Simple Poverty Scorecard® Poverty-Assessment Tool Pakistan

Mark Schreiner

2 November 2006

This document is at SimplePovertyScorecard.com.

Abstract

The Simple Poverty Scorecard-brand poverty-assessment tool uses 10 low-cost indicators from the 2001 Pakistan Integrated Household Survey to estimate the likelihood that a household has consumption below a given poverty line. Field workers can collect responses in about ten minutes. Accuracy is reported for a range of poverty lines. The scorecard is a practical way for pro-poor programs in Pakistan to measure poverty rates, to track changes in poverty rates over time, and to segment clients for differentiated treatment.

Acknowledgements

This paper was commissioned by Grameen Foundation and funded by the Consultative Group to Assist the Poorest under the CGAP-Ford Social Indicators Project. Data was provided by the Government of Pakistan, Federal Bureau of Statistics, with help from Sadaffe Abid, Gretel Guzmán, and Maryam Khan. I am grateful as well to Nigel Biggar, Syed Hashemi, Frances Sinha, and Jeff Toohig. "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

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

Indicator	Response	Points	Score
1. Do all children ages 6 to 17 attend school?	A. No, or five or more children	0	
	B. Yes, and three or four children	10	
	C. Yes, and two children	15	
	D. Yes, and one child	20	
	E. No children ages 6 to 17	23	
2. What is the household's main source of drinking water?	A. Hand pump	0	
	B. Other	7	
3. Does the household own a refrigerator or freezer?	A. No	0	
	B. Yes	15	
4. What type of toilet is used by the household?	A. Other	0	
	B. Flush connected to pit	7	
	C. Flush connected to public sewer	10	
5. Does the household own a cooking stove?	A. No	0	
	B. Yes	9	
6. How many household members have salaried employment?	A. None	0	
	B. One, or two	3	
	C. Three or more	9	
7. Does the household own any type of land?	A. No	0	
	B. Yes	7	
8. If the household is rural, then does it own any buffaloes?	A. Rural, but no buffaloes	0	
	B. Urban (regardless of buffaloes)	1	
	C. Rural, and has buffaloes	4	
9. Does the household own any motorcycles or scooters?	A. No	0	
	B. Yes	11	
10. Does the household own any radios or cassette players?	A. No	0	
	B. Yes	5	

Score:

Simple Poverty Scorecard® Poverty-Assessment Tool
Pakistan

1. Introduction

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

Indicators were derived from the 15,503 households in the 2001 Pakistan Integrated Household Survey (PIHS). Selection criteria included:

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

All points are positive integers, and scores range from 0 (most-likely "poor") to 100 (least-likely "poor"). The scorecard is easy to understand, and field workers can compute scores by hand, on paper, in real time.

A participant's score corresponds to a "poverty likelihood", that is, the probability of being poor. For a group, the overall poverty rate (the so-called "head-count index") is the average poverty likelihood of the individuals in the group. For a group over time, progress (or regress) is the change in its average poverty likelihood.

The scorecard should qualify for certification for the reporting required of USAID's microenterprise partners. In particular, it is highly practical to use. Also, it accurately and objectively estimates the likelihood of having income below the national poverty line. With 90-percent confidence, a household's estimated poverty likelihood is accurate within 10 percentage points, and a group's estimated overall poverty rate is accurate within ±1.1 percentage points (n = 16,384).

2. Data and poverty lines

The analysis uses the 15,503 households in the 2001 PIHS from Pakistan's Federal Bureau of Statistics. This is the best, most recent household consumption survey for Pakistan. This paper divides households into three random samples (Figure 1), with one-half used for constructing the scorecard, one-fourth used for associating scores with estimated poverty likelihoods, and the final one-fourth used for measuring the accuracy of estimates derived from the scorecard.

Pakistan's official poverty line in 2000–01 was Rs748.57 per adult equivalent per month (World Bank, 2004). Although derived from age- and sex-specific caloric guidelines (Figure 2), the poverty line does not presume that people would spend all of their first Rs748.57 on food. Rather, it is the amount of consumption (on both food and non-food) observed among people who just meet the caloric guidelines. Each household's poverty line was taken as the sum of the individual poverty lines of its members.

Applying the official poverty line to the 2001 PIHS gives an overall poverty rate of 40.3 percent, or 51.3 million people. The rural poverty rate is 46.6 percent, while urban is 24.5 percent. This paper presents a single scorecard for use anywhere in Pakistan, as studies of India and Mexico (Schreiner, 2006 and 2005a) found only small returns to segmenting scorecards by rural and urban.

3. Scorecard construction

About 400 potential poverty indicators were prepared, including:

- Household and housing characteristics (such as cooking fuel and type of floor)
- Individual characteristics (such as age and highest grade completed)
- Household consumption (such as milk and meat)
- Household durable goods (such as electric fans and stoves)

Each indicator's ability to predict poverty was tested first with the entropy-based "uncertainty coefficient" (Goodman and Kruskal, 1979). This resembles a correlation coefficient, but it is applied to categorical indicators (such as "type of floor") rather than continuous ones (such as "square meters of floor space"). About 150 indicators were then selected for further analysis. Figure 3 lists the top 73, ranked by uncertainty coefficient. Responses are ordered by strength of association with poverty.

Many indicators in Figure 3 are similar in terms of their link with poverty. For example, households with a flush toilet connected to a public sewer are also more likely than other households to have piped water. If a scorecard already includes "flush toilet connected to public sewerage", then "piped drinking water" is more or less superfluous. Thus, many indicators strongly linked with poverty are not included because similar indicators are already included.

The scorecard also aims to measure *changes* in poverty through time. Thus, some powerful indicators (such as education of the female head/spouse) that are unlikely to change as poverty changes were omitted in favor of slightly less-powerful indicators (such as the presence of a radio) that are more likely to change. No indicators of past consumption (such as “In the past two weeks, did anyone in the household eat any tomatoes”) were selected because they cannot be directly observed nor verified.

Finally, some indicators were not selected because they are difficult to collect (“Have you received or contributed to Zakat, Usher, or Nazrana?”), difficult to compute (“What is the ratio of adults to children in the household?”) or too sensitive (“Who decides whether the female head/spouse uses contraception?”).

The scorecard itself was constructed using Logit regression. Indicator selection combined statistics with the judgment of an analyst with expertise in scoring and development. Starting with a scorecard with no indicators, each candidate indicator was added, one-by-one, to a one-indicator scorecard, using Logit to derive weights. The improvement in accuracy for each indicator was recorded using the “c” statistic.[1]

After all indicators had been tested, one was selected based on several criteria (Schreiner *et al.*, 2004; Zeller, 2004). These included the improvement in accuracy, the likelihood of acceptance by users (determined by simplicity, cost of collection, and “face validity” in terms of experience, theory, and common sense), the ability of the indicator

[1] Higher “c” indicates greater ability to rank households by poverty status. For a Logit regression with a categorical outcome (such as poor/not poor), “c” is a general measure of explanatory power, much like R^2 in a least-squares regression on a continuous outcome. “c” is equal to the Mann-Whitney statistic (also known as the Wilcoxon rank-sum statistic) that indicates how much two distributions overlap (here, the distributions are of the estimated poverty likelihoods for poor and non-poor households). “c” is also equivalent to the area under an ROC curve—discussed in more detail later—that plots the share of poor and non-poor households versus all households ranked by score. Finally, “c” can also be seen as the share of all possible pairs of poor and non-poor households in which the poor household has a lower score. The more often the poor household has the lower score, the better the ranking by poverty status.

to change values as poverty status changes over time, variety vis-à-vis other indicators already in the scorecard, and ease of observation/verification.

The selected indicator was then added to the scorecard, and the previous steps were repeated until 10 indicators were selected. Finally, the Logit coefficients were transformed into non-negative integers such that the lowest possible score is 0 (most likely poor) and the highest is 100.

This statistical algorithm is the Logit analogue to the stepwise "MAXR" in, for example, Zeller, Alcaraz and Johannsen (2005) and IRIS (2005a and 2005b). The procedure here diverges from naïve stepwise in that expert judgment and non-statistical criteria are used to select from among the most-predictive indicators. This improves robustness and, more importantly, helps ensure that the indicators are simple and sensible, increasing the likelihood of acceptance by users.

4. Scorecard use

As explained in Schreiner (2005b), the central challenge is not to maximize accuracy but rather to maximize the likelihood of programs' using scoring appropriately. When scoring projects fail, the culprit is usually not inaccuracy but rather the failure of users to accept scoring and to use it properly (Schreiner, 2002). The challenge is not technical but human and organizational, not statistics but change management. Accuracy is easier—and less important—than practicality.

The scorecard here was designed to help users to understand and trust it (and thus use it properly). While accuracy matters, it must be balanced against simplicity, ease-of-use, and "face validity". In particular, programs are more likely to collect data, compute scores, and pay attention to the results if, in their view, scoring avoids creating "extra" work and if the whole process generally seems to make sense.

This practical focus naturally leads to a one-page scorecard that allows field workers to score households by hand in real time because it features:

- Only 10 indicators
- Only observable, categorical indicators ("flooring material", not "value of house")
- User-friendly weights (non-negative integers, no arithmetic beyond simple addition)

Among other things, this simplicity enables "rapid targeting", such as determining (in a day) who in a village qualifies for, say, work-for-food, or ration cards.

The scorecard is ready to be photocopied. A field agent collecting data and computing scores on paper would:

- Read each question off the scorecard
- Circle the response and the corresponding points
- Write the points in the far-right column
- Add up the points to get the total score
- Implement program policy based on the score

4.1 Scores and poverty likelihoods

A score is not a poverty likelihood (that is, the probability of being poor), but each score is associated with an estimated poverty likelihood via a simple table (Figure 5). For example, scores of 20–24 correspond to a poverty likelihood of 57.6 percent.

Scores (sums of weights) are associated with estimated poverty likelihoods (probabilities of being poor) via the "bootstrap" (Efron and Tibshirani, 1993):

- From the first one-fourth hold-out sample, draw a new sample of the same size *with replacement*
- For people in a given score range, compute the share who are poor
- Repeat the previous two steps 10,000 times
- For a given score range, define the poverty likelihood as the average of the shares of people who are poor in that score range across the 10,000 samples

These resulting poverty likelihoods are objective, that is, based on data. This process would produce objective poverty likelihoods *even if the scorecards themselves were constructed without data*. In fact, scorecards of objective, proven accuracy are often constructed *only* with qualitative judgment (Fuller, 2006; Caire, 2004; Schreiner *et al.*, 2004). Of course, the scorecard here uses data. While its construction—like any statistical analysis—was partially informed by the analyst's judgment, the explicit acknowledgment of this fact is irrelevant for the objectivity of the poverty likelihoods. After all, objectively depends on using data to associate scores with poverty likelihoods, not on pretending to avoid the use of judgment during scorecard construction.

Figure 6 depicts the precision of estimated poverty likelihoods as point estimates with 90-, 95-, and 99-percent confidence intervals. This is a standard, widely understood way to measure accuracy. The confidence intervals here were derived empirically from the 10,000 bootstrap samples described above. For a given score, the lower (upper) bound on the x-percent confidence interval is the value less (greater) than $(100–x)/2$ percent ($(100+x)/2$ percent) of the bootstrapped likelihoods.

For example, the average poverty rate across bootstrap samples for people with scores of 20–24 is 57.6 percent (this is the poverty likelihood in Figure 5). In 90 percent of samples, the poverty rate is between 52.8–62.4 percent (Figure 6). In 95 percent of samples, the share is 51.9–63.2; in 99 percent of samples, the share is 49.9–64.9.

For estimated and true poverty likelihoods, Figure 7 depicts mean absolute differences and confidence intervals from bootstrapping the second one-fourth hold-out sample from the 2001 PIHS. The mean absolute difference is 5.8 percentage points.

This discussion so far looks at whether estimated poverty likelihoods are close to true poverty likelihoods. There is another aspect of accuracy, one associated with targeting: how well the poor are concentrated in low scores. A perfect scorecard would assign all the lowest scores to poor people (and all the highest scores to non-poor people). In reality, no scorecard is perfect, so some poor people have high scores, and vice versa.

ROC curves are standard tools for showing how well the poor are concentrated in lower scores (Baulch, 2003; Wodon, 1997). They plot the share of poor and non-poor households against the share of all households ranked by score.

What does the ROC curve in Figure 8 mean? Suppose a program sets a cut-off so as to target the lowest-scoring x percent of people. The ROC curve then shows the share of the poor (northwest curve) and non-poor (southwest curve) targeted. Greater ability to rank-order—with less leakage and less undercoverage—is shown by curves that are closer to the northwest and southeast corners of the graph.

In Figure 8, the northwest (southeast) curve depicts accuracy among the poor (non-poor). As a benchmark, the external trapezoid shows the accuracy of a hypothetical perfect scorecard that assigns all of the lowest scores to poor people. The diagonal line represents random targeting.

The curves for the scorecard show, for example, that targeting the 20 percent of households with the lowest scores would target 37 percent of all the poor and 8 percent of all the non-poor. In contrast, randomly targeting 20 percent of cases would target 20 percent of the poor and 20 percent of the non-poor.

Figure 8 also reports two other common measures of rank-ordering. The first is the Kolmogorov-Smirnov (KS) statistic, defined as the maximum distance between the poor and non-poor curves (here 47.5). Higher KS implies better rank-ordering.

The second measure is the ratio of the area inside the ROC curves to the area inside the trapezoid of a hypothetical perfect scorecard (here 60.5). Again, greater area within the curves implies better rank-ordering.

4.2 Estimates of overall poverty rates

The estimated overall poverty rate is the average of the estimated poverty likelihoods of individuals.

For example, suppose a program has 3,000 participants on Jan. 1, 2006 and that 1,000 have scores of 20, 1,000 have scores of 30, and 1,000 have scores of 40. The poverty likelihoods that correspond to these scores are 57.6, 36.2, and 17.9 percent (Figure 5). The overall poverty rate is the participants' average poverty likelihood, that is, 1,000 x (57.6 + 36.2 + 17.9) ÷ 3,000 = 37.2 percent.

To test accuracy and precision, the scorecard was applied to 10,000 bootstrap replicates from the second one-fourth hold-out sample, comparing the estimated overall poverty rates with the true values. The mean difference was 4.3 percentage points, with a standard deviation of 0.68. The 90-percent confidence interval around the mean was ±1.1 percentage points, the 95-percent interval was ±1.3 percentage points, and the 99-percent interval was ±1.7 percentage points.

In practice, this means that subtracting 4.3 percentage points from a group's average poverty likelihood would produce an unbiased estimate that, in 99 of 100 cases, would be within ±1.7 percentage points of the true overall poverty rate.

4.3 Change over time

For a given group, change in poverty over time is estimated as the change in the average poverty likelihood.

Continuing the previous example, suppose that on Jan. 1, 2007, the same 3,000 people (some of whom may no longer be participants) are now in groups of 500 with scores of 20, 25, 30, 35, 40, and 45 (by Figure 5, poverty likelihoods of 57.6, 34.5, 36.2, 24.6, 17.9, and 13.5 percent). Their average poverty likelihood is now 30.7 percent, an improvement of 37.2 – 30.7 = 6.5 percentage points. In other words, 6.5 of every 100 in this group left poverty. Among those who were poor to start with, about one in six (6.5 ÷ 37.2 = 17.4 percent) left poverty.

Of course, the scorecard does not indicate what *caused* progress; it just measures the change, regardless of cause.

5. Setting targeting cut-offs

How can the scorecard be used for targeting? Potential participants with scores at or below a targeting cut-off are labeled *targeted* and treated—for program purposes—as if they were poor. Those with higher scores are *non-targeted* and treated—again, for program purposes—as if they were non-poor.

Poverty status (consumption below a poverty line) is distinct from *targeting status* (score below a cut-off). Poverty status is a fact whose determination requires an expensive survey. In contrast, targeting status is a policy choice whose determination requires a cut-off and an inexpensive estimate of poverty likelihood. Indeed, the purpose of scoring is to infer poverty status without incurring the cost of direct measurement.

No scorecard is perfect, so some of the truly poor will not be targeted, and some of the truly non-poor will be targeted. Targeting is accurate to the extent that poverty status matches targeting status. In turn, this depends on the selection of a targeting cut-offs and how it balances accuracy for the poor versus non-poor. The standard approach uses a *classification matrix* and a *net-benefit matrix* (Adams and Hand, 2000; Hoadley and Oliver, 1998; Greene, 1993).

5.1 Classification matrix

Given a targeting cut-off, there are four possible classification results:

A.	Truly poor	correctly	targeted	(score at or below the cut-off)
B.	Truly poor	mistakenly	non-targeted	(score above cut-off)
C.	Truly non-poor	mistakenly	targeted	(score at or below cut-off)
D.	Truly non-poor	correctly	non-targeted	(score above cut-off)

These four possibilities can be shown as a general classification matrix (Figure 9). Accuracy improves as there are more cases in A and D and fewer in B and C.

Figure 10 shows the number of people in each classification by score in the second one-fourth hold-out sample. For example, with a cut-off of 20–24, there are:

A.	32.8	truly poor	correctly	targeted
B.	10.9	truly poor	mistakenly	non-targeted
C.	13.2	truly non-poor	mistakenly	targeted
D.	43.1	truly non-poor	correctly	non-targeted

Targeting accuracy (and errors of undercoverage and leakage) depends on the cut-off. For example, if the cut-off were increased to 25–29, more poor (but less non-poor) are correctly targeted:

A.	35.9	truly poor	correctly	targeted
B.	7.8	truly poor	mistakenly	non-targeted
C.	19.0	truly non-poor	mistakenly	targeted
D.	37.3	truly non-poor	correctly	non-targeted

Whether a cut-off of 20–24 is preferred to 25–29 depends on net benefit.

5.2 Net-benefit matrix

Each of the four classification results is associated with a net benefit (Figure 11):

α. Benefit	per truly poor person	correctly	targeted
β. Cost (negative net benefit)	per truly poor person	mistakenly	non-targeted
γ. Cost (negative net benefit)	per truly non-poor person	mistakenly	targeted
δ. Benefit	per truly non-poor person	correctly	non-targeted

Each net benefit α, β, γ, and δ corresponds to one of the quadrants in the general classification matrix in Figure 9. For example, α is the net benefit associated with each truly poor person who is correctly targeted in quadrant A, and β is the cost (negative net benefit) associated with each truly poor person incorrectly targeted in quadrant B.

Given a net-benefit matrix and a classification matrix, total net benefit is the sum of the net benefit per person in each quadrant multiplied by the number of people in the quadrant, summed across all four quadrants:

$$\text{Total net benefit} = \alpha \cdot A + \beta \cdot B + \gamma \cdot C + \delta \cdot D.$$

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

- Select a net-benefit matrix based on its values and mission
- Compute total net benefits for each cut-off with the net-benefit matrix and Figure 10
- Select the cut-off with the highest total net benefit

The only non-trivial step is selecting a net-benefit matrix. Some common net-benefit matrices are discussed below. In general, however, each program should thoughtfully decide for itself how much it values successful targeting versus errors of undercoverage and leakage. Of course, any program that targets already uses (if only implicitly) a net-benefit matrix. It is healthy to go through a process of thinking explicitly and intentionally about the value of possible targeting outcomes.

For example, suppose a program places great importance on correctly targeting the poor, even at the cost of accidentally targeting more non-poor. It could reflect this valuation by increasing the weight on quadrant A (by increasing its net benefit α), and/or by decreasing the weight on quadrant B (by decreasing its net benefit β). The examples of net-benefit matrices discussed next represent different valuations of correctly/incorrectly targeting the poor/non-poor.

5.2.1 "Total Accuracy"

As an example, suppose a program selects the net-benefit matrix that corresponds to the "Total Accuracy" criterion (Figure 12, IRIS, 2005b). Then total net benefit is the number of people correctly classified:

$$\begin{aligned} \text{Total net benefit} \quad &= 1 \cdot \text{A} + 0 \cdot \text{B} + 0 \cdot \text{C} + 1 \cdot \text{D}, \\ &= \text{A} + \text{D}. \end{aligned}$$

This values correct classifications of the poor and non-poor equally. Grootaert and Braithwaite (1998) and Zeller, Alcaraz, and Johannsen (2005) use "Total Accuracy" to evaluate their poverty-assessment tools.

Figure 13 shows "Total Accuracy" for all cut-offs. Total net benefit is greatest (75.9) for a cut-off of 20–24; at that point, poverty segment matches poverty status for three out of four people.

"Total Accuracy" weighs the poor and non-poor the same. If most people are non-poor and/or if a scorecard is more accurate for the non-poor, then "Total Accuracy" might look good even if few poor people are correctly classified. Development programs, however, probably value correct targeting more for the poor than for the non-poor.

A simple, transparent way to reflect this valuation is to increase the relative net benefit α of correctly classifying the poor. For example, if a program values correctly targeting the poor twice as much as correctly not targeting the non-poor, then α should be set twice as high as δ in the net-benefit matrix. Then the new optimal cut-off is 30–34, the cut-off point where $\alpha.A + \delta.D = 2.A + D$ is highest.

5.2.2 "Poverty Accuracy"

A criterion that values only correctly classifying the poor is "Poverty Accuracy" (Figure 14, IRIS, 2005b):

$$\begin{aligned} \text{Total net benefit} \quad &= 1 \cdot A + 0 \cdot B + 0 \cdot C + 0 \cdot D, \\ &= A. \end{aligned}$$

Of course, correctly targeting the poor is rarely the sole criteria. In fact, Figure 13 shows that "Poverty Accuracy" is greatest with a cut-off of 95–100. While targeting everyone does ensure that all poor people are targeted and so minimizes *undercoverage* of the poor (second-to-last column of Figure 13), it also targets all the non-poor and so maximizes *leakage* (the last column of Figure 13).

5.2.3 "Non-poverty Accuracy"

"Non-poverty Accuracy" counts only correct classifications of the non-poor (total net benefit is D). This is maximized by setting a cut-off of 0–4 and thus not targeting anyone (minimum leakage but maximum undercoverage).

5.2.4 "BPAC"

IRIS (2005b) proposes a new measure of accuracy called the "Balanced Poverty Accuracy Criterion". BPAC balances two goals:

- Accuracy of the estimated overall poverty rate
- "Poverty Accuracy"

According to IRIS (2005b), the first goal is optimized when undercoverage B is balanced by leakage C, and the second goal is optimized by maximizing A. If B > C, then BPAC's net-benefit matrix is Figure 15. In essence, BPAC maximizes A while making B and C as close to each other as possible:

$$\begin{aligned} \text{Total net benefit} \quad &= 1 \cdot \text{A} + 1 \cdot \text{B} + (-1) \cdot \text{C} + 0 \cdot \text{D}, \\ &= \text{A} + (\text{B} - \text{C}). \end{aligned}$$

If C > B, then total net benefit under BPAC is A + (C – B).

BPAC was invented because IRIS does not estimate poverty likelihoods. Instead, IRIS estimates consumption and then labels as poor those households with estimated consumption less than the poverty line. In this set-up, the overall poverty rate is estimated as the share of people targeted, and this estimate is most accurate (that is, closest to the true value) when undercoverage B equals leakage C.

For a scorecard (like the one here) that estimates poverty likelihoods, however, BPAC is not meaningful. This is because the estimated overall poverty rate is the average of participants' estimated poverty likelihoods. These estimates are independent of whatever targeting cut-off a program might set. In contrast, the targeting errors of undercoverage B and leakage C depend directly on the cut-off. Thus, for scorecards that estimate poverty likelihoods, getting B close to C is not related to optimizing the accuracy of the estimated overall poverty rate and so is not related to BPAC's goals.

6. Training, quality-control, and MIS

The technical aspects of scorecard construction and accuracy just discussed are important, but gaining the trust and acceptance of managers and field workers is even more important (Schreiner, 2002).

In particular, the field workers who collect indicators must be trained. If they put garbage in, the scorecard will put garbage out. To prevent abuse, on-going quality control of data is required.

Programs should record in their MIS at least the poverty likelihood along with an identifier for each client. Ideally, they would also record the score, the indicators, and the values of the indicators. This will allow quick computation of average poverty likelihoods (as well as other analyses), both for a point in time and for changes through time (Matul and Kline, 2003).

7. Calibrating the scorecard for the very poor

The scorecard can be used to track outreach not only to the poor but also to the *very poor*, that is, the poorest half of the poor below the national poverty line. This is the relevant group for USAID certification.

7.1 Poverty likelihoods

As before, scores are associated with the probability of being very poor by bootstrapping 10,000 samples from first one-fourth hold-out sample from the 2001 PIHS. The poverty likelihood for a given score is then taken as the average of the shares of people with that score who are very poor across the 10,000 samples.

Columns 2–4 in Figure 16 are the poverty likelihoods for the three classes for all scores. For example, if a potential participant has a score of 10–14, the probability of being very poor is 44.5 percent, the probability of being poor is 24.6 percent, and the probability of being non-poor is 30.9 percent.

Columns 5–7 in Figure 16 are the share of targeted participants by poverty status and by cut-off. For example, for a cut-off of 10–14, 52.5 percent of those targeted would be very poor, 23.9 percent would be poor, and 23.6 percent would be non-poor.

Each person is associated with three poverty likelihoods. For example, a person with a score of 10 may be targeted as very poor, but the likelihood of truly being very poor is not 100 percent but rather 44.5 percent (from Figure 16). The same person has a 24.6-percent likelihood of being truly poor, and a 30.9-percent likelihood of being truly non-poor. Each person has one targeting status (for program purposes), one true poverty status (in reality), and three estimated poverty likelihoods (one for each possible poverty status).

As before, these poverty likelihoods are objective, that is, based on data. They are valid even though the scorecard was not constructed originally to predict the likelihood of being very poor. It works because the likelihood of being very poor is highly correlated with having a low score (high likelihood of being poor). A scorecard could be built specifically for the very poor, but it would add cost and complexity.

Figure 17 shows the precision of estimated poverty likelihoods for being very poor as point estimates with 90-, 95-, and 99-percent confidence intervals. For example, the average poverty rate (the poverty likelihood) across bootstrap samples for people with scores of 10–14 was 44.5 percent. In 90 percent of 10,000 bootstraps from the second one-fourth hold-out sample, the share was between 40.2–48.9 percent. In 95 percent of samples, the share was between 39.3–49.8, and in 99 percent of samples, the share was between 37.7–51.6.

For estimated and true poverty likelihoods, Figure 18 depicts mean absolute differences and confidence intervals from 10,000 bootstraps on the second one-fourth hold-out sample. Weighting by the people in a score range, the mean absolute difference is 4.9 percentage points, with a 90-percent interval of ±3.6 percentage points.

The other aspect of accuracy is how well the very poor are concentrated in low scores. Once again, an ROC curve is a useful way to look at this.

Figure 19 plots the share of the very poor against the share of the not very poor, ranked by score. For example, targeting the 25 percent of cases with the lowest scores would target 53 percent of all the very poor and 18 percent of all the not very poor.

In terms of the Kolmogorov-Smirnov statistic, the maximum distance between the curves is 46.6. In terms of the ratio of the area inside the scorecard curves to the area inside the trapezoid of a hypothetical perfect scorecard, the value is 57.4.

7.2 Overall poverty rates for the very poor

The average of estimated poverty likelihoods for a group is their estimated overall (very poor) poverty rate. To measure the accuracy and precision of this estimate, the scorecard was applied to 10,000 bootstrap replicates from the second one-fourth hold-out samples from the 2001 PIHS, and then the estimated overall poverty rates were compared with the true values. The mean difference was 5.7 percentage points, with a standard deviation of 0.60. The 90-percent confidence interval around the mean was ±1.0 percentage points, the 95-percent interval was ±1.2 percentage points, and the 99-percent interval was ±1.5 percentage points.

Thus, subtracting 5.7 percentage points to a group's average poverty likelihood would produce an unbiased estimate that, in 99 of 100 cases, would be within ±1.5 percentage points of the true overall (very poor) poverty rate. This estimate is both accurate and precise.

7.3 Targeting the very poor

As before, targeting involves using a classification matrix and a net-benefit matrix to select a cut-off. The wrinkle is that there are now three poverty statuses:

- Very poor: Poorest half of those with consumption at or below the poverty line
- Poor: Least-poor half of those with consumption at or below poverty
- Non-poor: Consumption above poverty

There are also three targeting segments:

- Very poor: Score at or below the very poor/poor cut-off
- Poor: Score above the very poor/poor cut-off and at or below the poor/non-poor cut-off
- Non-poor: Score above the poor/non-poor cut-off

There are two cut-offs (very poor/poor and poor/non-poor) and 9 classification results (Figure 20):

A.	Truly very poor	correctly targeted as	very poor
B.	Truly very poor	incorrectly targeted as	poor
C.	Truly very poor	incorrectly targeted as	non-poor
D.	Truly poor	incorrectly targeted as	very poor
E.	Truly poor	correctly targeted as	poor
F.	Truly poor	incorrectly targeted as	non-poor
G.	Truly non-poor	incorrectly targeted as	very poor
H.	Truly non-poor	incorrectly targeted as	poor
I.	Truly non-poor	correctly targeted as	non-poor

The general classification matrix (Figure 20) and the net-benefit matrix (Figure 21) are combined as before to define total net benefit:

Total net benefit = $\alpha \cdot A + \beta \cdot B + \gamma \cdot C + \delta \cdot D + \varepsilon \cdot E + \zeta \cdot F + \eta \cdot G + \theta \cdot H + \iota \cdot I$.

Figure 22 shows classification results for all possible pairs of cut-off scores in the second one-fourth hold-out sample. For example, suppose a program defined:

- Very poor/poor cut-off of 10–14 (so scores of 0–14 are targeted as very poor)
- Poor/non-poor cut-off of 20–24 (so scores of 15–24 are targeted as poor, and scores of 25–100 are targeted as non-poor)

As with any scorecard and cut-offs, there is both successful targeting and errors. For the example cut-offs of 10–14 and 20–24, targeting would be correct for 60 percent of the very poor, 25 percent of the poor, and 76 percent of the non-poor (Figure 23).

The program chooses a set of cut-offs to optimize the benefits of correct classifications, net of the costs (negative benefits) of incorrect classifications. For example, suppose the net-benefit matrix is Figure 24, representing one way to reflect:

- Greater importance of correctly targeting the very poor and poor
- Greater cost of gross errors such as targeting the truly very poor as non-poor

Given the classification results in Figure 23 and net benefits in Figure 24, total net benefit for the cut-off pair of 10–14 and 20–24 is +584 (Figure 25).

Is this the best pair of cut-offs? The answer requires applying the net-benefit matrix to the classification results for all 190 possible pairs (Figure 22). It turns out that total net benefit is highest for cut-offs 20–24 and 30–34, giving a net benefit of 952.

8. Conclusion

Pro-poor programs in Pakistan can use the scorecard to segment clients for differentiated services 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 between two points in time

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

The scorecard is built and tested using data on 15,503 households from the 2001 PIHS. The scorecard is calibrated to estimate the likelihood of being poor (consumption below the national poverty line) or very poor (poorest half of the poor).

Out-of-sample bootstrap tests show that the estimates are both accurate and precise. For individual poverty likelihoods (whether poor or very poor), estimates are within 10 percentage points of the true value with 90-percent confidence (n = 16,384). For a group's overall poverty rate (again, whether poor or very poor), estimates are within 1.1 percentage points of the true value.

For targeting, programs can use the classification results reported here to select the best cut-off for their particular values and mission.

Accuracy is important, but ease-of-use is even more important; a perfectly accurate scorecard is worthless if programs feel daunted by its complexity and so never even try to use it. For this reason, the scorecard here is kept simple, using 10 indicators that are inexpensive to collect and that are straightforward to observe and verify. Points are either zeros or positive integers, and scores range from 0 (most likely poor) to 100 (least likely poor). Scores are related to poverty likelihoods via a simple look-up table, and targeting cut-offs are also simple to apply. Thus, users can not only understand the scorecard, but they can also use it to compute scores in the field, by hand, in real time.

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

References

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

Baulch, Bob. (2003) "Poverty Monitoring and Targeting Using ROC Curves: Examples from Vietnam", IDS Working Paper No. 161, `www.ids.ac.uk/ids/bookshop/wp/wp161.pdf`.

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

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

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

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

Greene, William H. (1993) *Econometric Analysis: Second Edition*.

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

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

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

Matul, Michal; and Sean Kline. (2003) "Scoring Change: Prizma's Approach to Assessing Poverty", MFC Spotlight Note No. 4, www.mfc.org.pl/doc/Research/ImpAct/SN/MFC_SN04_eng.pdf.

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

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

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

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

_____; 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.

Wodon, Quentin T. (1997) "Targeting the Poor Using ROC Curves", *World Development*, Vol. 25, No. 12, pp. 2083–2092.

World Bank. (2004) "Pakistan: Joint Staff Assessment of the Poverty Reduction Strategy Paper", Report No. 27625–PK.

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

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

Figure 1: Households surveyed, people represented, and overall poverty rates

Sub-sample	Households	People	% poor
Constructing scorecards	7,715	63,942,568	40.2
Associating scores with likelihoods	3,939	32,844,834	41.5
Testing accuracy	3,849	30,385,549	39.4
Source: 2001 PIHS.	15,503	127,172,951	40.3

Figure 2: Equivalence factors for age/sex-specific official poverty lines

Sex/age	Calories/day	Equivalence factor
Children		
<1	1,010	0.4298
1-4	1,304	0.5549
5-9	1,768	0.7523
Males		
10-14	2,816	1.1983
15-19	3,087	1.3136
20-39	2,760	1.1745
40-49	2,640	1.1234
50-59	2,460	1.0468
60 or more	2,146	0.9132
Females		
10-14	2,464	1.0485
15-19	2,322	0.9881
20-39	2,080	0.8851
40-49	1,976	0.8409
50-59	1,872	0.7966
60 or more	1,632	0.6945
National average:	2,350	1.0000

Source: World Bank, 2004.

Figure 3: Poverty indicators ranked by uncertainty coefficient

Uncertainty coefficient		Indicator (Responses ordered starting from the one most-closely linked with poverty)
1.	95	Do all children of ages 6 to 17 attend school? (No; Yes; No children in this age range)
2.	84	Does the household own a refrigerator or freezer? (No; Yes)
3.	83	How many people of ages 0 to 17 live in the household? (7 or more; 6; 5; 4; 3; 2; 1; 0)
4.	77	What type of toilet is used by the household? (All others; Flush connected to pit; Flush connected to public sewerage)
5.	76	Does the household own a washing machine or clothes dryer? (No; Yes)
6.	74	Do all girls of ages 6 to 17 attend school? (No; Yes; No girls in this age range)
7.	71	Do all boys of ages 6 to 17 attend school? (No; Yes; No boys in this age range)
8.	68	What is the highest education level completed by any family member? (None or grades 1 to 3; Grades 4 to 7; Grades 8 to 9; Grade 10; Grade 11 or more)
9.	66	Does the household own any fans (ceiling, table, pedestal, or exhaust)? (0; 1; 2; 3; 4 or more)
10.	65	In the past month, did anyone in the household spend anything on telephone, telegraph, postal, fax, e-mail, internet, etc.? (No; Yes)
11.	63	What is the highest education level completed by the male head/spouse? (None or grades 1 to 3; Grades 4 to 7; Grades 8 to 9; Grade 10; Grade 11 or more)
12.	57	What is the household's main source of drinking water? (Hand pump; All other sources)
13.	57	Does the household own a television (No; Yes)
14.	56	In the past two weeks, did anyone in the household eat any chicken? (No; Yes)
15.	56	In the past month, did anyone in the household use any shampoo (No; Yes)
16.	55	Does the household own a cooking stove? (No; Yes)
17.	52	Does the household have a direct telephone connection (No; Yes)
18.	51	In the past two weeks, did anyone in the household eat any ginger? (No; Yes)
19.	50	In the past two weeks, did anyone in the household eat any curd or yoghurt? (No; Yes)
20.	49	Can the female head/spouse both read and write with understanding in some language? (No; Yes)
21.	46	Did the female head/spouse ever attend school? (No; Yes)
22.	47	Does the household own any agricultural land? (Rural, no; Urban, no or yes; Rural, yes)
23.	44	What kind of connection does the residence have with a drainage or sewerage system? (None; Open drain; Underground or covered drains)
24.	43	Do all girls of ages 12 to 17 attend school? (No; Yes; No girls in this age range)
25.	43	In the past two weeks, did anyone in the household eat any tomatoes? (No; Yes)
26.	43	In the past month, did the household use any piped gas? (No; Yes)

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

Uncertainty coefficient		Indicator (Responses ordered starting from the one most-closely linked with poverty)
27.	42	In the past month, did anyone in the household use any cooking oil? (No; Yes)
28.	42	Do all boys of ages 12 to 17 attend school? (No; Yes; No boys in this age range)
29.	42	Does the household have a direct electrical connection? (No; Yes)
30.	41	Does anyone in the household attend a private school or have a private tutor? (No; Yes)
31.	41	In the past two weeks, did anyone in the household consume any non-alcoholic beverages (carbonated beverages, non-medicated squashes and syrups, sugarcane juice and other fresh juices, packed fruit juices, mineral water, etc.)? (No; Yes)
32.	39	How many acres of agricultural land does the household own? (Rural, <8 acres; Urban, any amount; Rural, ≥8 acres)
33.	38	In the past month, did anyone in the household rent or buy a newspaper, magazine, novel, or book (not for school)? (No; Yes)
34.	37	Does the household keep 1 or more buffalos or camels, 2 or more cattle, 5 or more sheep or goats, 20 or more poultry birds, or fish in a fish farm? (Rural, no; Urban, no or yes; Rural, yes)
35.	36	Does the household raise any poultry? (Rural, no; Urban, no or yes; Rural, yes)
36.	35	In the past year, did the household pay any license fees for TV, VCR, dish antenna, etc.? (No; Yes)
37.	35	In the past two weeks, did anyone in the household consume any mutton? (No; Yes)
38.	34	In the past two weeks, did anyone in the household consume any apples? (No; Yes)
39.	34	Does the household own a buffalo? (Rural, no; Urban, no or yes; Rural, yes)
40.	33	Can the male head/spouse both read and write with understanding in some language? (No; Yes)
41.	32	In the past two weeks, did anyone in the household consume any bananas? (No; Yes)
42.	32	In the past months, did anyone in the household consume any biscuits (sweet or saltish)? (No; Yes)
43.	32	In the past two weeks, did anyone in the household consume any meat, poultry, or fish? (No; Yes)
44.	32	In what type of area does the household live? (Rural; Urban)
45.	32	Does the household rent-in any agricultural land? (Rural, no; Urban, no or yes; Rural, yes)
46.	32	Does the household own any cattle, buffalo, or camels? (Rural, no; Urban, no or yes; Rural, yes)
47.	32	Does the household own any horses, asses, or mules? (Rural, no; Urban, no or yes; Rural, yes)
48	32	Does the household own any cattle? (Rural, no; Urban, no or yes; Rural, yes)
49.	32	Does the household own any sheep or goats? (Rural, no; Urban, no or yes; Rural, yes)
50.	32	Does the household own any goats? (Rural, no; Urban, no or yes; Rural, yes)
51.	32	Does the household own any irrigated agricultural land? (Rural, no; Urban, no or yes; Rural, yes)
52.	32	How children aged 17 or younger are there per adult aged 18 or older? (>0.5; ≤0.5)
53.	31	In the past months, did anyone in the household consume any eggs? (No; Yes)
54.	31	Does the household own any sheep? (Rural, no; Urban, no or yes; Rural, yes)
55.	30	In the past year, did anyone in the household acquire any gold, silver, jewelry, stones, etc.? (No; Yes)
56.	29	Does the household own any sewing or knitting machines? (No; Yes)
57.	29	In the past two weeks, did anyone in the household consume any bread, buns, or Sheermal? (No; Yes)

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

58.	29	Does the household own a machine for sewing or knitting? (No; Yes)
59.	28	In the past two weeks, did anyone in the household consume any iodised salt? (No; Yes)
60.	28	How many rooms does the house have, including bedrooms and living rooms but excluding storage rooms, bathrooms, toilets, kitchens, and rooms for business? (1 or 2; 3; 4 or more)
61.	27	Does the household own a scooter or a motorcycle? (No; Yes)
62.	22	In the past year, did anyone in the household buy a Burka, Chadar, Ajrak, etc.? (No; Yes)
63.	19	Does the household own a radio or cassette player? (No; Yes)
64.	17	In the past two weeks, did anyone in the household consume any Desi ghee? (No; Yes)
65.	14	Does the household own a car or other motor vehicle? (No; Yes)
66.	12	Does the household own a VCR, VCP, receiver, or decoder? (No; Yes)
67.	9	Does the household own land of any type? (No; Yes)
68.	4	How many household members have salaried employment? (0; 1; 2 or more)
69.	3	In the past two weeks, did anyone in the household consume any eggs? (No; Yes)
70.	3	In the past two weeks, did anyone in the household consume any apples? (No; Yes)
71.	2	In the past month, did anyone in the household use any gas from a cylinder? (No; Yes)
72.	1	Does anyone in the household owe a debt on a loan? (Yes; No)
73.	1	Does the household own a bicycle? (No; Yes)

Source: Based on 2001 PIHS.

Figure 5: Scores and poverty likelihoods

Score	Poverty likelihood for people with score in range (%)	% of people <=score who are poor	% of people >score who are non-poor
0-4	87.3	87.3	59.9
5-9	77.1	81.6	64.2
10-14	69.1	76.4	70.0
15-19	67.3	74.6	74.4
20-24	57.6	71.3	79.8
25-29	34.5	65.4	82.6
30-34	36.2	61.2	87.4
35-39	24.6	57.2	90.8
40-44	17.9	54.0	93.3
45-49	13.5	51.6	95.3
50-54	8.2	49.4	96.5
55-59	7.8	48.1	97.8
60-64	1.0	46.2	97.0
65-69	3.8	45.4	97.4
70-74	0.5	44.5	95.2
75-79	8.7	44.1	100.0
80-84	0.0	43.8	100.0
85-89	0.0	43.8	100.0
90-94	0.0	43.7	100.0
95-100	0.0	43.7	100.0

Surveyed cases weighted to represent the Pakistani population.

Source: Based on the 2001 PIHS.

Figure 6: Confidence intervals for estimated poverty likelihoods associated

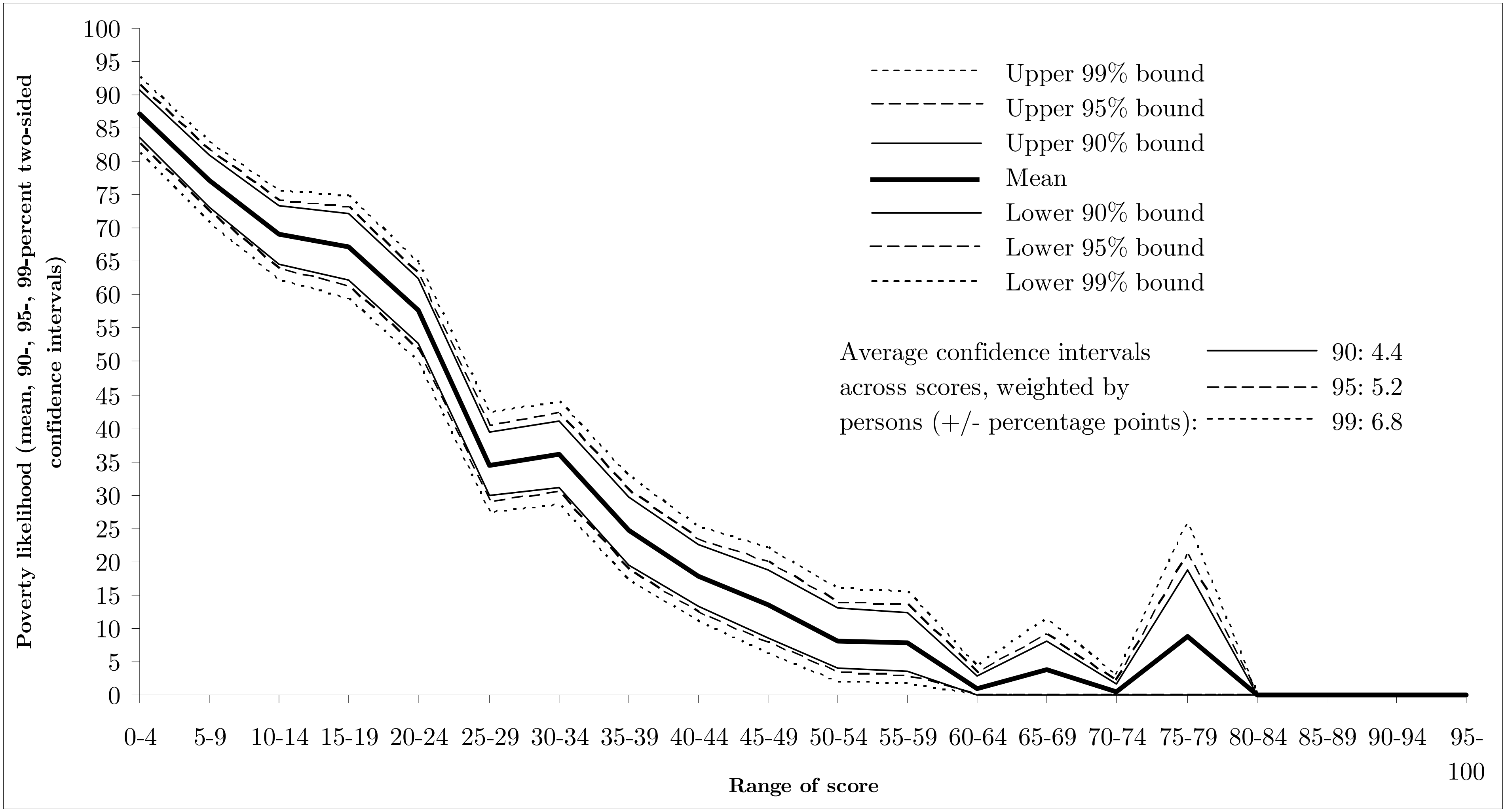

Figure 7: Differences between estimated and true poverty likelihoods

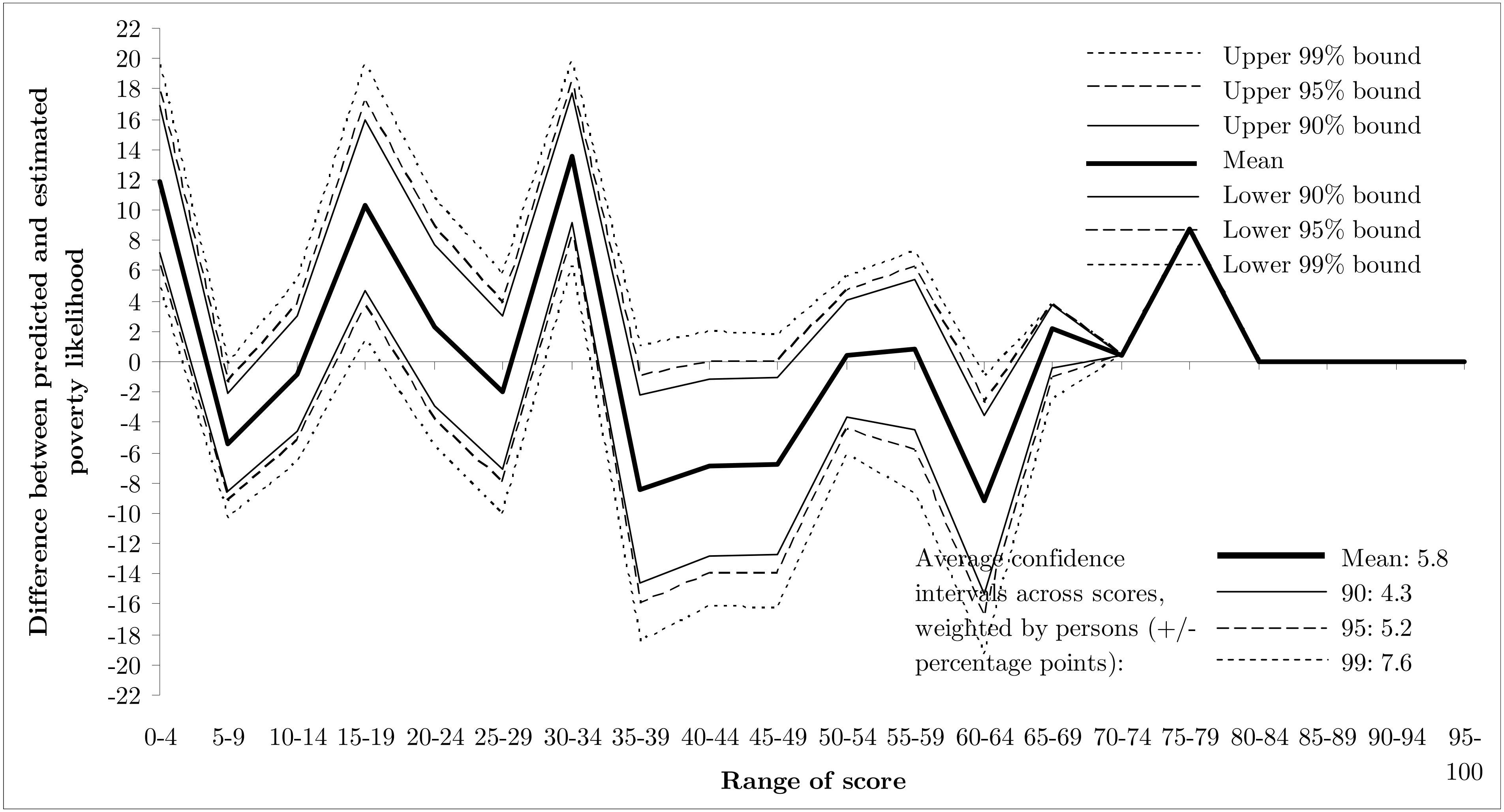

Figure 8: ROC curve of ability to rank-order households by poverty status

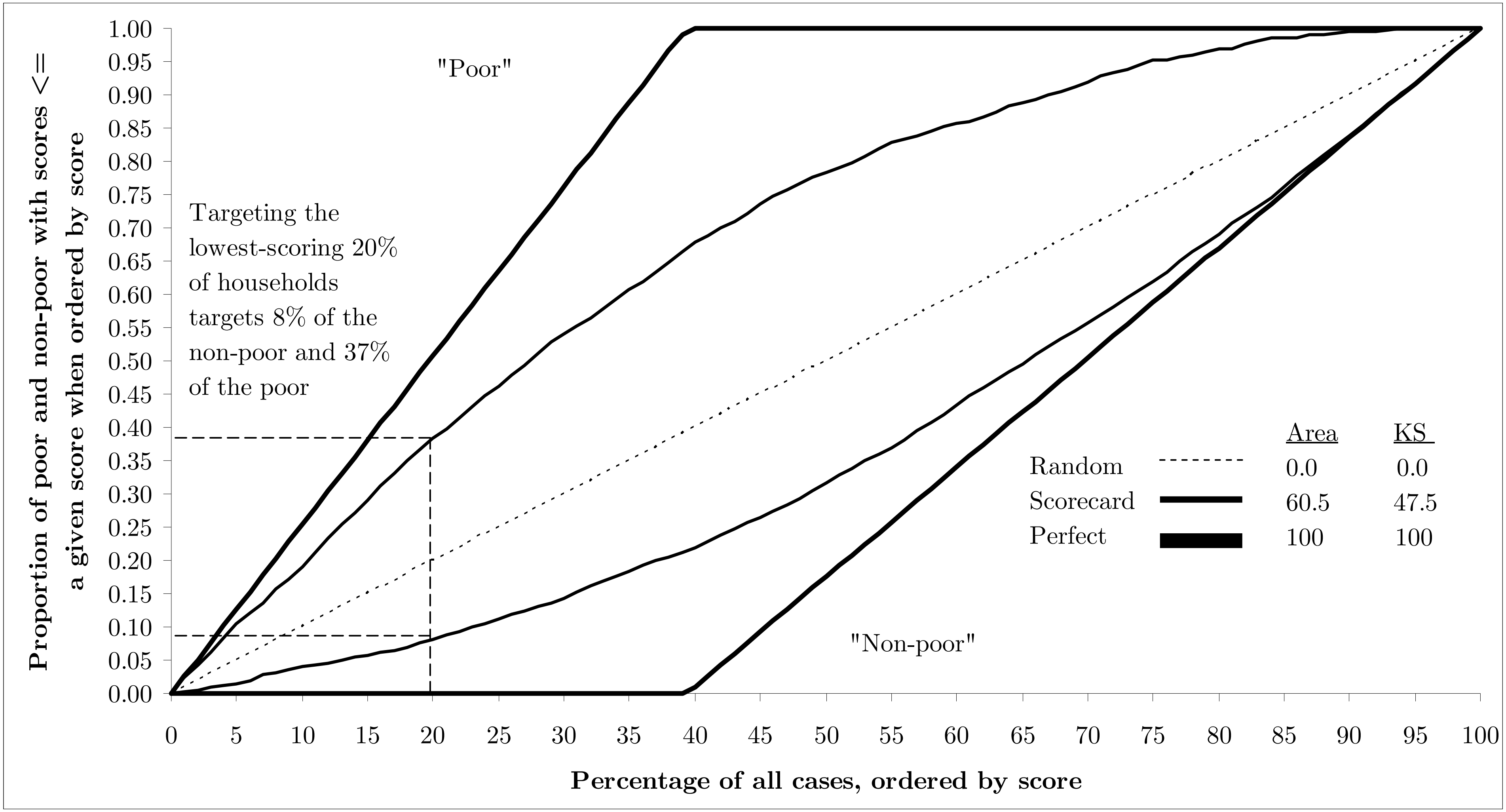

Figure 9: General classification matrix

		Targeting segment	
		Targeted	**Non-targeted**
True poverty status	**Poor**	**A.** Truly poor correctly targeted	**B.** Truly poor mistakenly non-targeted
	Non-poor	**C.** Truly non-poor mistakenly targeted	**D.** Truly non-poor correctly non-targeted

Figure 10: People by targeting classification and score

Score	A. Truly poor correctly targeted	B. Truly poor mistakenly non-targeted	C. Truly non-poor mistakenly targeted	D. Truly non-poor correctly non-targeted
0-4	6.7	37.0	1.0	55.3
5-9	14.1	29.6	3.2	53.1
10-14	22.6	21.1	7.0	49.3
15-19	27.6	16.1	9.4	46.9
20-24	32.8	10.9	13.2	43.1
25-29	35.9	7.8	19.0	37.3
30-34	39.2	4.5	24.9	31.5
35-39	41.1	2.6	30.7	25.6
40-44	42.2	1.5	35.9	20.4
45-49	42.9	0.8	40.2	16.1
50-54	43.3	0.4	44.4	11.9
55-59	43.5	0.2	47.0	9.3
60-64	43.5	0.2	50.6	5.7
65-69	43.6	0.1	52.4	3.9
70-74	43.6	0.1	54.4	1.9
75-79	43.7	0.0	55.4	0.9
80-84	43.7	0.0	56.1	0.2
85-89	43.7	0.0	56.1	0.2
90-94	43.7	0.0	56.3	0.0
95-100	43.7	0.0	56.3	0.0

Figures normalized to sum to 100.

Source: Based on the 2001 PIHS.

Figure 11: General net-benefit matrix

		Targeting segment	
		Targeted	Non-targeted
True poverty status	Poor	α	β
	Non-poor	γ	δ

Figure 12: "Total Accuracy" net-benefit matrix

		Targeting segment	
		Targeted	Non-targeted
True poverty status	Poor	1	0
	Non-poor	0	1

Figure 13: Total net benefit for some common net-benefit matrices

<table>
<tr><th></th><th colspan="2">Total Accuracy</th><th colspan="2">Poverty Accuracy</th><th colspan="2">Non-poverty Accuracy</th><th colspan="2">Undercoverage</th><th colspan="2">Leakage</th></tr>
<tr><th></th><th colspan="2">(A + B)</th><th colspan="2">100*A / (A+B)</th><th colspan="2">100*D / (C+D)</th><th colspan="2">100*B / (A+B)</th><th colspan="2">100*C / (A+C)</th></tr>
<tr><th></th><th>1</th><th>0</th><th>1</th><th>0</th><th>0</th><th>0</th><th>0</th><th>-1</th><th>0</th><th>0</th></tr>
<tr><th>Score</th><th>0</th><th>1</th><th>0</th><th>0</th><th>0</th><th>1</th><th>0</th><th>0</th><th>-1</th><th>0</th></tr>
<tr><td>0-4</td><td colspan="2">62.0</td><td colspan="2">15.3</td><td colspan="2">98.3</td><td colspan="2">84.7</td><td colspan="2">12.7</td></tr>
<tr><td>5-9</td><td colspan="2">67.3</td><td colspan="2">32.3</td><td colspan="2">94.3</td><td colspan="2">67.7</td><td colspan="2">18.4</td></tr>
<tr><td>10-14</td><td colspan="2">71.9</td><td colspan="2">51.7</td><td colspan="2">87.6</td><td colspan="2">48.3</td><td colspan="2">23.6</td></tr>
<tr><td>15-19</td><td colspan="2">74.5</td><td colspan="2">63.2</td><td colspan="2">83.3</td><td colspan="2">36.8</td><td colspan="2">25.4</td></tr>
<tr><td>20-24</td><td colspan="2">75.9</td><td colspan="2">75.1</td><td colspan="2">76.5</td><td colspan="2">24.9</td><td colspan="2">28.7</td></tr>
<tr><td>25-29</td><td colspan="2">73.2</td><td colspan="2">82.1</td><td colspan="2">66.3</td><td colspan="2">17.9</td><td colspan="2">34.6</td></tr>
<tr><td>30-34</td><td colspan="2">70.6</td><td colspan="2">89.7</td><td colspan="2">55.9</td><td colspan="2">10.3</td><td colspan="2">38.8</td></tr>
<tr><td>35-39</td><td colspan="2">66.7</td><td colspan="2">94.0</td><td colspan="2">45.4</td><td colspan="2">6.0</td><td colspan="2">42.8</td></tr>
<tr><td>40-44</td><td colspan="2">62.6</td><td colspan="2">96.6</td><td colspan="2">36.2</td><td colspan="2">3.4</td><td colspan="2">46.0</td></tr>
<tr><td>45-49</td><td colspan="2">59.0</td><td colspan="2">98.2</td><td colspan="2">28.6</td><td colspan="2">1.8</td><td colspan="2">48.4</td></tr>
<tr><td>50-54</td><td colspan="2">55.2</td><td colspan="2">99.0</td><td colspan="2">21.2</td><td colspan="2">1.0</td><td colspan="2">50.6</td></tr>
<tr><td>55-59</td><td colspan="2">52.8</td><td colspan="2">99.5</td><td colspan="2">16.5</td><td colspan="2">0.5</td><td colspan="2">51.9</td></tr>
<tr><td>60-64</td><td colspan="2">49.2</td><td colspan="2">99.6</td><td colspan="2">10.1</td><td colspan="2">0.4</td><td colspan="2">53.8</td></tr>
<tr><td>65-69</td><td colspan="2">47.5</td><td colspan="2">99.8</td><td colspan="2">7.0</td><td colspan="2">0.2</td><td colspan="2">54.6</td></tr>
<tr><td>70-74</td><td colspan="2">45.5</td><td colspan="2">99.8</td><td colspan="2">3.3</td><td colspan="2">0.2</td><td colspan="2">55.5</td></tr>
<tr><td>75-79</td><td colspan="2">44.6</td><td colspan="2">100.0</td><td colspan="2">1.6</td><td colspan="2">0.0</td><td colspan="2">55.9</td></tr>
<tr><td>80-84</td><td colspan="2">43.9</td><td colspan="2">100.0</td><td colspan="2">0.4</td><td colspan="2">0.0</td><td colspan="2">56.2</td></tr>
<tr><td>85-89</td><td colspan="2">43.9</td><td colspan="2">100.0</td><td colspan="2">0.3</td><td colspan="2">0.0</td><td colspan="2">56.2</td></tr>
<tr><td>90-94</td><td colspan="2">43.7</td><td colspan="2">100.0</td><td colspan="2">0.1</td><td colspan="2">0.0</td><td colspan="2">56.3</td></tr>
<tr><td>95-100</td><td colspan="2">43.7</td><td colspan="2">100.0</td><td colspan="2">0.0</td><td colspan="2">0.0</td><td colspan="2">56.3</td></tr>
</table>

All figures in percentage units.

Figure 14: "Poverty Accuracy" net-benefit matrix

True poverty status		Targeting segment: Targeted	Targeting segment: Non-targeted
	Poor	1	0
	Non-poor	0	0

Figure 15: Net-benefit matrix for BPAC

True poverty status	Targeting segment: Targeted	Targeting segment: Non-targeted
Poor	1	1
Non-poor	-1	0

Figure 16: Poverty likelihoods for the very poor, poor, and non-poor by score

	Poverty likelihood in score range			Share of cases <= score		
Score	Very Poor	Poor	Non-poor	Very Poor	Poor	Non-poor
0-4	61.3	26.0	12.7	61.3	26.0	12.7
5-9	55.8	21.4	22.9	58.2	23.4	18.4
10-14	44.5	24.6	30.9	52.5	23.9	23.6
15-19	45.7	21.6	32.7	51.2	23.4	25.4
20-24	27.3	30.4	42.4	46.4	24.8	28.7
25-29	14.8	19.8	65.5	41.4	24.0	34.6
30-34	13.2	22.9	63.8	37.3	23.8	38.8
35-39	14.3	10.3	75.4	34.8	22.4	42.8
40-44	3.9	14.0	82.2	32.3	21.7	46.0
45-49	7.6	6.0	86.5	30.9	20.8	48.4
50-54	5.2	3.0	91.8	29.5	19.8	50.6
55-59	1.4	6.4	92.3	28.6	19.4	51.9
60-64	1.0	0.0	99.0	27.6	18.7	53.8
65-69	3.3	0.5	96.2	27.1	18.3	54.6
70-74	0.0	0.5	99.6	26.5	17.9	55.5
75-79	0.0	8.7	91.3	26.2	17.8	55.9
80-84	0.0	0.0	100.0	26.1	17.7	56.2
85-89	0.0	0.0	100.0	26.1	17.7	56.2
90-94	0.0	0.0	100.0	26.0	17.7	56.3
95-100	0.0	0.0	100.0	26.0	17.7	56.3

Figure 17: Confidence intervals for estimated poverty likelihoods for being very poor associated with scores

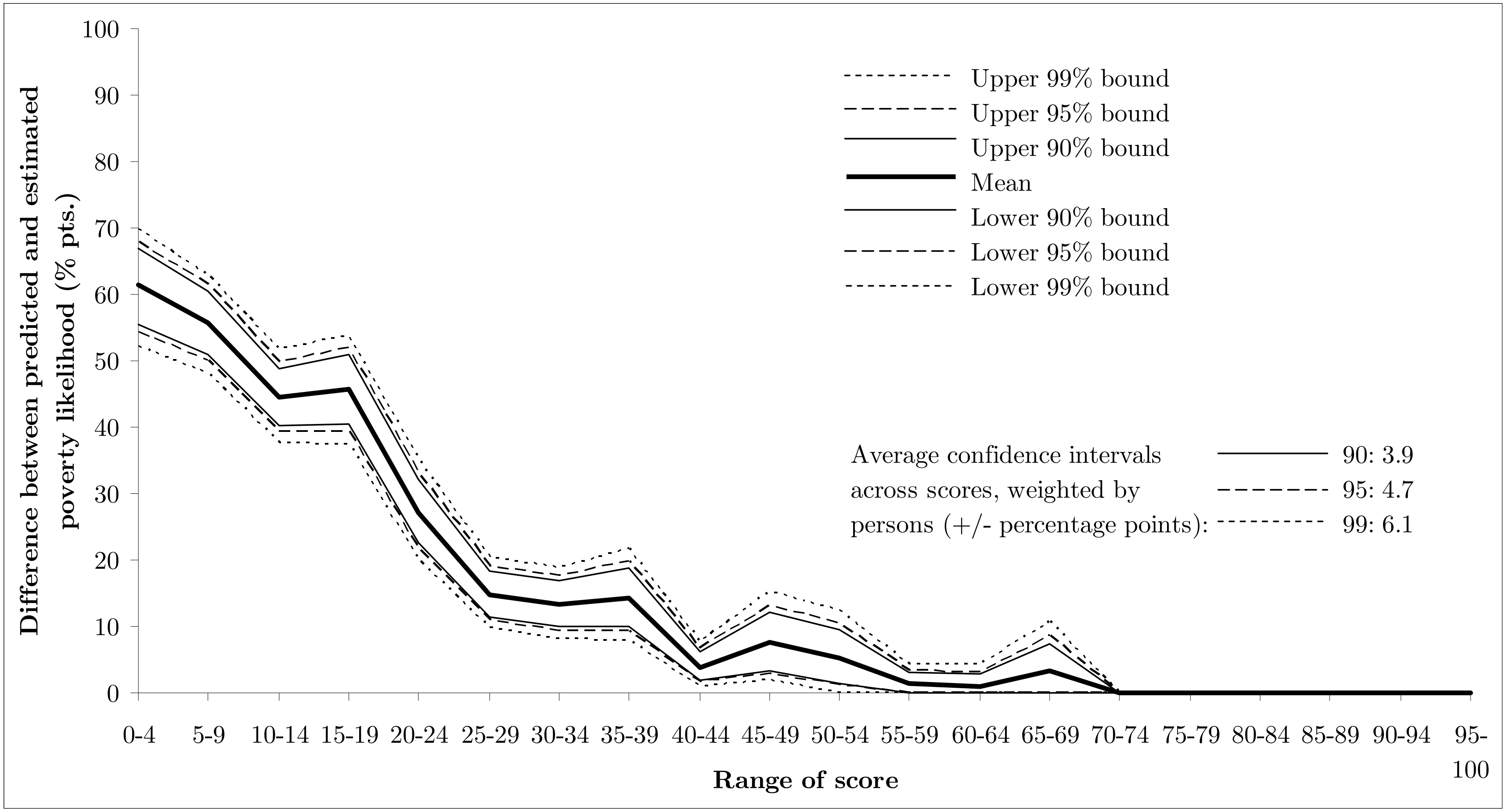

Figure 18: Differences between estimated and true poverty likelihoods for the very poor

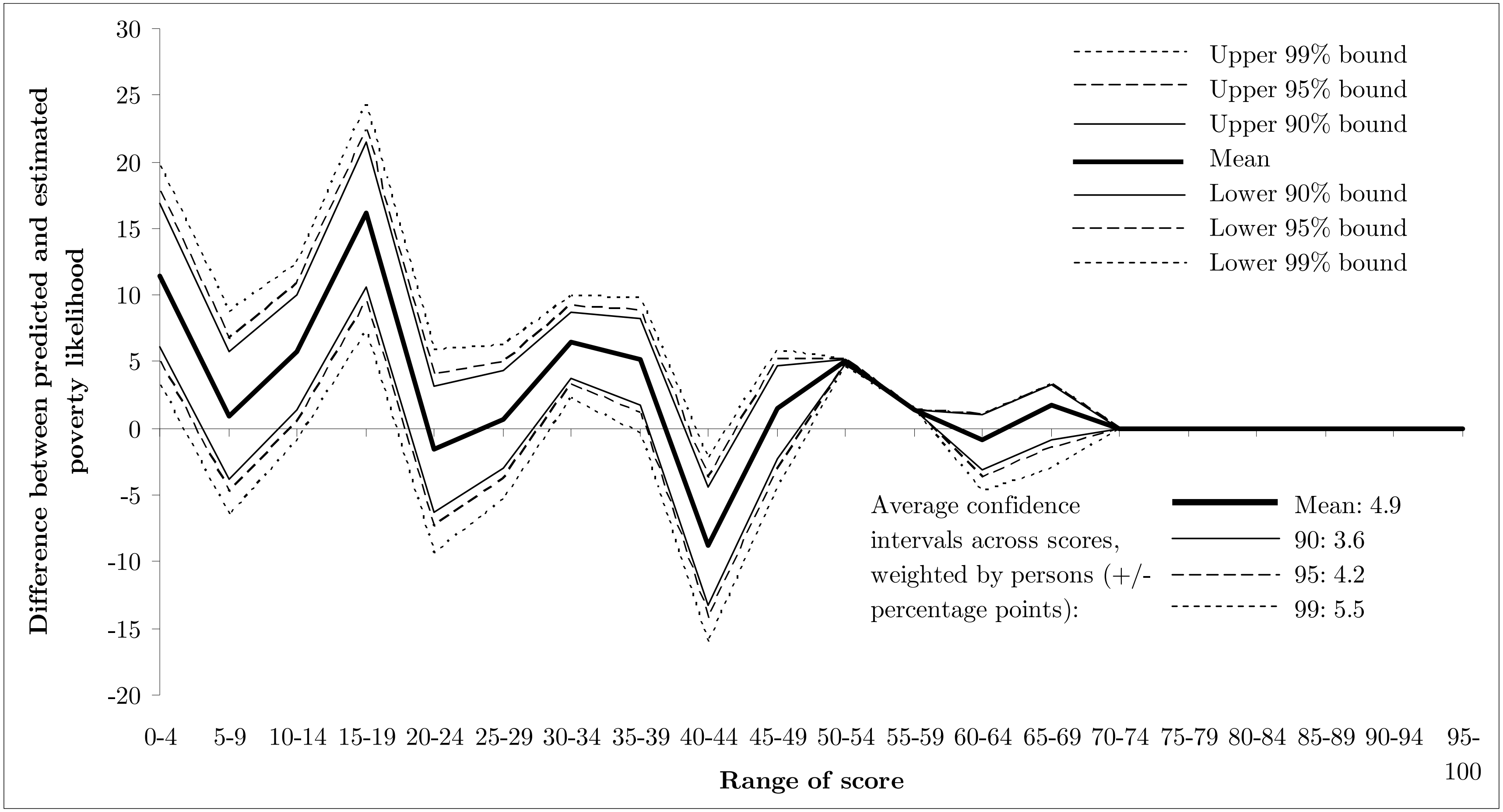

Figure 19: ROC curve of ability to rank-order households by very poor versus not very poor poverty status

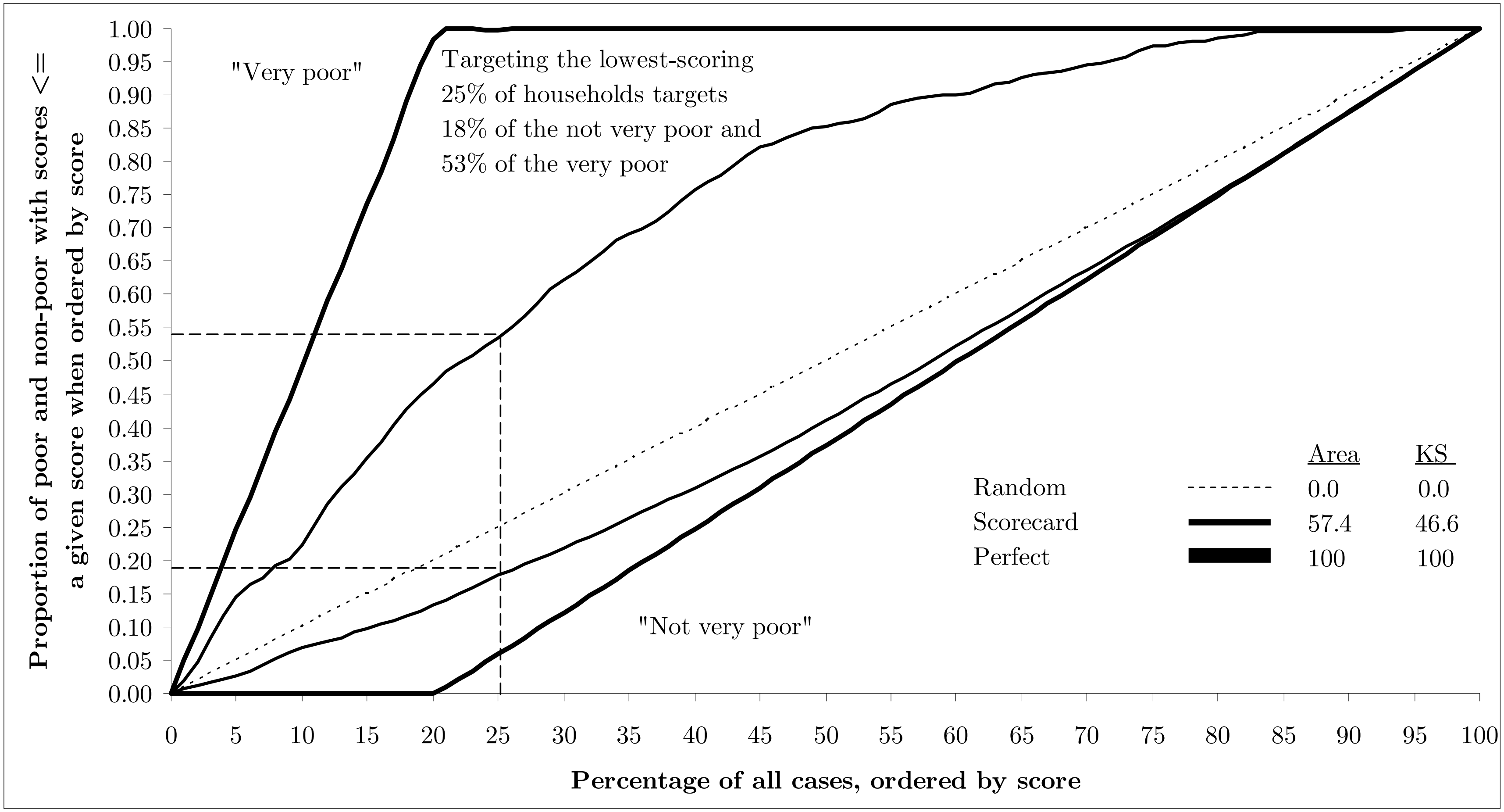

Figure 20: Classification matrix, three segments

True poverty status	Targeting segment: Very Poor	Targeting segment: Poor	Targeting segment: Non-poor
Very Poor	**A.** Truly very poor correctly targeted as very poor	**B.** Truly very poor incorrectly targeted as poor	**C.** Truly very poor incorrectly targeted as non-poor
Poor	**D.** Truly poor incorrectly targeted as very poor	**E.** Truly poor correctly targeted as poor	**F.** Truly poor incorrectly targeted as non-poor
Non-poor	**G.** Truly non-poor incorrectly targeted as very poor	**H.** Truly non-poor incorrectly targeted as poor	**I.** Truly non-poor correctly targeted as poor

Figure 21: Net-benefit matrix, three segments

		Targeting segment		
		Very Poor	Poor	Non-poor
True poverty status	Very Poor	α	β	γ
	Poor	δ	ε	ζ
	Non-poor	η	θ	ι

Figure 22: Classification results, very poor/poor cut-offs from 0 to 44 and poor/non-poor cut-offs from 5 to 49

Upper bound, very poor segment	Upper bound, poor segment																										
	5-9			10-14			15-19			20-24			25-29			30-34			35-39			40-44			45-49		
0-4	143	164	484	143	329	319	143	432	216	143	508	140	143	547	101	143	584	64	143	618	30	143	625	23	143	637	11
	60	63	414	60	154	323	60	203	274	60	287	190	60	340	137	60	404	73	60	428	49	60	455	22	60	464	13
	30	67	1,614	30	182	1,499	30	256	1,425	30	373	1,308	30	548	1,134	30	726	956	30	904	777	30	1,062	619	30	1,192	489
5-9				306	166	319	306	269	216	306	344	140	306	383	101	306	420	64	306	454	30	306	461	23	306	473	11
				123	91	323	123	140	274	123	224	190	123	277	137	123	341	73	123	365	49	123	392	22	123	401	13
				97	115	1,499	97	189	1,425	97	306	1,308	97	480	1,134	97	658	956	97	837	777	97	995	619	97	1,125	489
10-14							472	103	216	472	178	140	472	218	101	472	255	64	472	289	30	472	296	23	472	307	11
							215	49	274	215	133	190	215	185	137	215	249	73	215	274	49	215	301	22	215	310	13
							212	74	1,425	212	191	1,308	212	365	1,134	212	543	956	212	722	777	212	880	619	212	1,010	489
15-19										575	75	140	575	115	101	575	151	64	575	185	30	575	193	23	575	204	11
										263	84	190	263	137	137	263	201	73	263	225	49	263	252	22	263	261	13
										286	117	1,308	286	292	1,134	286	470	956	286	648	777	286	806	619	286	936	489
20-24													650	39	101	650	76	64	650	110	30	650	118	23	650	129	11
													347	53	137	347	117	73	347	141	49	347	168	22	347	177	13
													402	175	1,134	402	353	956	402	531	777	402	689	619	402	820	489
25-29																690	37	64	690	71	30	690	78	23	690	90	11
																400	64	73	400	88	49	400	115	22	400	124	13
																577	178	956	577	356	777	577	514	619	577	645	489
30-34																			727	34	30	727	41	23	727	53	11
																			464	24	49	464	51	22	464	60	13
																			755	178	777	755	336	619	755	467	489
35-39																						760	7	23	760	19	11
																						488	27	22	488	36	13
																						933	158	619	933	289	489
40-44																									768	11	11
																									515	9	13
																									1,091	131	489
45-49																											

Figures in units of 10,000 people.

Figure 22 (cont.): Classification results, very poor/poor cut-offs from 0 to 49 and poor/non-poor cut-offs from 50 to 100

		Upper bound, poor segment																													
		50-54			55-59			60-64			65-69			70-74			75-79			80-84			85-89			90-94			95-100		
Upper bound, very poor segment		143	644	4	143	645	3	143	646	2	143	648	0	143	648	0	143	648	0	143	648	0	143	648	0	143	648	0	143	648	0
	0-4	60	468	9	60	473	3	60	473	3	60	474	3	60	474	3	60	477	0	60	477	0	60	477	0	60	477	0	60	477	0
		30	1,319	362	30	1,399	283	30	1,508	173	30	1,562	119	30	1,625	57	30	1,655	27	30	1,675	6	30	1,676	5	30	1,680	2	30	1,681	0
		306	480	4	306	481	3	306	482	2	306	484	0	306	484	0	306	484	0	306	484	0	306	484	0	306	484	0	306	484	0
	5-9	123	405	9	123	411	3	123	411	3	123	411	3	123	411	3	123	414	0	123	414	0	123	414	0	123	414	0	123	414	0
		97	1,252	362	97	1,332	283	97	1,441	173	97	1,495	119	97	1,557	57	97	1,587	27	97	1,608	6	97	1,609	5	97	1,613	2	97	1,614	0
		472	315	4	472	316	3	472	317	2	472	319	0	472	319	0	472	319	0	472	319	0	472	319	0	472	319	0	472	319	0
	10-14	215	314	9	215	319	3	215	319	3	215	319	3	215	320	3	215	323	0	215	323	0	215	323	0	215	323	0	215	323	0
		212	1,137	362	212	1,217	283	212	1,326	173	212	1,380	119	212	1,442	57	212	1,472	27	212	1,493	6	212	1,494	5	212	1,498	2	212	1,499	0
		575	211	4	575	213	3	575	214	2	575	216	0	575	216	0	575	216	0	575	216	0	575	216	0	575	216	0	575	216	0
	15-19	263	265	9	263	270	3	263	270	3	263	271	3	263	271	3	263	274	0	263	274	0	263	274	0	263	274	0	263	274	0
		286	1,063	362	286	1,143	283	286	1,252	173	286	1,306	119	286	1,369	57	286	1,399	27	286	1,419	6	286	1,420	5	286	1,424	2	286	1,425	0
		650	136	4	650	137	3	650	139	2	650	140	0	650	140	0	650	140	0	650	140	0	650	140	0	650	140	0	650	140	0
	20-24	347	181	9	347	187	3	347	187	3	347	187	3	347	187	3	347	190	0	347	190	0	347	190	0	347	190	0	347	190	0
		402	946	362	402	1,026	283	402	1,135	173	402	1,189	119	402	1,252	57	402	1,282	27	402	1,302	6	402	1,304	5	402	1,307	2	402	1,308	0
		690	97	4	690	98	3	690	99	2	690	101	0	690	101	0	690	101	0	690	101	0	690	101	0	690	101	0	690	101	0
	25-29	400	128	9	400	134	3	400	134	3	400	134	3	400	134	3	400	137	0	400	137	0	400	137	0	400	137	0	400	137	0
		577	772	362	577	851	283	577	960	173	577	1,015	119	577	1,077	57	577	1,107	27	577	1,127	6	577	1,129	5	577	1,132	2	577	1,134	0
		727	60	4	727	61	3	727	62	2	727	64	0	727	64	0	727	64	0	727	64	0	727	64	0	727	64	0	727	64	0
	30-34	464	64	9	464	70	3	464	70	3	464	70	3	464	70	3	464	73	0	464	73	0	464	73	0	464	73	0	464	73	0
		755	594	362	755	673	283	755	782	173	755	837	119	755	899	57	755	929	27	755	949	6	755	951	5	755	954	2	755	956	0
		760	26	4	760	27	3	760	28	2	760	30	0	760	30	0	760	30	0	760	30	0	760	30	0	760	30	0	760	30	0
	35-39	488	40	9	488	46	3	488	46	3	488	46	3	488	46	3	488	49	0	488	49	0	488	49	0	488	49	0	488	49	0
		933	415	362	933	495	283	933	604	173	933	658	119	933	721	57	933	751	27	933	771	6	933	773	5	933	776	2	933	777	0
		768	19	4	768	20	3	768	21	2	768	23	0	768	23	0	768	23	0	768	23	0	768	23	0	768	23	0	768	23	0
	40-44	515	13	9	515	19	3	515	19	3	515	19	3	515	19	3	515	22	0	515	22	0	515	22	0	515	22	0	515	22	0
		1,091	257	362	1,091	337	283	1,091	446	173	1,091	500	119	1,091	563	57	1,091	593	27	1,091	613	6	1,091	615	5	1,091	618	2	1,091	619	0
		779	7	4	779	8	3	779	9	2	779	11	0	779	11	0	779	11	0	779	11	0	779	11	0	779	11	0	779	11	0
	45-49	524	4	9	524	10	3	524	10	3	524	10	3	524	10	3	524	13	0	524	13	0	524	13	0	524	13	0	524	13	0
		1,222	127	362	1,222	206	283	1,222	315	173	1,222	370	119	1,222	432	57	1,222	462	27	1,222	482	6	1,222	484	5	1,222	487	2	1,222	489	0

Figures in units of 10,000 people.

Figure 22 (cont.): Classification results, very poor/poor cut-offs from 50 to 94 and poor/non-poor cut-offs from 55 to 100

			55-59			60-64			65-69			70-74			75-79			80-84			85-89			90-94			95-100		
Upper bound, very poor segment	50-54		786	1	3	786	2	2	786	4	0	786	4	0	786	4	0	786	4	0	786	4	0	786	4	0	786	4	0
			528	5	3	528	5	3	528	6	3	528	6	3	528	9	0	528	9	0	528	9	0	528	9	0	528	9	0
			1,349	80	283	1,349	189	173	1,349	243	119	1,349	305	57	1,349	335	27	1,349	356	6	1,349	357	5	1,349	361	2	1,349	362	0
	55-59					788	1	2	788	3	0	788	3	0	788	3	0	788	3	0	788	3	0	788	3	0	788	3	0
						534	0	3	534	0	3	534	1	3	534	3	0	534	3	0	534	3	0	534	3	0	534	3	0
						1,428	109	173	1,428	164	119	1,428	226	57	1,428	256	27	1,428	276	6	1,428	278	5	1,428	281	2	1,428	283	0
	60-64								789	2	0	789	2	0	789	2	0	789	2	0	789	2	0	789	2	0	789	2	0
									534	0	3	534	1	3	534	3	0	534	3	0	534	3	0	534	3	0	534	3	0
									1,537	54	119	1,537	117	57	1,537	147	27	1,537	167	6	1,537	169	5	1,537	172	2	1,537	173	0
	65-69											791	0	0	791	0	0	791	0	0	791	0	0	791	0	0	791	0	0
												534	0	3	534	3	0	534	3	0	534	3	0	534	3	0	534	3	0
												1,592	62	57	1,592	92	27	1,592	113	6	1,592	114	5	1,592	117	2	1,592	119	0
	70-74														791	0	0	791	0	0	791	0	0	791	0	0	791	0	0
															534	3	0	534	3	0	534	3	0	534	3	0	534	3	0
															1,654	30	27	1,654	50	6	1,654	52	5	1,654	55	2	1,654	57	0
	75-79																	791	0	0	791	0	0	791	0	0	791	0	0
																		537	0	0	537	0	0	537	0	0	537	0	0
																		1,684	20	6	1,684	22	5	1,684	25	2	1,684	27	0
	80-84																				791	0	0	791	0	0	791	0	0
																					537	0	0	537	0	0	537	0	0
																					1,704	2	5	1,704	5	2	1,704	6	0
	85-89																							791	0	0	791	0	0
																								537	0	0	537	0	0
																								1,706	3	2	1,706	5	0
	90-94																										791	0	0
																											537	0	0
																											1,709	2	0

Figures in units of 10,000 people.

Figure 23: Classification results, very poor 0–14, poor 15–24, and non-poor 25–100

		People with score in range					
Segment	**Score**	**Very Poor**		**Poor**		**Non-poor**	
Very poor	0-4	**472**	143	**215**	60	**212**	30
0-24	5-9	**60%**	164	**40%**	63	**12%**	67
	10-14		166		91		115
Poor	15-19	**178**	103	**133**	49	**191**	74
25-34	20-24	**23%**	75	**25%**	84	**11%**	117
	25-29		39		53		175
	30-34		37		64		178
	35-39		34		24		178
	40-44		7		27		158
	45-49		11		9		131
	50-54		7		4		127
Non-poor	55-59	**140**	1	**190**	5	**1,308**	80
35-100	60-64	**18%**	1	**35%**	0	**76%**	109
	65-69		2		0		54
	70-74		0		0		62
	75-79		0		3		30
	80-84		0		0		20
	85-89		0		0		2
	90-94		0		0		3
	95-100		0		0		2
	Total:		791		537		1,711

Counts of people are in units of 10,000.

Figure 24: An example net-benefit matrix reflecting common values

True poverty status	Targeting segment: Very Poor	Targeting segment: Poor	Targeting segment: Non-poor
Very Poor	+3	-2	-6
Poor	-1	+2	-2
Non-poor	-2	-1	+1

Note: This is an example. Each program should define its own net-benefit matrix.

Figure 25: Computation of total net benefit for a cut-off pair of 10–14 and 20–24

Cell			Persons	Net benefit/person	Net benefit
A.	Truly very poor	as very poor	472	+3	+1,416
B.	Truly very poor	as poor	178	–2	–356
C.	Truly very poor	as non-poor	140	–6	–840
D.	Truly poor	as very poor	215	–1	–215
E.	Truly poor	as poor	133	+2	+266
F.	Truly poor	as non-poor	190	–2	–380
G.	Truly non-poor	as very poor	212	–2	–424
H.	Truly non-poor	as poor	191	–1	–191
I.	Truly non-poor	as non-poor	1,308	+1	+1,308
				Total net benefit:	+584

Note: Persons are counted in units of 10,000.