe80_90`
  - `dweaft90`
  - `stooven`
  - `owndwe`
  - `areles`
  - `area40_69`
  - `areamo70`
  - `prwork`
  - `hh_pri`

[25] Another apparent difference is that the developers of poverty mapping (Elbers, Lanjouw, and Lanjouw, 2003; Demombynes *et al.*, 2004) say that poverty mapping is too inaccurate to be used for targeting at the household level. In contrast, Schreiner (2008b) supports household-level targeting as a legitimate, potentially useful application of the scorecard. In Elbers *et al.* (2007), the developers of poverty mapping seem to take a small step away from their original position.

[26] Haslett and Jones (2006, p. 61) note that "the benefits of the [poverty-mapping] methodology accrue when interest is in several non-linear functions of the same target variable [such as the poverty gap or the Gini coefficient in Baschieri and Falkingham] . . . or in distributional properties. If only a single measure were of interest, it might be worthwhile to consider direct modelling of this. For example, small-area estimates of poverty incidence could be derived by estimating a logistic regression model for incidence in the survey data". This is what the scorecard does.

    - `hh_sec`
    - `propsec`
    - `prophigh`
    - `prop15`
    - `prop1560`
    - `propab60`
    - `tthh4`
    - `tthh5_7`
    - `tthh8_14`
    - `tthh25_4`
    - `tthh60`
- Indicators at the settlement-level:
    - Average number of rooms in the residence
    - Share of people who:
        - Are 14-years-old or younger
        - Have no education
        - Have a primary education
        - Are economically active
        - Are employed in a *dekhkan* farm
        - Are employed on their own "ancillary farms"
        - Live in a household with a head who is:
            - Younger than 15 or older than 60
            - Married
            - Divorced/separated
            - Widowed
            - Female
            - Working
        - Live in a household with a head who whose education is:
            - Primary
            - Higher
    - Share of households who live in a residence:
        - With an area of less than 40m^2
        - Detached
        - Shared
        - Non-shared apartment
        - Built before 1960
        - Built between 1960 and 1980
        - Built between 1981 and 1990
    - Share of households who have a:
        - Electric stove
        - Piped water
        - Telephone

- *Jamoat*-level indicators:
    - — Number of households
    - — Share of members 15-years-old or older who are:
        - Economically active
        - "Working on individual basic"
        - Working as employees
        - Working as employees in an enterprise, institution, or organization
        - Employed on their own "ancillary farms"
- Geographic indicators:
    - — Average distance to the nearest road
    - — Average altitude in a 200m radius of settlement
    - — Average altitude in a 200m radius of *jamoat*
    - — Share of land sloped from 0–5 degrees in a 500m radius of settlement
    - — Share of land sloped from 5–20 degrees in a 500m radius of settlement
    - — Share of crop land that is not irrigated in a 1000m radius of *jamoat*

Baschieri and Falkingham's tools are complex and were not designed for use by local, pro-poor organizations. For example, there are nine tools, complicating administration unless an organization happens to work only in a single survey stratum. Also, field workers cannot compute scores, and an organization's back-office must match up a household and its location with average values from the census for its settlement and for its *jamoat*.

Because Tajikistan's 2000 census does not measure consumption, Baschieri and Falkingham cannot test accuracy out-of-sample with the census, and they do not test accuracy (out-of-sample nor in-sample) with the 2003 TLSS. Thus, a comparison of bias for estimated poverty rates between poverty mapping and the scorecard is not possible. Nevertheless, a comparison of precision is possible because Baschieri and Falkingham—unlike most poverty maps—report both standard errors and sample sizes for estimated poverty rates with the census data, implying an all-Tajikistan α of 1.28. When the

scorecard here is applied to the 2007 TLSS with a line that gives a person-level poverty rate matching the 63.0 percent of Baschieri and Falkingham's $2.15/day 1993 PPP line, out-of-sample alpha is 0.94. Of course, this comparison is imperfect, as the two tools are applied to different data.

9.2 IRIS Center

IRIS Center (2006) was commissioned to build a poverty-assessment tool (PAT) for use by USAID's microenterprise partners in Tajikistan for reporting the share of their participants who are "very poor". Although IRIS does not report it, they used data on the 2,000 households in Tajikistan's Living Standards Measurement Survey fielded in May and June of 1999. The "very poor" are defined as those having per-capita consumption less than the median line, which gives a household-level poverty rate of 47.3 percent. In deriving this, IRIS used the single national poverty line of about 1,050 Tajikistani rubles per person per month that gives a person-level poverty rate of 96.0 percent (Falkingham, 2000).

After comparing several statistical approaches,[27] IRIS settles on a two-step linear-probability regression that estimates the likelihood of having per-capita household consumption below the median line. A household is counted as *very poor* if its estimated probability in the second step is more than 49 percent.

[27] Thanks to the "flat maximum", all methods have similar "Total Accuracy".

In general, the PAT is like the scorecard here, except that it:

- Uses older data (1999 rather than 2007)
- Has a more indicators (18 rather than 10)
- When estimating poverty rates, replaces a household's estimated poverty likelihood with either 0 (if the likelihood is 49 percent or less) or 100 percent, rather then averaging the original estimated poverty likelihoods

The PAT's 18 indicators are simple and verifiable:

- Residence:
  - Region
  - Urban/rural
- Demographics:
  - Number of household members
  - Age of the head of the household
  - Sex of the head of the household
- Highest level of education completed:
  - For the head, secondary school
  - Number of members other than the head, secondary school
  - Number of members other than the head, professional or technical school
  - Number of members other than the head, university
- Ownership of consumer durables:
  - House
  - Gas or electric stove
  - Refrigerator
  - Air conditioner
  - Sewing machine
  - Bicycle
  - Television
- Agricultural assets:
  - Number of cattle
  - Number of donkeys/mules

In terms of accuracy, IRIS does not report precision. It does report in-sample results—that is, based on applying the PAT to the same data that was used to construct the PAT in the first place—in terms of:

- Bias of estimated poverty rates at a point in time[28]
- Targeting (inclusion, undercoverage, leakage, and exclusion)
- The Balanced Poverty Accuracy Criterion, USAID's standard for certifying PATs

IRIS Center (2005) introduced BPAC. It considers accuracy in terms of inclusion and in terms of the absolute difference between undercoverage and leakage (which, under the PAT approach, is equal to the bias of the estimated poverty rate). The formula is $\text{BPAC} = 100 \cdot \left(\frac{\text{Inclusion} - \mid \text{Undercoverage} - \text{Leakage} \mid}{\text{Inclusion} + \text{Undercoverage}} \right)$.

Because bias (in the PAT approach) is the difference between undercoverage and leakage, and because the normalization term $\frac{100}{\text{Inclusion} + \text{Undercoverage}}$ matters only when comparing poverty-assessment tools across populations with different poverty rates (an irrelevant consideration when selecting among alternative tools for a given country such as Tajikistan), the formula boils down to $\text{BPAC} = \text{Inclusion} - \mid \text{Bias} \mid$.

Expressing BPAC as $\text{Inclusion} - \mid \text{Bias} \mid$ helps to show why BPAC is not useful for comparing the PAT with the scorecard (Schreiner, 2014). Regardless of whether undercoverage differs from leakage (and given the assumptions discussed earlier in this paper), the scorecard produces unbiased estimates of poverty rates. While BPAC can be

[28] IRIS (2005) calls bias the "Poverty Incidence Error" (PIE). In the consumption-estimation approach, bias is the absolute difference between undercoverage and leakage.

used to compare alternative poverty-assessment tools under the PAT's approach, it does not make sense to apply it to the scorecard's poverty-likelihood approach. This is because, when estimating poverty rates, the scorecard does not use a cut-off to classify households as either 100-percent poor or 100-percent non-poor. Instead, households have an estimated poverty likelihood somewhere between 0 to 100 percent. If a user of a scorecard sets a targeting cut-off, then that cut-off matters only for targeting, and it does not affect the estimation of poverty rates at all.

In any case, both the PAT and the scorecard give unbiased estimates of poverty rates (after subtracting off known bias), so any distinction between their accuracy must hinge on targeting or on the precision of estimated poverty rates. A clean comparison along these dimensions for Tajikistan is difficult, as IRIS uses 1999 data (versus 2007 data here), in-sample tests (versus out-of-sample here), and does not report precision at all (this paper does report precision).

An imperfect comparison of targeting accuracy is possible for IRIS with its median line (with a household-level poverty rate in the 1999 LSMS of 47.3 percent) and for the scorecard with the national line (with a household-level poverty rate in the 2007 TLSS of 47.7 percent). When the PAT targets the lowest-scoring 47.1 percent of households, inclusion is 33.1 percent, exclusion is 38.6 percent, and "Total Accuracy" is 71.7 percent. When the scorecard here targets this same share of households, inclusion is 31.6 percent, exclusion is 36.7 percent, and "Total Accuracy" is 68.3 percent. In sum, the PAT targets better, but the difference (about 3 household per 100) may be due to

the PAT's in-sample testing, the differences in underlying poverty rates, and to the eight-year data gap.

Although IRIS reports the PAT's targeting accuracy and although the BPAC formula considers targeting accuracy, IRIS says that the PAT should not be used for targeting.[29] IRIS also doubts that the PAT can be useful for measuring change, noting that "it is unclear that the tools will be able to identify real changes in poverty over time due to their inherent measurement errors. Unless the changes in the poverty rate are exceptionally large and the tools exceptionally accurate, the changes identified are likely to be contained within the margin of error."[30]

Targeting and estimating changes over time are possible uses that are supported for the scorecard. This paper reports targeting accuracy as well as formula for standard errors for measures of change over time so that users can decide for themselves whether accuracy is adequate for their purposes.

[29] `povertytools.org/faq/faq.html#11`, retrieved 19 February 2009.
[30] `povertytools.org/faq/faq2.html`, retrieved 7 December 2012.

10. Conclusion

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

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

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

The scorecard is constructed with half of the data from Tajikistan's 2007 TLSS, calibrated to eight poverty lines, and tested on data from the other half of the 2007 TLSS. Bias and precision are reported for estimates of households' poverty likelihoods, groups' poverty rates at a point in time, and changes in groups' poverty rates over time. Of course, the scorecard's estimates of change are not the same as estimates of program impact. Targeting accuracy is also reported.

When the scorecard is applied to the validation sample, the maximum absolute difference between estimates versus true poverty rates for groups of households at a point in time is 0.9 percentage points. The average absolute bias across the eight poverty lines is about 0.4 percentage points. Unbiased estimates may be had by subtracting the known bias for a given poverty line from the original estimates. For $n =$ 16,384 and 90-percent confidence, the precision of these differences is ± 0.6 percentage points or better.

If an organization wants to use the scorecard for targeting, then the results here provide useful information for selecting a cut-off that fits its values and mission.

Although the statistical technique is innovative, and although technical accuracy is important, the design of the scorecard focuses on transparency and ease-of-use. After all, accuracy is irrelevant if an organization feels so daunted by a scorecard's complexity or its cost that it does not even try to use it.

For this reason, the scorecard is kept simple, using ten indicators that are straightforward, low-cost, and verifiable. Points are all zeros or positive integers, and scores range from 0 (most likely below a poverty line) to 100 (least likely below a poverty line). Scores are converted to poverty likelihoods via simple look-up tables, and targeting cut-offs are likewise straightforward to apply. The design attempts to facilitate voluntary adoption by helping managers to understand and trust scoring and by allowing non-specialists to add up scores quickly in the field.

In summary, the scorecard is a practical, objective way for pro-poor programs in Tajikistan to estimate consumption-based poverty rates, track changes in poverty rates over time, and target services. The same approach can be applied to any country with similar data.

References

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

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

Baschieri, Angela; and Jane Falkingham. (2005) "Developing a Poverty Map of Tajikistan: A Technical Note", Southampton Statistical Sciences Research Institute Applications and Policy Working Paper No. A05/11, `eprints.soton.ac.uk/19128/`, retrieved 20 October 2014.

Caire, Dean. (2004) "Building Credit Scorecards for Small-Business Lending in Developing Markets", `microfinance.com/English/Papers/Scoring_SMEs_Hybrid.pdf`, retrieved 20 October 2014.

_____; and Mark Schreiner. (2012) "Cross-Tab Weighting for Credit Scorecards in Developing Markets", `dean_caire@hotmail.com`.

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

Chen, Shiyuan; and Mark Schreiner. (2009) "Simple Poverty Scorecard Poverty-Assessment Tool: Vietnam", `simplepovertyscorecard.com/VNM_2006_ENG.pdf`, retrieved 20 October 2014.

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

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

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

Demombynes, Gabriel; Elbers, Chris; Lanjouw, Jenny; Lanjouw, Peter; Mistiaen, Johan; and Berk Özler. (2004) "Producing an Improved Geographic Profile of Poverty: Methodology and Evidence from Three Developing Countries", pp. 154–176 in Anthony Shorrocks and Rolph van der Hoeven (eds.) *Growth, Inequality, and Poverty*.

Elbers, Chris; Fujii, Tomoki; Lanjouw, Peter; Özler, Berk; and Wesley Yin. (2007) "Poverty Alleviation through Geographic Targeting: How Much Does Disaggregation Help?", *Journal of Development Economics*, Vol. 83, pp. 198–213.

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

Falkingham, Jane. (2000) "A Profile of Poverty in Tajikistan", Centre for Analysis of Social Exclusion Paper No. 39, `sticerd.lse.ac.uk/dps/case/cp/CASEpaper39.pdf`, retrieved 20 October 2014.

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 20 October 2014.

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 20 October 2014.

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 20 October 2014.

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

Haslett, Stephen. (2012) "Practical Guidelines for the Design and Analysis of Sample Surveys for Small-Area Estimation", *Journal of the Indian Society of Agricultural Statistics*, Vol. 66, No. 1, pp. 203–212.

_____; and Geoffrey Jones. (2006) "Small-Area Estimation of Poverty, Caloric Intake, and Malnutrition in Nepal", `un.org.np/node/10501`, retrieved 20 October 2014.

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

IRIS Center. (2011) "Accuracy Results for 31 Poverty-Assessment Tool Countries", `povertytools.org/USAID_documents/Accuracy/PAT_country_accuracy_results_06282011.pdf`, retrieved 20 October 2014.

_____. (2007a) "Manual for the Implementation of USAID Poverty Assessment Tools", `povertytools.org/training_documents/Manuals/USAID_PAT_Manual_Eng.pdf`, retrieved 20 October 2014.

_____. (2007b) "Introduction to Sampling for the Implementation of PATs", `povertytools.org/training_documents/Sampling/Introduction_Sampling.ppt`, retrieved 20 October 2014.

_____. (2006) "Poverty Assessment Tool: Tajikistan", `povertytools.org/countries/Tajikistan/Tajikistan.html`, retrieved 20 October 2014.

_____. (2005) "Notes on Assessment and Improvement of Tool Accuracy", `povertytools.org/other_documents/AssessingImproving_Accuracy.pdf`, retrieved 20 October 2014.

Johnson, Glenn. (2007) "Lesson 3: Two-Way Tables—Dependent Samples", `onlinecourses.science.psu.edu/stat504/node/96`, retrieved 20 October 2014.

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

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

Martinelli, César; and Susan W. Parker. (2007) "Deception and Misreporting in a Social Program", *Journal of the European Economic Association*, Vol. 4, No. 6, pp. 886–908.

Matul, Michal; and Sean Kline. (2003) "Scoring Change: Prizma's Approach to Assessing Poverty", Microfinance Centre for Central and Eastern Europe and the New Independent States Spotlight Note No. 4, `mfc.org.pl/sites/mfc.org.pl/files/spotlight4.PDF`, retrieved 20 October 2014.

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 20 October 2014.

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

Ravallion, Martin. (1998) "Poverty Lines in Theory and Practice", World Bank LSMS Working Paper No. 133, go.worldbank.org/8P3IBJPQS1, retrieved 20 October 2014.

SAS Institute Inc. (2004) "The LOGISTIC Procedure: Rank Correlation of Observed Responses and Predicted Probabilities", in *SAS/STAT User's Guide, Version 9*, support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug_logistic_sect035.htm, retrieved 20 October 2014.

Schreiner, Mark. (2014) "How Do the Simple Poverty Scorecard: and the PAT Differ?", microfinance.com/English/Papers/Scorecard_versus_PAT.pdf, retrieved 20 October 2014.

_____. (2013a) "Simple Poverty Scorecard Poverty-Assessment Tool: Bangladesh", simplepovertyscorecard.com/BGD_2010_ENG.pdf, retrieved 11 July 2016.

_____. (2013b) "Simple Poverty Scorecard Poverty-Assessment Tool: Nicaragua", simplepovertyscorecard.com/NIC_2009_ENG.pdf, retrieved 11 July 2016.

_____. (2012a) "An Expert-Based Poverty Scorecard for Rural China", microfinance.com/English/Papers/Scoring_Poverty_China_EN.pdf, retrieved 11 July 2016.

_____. (2012b) "Simple Poverty Scorecard Poverty-Assessment Tool: Colombia", simplepovertyscorecard.com/COL_2009_ENG.pdf, retrieved 11 July 2016.

_____. (2012c) "Simple Poverty Scorecard Poverty-Assessment Tool: Peru", simplepovertyscorecard.com/PER_2010_ENG.pdf, retrieved 11 July 2016.

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

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

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

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

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

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

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

_____. (2008b) "Simple Poverty Scorecard Poverty-Assessment Tool: Ecuador", simplepovertyscorecard.com/ECU_2005_ENG.pdf, retrieved 11 July 2016.

_____. (2006) "Is One Simple Poverty Scorecard Poverty-Assessment Tool Enough for India?", microfinance.com/English/Papers/Scoring_Poverty_India_Segments.pdf, retrieved 20 October 2014.

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

_____. (2005b) "IRIS Questions on the Simple Poverty Scorecard Poverty-Assessment Tool", microfinance.com/English/Papers/Scoring_Poverty_Response_to_IRIS.pdf, retrieved 20 October 2014.

_____. (2002) *Scoring: The Next Breakthrough in Microfinance?* CGAP Occasional Paper No. 7, microfinance.com/English/Papers/Scoring_Breakthrough_CGAP.pdf, retrieved 20 October 2014.

_____; Matul, Michal; Pawlak, Ewa; and Sean Kline. (2004) "Poverty Scoring: Lessons from a Microlender in Bosnia-Herzegovina", microfinance.com/English/Papers/Scoring_Poverty_in_BiH_Short.pdf, retrieved 20 October 2014.

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

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

Sharif, Iffath Anwar. (2009) "Building a Targeting System for Bangladesh Based on Proxy-Means Testing", World Bank Social Protection Discussion Paper No. 0914, siteresources.worldbank.org/SOCIALPROTECTION/Resources/SP-Discussion-papers/Safety-Nets-DP/0914.pdf, retrieved 20 October 2014.

Sillers, Don. (2006) "National and International Poverty Lines: An Overview", pdf.usaid.gov/pdf_docs/Pnadh069.pdf, retrieved 20 October 2014.

State Committee on Statistics. (2009) "Tajikistan Living Standards Measurement Survey 2007", stat.tj/en/img/e4f46968ef1143d0ddb14e16b8d9d6d9_1280833182.pdf, retrieved 20 October 2014.

Steele, Diane. (2008) "Basic Information Document: Tajikistan Living Standards Measurement Survey, 2007", microdata.worldbank.org/index.php/catalog/72/download/11528, retrieved 20 October 2014.

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

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

Toohig, Jeff. (2008) "PPI Pilot Training Guide", microfinancegateway.org/sites/default/files/mfg-en-paper-progress-out-of-poverty-index-ppi-pilot-training-mar-2008.pdf, retrieved 20 October 2014.

United States Congress. (2004) "Microenterprise Results and Accountability Act of 2004 (HR 3818 RDS)", November 20, smith4nj.com/laws/108-484.pdf, retrieved 20 October 2014.

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

World Bank. (2012) *Targeting Poor and Vulnerable Households in Indonesia*, ausaid.gov.au/Publications/Pages/report-targeting-poor-households-indonesia.aspx, retrieved 20 October 2014.

_____. (2009) "Republic of Tajikistan: Poverty Assessment", Report No. 51341-TJ, Washington, D.SC. documents.worldbank.org/curated/en/2009/12/11664677/tajikistan-poverty-assessment, retrieved 20 October 2014.

_____. (2008) "International Comparison Project: Tables of Results", siteresources.worldbank.org/ICPINT/Resources/icp-final-tables.pdf, retrieved 20 October 2014.

Zeller, Manfred. (2004) "Review of Poverty Assessment Tools", pdf.usaid.gov/pdf_docs/PNADH120.pdf, retrieved 20 October 2014.

Guidelines for the Interpretation of Scorecard Indicators

The following comes from:

GosKomStat. (2007) *Tajikistan 2007 Living Standards Survey: Manual for Interviewers*, `go.worldbank.org/6TUMCB3K30`, retrieved 20 October 2014. [the *Manual*]

and

GosKomStat. (2007) "Tajikistan 2007 Living Standards Survey: Main Questionnaire", `go.worldbank.org/6TUMCB3K30`, retrieved 20 October 2014. [the *Questionnaire*]

General Guidelines:

According to p. 7 of the *Manual*, "[This *Manual*] should be your first reference if you encounter any problems in administering . . . the questionnaire."

Confidentiality
According to p. 1 of the *Manual*, "The information being collected in this survey is confidential and must therefore not be divulged to any unauthorized person."

According to p. 9–10 of the *Manual*, "The setting of the questionnaire administration should be relatively private. Some of the questions being asked are of a personal and private nature. You should respect the desire of the respondents for privacy. . . .

"No person except your supervisor or people from the project's staff should come with you when you interview. If project staff do accompany you to an interview, you should always be sure to introduce the staff member to the respondent, making clear to the respondent the purpose of the presence of the project staff member.

"Any other persons not connected to the [survey] or to the household should not be present when you are administering the questionnaire. If any such individuals are present when you begin your interview, you must politely request them to leave in order to respect the privacy of the survey household. If they cannot leave at that time, you should schedule the interview for a later time, when greater privacy can be assured."

The respondent

According to p. 4 of the *Manual*, questions that concern the household as a whole should be asked of the head of the household, "assisted by other household members if necessary".

According to p. 14 of the *Manual*, the respondent should be the household head, if available. If he or she is not available, then the respondent should be the most senior member of the household who is present.

According to p. 10 of the *Manual*, "Do not try to obtain data directly from young children [those under age 10]."

How to interact with the respondent

According to p. 10 of the *Manual,* "Above all, your attitude towards the respondents in the survey households must be one of respect. You must always be courteous and patient towards household members. Be business-like in your conduct—never bullying, demanding, nor rude. Always act in a way that warrants respect and cooperation from the respondent. You will find work more pleasant if you remain polite and friendly to everyone at all times.

"Be willing to answer any questions that the respondents ask you about why you are asking the questions you are asking."

Verifying responses

According to p. 10–11 of the *Manual*, "In conducting an interview, if it is clear that the respondent has understood the question you have asked, you must accept whatever response the respondent provides you. You can probe to make sure the respondent understands the key elements of the question being asked. However, you must never second-guess the respondent nor make the assumption that you have a better understanding of the condition of the individual or household than the respondent does. The function of the interviewer is not to verify that the information provided is correct. The [survey is] interested in what the respondent actually says. It is always possible that the respondent will lie to you or provide inaccurate information, but you, as the interviewer, should not make any judgments on the information provided.

"There are exceptions, of course. If the respondent says that he or she has no livestock and there are chickens pecking at your feet or goats tied up nearby, you should inquire about these animals. However, you should not probe excessively after seeking initial clarification. In any case, you should never go [to someone] outside of the household to get information. This is beyond the scope of your work.

"Ultimately, the question of 'wrong' or 'right' answers does not apply in administering the [survey]. The questionnaire is being administered to the survey household members because we rightly expect that they will be able to provide the best information about their own living conditions."

Dealing with reluctant respondents

According to pp. 11–12 of the *Manual*, "If a respondent refuses to answer, the interviewer should remind him or her of the importance and confidentiality of the survey. It is very important that an answer is collected for each question.

"Sometimes the respondents do not understand and misinterpret the meaning of questions. In this case, the interviewer's response is to re-read the question in its entirety, emphasizing the words or part of the question that the respondent missed the first time. If the question contains a term or concept that is not clear to the respondent, the interviewer cannot use his/her discretion, and the respondent should answer according to the way in which he/she seems to understand the term.

"Sometimes the interviewers have to stimulate respondents to amplify, clarify, or in some other way modify their original answer to be complete and to meet question objectives. . . .

"When the respondent says 'I don't know', the interviewer's first task is to attempt to diagnose the origin of the problem:

- If 'I don't know' is considered to be an accurate, thoughtful answer to an information question, [then ask another household member]. Be careful, however, not to do this too often. [If the final answer is 'I don't know', write "DOES NOT KNOW' in the 'Points' column]
- If it is a delaying response style, then give the respondent time to think about the answer. You may want to repeat the question to help the respondent think it through
- If the respondent has not thought about the question, then encourage the respondent to think about the question, emphasizing that he/she is uniquely qualified to provide information on the topic. Then repeat the question
- If the respondent is not sure about the quality or precision of the answer, then be reassuring. . . . The respondent's own best estimate will be better than not having any information at all
- Alternatively, ask to speak to someone who does know"

Guidelines relating to specific indicators in the scorecard

1. How many members does the household have?
    A. Eight or more
    B. Seven
    C. Six
    D. Five
    E. Four
    F. Three
    G. One or two

According to p. 4–6 of the *Manual*, "A *household* may be either a person living alone or a group of people, either related or unrelated, who live together as a single unit in the sense that they have common housekeeping arrangements (that is, share or are supported by a common budget). A standard definition of a *household* is 'a group of people who live together, pool their money, and eat at least one meal together each day'. In most cases, someone who does not live with the household during the survey period is not a current member of the household.

"Members of a household need not necessarily be related by blood or by marriage. On the other hand, not all those who are related and are living in the same compound or dwelling are necessarily members of the same household. [For example], two brothers who live in the same dwelling with their own wives and children may or may not have a common housekeeping arrangement. If they do not, they should be considered separate households.

"One should make a distinction between *family* and *household*. The first reflects social relationships, blood descent, and marriage. The second is used here to identify an economic unit. While families and households are often the same, this is not necessarily the case. You must be cautious and use the criteria provided on household membership to determine which individuals make up a particular household.

"In the case of extended-family systems, household members may be distributed over two or more dwellings. If these dwelling units are in the same compound or nearby . . ., then the residents of these separate dwelling units should be treated as one household. . . .

"Having identified a social unit that shares a common housekeeping arrangement—that is, a *household*—it then becomes necessary to determine who is and who is not a member of that household. . . . We will consider as *members* all individuals who have been absent from the household for less than 12 of the last 12 months."

However, there are several exceptions to this rule:

- The household head is a member regardless of how long he/she has been absent
- Young infants less than 12-months-old [are counted as members]
- New spouses who have recently come into the household and are now residing with the household [are counted as members]
- Household members residing in an institution elsewhere, but still dependent on the household [are counted as members]. This principally includes boarding-school students. However, it does not include military personnel, prisoners, or other individuals who are not primarily dependent on the household for their welfare

"Non-relatives who are resident in the household for 12 months and are included in a common housekeeping arrangement under the head of household are to be considered as household members. However, servants, other hired workers, and lodgers (individuals who pay to reside in the dwelling of the household) should not be considered to be household members if they have their own household elsewhere which they head or upon which they are dependent.

"The household head is a member of the household regardless of how long he/she has been away from the household. . . .

"You should be very careful when dealing with this rather complex task in the survey administration of determining who should be included and who should not be included as a member of a survey household."

2. Do all household members 8- to 17-years old currently attend (or plan to attend) an educational institution in this academic year?
    A. No
    B. Yes
    C. No one is 8- to 17-years-old

According to p. 14 of the *Manual*, consider age in terms of completed years. This is the same as the age of the person on his/her most recent birthday.

According to pp. 25 of the *Manual*, this question asks about current attendance or, if school is not currently in session, about intended attendance once school is back in session.

3. What is the highest educational level that the female head/spouse has achieved (do not include incomplete levels)?
    A. None, primary (grades 1–4), or basic (brades 1–8(9))
    B. Secondary general (grades 9–10(11))
    C. Secondary special, secondary technical, higher education, or graduate school/*aspirantura*
    D. No female head/spouse

According to p. 23 of the *Manual*, "The highest [educational level] achieved requires completing the grade level. [For example,] if a respondent completed grade 7, then the highest diploma achieved would be 'Primary' because that respondent has not completed the 'Basic' grades."

According to p. 5 of the *Manual*, "The *head of household* is the person commonly regarded by the household as their head. The head would usually be the main income earner and decision maker for the household, but you should accept the decision of the household members as to who is their head. The head of household may not be currently residing with the household. There must be one and only one head in the household. If more than one individual in a potential household claims headship, or if individuals within a potential household give conflicting statements as to who is the head of household, it is very likely that you are dealing with two or more households, rather than one. In such cases, it is extremely important that you apply the criteria provided to delimit membership in the survey household."

According to p. 14 of the *Manual*, the head of the household should be an adult.

For the purposes of the scorecard:, the *female head/spouse* is defined as:

- The household head, if the head is female
- The spouse/partner/companion of the household head, if the head is male
- Non-existent, if the head is male and if he does not have a spouse/partner/companion who is also a member of the household

4. How many rooms does the household occupy (excluding any kitchens, balconies, or hallways)?
    A. One
    B. Two
    C. Three
    D. Four
    E. Five or more

The *Manual* has no additional guidelines for this indicator.

5. What is the major construction material of the external walls of the residence?
    A. Adobe, mud bricks, wood, logs, tin, or other
    B. Mud, stone, baked bricks, or concrete

The *Manual* has no additional guidelines for this indicator.

6. Does the residence have a separate bath/shower?
    A. No
    B. Yes

The *Manual* has no additional guidelines for this indicator.

7. Does the household own and use a gas oven, electric oven, or a microwave oven?
    A. No
    B. Yes

The *Manual* has no additional guidelines for this indicator.

8. Does the household own and use an electric iron?
    A. No
    B. Yes

The *Manual* has no additional guidelines for this indicator.

9. Does the household own and use a color television?
    A. No
    B. Yes

The *Manual* has no additional guidelines for this indicator.

10. Does the household currently own any cattle (beef or dairy), or any sheep or goats?
    A. No
    B. Only sheep or goats
    C. Only cattle
    D. Both

The *Manual* has no additional guidelines for this indicator.

Figure 1: Poverty lines and poverty rates for Tajikistan overall, by construction/validation samples, by poverty line, and by households and people

				Poverty rates (% with expenditure less than a poverty line) and poverty lines (TJS per day per person)							
				National					Intl. 2005 PPP		
Sample	**Line or rate**	**Level**	***n***	**Food**	**100%**	**150%**	**200%**	**Median**	**$1.25**	**$2.00**	**$2.50**
All Tajikistan	Line	People	4,860	2.92	4.56	6.84	9.12	3.37	1.55	2.48	3.10
	Rate	Households		14.6	47.7	77.3	89.1	22.8	0.9	8.2	18.3
		People		17.1	53.5	83.1	93.2	26.8	1.0	9.9	21.3
Construction and calibration											
Selecting indicators and points, and associating scores with likelihoods	Rate	Households	2,419	14.6	47.7	76.9	89.1	22.7	0.9	8.2	18.1
	Rate	People		16.9	53.3	82.8	93.0	26.7	1.0	9.6	20.9
Validation											
Measuring accuracy	Rate	Households	2,441	14.6	47.8	77.8	89.1	22.8	0.9	8.2	18.6
	Rate	People		17.3	53.7	83.5	93.4	26.8	1.0	10.1	21.6

Source: 2007 TLSS. Poverty lines in average prices for all of Tajikistan from 29 October to 10 November 2007.

Figure 2: Poverty lines and poverty rates for Tajikistan overall and for its four divisions, by urban/rural, by poverty line, and by households and people

Division	Line or rate	Level	n	Poverty rates (% with expenditure less than a poverty line) and poverty lines (TJS per day per person)							
				National					Intl. 2005 PPP		
				Food	100%	150%	200%	Median	$1.25	$2.00	$2.50
All Tajikistan	Line			2.92	4.56	6.84	9.12	3.37	1.55	2.48	3.10
	Rate	Households	4,860	14.6	47.7	77.3	89.1	22.8	0.9	8.2	18.3
	Rate	People		17.1	53.5	83.1	93.2	26.8	1.0	9.9	21.3
Dushanbe	Line			3.21	5.01	7.52	10.02	3.55	1.71	2.73	3.41
	Rate	Households	900	11.6	34.7	63.8	79.6	15.5	0.9	7.4	14.4
	Rate	People		16.4	43.3	72.8	87.0	21.6	1.3	10.6	19.9
Sughd (urban)	Line			3.02	4.72	7.08	9.43	3.15	1.61	2.57	3.21
	Rate	Households	324	16.7	41.7	63.0	77.9	18.9	2.7	10.4	20.1
	Rate	People		24.5	53.6	73.9	86.3	26.9	4.0	16.4	28.5
Sughd (rural)	Line			2.91	4.55	6.83	9.11	3.04	1.55	2.48	3.10
	Rate	Households	684	28.5	68.3	87.8	93.4	32.3	1.6	16.6	34.4
	Rate	People		33.3	74.0	90.6	95.5	37.0	1.6	20.6	39.5

Source: 2007 TLSS. Poverty lines in average prices for all of Tajikistan from 29 October to 10 November 2007.

Figure 2 (cont.): Poverty lines and poverty rates for Tajikistan overall and for its four divisions, by urban/rural, by poverty line, and by households and people

Division	Line or rate	Level	n	Poverty rates (% with expenditure less than a poverty line) and poverty lines (TJS per day per person)							
				National					Intl. 2005 PPP		
				Food	100%	150%	200%	Median	$1.25	$2.00	$2.50
Khatlon (urban)	Line		216	2.79	4.36	6.54	8.72	3.27	1.48	2.37	2.97
	Rate	Households		10.4	43.3	75.3	84.8	19.8	0.6	4.2	13.7
	Rate	People		14.4	52.5	82.1	91.3	26.2	1.3	6.9	18.6
Khatlon (rural)	Line		1,062	2.77	4.33	6.50	8.67	3.50	1.47	2.36	2.95
	Rate	Households		6.1	42.8	82.1	94.6	21.2	0.2	2.5	9.4
	Rate	People		6.8	46.2	85.9	96.4	23.1	0.2	2.6	10.7
RRS (urban)	Line		162	2.99	4.67	7.00	9.34	3.21	1.59	2.54	3.18
	Rate	Households		21.5	54.0	75.7	89.9	25.1	0.4	9.5	24.0
	Rate	People		24.8	56.8	79.1	91.5	28.6	0.4	11.3	27.1
RRS (rural)	Line		810	2.93	4.58	6.87	9.16	3.53	1.56	2.49	3.12
	Rate	Households		12.0	44.1	76.3	90.1	22.0	0.4	7.2	14.8
	Rate	People		12.5	47.6	80.3	92.4	23.9	0.4	7.6	15.8
GBAO (urban)	Line		108	3.25	5.08	7.62	10.15	3.87	1.73	2.76	3.46
	Rate	Households		1.0	14.9	45.5	74.3	6.2	0.0	0.0	3.6
	Rate	People		0.8	18.4	50.2	78.9	8.7	0.0	0.0	5.5
GBAO (rural)	Line		594	3.34	5.21	7.82	10.42	3.94	1.77	2.84	3.55
	Rate	Households		9.0	41.2	80.6	91.8	19.2	0.7	3.4	11.8
	Rate	People		11.2	47.2	85.6	95.2	23.6	0.7	4.1	14.8

Source: 2007 TLSS. Poverty lines in average prices for all of Tajikistan from 29 October to 10 November 2007.

Tables for
100% of the National Poverty Line

(and Tables Pertaining to All Eight Poverty Lines)

Figure 3 (100% of the national line): Estimated poverty likelihoods associated with scores

If a household's score is . . .	. . . then the likelihood (%) of being below the poverty line is:
0–4	100.0
5–9	100.0
10–14	100.0
15–19	100.0
20–24	84.8
25–29	77.9
30–34	73.3
35–39	63.3
40–44	51.5
45–49	45.3
50–54	37.5
55–59	23.4
60–64	17.1
65–69	13.2
70–74	12.8
75–79	4.8
80–84	0.0
85–89	0.0
90–94	0.0
95–100	0.0

Figure 4 (100% of the national line): Derivation of estimated poverty likelihoods associated with scores

Score	Households at score and < poverty line		All households at score		Poverty likelihood (%)
0–4	0	÷	0	=	100.0
5–9	15	÷	15	=	100.0
10–14	0	÷	0	=	100.0
15–19	1,210	÷	1,210	=	100.0
20–24	2,019	÷	2,381	=	84.8
25–29	4,842	÷	6,218	=	77.9
30–34	6,757	÷	9,213	=	73.3
35–39	8,852	÷	13,988	=	63.3
40–44	8,187	÷	15,891	=	51.5
45–49	6,941	÷	15,322	=	45.3
50–54	4,859	÷	12,971	=	37.5
55–59	1,913	÷	8,161	=	23.4
60–64	938	÷	5,485	=	17.1
65–69	453	÷	3,436	=	13.2
70–74	337	÷	2,626	=	12.8
75–79	69	÷	1,427	=	4.8
80–84	0	÷	1,520	=	0.0
85–89	0	÷	137	=	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 5: Probability that a given household's consumption is in a range demarcated by two adjacent poverty lines

	Likelihood (%) of having daily per-capita expenditure in a range demarcated by adjacent poverty lines								
	<$1.25/day	**≥$1.25/day and <$2.00/day**	**≥$2.00/day and <Food**	**≥Food and <$2.50/day**	**≥$2.50/day and <Median**	**≥Median and <100% Natl.**	**≥100% Natl. and <150% Natl.**	**≥150% Natl. and <200% Natl.**	**≥$200% Natl.**
Score	**<TJS1.55**	**≥TJS1.55 and <TJS2.48**	**≥TJS2.48 and <TJS2.92**	**≥TJS2.92 and <TJS3.10**	**≥TJS3.10 and <TJS3.37**	**≥TJS3.37 and <TJS4.56**	**≥TJS4.56 and <TJS6.84**	**≥TJS6.84 and <TJS9.12**	**≥TJS9.12**
0–4	18.2	15.8	14.3	6.1	16.6	29.1	0.0	0.0	0.0
5–9	18.2	15.8	14.3	6.1	16.6	29.1	0.0	0.0	0.0
10–14	13.6	19.7	14.8	6.1	13.8	31.9	0.0	0.0	0.0
15–19	3.6	25.6	19.1	6.1	7.8	37.9	0.0	0.0	0.0
20–24	3.6	24.8	10.6	5.7	9.8	30.3	14.5	0.7	0.0
25–29	2.1	18.0	17.7	5.9	5.0	29.2	18.9	2.8	0.5
30–34	2.0	14.4	11.9	6.2	8.3	30.5	21.6	4.0	1.1
35–39	0.7	9.6	9.0	5.6	7.2	31.3	28.1	6.7	1.9
40–44	0.7	6.8	5.9	4.1	5.7	28.4	33.9	10.3	4.3
45–49	0.7	3.7	5.3	1.4	4.6	29.6	36.5	11.2	7.0
50–54	0.5	3.1	4.7	2.8	2.4	23.9	35.1	15.4	12.0
55–59	0.3	2.2	1.3	2.6	2.2	14.8	32.9	23.2	20.4
60–64	0.2	2.3	1.1	0.3	0.0	13.2	29.9	28.4	24.5
65–69	0.0	2.4	0.9	0.3	0.1	9.5	26.6	26.8	33.4
70–74	0.0	1.3	0.0	0.4	0.3	10.9	16.9	16.8	53.6
75–79	0.0	0.7	0.0	0.3	0.0	3.9	18.3	15.1	61.9
80–84	0.0	0.0	0.0	0.0	0.0	0.0	22.4	15.7	61.9
85–89	0.0	0.0	0.0	0.0	0.0	0.0	0.0	8.7	91.3
90–94	0.0	0.0	0.0	0.0	0.0	0.0	0.0	8.7	91.3
95–100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	8.7	91.3

Figure 6 (100% of the national line): Average differences by score range between estimated and true poverty likelihoods for households, with confidence intervals, from 1,000 bootstraps of n = 16,384, 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	+100.0	0.0	0.0	0.0
10–14	+0.0	0.0	0.0	0.0
15–19	+13.4	4.0	4.9	6.1
20–24	–1.9	2.8	3.3	4.1
25–29	–4.2	3.2	3.3	3.6
30–34	–1.5	2.0	2.3	3.0
35–39	–2.7	2.2	2.4	2.8
40–44	–2.7	2.2	2.3	2.7
45–49	+4.2	1.7	2.0	2.8
50–54	+3.9	1.9	2.2	2.9
55–59	–3.8	3.0	3.2	3.5
60–64	+0.2	2.2	2.6	3.5
65–69	+4.1	2.1	2.4	3.2
70–74	–6.3	5.0	5.2	6.0
75–79	–16.3	10.5	10.8	11.7
80–84	–8.1	5.6	5.8	6.6
85–89	+0.0	0.0	0.0	0.0
90–94	+0.0	0.0	0.0	0.0
95–100	+0.0	0.0	0.0	0.0

Figure 7 (100% of the national line): Average differences between estimated poverty rates and true values for a group at a point in time, with confidence intervals from 1,000 bootstraps of various sample sizes, scorecard applied to the validation sample

Sample Size	Difference between estimate and true value			
		Confidence interval (±percentage points)		
n	Diff.	90-percent	95-percent	99-percent
1	+0.9	69.9	78.1	90.0
4	+0.4	39.8	44.5	55.0
8	–0.5	29.3	33.7	45.7
16	–0.5	19.8	23.0	30.9
32	–0.6	14.0	16.8	21.2
64	–0.4	9.9	12.1	16.6
128	–0.5	7.0	8.4	10.5
256	–0.5	5.0	5.7	7.3
512	–0.5	3.5	4.2	5.5
1,024	–0.5	2.4	2.9	3.8
2,048	–0.5	1.8	2.0	2.6
4,096	–0.6	1.2	1.5	1.8
8,192	–0.6	0.8	1.0	1.3
16,384	–0.6	0.6	0.7	0.9

Figure 8 (All poverty lines): Average differences between estimates and true values for poverty rates of a group of households at a point in time, precision, and the α factor for precision, scorecard applied to the validation sample

	Poverty line							
		National				Intl. 2005 PPP		
	Food	**100%**	**150%**	**200%**	**Median**	**$1.25**	**$2.00**	**$2.50**
Estimate minus true value	–0.5	–0.6	–0.8	–0.0	+0.2	–0.0	–0.5	–0.9
Precision of difference	0.5	0.6	0.5	0.4	0.5	0.1	0.4	0.5
α factor for precision	1.08	0.95	0.91	0.95	1.02	1.06	1.10	1.07

Precision is measured as 90-percent confidence intervals in units of ± percentage points.
Differences and precision estimated from 1,000 bootstraps with 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 9 (All poverty lines): Possible targeting outcomes

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

Figure 10 (100% of the national line): Shares of households by cut-off score and targeting classification, 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
≤4	0.0	47.8	0.0	52.2	52.2	–100.0
≤9	0.0	47.8	0.0	52.2	52.2	–100.0
≤14	0.0	47.8	0.0	52.2	52.2	–100.0
≤19	1.0	46.7	0.2	52.0	53.1	–95.3
≤24	3.0	44.7	0.6	51.7	54.7	–86.1
≤29	8.0	39.7	1.8	50.4	58.5	–62.6
≤34	14.8	32.9	4.2	48.0	62.9	–29.1
≤39	23.8	24.0	9.2	43.0	66.8	+19.0
≤44	32.6	15.2	16.3	35.9	68.4	+65.8
≤49	38.9	8.8	25.3	26.9	65.9	+47.1
≤54	43.4	4.4	33.8	18.4	61.8	+29.2
≤59	45.5	2.2	39.8	12.4	57.9	+16.6
≤64	46.5	1.3	44.4	7.9	54.4	+7.1
≤69	46.9	0.9	47.4	4.8	51.7	+0.8
≤74	47.4	0.4	49.6	2.7	50.0	–3.7
≤79	47.6	0.1	50.7	1.5	49.2	–6.1
≤84	47.8	0.0	52.1	0.1	47.9	–9.0
≤89	47.8	0.0	52.2	0.0	47.8	–9.3
≤94	47.8	0.0	52.2	0.0	47.8	–9.3
≤100	47.8	0.0	52.2	0.0	47.8	–9.3

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

Figure 11 (100% of the national line): Share of all households who are targeted (that is, score at or below the cut-off), the share of targeted households who are poor (that is, have consumption below the poverty line), the share of poor households who are targeted, and the number of poor households who are successfully targeted (inclusion) per non-poor household mistakenly targeted (leakage), scorecard applied to the validation sample

Targeting cut-off	% all HHs who are targeted	% targeted HHs who are poor	% poor HHs who are targeted	Poor HHs targeted per non-poor HH targeted
≤4	0.0	100.0	0.0	Only poor targeted
≤9	0.0	0.0	0.0	0.0:1
≤14	0.0	0.0	0.0	0.0:1
≤19	1.2	84.4	2.2	5.4:1
≤24	3.6	84.3	6.4	5.4:1
≤29	9.8	81.7	16.8	4.5:1
≤34	19.0	77.9	31.1	3.5:1
≤39	33.0	72.1	49.8	2.6:1
≤44	48.9	66.6	68.2	2.0:1
≤49	64.2	60.6	81.5	1.5:1
≤54	77.2	56.2	90.8	1.3:1
≤59	85.4	53.3	95.3	1.1:1
≤64	90.9	51.2	97.3	1.0:1
≤69	94.3	49.7	98.1	1.0:1
≤74	96.9	48.9	99.1	1.0:1
≤79	98.3	48.4	99.7	0.9:1
≤84	99.9	47.8	100.0	0.9:1
≤89	100.0	47.8	100.0	0.9:1
≤94	100.0	47.8	100.0	0.9:1
≤100	100.0	47.8	100.0	0.9:1

Tables for

the Food Poverty Line

Figure 3 (Food line): Estimated poverty likelihoods associated with scores

If a household's score is . . .	. . . then the likelihood (%) of being below the poverty line is:
0–4	48.2
5–9	48.2
10–14	48.2
15–19	48.2
20–24	39.0
25–29	37.8
30–34	28.3
35–39	19.2
40–44	13.3
45–49	9.7
50–54	8.2
55–59	3.9
60–64	3.6
65–69	3.3
70–74	1.3
75–79	0.7
80–84	0.0
85–89	0.0
90–94	0.0
95–100	0.0

Figure 6 (Food line): Average differences by score range between estimated and true poverty likelihoods for households, with confidence intervals, from 1,000 bootstraps of n = 16,384, 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	+48.2	0.0	0.0	0.0
10–14	+0.0	0.0	0.0	0.0
15–19	–13.9	9.9	10.5	11.4
20–24	–8.9	6.7	7.2	8.0
25–29	–0.7	2.7	3.1	4.2
30–34	+5.7	1.9	2.2	2.8
35–39	+0.6	1.4	1.6	2.0
40–44	–2.4	1.9	2.0	2.2
45–49	–1.1	1.2	1.4	1.7
50–54	–1.2	1.2	1.4	1.8
55–59	+0.8	0.8	1.0	1.3
60–64	+2.0	0.7	0.9	1.1
65–69	+2.0	0.6	0.7	1.0
70–74	–5.5	3.9	4.2	4.6
75–79	–1.5	1.5	1.7	2.2
80–84	+0.0	0.0	0.0	0.0
85–89	+0.0	0.0	0.0	0.0
90–94	+0.0	0.0	0.0	0.0
95–100	+0.0	0.0	0.0	0.0

Figure 7 (Food line): Average differences between estimated poverty rates and true values for a group at a point in time, with confidence intervals from 1,000 bootstraps of various sample sizes, scorecard applied to the validation sample

Sample Size	Difference between estimate and true value			
		Confidence interval (±percentage points)		
n	Diff.	90-percent	95-percent	99-percent
1	+0.1	62.2	64.0	72.2
4	+0.1	28.5	34.2	44.5
8	–0.2	20.9	24.6	31.3
16	–0.2	14.6	16.9	23.3
32	–0.3	11.0	12.7	15.6
64	–0.3	7.5	9.1	12.6
128	–0.5	5.3	6.5	8.7
256	–0.5	3.9	4.7	6.2
512	–0.4	2.8	3.3	4.5
1,024	–0.5	2.0	2.4	3.0
2,048	–0.5	1.4	1.6	2.1
4,096	–0.5	1.0	1.2	1.6
8,192	–0.5	0.7	0.8	1.1
16,384	–0.5	0.5	0.6	0.7

Figure 10 (Food line): Shares of households by cut-off score and targeting classification, 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
≤4	0.0	14.6	0.0	85.4	85.4	–100.0
≤9	0.0	14.6	0.0	85.4	85.4	–99.9
≤14	0.0	14.6	0.0	85.4	85.4	–99.9
≤19	0.7	13.9	0.6	84.8	85.5	–87.0
≤24	1.8	12.8	1.8	83.6	85.4	–63.0
≤29	4.0	10.6	5.8	79.6	83.7	–5.1
≤34	6.2	8.4	12.8	72.6	78.8	+12.5
≤39	9.0	5.6	24.0	61.4	70.4	–64.2
≤44	11.4	3.2	37.5	47.9	59.3	–156.6
≤49	12.9	1.7	51.4	34.0	46.9	–251.4
≤54	14.0	0.6	63.2	22.2	36.2	–332.3
≤59	14.2	0.4	71.1	14.3	28.5	–386.6
≤64	14.3	0.3	76.5	8.9	23.2	–423.5
≤69	14.4	0.2	79.9	5.5	19.9	–446.5
≤74	14.6	0.0	82.3	3.0	17.6	–463.3
≤79	14.6	0.0	83.7	1.7	16.3	–472.8
≤84	14.6	0.0	85.2	0.1	14.8	–483.2
≤89	14.6	0.0	85.4	0.0	14.6	–484.1
≤94	14.6	0.0	85.4	0.0	14.6	–484.1
≤100	14.6	0.0	85.4	0.0	14.6	–484.1

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

Figure 11 (Food line): Share of all households who are targeted (that is, score at or below the cut-off), the share of targeted households who are poor (that is, have consumption below the poverty line), the share of poor households who are targeted, and the number of poor households who are successfully targeted (inclusion) per non-poor household mistakenly targeted (leakage), scorecard applied to the validation sample

Targeting cut-off	% all HHs who are targeted	% targeted HHs who are poor	% poor HHs who are targeted	Poor HHs targeted per non-poor HH targeted
≤4	0.0	0.0	0.0	0.0:1
≤9	0.0	0.0	0.0	0.0:1
≤14	0.0	0.0	0.0	0.0:1
≤19	1.2	54.9	4.6	1.2:1
≤24	3.6	50.0	12.3	1.0:1
≤29	9.8	41.2	27.7	0.7:1
≤34	19.0	32.8	42.7	0.5:1
≤39	33.0	27.3	61.8	0.4:1
≤44	48.9	23.3	78.0	0.3:1
≤49	64.2	20.0	88.0	0.3:1
≤54	77.2	18.1	95.9	0.2:1
≤59	85.4	16.7	97.4	0.2:1
≤64	90.9	15.8	98.0	0.2:1
≤69	94.3	15.3	98.6	0.2:1
≤74	96.9	15.0	99.7	0.2:1
≤79	98.3	14.9	100.0	0.2:1
≤84	99.9	14.6	100.0	0.2:1
≤89	100.0	14.6	100.0	0.2:1
≤94	100.0	14.6	100.0	0.2:1
≤100	100.0	14.6	100.0	0.2:1

Tables for

150% of the National Poverty Line

Figure 3 (150% of the national line): Estimated poverty likelihoods associated with scores

If a household's score is . . .	. . . then the likelihood (%) of being below the poverty line is:
0–4	100.0
5–9	100.0
10–14	100.0
15–19	100.0
20–24	99.3
25–29	96.7
30–34	95.0
35–39	91.4
40–44	85.4
45–49	81.8
50–54	72.6
55–59	56.4
60–64	47.0
65–69	39.8
70–74	29.7
75–79	23.1
80–84	22.4
85–89	0.0
90–94	0.0
95–100	0.0

Figure 6 (150% of the national line): Average differences by score range between estimated and true poverty likelihoods for households, with confidence intervals, from 1,000 bootstraps of n = 16,384, 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	+0.0	0.0	0.0	0.0
10–14	+0.0	0.0	0.0	0.0
15–19	+13.4	4.0	4.9	6.1
20–24	+4.6	1.9	2.2	2.9
25–29	–2.1	1.3	1.3	1.4
30–34	–0.6	1.0	1.2	1.5
35–39	–0.2	1.0	1.2	1.6
40–44	–4.1	2.5	2.6	2.9
45–49	+5.0	1.4	1.8	2.3
50–54	+3.0	1.8	2.1	2.8
55–59	–4.7	3.5	3.7	4.1
60–64	–13.0	8.0	8.2	8.8
65–69	+2.9	3.6	4.2	5.4
70–74	–12.4	8.4	8.8	9.8
75–79	–9.6	7.4	8.0	9.2
80–84	+1.6	4.2	4.9	6.4
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 7 (150% of the national line): Average differences between estimated poverty rates and true values for a group at a point in time, with confidence intervals from 1,000 bootstraps of various sample sizes, scorecard applied to the validation sample

Sample Size	Difference between estimate and true value			
		Confidence interval (±percentage points)		
n	Diff.	90-percent	95-percent	99-percent
1	+0.9	67.4	75.8	86.3
4	–0.3	32.7	38.2	47.3
8	–0.8	22.6	26.9	33.2
16	–0.7	15.7	18.7	25.2
32	–0.7	11.6	13.6	17.4
64	–0.7	8.4	10.0	12.7
128	–0.6	5.7	6.7	9.1
256	–0.7	4.0	4.6	5.8
512	–0.8	2.7	3.3	4.5
1,024	–0.8	1.9	2.2	3.0
2,048	–0.7	1.4	1.7	2.1
4,096	–0.8	1.0	1.2	1.6
8,192	–0.8	0.7	0.8	1.1
16,384	–0.8	0.5	0.6	0.8

Figure 10 (150% of the national line): Shares of households by cut-off score and targeting classification, 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
≤4	0.0	77.8	0.0	22.2	22.2	–100.0
≤9	0.0	77.8	0.0	22.2	22.2	–100.0
≤14	0.0	77.8	0.0	22.2	22.2	–100.0
≤19	1.0	76.7	0.2	22.1	23.1	–97.1
≤24	3.3	74.5	0.3	21.9	25.2	–91.1
≤29	9.4	68.3	0.4	21.8	31.2	–75.3
≤34	18.2	59.5	0.8	21.4	39.7	–52.1
≤39	31.0	46.8	2.1	20.2	51.1	–17.7
≤44	45.2	32.6	3.7	18.5	63.7	+21.0
≤49	57.0	20.8	7.3	15.0	71.9	+55.9
≤54	66.1	11.6	11.1	11.2	77.3	+84.3
≤59	71.2	6.6	14.2	8.1	79.3	+81.8
≤64	74.3	3.4	16.5	5.7	80.0	+78.7
≤69	75.8	2.0	18.5	3.7	79.5	+76.2
≤74	76.9	0.8	20.0	2.3	79.2	+74.3
≤79	77.4	0.4	20.9	1.3	78.7	+73.1
≤84	77.8	0.0	22.1	0.1	77.9	+71.6
≤89	77.8	0.0	22.2	0.0	77.8	+71.4
≤94	77.8	0.0	22.2	0.0	77.8	+71.4
≤100	77.8	0.0	22.2	0.0	77.8	+71.4

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

Figure 11 (150% of the national line): Share of all households who are targeted (that is, score at or below the cut-off), the share of targeted households who are poor (that is, have consumption below the poverty line), the share of poor households who are targeted, and the number of poor households who are successfully targeted (inclusion) per non-poor household mistakenly targeted (leakage), scorecard applied to the validation sample

Targeting cut-off	% all HHs who are targeted	% targeted HHs who are poor	% poor HHs who are targeted	Poor HHs targeted per non-poor HH targeted
≤4	0.0	100.0	0.0	Only poor targeted
≤9	0.0	100.0	0.0	Only poor targeted
≤14	0.0	100.0	0.0	Only poor targeted
≤19	1.2	85.6	1.3	6.0:1
≤24	3.6	91.7	4.3	11.1:1
≤29	9.8	95.9	12.1	23.3:1
≤34	19.0	95.8	23.5	23.0:1
≤39	33.0	93.7	39.8	15.0:1
≤44	48.9	92.3	58.1	12.1:1
≤49	64.2	88.7	73.3	7.8:1
≤54	77.2	85.7	85.0	6.0:1
≤59	85.4	83.4	91.6	5.0:1
≤64	90.9	81.8	95.6	4.5:1
≤69	94.3	80.4	97.5	4.1:1
≤74	96.9	79.4	98.9	3.9:1
≤79	98.3	78.7	99.5	3.7:1
≤84	99.9	77.9	100.0	3.5:1
≤89	100.0	77.8	100.0	3.5:1
≤94	100.0	77.8	100.0	3.5:1
≤100	100.0	77.8	100.0	3.5:1

Tables for

200% of the National Poverty Line

Figure 3 (200% of the national line): Estimated poverty likelihoods associated with scores

If a household's score is . . .	. . . then the likelihood (%) of being below the poverty line is:
0–4	100.0
5–9	100.0
10–14	100.0
15–19	100.0
20–24	100.0
25–29	99.5
30–34	98.9
35–39	98.1
40–44	95.7
45–49	93.0
50–54	88.0
55–59	79.6
60–64	75.5
65–69	66.6
70–74	46.4
75–79	38.1
80–84	38.1
85–89	8.7
90–94	8.7
95–100	8.7

Figure 6 (200% of the national line): Average differences by score range between estimated and true poverty likelihoods for households, with confidence intervals, from 1,000 bootstraps of n = 16,384, 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	+0.0	0.0	0.0	0.0
10–14	+0.0	0.0	0.0	0.0
15–19	+0.0	0.0	0.0	0.0
20–24	+0.0	0.0	0.0	0.0
25–29	+0.3	0.4	0.5	0.6
30–34	–0.7	0.4	0.4	0.5
35–39	+1.8	0.7	0.8	1.1
40–44	+0.1	0.7	0.9	1.1
45–49	+0.3	0.9	1.1	1.4
50–54	–0.0	1.3	1.5	1.9
55–59	+1.0	2.1	2.5	3.2
60–64	–2.5	2.5	3.0	3.8
65–69	–0.8	3.4	4.1	5.2
70–74	–9.0	6.6	7.0	8.1
75–79	–12.1	9.0	9.5	10.2
80–84	+12.9	4.2	5.1	7.2
85–89	+3.0	5.1	6.7	8.7
90–94	+0.0	0.0	0.0	0.0
95–100	+0.0	0.0	0.0	0.0

Figure 7 (200% of the national line): Average differences between estimated poverty rates and true values for a group at a point in time, with confidence intervals from 1,000 bootstraps of various sample sizes, scorecard applied to the validation sample

Sample Size	Difference between estimate and true value			
		Confidence interval (±percentage points)		
n	Diff.	90-percent	95-percent	99-percent
1	+1.2	52.1	61.9	80.0
4	–0.3	24.3	29.8	39.2
8	–0.2	16.6	20.3	26.6
16	+0.1	11.3	13.8	19.2
32	+0.1	8.2	10.1	13.4
64	+0.0	6.2	7.1	10.0
128	+0.1	4.2	5.2	6.6
256	+0.1	3.0	3.5	4.5
512	+0.0	2.1	2.5	3.2
1,024	–0.0	1.5	1.8	2.3
2,048	+0.0	1.1	1.3	1.7
4,096	+0.0	0.8	1.0	1.2
8,192	–0.0	0.5	0.7	0.9
16,384	–0.0	0.4	0.5	0.6

Figure 10 (200% of the national line): Shares of households by cut-off score and targeting classification, 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
≤4	0.0	89.1	0.0	10.9	10.9	–100.0
≤9	0.0	89.0	0.0	10.9	11.0	–100.0
≤14	0.0	89.0	0.0	10.9	11.0	–100.0
≤19	1.2	87.8	0.0	10.9	12.2	–97.2
≤24	3.6	85.5	0.0	10.9	14.5	–91.9
≤29	9.8	79.3	0.1	10.9	20.6	–78.0
≤34	18.9	70.2	0.1	10.8	29.7	–57.4
≤39	32.4	56.7	0.6	10.3	42.7	–26.6
≤44	47.7	41.4	1.3	9.7	57.3	+8.4
≤49	61.9	27.2	2.4	8.6	70.5	+41.6
≤54	73.3	15.7	3.9	7.0	80.4	+69.0
≤59	79.8	9.3	5.6	5.4	85.2	+85.5
≤64	84.1	5.0	6.8	4.2	88.3	+92.4
≤69	86.4	2.7	7.9	3.0	89.4	+91.1
≤74	87.8	1.2	9.1	1.9	89.7	+89.8
≤79	88.6	0.5	9.8	1.2	89.7	+89.0
≤84	89.0	0.0	10.8	0.1	89.2	+87.9
≤89	89.1	0.0	10.9	0.0	89.1	+87.7
≤94	89.1	0.0	10.9	0.0	89.1	+87.7
≤100	89.1	0.0	10.9	0.0	89.1	+87.7

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

Figure 11 (200% of the national line): Share of all households who are targeted (that is, score at or below the cut-off), the share of targeted households who are poor (that is, have consumption below the poverty line), the share of poor households who are targeted, and the number of poor households who are successfully targeted (inclusion) per non-poor household mistakenly targeted (leakage), scorecard applied to the validation sample

Targeting cut-off	% all HHs who are targeted	% targeted HHs who are poor	% poor HHs who are targeted	Poor HHs targeted per non-poor HH targeted
≤4	0.0	100.0	0.0	Only poor targeted
≤9	0.0	100.0	0.0	Only poor targeted
≤14	0.0	100.0	0.0	Only poor targeted
≤19	1.2	100.0	1.4	Only poor targeted
≤24	3.6	100.0	4.0	Only poor targeted
≤29	9.8	99.4	11.0	161.0:1
≤34	19.0	99.4	21.2	155.1:1
≤39	33.0	98.1	36.4	50.4:1
≤44	48.9	97.4	53.5	37.9:1
≤49	64.2	96.3	69.5	26.3:1
≤54	77.2	95.0	82.3	18.9:1
≤59	85.4	93.5	89.6	14.4:1
≤64	90.9	92.6	94.4	12.4:1
≤69	94.3	91.6	97.0	10.9:1
≤74	96.9	90.6	98.6	9.7:1
≤79	98.3	90.1	99.4	9.1:1
≤84	99.9	89.2	100.0	8.2:1
≤89	100.0	89.1	100.0	8.1:1
≤94	100.0	89.1	100.0	8.1:1
≤100	100.0	89.1	100.0	8.1:1

Tables for

the Median Poverty Line

Figure 3 (Median line): Estimated poverty likelihoods associated with scores

If a household's score is . . .	. . . then the likelihood (%) of being below the poverty line is:
0–4	70.9
5–9	70.9
10–14	68.1
15–19	62.1
20–24	54.5
25–29	48.7
30–34	42.8
35–39	32.0
40–44	23.1
45–49	15.7
50–54	13.5
55–59	8.6
60–64	3.9
65–69	3.7
70–74	1.9
75–79	1.0
80–84	0.0
85–89	0.0
90–94	0.0
95–100	0.0

Figure 6 (Median line): Average differences by score range between estimated and true poverty likelihoods for households, with confidence intervals, from 1,000 bootstraps of n = 16,384, 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	+70.9	0.0	0.0	0.0
10–14	+0.0	0.0	0.0	0.0
15–19	–10.3	7.8	8.3	9.0
20–24	+3.5	4.3	5.4	6.8
25–29	+2.5	2.7	3.2	4.2
30–34	+7.0	2.1	2.6	3.4
35–39	+2.4	1.7	1.9	2.3
40–44	–0.2	1.5	1.7	2.2
45–49	–3.9	2.7	2.8	3.3
50–54	–1.5	1.4	1.7	2.0
55–59	+1.1	1.2	1.4	1.9
60–64	+0.4	1.1	1.3	1.8
65–69	+1.8	0.8	0.9	1.2
70–74	–8.1	5.4	5.8	6.3
75–79	–1.2	1.4	1.7	2.2
80–84	+0.0	0.0	0.0	0.0
85–89	+0.0	0.0	0.0	0.0
90–94	+0.0	0.0	0.0	0.0
95–100	+0.0	0.0	0.0	0.0

Figure 7 (Median line): Average differences between estimated poverty rates and true values for a group at a point in time, with confidence intervals from 1,000 bootstraps of various sample sizes, scorecard applied to the validation sample

Sample Size	Difference between estimate and true value			
		Confidence interval (±percentage points)		
n	Diff.	90-percent	95-percent	99-percent
1	+0.1	59.8	67.6	79.1
4	+1.4	33.4	38.2	50.6
8	+0.0	24.4	28.9	37.5
16	+0.3	16.8	19.6	26.6
32	+0.2	12.6	14.8	18.7
64	+0.3	8.9	11.0	14.5
128	+0.3	6.2	7.5	10.4
256	+0.3	4.3	5.0	7.1
512	+0.3	3.1	3.7	4.9
1,024	+0.3	2.2	2.5	3.5
2,048	+0.3	1.5	1.8	2.4
4,096	+0.2	1.1	1.3	1.8
8,192	+0.3	0.8	0.9	1.2
16,384	+0.2	0.5	0.6	0.9

Figure 10 (Median line): Shares of households by cut-off score and targeting classification, 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
≤4	0.0	22.7	0.0	77.1	77.1	–100.0
≤9	0.0	22.7	0.0	77.1	77.1	–99.9
≤14	0.0	22.7	0.0	77.1	77.1	–99.9
≤19	0.8	21.8	0.4	76.7	77.5	–91.0
≤24	2.0	20.6	1.5	75.6	77.6	–75.3
≤29	4.9	17.8	4.9	72.2	77.1	–35.3
≤34	8.3	14.3	10.6	66.5	74.8	+20.4
≤39	12.7	10.0	20.2	56.9	69.6	+10.6
≤44	16.5	6.1	32.2	44.9	61.4	–42.2
≤49	19.5	3.1	44.6	32.5	52.1	–96.7
≤54	21.4	1.2	55.6	21.5	43.0	–145.3
≤59	22.1	0.6	63.1	14.0	36.1	–178.4
≤64	22.3	0.4	68.4	8.7	31.0	–201.7
≤69	22.4	0.3	71.7	5.4	27.8	–216.4
≤74	22.6	0.0	74.1	3.0	25.7	–226.9
≤79	22.7	0.0	75.4	1.7	24.3	–233.0
≤84	22.7	0.0	77.0	0.1	22.8	–239.7
≤89	22.7	0.0	77.1	0.0	22.7	–240.3
≤94	22.7	0.0	77.1	0.0	22.7	–240.3
≤100	22.7	0.0	77.1	0.0	22.7	–240.3

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

Figure 11 (Median line): Share of all households who are targeted (that is, score at or below the cut-off), the share of targeted households who are poor (that is, have consumption below the poverty line), the share of poor households who are targeted, and the number of poor households who are successfully targeted (inclusion) per non-poor household mistakenly targeted (leakage), scorecard applied to the validation sample

Targeting cut-off	% all HHs who are targeted	% targeted HHs who are poor	% poor HHs who are targeted	Poor HHs targeted per non-poor HH targeted
≤4	0.0	100.0	0.0	Only poor targeted
≤9	0.0	0.0	0.0	0.0:1
≤14	0.0	0.0	0.0	0.0:1
≤19	1.2	66.3	3.6	2.0:1
≤24	3.6	56.5	9.0	1.3:1
≤29	9.8	49.7	21.5	1.0:1
≤34	19.0	43.7	36.7	0.8:1
≤39	33.0	38.4	56.0	0.6:1
≤44	48.9	33.8	73.0	0.5:1
≤49	64.2	30.4	86.2	0.4:1
≤54	77.2	27.7	94.5	0.4:1
≤59	85.4	25.8	97.4	0.3:1
≤64	90.9	24.5	98.2	0.3:1
≤69	94.3	23.7	98.7	0.3:1
≤74	96.9	23.3	99.8	0.3:1
≤79	98.3	23.0	100.0	0.3:1
≤84	99.9	22.7	100.0	0.3:1
≤89	100.0	22.7	100.0	0.3:1
≤94	100.0	22.7	100.0	0.3:1
≤100	100.0	22.7	100.0	0.3:1

Tables for

the $1.25/day 2005 PPP Poverty Line

Figure 3 ($1.25/day line): Estimated poverty likelihoods associated with scores

If a household's score is . . .	. . . then the likelihood (%) of being below the poverty line is:
0–4	18.2
5–9	18.2
10–14	13.6
15–19	3.6
20–24	3.6
25–29	2.1
30–34	2.0
35–39	0.7
40–44	0.7
45–49	0.7
50–54	0.5
55–59	0.3
60–64	0.2
65–69	0.0
70–74	0.0
75–79	0.0
80–84	0.0
85–89	0.0
90–94	0.0
95–100	0.0

Figure 6 ($1.25/day line): Average differences by score range between estimated and true poverty likelihoods for households, with confidence intervals, from 1,000 bootstraps of n = 16,384, 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	+18.2	0.0	0.0	0.0
10–14	+0.0	0.0	0.0	0.0
15–19	+2.2	1.0	1.2	1.6
20–24	–4.2	3.2	3.4	3.7
25–29	+1.3	0.4	0.5	0.6
30–34	+0.0	0.6	0.7	0.8
35–39	–0.4	0.3	0.4	0.5
40–44	–1.2	0.8	0.9	1.0
45–49	+0.7	0.0	0.0	0.0
50–54	+0.5	0.0	0.0	0.0
55–59	+0.3	0.0	0.0	0.0
60–64	+0.2	0.0	0.0	0.0
65–69	+0.0	0.0	0.0	0.0
70–74	+0.0	0.0	0.0	0.0
75–79	+0.0	0.0	0.0	0.0
80–84	+0.0	0.0	0.0	0.0
85–89	+0.0	0.0	0.0	0.0
90–94	+0.0	0.0	0.0	0.0
95–100	+0.0	0.0	0.0	0.0

Figure 7 ($1.25/day line): Average differences between estimated poverty rates and true values for a group at a point in time, with confidence intervals from 1,000 bootstraps of various sample sizes, scorecard applied to the validation sample

Sample Size	Difference between estimate and true value			
		Confidence interval (±percentage points)		
n	Diff.	90-percent	95-percent	99-percent
1	+0.0	1.0	1.8	50.4
4	–0.1	0.8	11.7	17.7
8	–0.2	6.4	7.7	10.3
16	–0.0	3.8	4.4	6.4
32	+0.0	2.3	2.8	4.4
64	–0.1	1.8	2.1	3.0
128	–0.1	1.4	1.7	2.0
256	–0.1	1.0	1.2	1.4
512	–0.0	0.7	0.9	1.0
1,024	–0.0	0.5	0.6	0.8
2,048	–0.0	0.4	0.4	0.6
4,096	–0.0	0.3	0.3	0.4
8,192	–0.0	0.2	0.2	0.3
16,384	–0.0	0.1	0.2	0.2

Figure 10 ($1.25/day line): Shares of households by cut-off score and targeting classification, 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
≤4	0.0	0.9	0.0	99.1	99.1	–100.0
≤9	0.0	0.9	0.0	99.1	99.1	–98.3
≤14	0.0	0.9	0.0	99.1	99.1	–98.3
≤19	0.0	0.9	1.2	97.9	97.9	–32.0
≤24	0.3	0.6	3.3	95.7	96.0	–270.7
≤29	0.3	0.6	9.5	89.6	89.9	–953.4
≤34	0.5	0.4	18.5	80.6	81.1	–1,950.6
≤39	0.7	0.2	32.4	66.7	67.4	–3,482.8
≤44	0.9	0.0	48.0	51.1	52.0	–5,216.4
≤49	0.9	0.0	63.3	35.8	36.7	–6,913.0
≤54	0.9	0.0	76.3	22.8	23.7	–8,349.2
≤59	0.9	0.0	84.5	14.6	15.5	–9,252.9
≤64	0.9	0.0	90.0	9.1	10.0	–9,860.3
≤69	0.9	0.0	93.4	5.7	6.6	–10,240.7
≤74	0.9	0.0	96.0	3.1	4.0	–10,531.4
≤79	0.9	0.0	97.4	1.7	2.6	–10,689.4
≤84	0.9	0.0	99.0	0.1	1.0	–10,857.7
≤89	0.9	0.0	99.1	0.0	0.9	–10,872.9
≤94	0.9	0.0	99.1	0.0	0.9	–10,872.9
≤100	0.9	0.0	99.1	0.0	0.9	–10,872.9

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

Figure 11 ($1.25/day line): Share of all households who are targeted (that is, score at or below the cut-off), the share of targeted households who are poor (that is, have consumption below the poverty line), the share of poor households who are targeted, and the number of poor households who are successfully targeted (inclusion) per non-poor household mistakenly targeted (leakage), scorecard applied to the validation sample

Targeting cut-off	% all HHs who are targeted	% targeted HHs who are poor	% poor HHs who are targeted	Poor HHs targeted per non-poor HH targeted
≤4	0.0	100.0	0.0	Only poor targeted
≤9	0.0	0.0	0.0	0.0:1
≤14	0.0	0.0	0.0	0.0:1
≤19	1.2	2.7	3.7	0.0:1
≤24	3.6	7.2	28.6	0.1:1
≤29	9.8	3.2	34.4	0.0:1
≤34	19.0	2.7	57.3	0.0:1
≤39	33.0	2.0	74.0	0.0:1
≤44	48.9	1.8	100.0	0.0:1
≤49	64.2	1.4	100.0	0.0:1
≤54	77.2	1.2	100.0	0.0:1
≤59	85.4	1.1	100.0	0.0:1
≤64	90.9	1.0	100.0	0.0:1
≤69	94.3	1.0	100.0	0.0:1
≤74	96.9	0.9	100.0	0.0:1
≤79	98.3	0.9	100.0	0.0:1
≤84	99.9	0.9	100.0	0.0:1
≤89	100.0	0.9	100.0	0.0:1
≤94	100.0	0.9	100.0	0.0:1
≤100	100.0	0.9	100.0	0.0:1

Tables for

the $2.00/day 2005 PPP Poverty Line

Figure 3 ($2.00/day line): Estimated poverty likelihoods associated with scores

If a household's score is . . .	. . . then the likelihood (%) of being below the poverty line is:
0–4	34.0
5–9	34.0
10–14	33.4
15–19	29.2
20–24	28.4
25–29	20.0
30–34	16.4
35–39	10.3
40–44	7.5
45–49	4.4
50–54	3.6
55–59	2.5
60–64	2.5
65–69	2.4
70–74	1.3
75–79	0.7
80–84	0.0
85–89	0.0
90–94	0.0
95–100	0.0

Figure 6 ($2.00/day line): Average differences by score range between estimated and true poverty likelihoods for households, with confidence intervals, from 1,000 bootstraps of n = 16,384, 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	+34.0	0.0	0.0	0.0
10–14	+0.0	0.0	0.0	0.0
15–19	+5.8	5.1	6.2	7.8
20–24	–6.5	5.4	5.6	6.5
25–29	–8.6	5.6	5.8	6.3
30–34	+1.1	1.6	1.9	2.5
35–39	–1.3	1.2	1.3	1.7
40–44	–1.0	1.0	1.2	1.5
45–49	–0.1	0.8	1.0	1.2
50–54	+1.7	0.5	0.6	0.7
55–59	+0.7	0.6	0.7	1.0
60–64	+2.5	0.0	0.0	0.0
65–69	+1.2	0.6	0.7	0.9
70–74	+1.3	0.0	0.0	0.0
75–79	–1.5	1.5	1.7	2.2
80–84	+0.0	0.0	0.0	0.0
85–89	+0.0	0.0	0.0	0.0
90–94	+0.0	0.0	0.0	0.0
95–100	+0.0	0.0	0.0	0.0

Figure 7 ($2.00/day line): Average differences between estimated poverty rates and true values for a group at a point in time, with confidence intervals from 1,000 bootstraps of various sample sizes, scorecard applied to the validation sample

Sample Size	Difference between estimate and true value			
		Confidence interval (±percentage points)		
n	Diff.	90-percent	95-percent	99-percent
1	–0.3	51.8	59.1	62.8
4	–0.2	21.7	27.3	38.1
8	–0.1	15.5	17.7	24.9
16	–0.2	11.9	13.6	17.5
32	–0.2	8.3	10.0	13.6
64	–0.3	6.2	7.7	9.8
128	–0.5	4.3	5.0	6.5
256	–0.5	3.1	3.9	4.8
512	–0.4	2.2	2.6	3.4
1,024	–0.5	1.6	1.9	2.5
2,048	–0.5	1.1	1.3	1.7
4,096	–0.5	0.8	0.9	1.2
8,192	–0.5	0.5	0.6	0.9
16,384	–0.5	0.4	0.5	0.6

Figure 10 ($2.00/day line): Shares of households by cut-off score and targeting classification, 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
≤4	0.0	8.2	0.0	91.8	91.8	–100.0
≤9	0.0	8.2	0.0	91.8	91.8	–99.8
≤14	0.0	8.2	0.0	91.8	91.8	–99.8
≤19	0.3	7.9	0.9	90.9	91.1	–81.6
≤24	1.1	7.1	2.5	89.3	90.4	–42.3
≤29	2.7	5.5	7.1	84.7	87.4	+13.6
≤34	4.2	4.0	14.8	77.0	81.2	–80.5
≤39	5.9	2.3	27.2	64.6	70.5	–231.1
≤44	7.1	1.1	41.8	50.0	57.1	–409.7
≤49	7.6	0.6	56.6	35.2	42.9	–589.9
≤54	7.9	0.3	69.3	22.5	30.5	–744.5
≤59	8.1	0.1	77.3	14.5	22.6	–842.1
≤64	8.1	0.1	82.8	9.0	17.1	–908.9
≤69	8.2	0.0	86.1	5.7	13.8	–950.0
≤74	8.2	0.0	88.8	3.0	11.2	–982.1
≤79	8.2	0.0	90.1	1.7	9.9	–999.0
≤84	8.2	0.0	91.7	0.1	8.3	–1,017.5
≤89	8.2	0.0	91.8	0.0	8.2	–1,019.2
≤94	8.2	0.0	91.8	0.0	8.2	–1,019.2
≤100	8.2	0.0	91.8	0.0	8.2	–1,019.2

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

Figure 11 ($2.00/day line): Share of all households who are targeted (that is, score at or below the cut-off), the share of targeted households who are poor (that is, have consumption below the poverty line), the share of poor households who are targeted, and the number of poor households who are successfully targeted (inclusion) per non-poor household mistakenly targeted (leakage), scorecard applied to the validation sample

Targeting cut-off	% all HHs who are targeted	% targeted HHs who are poor	% poor HHs who are targeted	Poor HHs targeted per non-poor HH targeted
≤4	0.0	0.0	0.0	0.0:1
≤9	0.0	0.0	0.0	0.0:1
≤14	0.0	0.0	0.0	0.0:1
≤19	1.2	23.1	3.4	0.3:1
≤24	3.6	31.2	13.7	0.5:1
≤29	9.8	27.9	33.4	0.4:1
≤34	19.0	22.2	51.6	0.3:1
≤39	33.0	17.8	71.5	0.2:1
≤44	48.9	14.5	86.6	0.2:1
≤49	64.2	11.9	93.3	0.1:1
≤54	77.2	10.3	96.8	0.1:1
≤59	85.4	9.5	98.7	0.1:1
≤64	90.9	8.9	98.7	0.1:1
≤69	94.3	8.7	99.5	0.1:1
≤74	96.9	8.4	99.5	0.1:1
≤79	98.3	8.3	100.0	0.1:1
≤84	99.9	8.2	100.0	0.1:1
≤89	100.0	8.2	100.0	0.1:1
≤94	100.0	8.2	100.0	0.1:1
≤100	100.0	8.2	100.0	0.1:1

Tables for

the $2.50/day 2005 PPP Poverty Line

Figure 3 ($2.50/day line): Estimated poverty likelihoods associated with scores

If a household's score is . . .	. . . then the likelihood (%) of being below the poverty line is:
0–4	54.3
5–9	54.3
10–14	54.3
15–19	54.3
20–24	44.7
25–29	43.7
30–34	34.5
35–39	24.8
40–44	17.4
45–49	11.1
50–54	11.1
55–59	6.5
60–64	3.9
65–69	3.6
70–74	1.6
75–79	1.0
80–84	0.0
85–89	0.0
90–94	0.0
95–100	0.0

Figure 6 ($2.50/day line): Average differences by score range between estimated and true poverty likelihoods for households, with confidence intervals, from 1,000 bootstraps of n = 16,384, 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	+54.3	0.0	0.0	0.0
10–14	+0.0	0.0	0.0	0.0
15–19	–11.8	8.8	9.3	10.4
20–24	–4.7	4.6	5.3	6.8
25–29	–0.1	2.7	3.1	4.2
30–34	+6.4	2.0	2.4	3.2
35–39	–0.7	1.6	1.9	2.4
40–44	–3.1	2.3	2.4	2.6
45–49	–3.6	2.4	2.6	2.9
50–54	–1.0	1.2	1.4	1.9
55–59	+1.4	1.0	1.2	1.6
60–64	+2.0	0.8	1.0	1.2
65–69	+2.3	0.6	0.7	1.0
70–74	–8.4	5.5	5.9	6.5
75–79	–1.2	1.4	1.7	2.2
80–84	+0.0	0.0	0.0	0.0
85–89	+0.0	0.0	0.0	0.0
90–94	+0.0	0.0	0.0	0.0
95–100	+0.0	0.0	0.0	0.0

Figure 7 ($2.50/day line): Average differences between estimated poverty rates and true values for a group at a point in time, with confidence intervals from 1,000 bootstraps of various sample sizes, scorecard applied to the validation sample

Sample Size	Difference between estimate and true value			
		Confidence interval (±percentage points)		
n	Diff.	90-percent	95-percent	99-percent
1	+0.3	63.1	66.3	73.9
4	+0.0	30.8	36.3	46.8
8	–0.7	22.9	26.4	36.0
16	–0.6	16.4	18.8	24.5
32	–0.7	11.7	14.0	17.6
64	–0.7	8.3	10.1	14.6
128	–0.8	5.9	6.9	9.5
256	–0.8	4.2	5.0	6.4
512	–0.8	3.0	3.5	4.6
1,024	–0.8	2.1	2.5	3.2
2,048	–0.9	1.5	1.8	2.4
4,096	–0.9	1.1	1.3	1.6
8,192	–0.9	0.8	0.9	1.1
16,384	–0.9	0.5	0.6	0.8

Figure 10 ($2.50/day line): Shares of households by cut-off score and targeting classification, 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
≤4	0.0	18.6	0.0	81.4	81.4	–100.0
≤9	0.0	18.6	0.0	81.4	81.4	–99.9
≤14	0.0	18.6	0.0	81.4	81.4	–99.9
≤19	0.7	17.8	0.5	80.9	81.7	–89.5
≤24	1.9	16.7	1.7	79.7	81.6	–70.3
≤29	4.5	14.0	5.3	76.1	80.7	–22.7
≤34	7.3	11.3	11.8	69.7	76.9	+36.6
≤39	10.9	7.7	22.1	59.3	70.2	–19.1
≤44	14.1	4.5	34.9	46.6	60.6	–87.8
≤49	16.1	2.4	48.1	33.3	49.4	–159.1
≤54	17.7	0.9	59.5	21.9	39.6	–220.6
≤59	18.1	0.5	67.3	14.1	32.2	–262.4
≤64	18.2	0.4	72.7	8.8	27.0	–291.3
≤69	18.3	0.3	76.0	5.4	23.7	–309.4
≤74	18.5	0.0	78.4	3.0	21.6	–322.2
≤79	18.6	0.0	79.8	1.7	20.2	–329.7
≤84	18.6	0.0	81.3	0.1	18.7	–337.8
≤89	18.6	0.0	81.4	0.0	18.6	–338.6
≤94	18.6	0.0	81.4	0.0	18.6	–338.6
≤100	18.6	0.0	81.4	0.0	18.6	–338.6

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

Figure 11 ($2.50/day line): Share of all households who are targeted (that is, score at or below the cut-off), the share of targeted households who are poor (that is, have consumption below the poverty line), the share of poor households who are targeted, and the number of poor households who are successfully targeted (inclusion) per non-poor household mistakenly targeted (leakage), scorecard applied to the validation sample

Targeting cut-off	% all HHs who are targeted	% targeted HHs who are poor	% poor HHs who are targeted	Poor HHs targeted per non-poor HH targeted
≤4	0.0	0.0	0.0	0.0:1
≤9	0.0	0.0	0.0	0.0:1
≤14	0.0	0.0	0.0	0.0:1
≤19	1.2	59.5	3.9	1.5:1
≤24	3.6	52.8	10.3	1.1:1
≤29	9.8	46.1	24.4	0.9:1
≤34	19.0	38.1	39.1	0.6:1
≤39	33.0	33.0	58.7	0.5:1
≤44	48.9	28.7	75.7	0.4:1
≤49	64.2	25.1	86.8	0.3:1
≤54	77.2	22.9	95.2	0.3:1
≤59	85.4	21.2	97.4	0.3:1
≤64	90.9	20.0	98.0	0.3:1
≤69	94.3	19.4	98.4	0.2:1
≤74	96.9	19.1	99.8	0.2:1
≤79	98.3	18.9	100.0	0.2:1
≤84	99.9	18.6	100.0	0.2:1
≤89	100.0	18.6	100.0	0.2:1
≤94	100.0	18.6	100.0	0.2:1
≤100	100.0	18.6	100.0	0.2:1