Philippine Journal of Development
Number 63, Second Semester 2007
Volume XXXIV, Number 2

A Simple Poverty Scorecard for the Philippines

MARK SCHREINER[1]

ABSTRACT

How poor are participants of development projects in the Philippines? This paper uses the 2002 Annual Poverty Indicators Survey to construct an easy-to-use objective poverty scorecard that estimates the likelihood that a participant has income below the national poverty line. The scorecard uses 10 simple indicators that field workers can quickly collect and verify. Scores can be computed by hand on paper in real time. With 99 percent confidence, estimates of a group's overall poverty rate are accurate to within +/–1.0 percentage point. The poverty scorecard can help development programs to target services to clients, track changes in poverty over time, and report on poverty rates.

INTRODUCTION

This paper presents an easy-to-use, objective poverty scorecard to help development programs in the Philippines to target services to clients, track changes in poverty over time, and report clients' poverty rates.

Indicators in the scorecard were derived from the 38,014 households surveyed in the 2002 Annual Poverty Indicators Survey (APIS). The criteria for the selection of indicators included:

---

[1] This research was funded by Grameen Foundation USA. Data were provided by the National Statistics Office (NSO) of the Republic of the Philippines. I am grateful for the help of Nigel Biggar, Cecilia del Castillo, Ron Chua, Gretel Guzmán, Syed Hashemi, Gilbert Maramba, Luca David Opramolla, Jeff Toohig, and the editor and referee.

- Inexpensive to collect, amenable to quick answers, and simple to verify
- Liable to change over time as poverty status changes
- Strongly correlated with poverty

All scorecard weights are non-negative integers, and scores range from 0 (most-likely "poor") to 100 (least-likely "poor"). The scorecard is easy for users 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 is highly practical, and it accurately and objectively estimates the likelihood of having income below the national poverty line. In particular, a group's estimated overall poverty rate is accurate within +/–1.0 percentage points with 99 percent confidence.

DATA AND POVERTY LINES

The analysis uses the 38,014 households in the 2002 APIS from the National Statistics Office (NSO). At the time the scorecard was built, this was the most extensive and most recent national household survey available that includes income or expenditure data.

This paper divides the households into three random samples (Figure 1). It employs half of the households for constructing the scorecard, a fourth for associating scores with estimated poverty likelihoods, and the last one-fourth for measuring the accuracy of estimates derived from the scorecard.

APIS measures income but not expenditure. The official poverty lines are in terms of income, and the Philippine government applies them only to a larger, more detailed survey, the triennial Family Income and Expenditure Survey (FIES). The 2003 FIES is not available, but Ericta (2005) reports that it gives a poverty rate of 30.4 percent.

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

Sub-sample	Households	People	% poor
Constructing scorecards	18,846	39,459,467	32.0
Associating scores with likelihoods	9,665	20,407,790	31.4
Testing accuracy	9,503	19,760,021	31.9
Source: 2002 APIS.	38,014	79,627,278	31.8

This paper applies the official poverty lines (Figure 2) to the income meaure in the 2002 APIS. While APIS uses different questions than FIES to measure income, the resulting overall poverty rate is 31.8 percent, remarkably close to the FIES rate.

The rural poverty rate based on APIS is 46.4 percent, while the urban rate is 17.3 percent. This paper presents a single scorecard for use anywhere in the Philippines, as evidence from India and Mexico (Schreiner 2006 and 2005a) suggests that there are only small returns to segmenting scorecards by rural and urban.

SCORECARD CONSTRUCTION

About 500 potential poverty indicators were prepared, including, for example:

- ✦ 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 spending on non-alcoholic drinks)
- ✦ Household durable goods (such as electric fans and telephones)

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 applied to categorical indicators (such as "type of floor") rather than continuous ones (such as "square meters of floor space"). About 120 indicators were then selected for further analysis. Figure 3 lists the top 50, ranked by uncertainty coefficient. Responses are then ordered by strength of association with poverty.

Many indicators in Figure 3 are similar in terms of their link with poverty. For example, most households who have a television also have electricity. If a scorecard already includes "has a television," then "has electricity" is superfluous. Thus, many indicators strongly linked with poverty are not in the scorecard 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 number of radios) that are more likely to change. No consumption indicators (such as "In the past six months, how much on average per week did the household spend on dairy products and eggs") were selected because they cannot be directly observed nor verified.

Figure 2. Official poverty lines, pesos/person/year

Province	Urban	Rural	Province	Urban	Rural	Province	Urban	Rural	Province	Urban	Rural
NCR			Region IV			Region VIII			Region XII		
1st District	16,496	N/A	Batangas	15,993	15,002	Eastern Samar	10,617	9,690	Lanao del Norte	12,393	11,630
2nd District	16,007	N/A	Cavite	14,851	16,240	Leyte	10,639	10,460	North Cotabato	11,172	9,761
3rd District	15,256	N/A	Laguna	14,147	12,312	Northern Samar	9,726	9,503	Sultan Kudarat	11,940	10,565
4th District	16,654	N/A	Marinduque	12,301	11,639	Western Samar	10,868	10,523			
			Occidental Mindoro	12,271	12,327	Southern Leyte	11,033	9,921			
			Oriental Mindoro	15,095	13,938	Biliran	10,218	10,644			
			Palawan	13,541	10,729						
			Quezon	13,430	12,605						
			Rizal	14,264	13,561						
			Romblon	12,770	11,234						
			Aurora	12,121	11,469						
Region I			Region V			Region IX			CAR		
Ilocos Norte	13,175	13,688	Albay	15,239	11,763	Basilan	11,891	9,350	Abra	13,201	13,928
Ilocos Sur	12,768	14,368	Camarines Norte	13,931	11,259	Zamboanga del Norte	11,715	9,377	Benguet	15,300	13,309
La Union	13,415	13,183	Camarines Sur	13,049	10,389	Zamboanga del Sur	10,676	9,385	Ifugao	13,353	12,330
Pangasinan	13,449	12,737	Catanduanes	13,523	10,653				Kalinga	12,128	11,469
			Masbate	13,784	10,903				Mt. Province	17,044	14,863
			Sorsogon	13,551	11,264				Apayao	11,030	11,200

Figure 2 continued

Province	Urban	Rural	Province	Urban	Rural	Province	Urban	Rural	Province	Urban	Rural
Region II			Region VI			Region X			ARMM		
Batanes	15,490	12,386	Aklan	12,581	11,938	Bukidnon	11,125	9,649	Lanao del Sur	13,459	14,725
Cagayan	12,507	10,127	Antique	11,981	10,969	Camiguin	14,228	11,943	Maguindanao	14,247	11,996
Isabela	14,883	11,317	Capiz	12,354	10,781	Misamis Occidental	11,898	10,081	Sulu	13,487	12,602
Nueva Vizcaya	13,707	10,730	Iloilo	12,948	12,328	Misamis Oriental	12,649	11,508	Tawi-tawi	13,192	13,259
Quirino	12,072	10,670	Negros Occidental	11,507	11,463						
			Guimaras	12,293	11,469						
Region III			Region VII			Region XI			Others		
Bataan	13,344	11,706	Bohol	11,070	10,060	Davao del Norte	11,648	11,401	Agusan del Norte	12,767	10,594
Bulacan	14,822	13,265	Cebu	10,950	9,817	Davao del Sur	12,457	9,912	Agusan del Sur	12,355	11,104
Nueva Ecija	16,048	14,182	Negros Oriental	11,587	8,358	Davao Oriental	12,624	10,289	Surigao del Norte	13,813	11,261
Pampanga	15,459	14,111	Siquijor	11,823	9,361	South Cotabato	12,803	11,659	Surigao del Sur	12,422	10,694
Tarlac	13,994	12,409				Saranggani	12,674	11,719			
Zambales	14,715	12,701									

Source: National Statistic Coordination Board (2004).

Figure 3. Poverty indicators ranked by uncertainty coefficient

	Uncertainty coefficient	Indicator (Value for those most likely “poor”; Value for those least likely “poor”)
1.	247	During the past six months, how much on average did the household spend per month on fuel, light, and water (charcoal, firewood, LPG, kerosene/ gas, electricity, candles, oils, water, etc.) (<P83/person; ≥P83/person)
2.	214	During the past six months, how much on average did the household spend per week on meat and meat preparations (fresh chicken, beef, pork, carabeef, goat's meat, corned beef, luncheon meat, meatloaf, Vienna sausage, longaniza, chorizo, hot dogs, tocino, tapa, etc.) (<P19/person; ≥P19/person)
3.	214	During the past six months, how much on average did the household spend per month on personal care and effects (cleansing cream, body deodorant, lotion, baby oil, toilet/bath soap, tissue paper, toothpaste, sanitary napkins, shampoo, jewelry, handbags, wallets, wristwatches, haircuts, manicure or pedicure, etc.) (<P30/person; ≥P30/person)
4.	195	During the past six months, how much on average did the household spend per month on transportation and communication (bus, jeepney, tricycle, air or water transport fare, gasoline/diesel, driver's salary, telephone bills, postage stamps, telegrams, driving lessons, feeds for animals used for transport, etc.) (<P38/person; ≥P38/person)
5.	190	Does the family own a gas stove or gas range? (No; Yes)
6.	183	Does the family own a refrigerator? (No; Yes)
7.	181	How many television sets does the family own? (0; 1; 2 or more)
8.	162	During the past six months, how much on average did the household spend per week on non-alcoholic beverages (soft drinks, pineapple juice, orange juice, ice candy, ice drop, ice buko, etc.) (<P3/person; ≥P3/person)
9.	152	During the past six months, how much on average did the household spend per week on dairy products and eggs (milk, ice cream, butter, cheese, fresh eggs, balut, salted eggs) (<P6/person; ≥P6/person)
10.	145	How many telephones and/or cell phones does the family own? (0; 1; 2 or more)
11.	140	During the past six months, how much on average did the household spend per week on food regularly consumed outside the home (meals at school, place of work, restaurants, merienda or snacks, etc.) (<P11/person; ≥P11/person)
12.	136	Does the family own a sala set? (No; Yes)
13.	127	Does the family own a washing machine? (No; Yes)
14.	126	Is there any electricity in the building/house? (No; Yes)

Figure 3 continued

	Uncertainty coefficient	Indicator (Value for those most likely "poor"; Value for those least likely "poor")
15.	123	Does the family own a dining set? (No; Yes)
16.	121	During the past six months, how much did the household spend on clothing, footwear, and other wear (clothing and ready-made apparel, footwear, sewing materials, accessories, service fees, etc.) (<P200/person; ≥P200/person)
17.	113	What is the highest grade completed by a household member? (Graduated secondary or less; 1 or more years of post-secondary)
18.	105	What is the primary occupation of the male head/spouse? (Farmers and laborers; Clerks, trades, special occupations, and occupations not elsewhere classified; Clerks, and plant and machine operators; Technicians, and officials of government and special-interest organizations; Professionals)
19.	104	What is the house's main source of water supply? (Spring, river, stream, dug well, or rain; Peddler or others; Own-use or shared-use from a tubed/piped well or a community water system)
20.	102	Does the family have health insurance from Philhealth, a Health Maintenance Organization, a private health-insurance company, or a community or cooperative? (No; Yes)
21.	98	Does the family have health insurance with Philhealth? (No; Yes)
22.	95	On a ladder with 10 steps going from lowest/poorest to highest/richest, on which step would you be? (1; 2; 3; 4; 5; 6 or more)
23.	94	How many people in the family are aged 0 to 17? (5 or more; 4; 2 or 3; 0 or 1)
24.	89	What are the house's outer walls made of? (Salvaged or makeshift materials or all light or predominantly light materials such as cogon, nipa, or sawali, bamboo, or anahaw; All strong or predominantly strong materials such as iron, aluminum, tile, concrete, brick, stone, wood, or asbestos)
25.	86	What kind of toilet facility does the family have in the house? (None, open pit, or others; Closed pit; Water sealed)
26.	80	How many radios does the family own? (0; 1; 2 or more)
27.	78	What type of construction materials is the house's roof made of? (Salvaged or makeshift materials or all light or predominantly light materials such as cogon, nipa, or anahaw; All strong or predominantly strong materials such as galvanized iron, aluminum tile, concrete, brick, stone, or asbestos)
28.	77	What is the highest grade completed by the male head/spouse? (Grades I to V of elementary; Did not complete any grade; Grade VI of elementary to graduate of secondary; 1 or more years of post-secondary)
29.	77	How many people are there in the family from ages 0 to 17? (3 or more; 2; 1; 0)

Figure 3 continued

	Uncertainty coefficient	Indicator (Value for those most likely "poor"; Value for those least likely "poor")
30.	77	What is the highest grade completed by the female head/spouse? (I to V of elementary; None; VI of elementary to graduate of secondary; 1 or more years of post-secondary)
31.	69	What is the primary occupation of the female head/spouse? (Farmers and laborers; Clerks, trades, special occupations, and occupations not elsewhere classified; Clerks, and plant and machine operators; Technicians, and officials of government and special-interest organizations; Professionals)
32.	61	Does the family live in an urban area? (No; Yes)
33.	58	How many children are there aged 17 or younger per adult aged 18 or older? (1 or more; <1)
34.	58	Does the family engage in crop farming or gardening? (Yes; No)
35.	53	Does the family own a vehicle? (No; Yes)
36.	48	What is the floor area of house in square meters? (50 or less; $\geq$50)
37.	47	Does the household have some type of health insurance? (No; Yes)
38.	41	During the past six months, did the household regularly consume food outside the home (meals at school, place of work, restaurants, merienda or snacks, etc.)? (No; Yes)
39.	40	During the past six months, did the household buy non-alcoholic beverages (soft drinks, pineapple juice, orange juice, ice candy, ice drop, ice buko, etc.)? (No; Yes)
40.	39	How many children are there aged 11 or younger per adult aged 18 or older? (0.5 or more; <0.5)
41.	38	During the past six months, did the household make any deposits in banks? (No; Yes)
42.	35	During the past six months, did anyone in the family receive cash, gifts, support, or relief from abroad (including pensions retirement, workmen's compensation, dividends from investments, etc.)? (No; Yes)
43.	31	Is the male head/spouse self-employed? (No; Yes)
44.	22	During the past six months, did the household buy dairy products and eggs (milk, ice cream, butter, cheese, fresh eggs, balut, salted eggs)? (No; Yes)
45.	22	Do all children in the family of ages 6 to 11 go to school? (No; Yes)
46.	22	Do any family members assist in the family business? (No; Yes)

Figure 3 continued

	Uncertainty coefficient	Indicator (Value for those most likely "poor"; Value for those least likely "poor")
47.	21	During the past six months, did anyone in the family make a deposit in a bank, receive interest on savings, or make a withdrawal from a savings account? (No; Yes)
48.	21	Do any family members have salaried employment? (No; Yes)
49.	18	In the next 12 months, do you expect your household's economic conditions to worsen, stay the same, or improve? (Worsen; Stay the Same; Improve)
50.	17	Does the family have any income from entrepreneurial activities? (Yes; No)
51.	16	Is your household's situation at present compared with the last 12 months worse, about the same, or better? (Worse; About the same; Better)
52.	16	Are any household members involved with self-employment? (No; Yes)

Source: Based on 2002 APIS.

The scorecard itself was constructed using Logit regression. Indicator selection combined statistics with the judgment of an analyst who has 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.[2]

After all indicators had been tested, one was selected based on several factors (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 to change values as poverty status changes over time, variety vis-à-vis other indicators already in the scorecard, and ease of observation and 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 zero (most likely poor) and the highest is 100. The final poverty scorecard appears in Figure 4.

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

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

[2] A higher *c* statistic indicates greater ability to rank households by poverty status. For a Logit regression with a categorical outcome (such as poor/not poor), the *c* statistic is a general measure of explanatory power, much like R^2 in a least-squares regression on a continuous outcome. The *c* statistic 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). The *c* statistic is also equivalent to the area under an ROC curve—discussed in more detail later—that plots the share of poor and nonpoor households versus all households ranked by score. Finally, the *c* statistic can also be seen as the share of all possible pairs of poor and nonpoor 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.

Figure 4. A simple poverty scorecard for the Philippines

Indicator	Values				Points
1. How many people in the family are aged 0 to 17?	5 or more	3 or 4	1 or 2	Zero	
	0	7	16	27	
2. Does the family own a gas stove or gas range?			No	Yes	
			0	13	
3. How many television sets does the family		Zero	1	2 or more	
		0	9	18	
4. What are the house's outer walls made of?		Light (cogon, nipa, or sawali, bamboo, anahaw)		Strong (iron, aluminum, tile, concrete, brick, stone, wood, asbestos)	
		0		4	
5. How many radios does the family own?		Zero	1	2 or more	
		0	3	10	
6. Does the family own a sala set?		No	Yes		
		0	9		
7. What is the house's roof made of?		Light (Salvaged, makeshift, cogon, nipa, or anahaw)		Strong (Galvanized iron, aluminum tile, concrete, brick, stone, or asbestos)	
		0		2	
8. What kind of toilet facility does the family have?		None, open pit, closed pit, or other		Water sealed	
		0		3	
9. Do all children in the family of ages 6 to 11 go to school?		No	Yes	No children ages 6-11	
		0	4	6	
10. Do any family members have salaried employment?		No	Yes		
		0	6		
					Total:

Source: Calculations based on the 2002 APIS.

2002). The challenge is not technical but human and organizational; not statistics but change management. Accuracy is easier to achieve—but less important—than practicality.

The scorecard here is designed to be easy to use so that users will 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 (Figure 4) 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"), and
- 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 in Figure 4 can be photocopied for immediate use. It can also serve as a template for data-entry screens with database software that records indicators, indicator values, scores, and poverty likelihoods.

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

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 based on a simple table (Figure 5). For example, scores of 25–29 correspond to a poverty likelihood of 76.8 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

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	99.3	99.3	68.4
5-9	92.5	94.6	69.4
10-14	91.9	93.2	70.9
15-19	93.4	93.3	72.5
20-24	77.6	88.3	74.3
25-29	76.8	84.8	77.0
30-34	77.7	82.9	80.6
35-39	48.6	74.0	83.4
40-44	48.3	68.7	86.8
45-49	33.6	63.0	89.1
50-54	34.4	58.4	92.7
60-64	10.1	49.0	94.9
65-69	10.2	43.7	96.6
70-74	6.9	40.1	97.9
75-79	3.8	37.1	98.8
80-84	2.1	34.1	100.0
85-89	0.0	32.9	100.0
90-94	0.0	32.5	100.0
95-100	0.0	32.1	100.0

Surveyed cases weighted to represent the Filipino population.
Source: Based on the 2002 APIS.

- 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, they are based on data. The 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 to 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 the standard way to measure accuracy, and it is widely understood by lay people. 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 25–29 is 76.8 percent (this is the poverty likelihood in Figure 5). In 90 percent of samples, the poverty rate is between 73.1–80.4 percent (Figure 6). In 95 percent of samples, the rate is 72.4–81.0 percent, and in 99 percent of samples, the rate is 70.8–82.3 percent.

For estimated and true poverty likelihoods, Figure 7 depicts average absolute differences and confidence intervals from bootstrapping the second one-fourth hold-out sample from the 2002 APIS. The mean absolute difference is 3.6 percentage points.

This discussion so far looks at whether estimated poverty likelihoods are close to true poverty likelihoods (and indeed they are). 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 nonpoor 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 nonpoor 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 nonpoor (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 south-east corners of the graph.

In Figure 8, the northwest (southeast) curve depicts accuracy among the poor (nonpoor). 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 51percent of all the poor and 6 percent of all the nonpoor. In contrast, randomly targeting 20 percent of cases would target 20 percent of the poor and 20 percent of the nonpoor.

Figure 6. Confidence intervals for estimated poverty likelihoods associated with scores

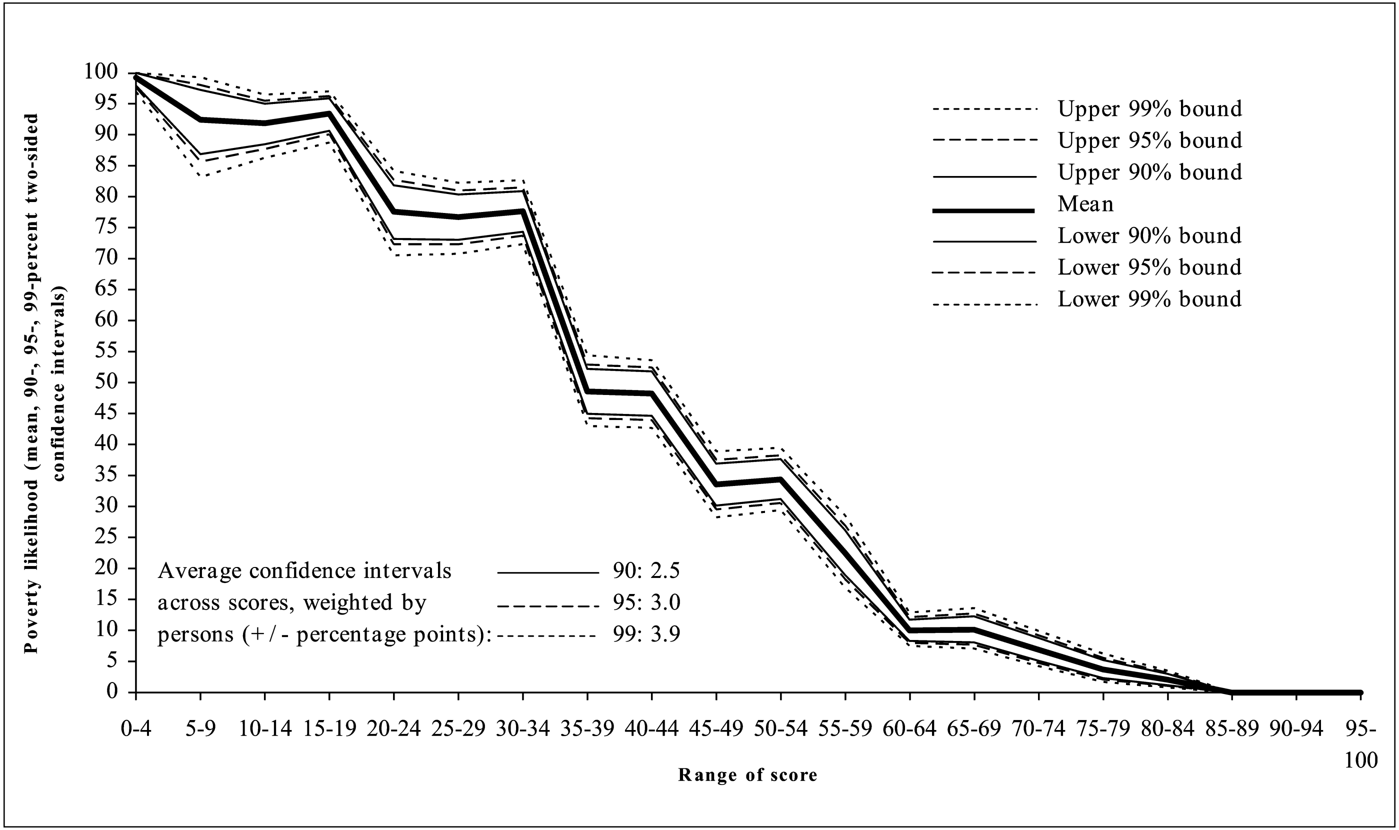

Figure 7. Differences between estimated and true poverty likelihoods

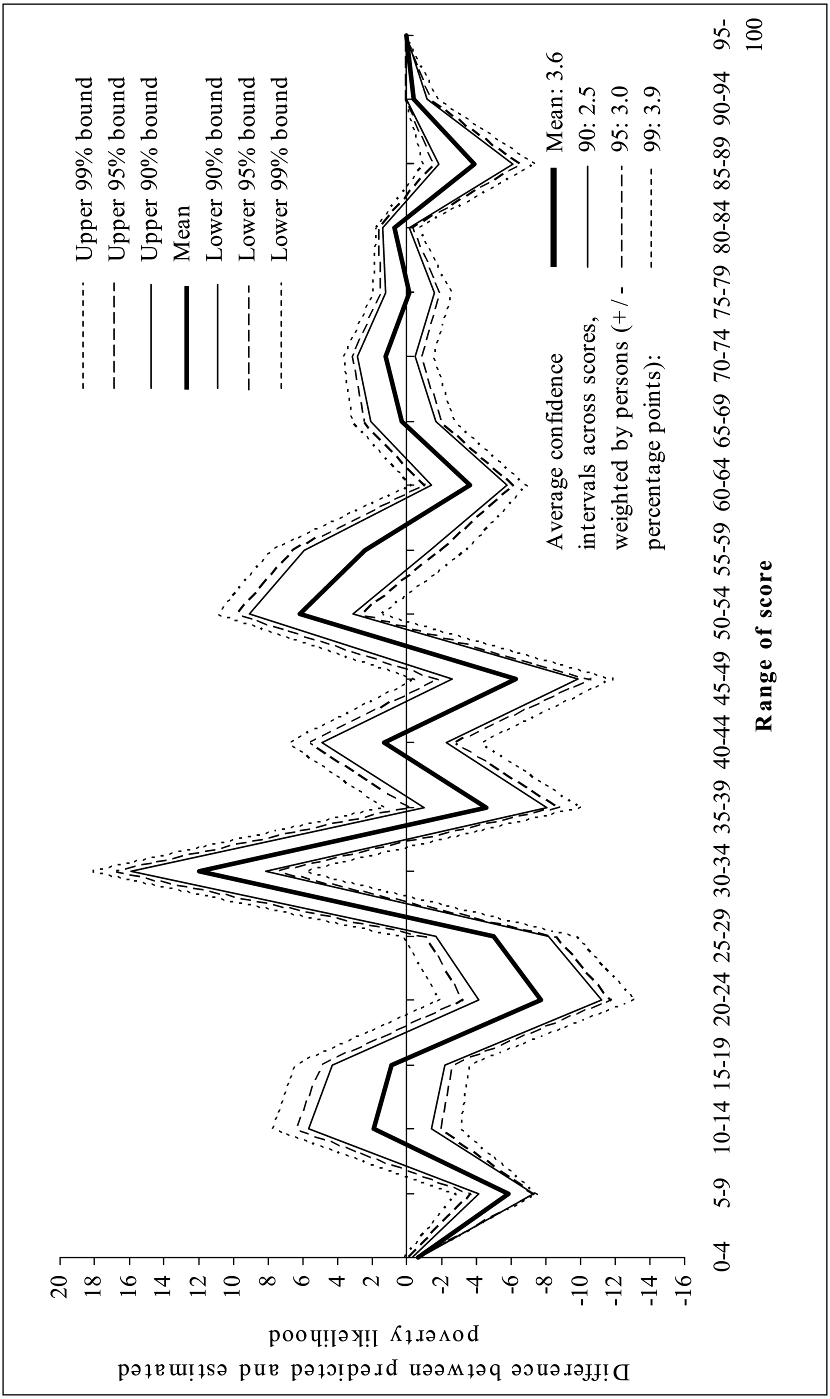

Figure 8. ROC curve of ability to rank-order households by poverty status

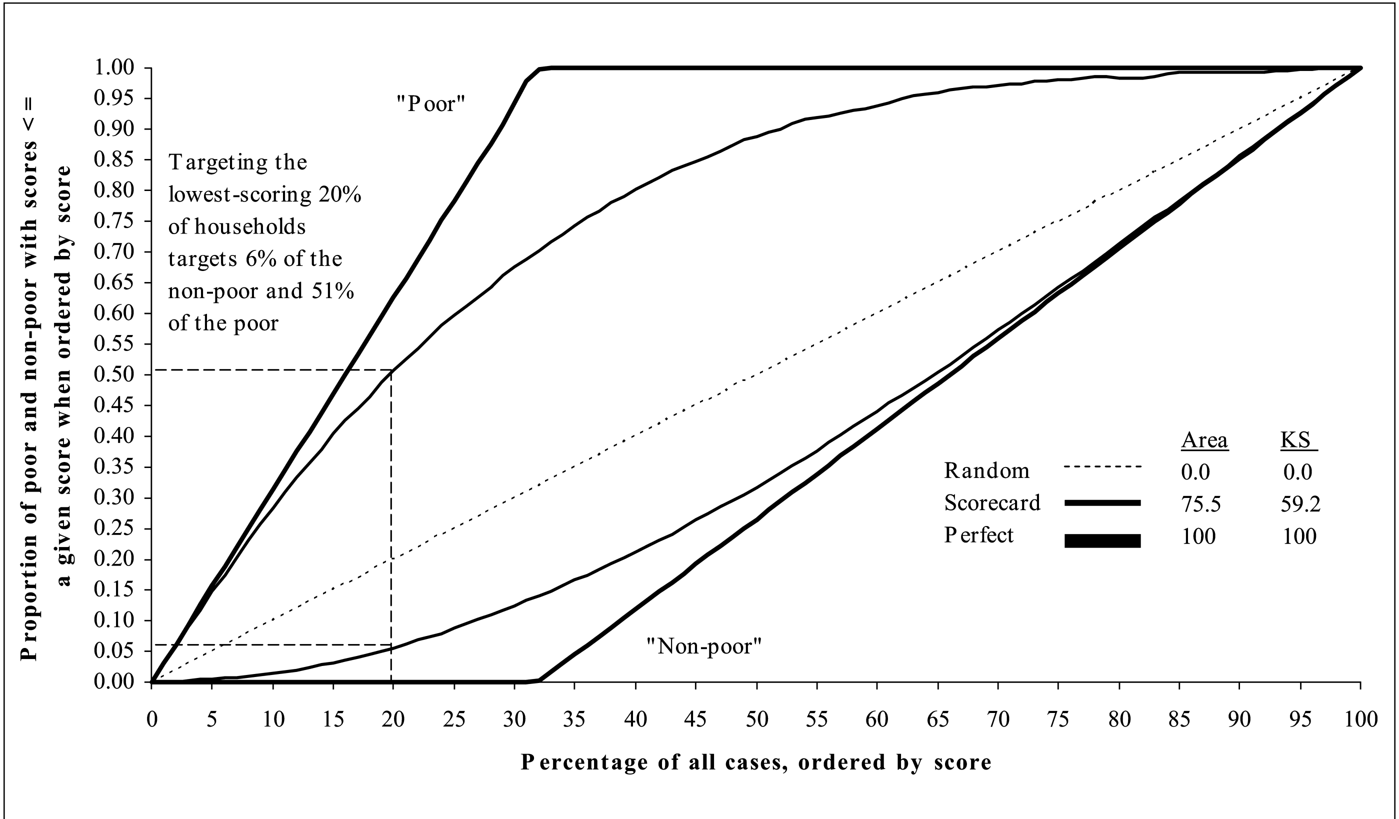

Figure 8 also reports two other common measures of rank-ordering. The first is the Kolmogorov-Smirnov (KS) statistic, defined as the maximum distance between the poor and nonpoor curves (or 59.2 in Figure 8). Higher KS implies better rank-ordering.

The second measure is the ratio of the area inside the ROC curves to the area inside the trapezoid of a hypothetical perfect scorecard (or 75.5 in Figure 8). Again, greater area within the curves implies better rank-ordering.

Is this scorecard accurate enough to be used for targeting? Errors due to scorecard inaccuracy are probably small relative to errors due to other sources (such as mistakes in data collection or fraud) and relative to the accuracy of other feasible targeting tools. All in all, Figures 6–8 suggest that the estimated likelihoods of being poor are estimated both accurately and precisely.

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 January 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 77.6, 77.7 and 48.3 percent (Figure 5). The overall poverty rate is the participants' average poverty likelihood, that is, 1,000 x (77.6 + 77.7 + 48.3) ÷ 3,000 = 67.9 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 0.1 percentage points, with a 0.37 standard deviation. The 90 percent confidence interval around the mean was +/–0.6 percentage points, the 95 percent interval was +/–0.7 percentage points, and the 99 percent interval was +/–1.0 percentage points. The estimated overall poverty rate is thus unbiased and highly precise.

Progress out of poverty over time

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

Continuing the previous example, suppose that on January 1, 2007, the same 3,000 people (some of whom may no longer be participants) are now in six groups of 500 with scores of 20, 25, 30, 35, 40, and 45 (by Figure 5, poverty likelihoods of 77.6, 76.8, 77.7, 48.6, 48.3, and 33.6 %). Their average poverty likelihood is now 60.4 percent, an improvement of 7.5 percentage points over the previous 67.9 percent. In other words, 7.5 of every 100 in this group left poverty. Among those who were poor to start with, one in nine (7.5 ÷ 67.9 = 11.1%) left poverty.

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

SETTING TARGETING CUT-OFFS

How would the poverty 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 nontargeted and treated—again, for program purposes—as if they were nonpoor.

Poverty status (expenditure below a poverty line) is distinct from targeting status (score below a cut-off). Determining poverty status requires an expensive survey. In contrast, determining targeting status 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 nonpoor will be targeted. Targeting is accurate to the extent that poverty status matches targeting status. In turn, this depends on the selection of targeting cut-offs and how it balances accuracy for the poor versus nonpoor. The standard approach uses a classification matrix and a net-benefit matrix (Adams and Hand 2000, Hoadley and Oliver 1998, Greene 1993).

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	nontargeted	(score above cut-off)
C. Truly nonpoor	mistakenly	targeted	(score at or below cut-off)
D. Truly nonpoor	correctly	nontargeted	(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 25–29, there are:

A. 12.4	truly poor	correctly	targeted
B. 19.7	truly poor	mistakenly	nontargeted
C. 2.2	truly nonpoor	mistakenly	targeted
D. 65.7	truly nonpoor	correctly	nontargeted

Targeting accuracy (and errors of undercoverage and leakage) depends on the cut-off. For example, if the cut-off were increased to 30–34, more poor (and less nonpoor) are correctly targeted:

A. 16.6	truly poor	correctly	targeted
B. 15.5	truly poor	mistakenly	nontargeted
C. 3.4	truly nonpoor	mistakenly	targeted
D. 64.5	truly nonpoor	correctly	nontargeted

Whether a cut-off of 30–34 is preferred to 25–29 depends on net benefit.

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	0.7	31.4	0.0	67.9
5-9	2.1	30.0	0.1	67.8
10-14	4.3	27.8	0.3	67.6
15-19	6.4	25.6	0.5	67.5
20-24	9.0	23.1	1.2	66.7
25-29	12.4	19.7	2.2	65.7
30-34	16.6	15.5	3.4	64.5
35-39	20.0	12.1	7.0	60.9
40-44	23.4	8.7	10.6	57.3
45-49	25.6	6.5	15.0	52.9
50-54	28.3	3.8	20.2	47.8
55-59	29.2	2.8	23.4	44.5
60-64	30.1	2.0	31.3	36.7
65-69	31.1	1.0	40.1	27.8
70-74	31.6	0.4	47.2	20.7
75-79	31.9	0.2	54.0	14.0
80-84	32.1	0.0	62.0	5.9
85-89	32.1	0.0	5.5	2.5
90-94	32.1	0.0	66.6	1.3
95-100	32.1	0.0	67.9	0.0

Figures normalized to sum to 100.

Source: Based on the 2002 APIS.

Net-benefit matrix

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

á Benefit	per truly poor person	correctly	targeted
â Cost (negative net benefit)	per truly poor person	mistakenly	nontargeted
ã Cost (negative net benefit)	per truly nonpoor person	mistakenly	targeted
ä Benefit	per truly nonpoor person	correctly	nontargeted

Each net benefit a, b, c, and d corresponds to a quadrant in the general classification matrix in Figure 9. For example, a is the net benefit associated with each truly poor person correctly targeted (quadrant A), and b is the cost (negative net benefit) associated with each truly poor person incorrectly targeted (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:

$$\text{Total net benefit} = a\cdot A + b\cdot B + c\cdot C + d\cdot D.$$

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 (Figure 10)
- ✦ Select the cut-off with the highest total net benefit

The only nontrivial step is selecting a net-benefit matrix. Some common net-benefit matrices are discussed below. In general, however, each program should thoughtfully decide how much value should be placed on successful targeting vis-à-vis errors of undercoverage and leakage. Of course, all programs already use (if only implicitly) a net-benefit matrix, and 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 nonpoor. It could reflect this valuation by increasing the weight on quadrant A (by increasing its net benefit a), and/or by decreasing the weight on quadrant B (by decreasing its net benefit a). The examples of net-benefit matrices discussed next represent different valuations of correctly/incorrectly targeting the poor/nonpoor.

"Total Accuracy"

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

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

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

This values correct classifications of the poor and nonpoor equally. Grootaert and Braithwaite (1998) and Zeller et al. (2005) use Total Accuracy to evaluate their poverty scorecards.

Figure 13 shows Total Accuracy for all cut-offs. Total net benefit is greatest (81.1) for a cut-off of 30–34; at that point, poverty segment matches poverty status for four out of five people.

The Total Accuracy criterion weighs the poor and nonpoor the same. If most people are nonpoor and/or if a scorecard is more accurate for the nonpoor, 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 nonpoor.

A simple, transparent way to reflect this valuation is to increase the relative net benefit a of correctly classifying the poor. For example, if a program values correctly targeting the poor twice as much as correctly not targeting the nonpoor, then a should be set twice as high as d in the net-benefit matrix. Then the new optimal cut-off is 50–54, the cut-off point where $a \cdot A + d \cdot D = 2 \cdot A + D$ is highest.

"Poverty Accuracy"

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

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

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 nonpoor and so maximizes leakage (the last column of Figure 13).

"Nonpoverty Accuracy"

"Nonpoverty Accuracy" counts only correct classifications of the nonpoor (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).

"BPAC"

IRIS (2005b) proposes a new measure of accuracy called the "Balanced Poverty Accuracy Criterion" or BPAC. This criterion 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 uses Figure 15. In essence, BPAC maximizes A while making B and C as close to each other as possible:

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

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 expenditure and then labels as poor those households with estimated expenditure 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.

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

Score	Total Accuracy (A + B)	Poverty Accuracy 100*A / (A + B)	Non-poverty Accuracy 100*D / (C + D)	Undercoverage 100*B / (A + B)	Leakage 100*C / (A + C)
	1 0 0 1	1 0 0 0	0 0 0 1	0 -1 0 0	0 0 -1 0
0-4	68.6	2.2	100.0	97.8	0.8
5-9	69.9	6.6	99.8	93.4	5.4
10-14	71.9	13.4	99.5	86.6	6.8
15-19	73.9	20.1	99.3	79.9	6.7
20-24	75.7	27.9	98.3	72.1	11.7
25-29	78.1	38.6	96.7	61.4	15.2
30-34	81.1	51.7	95.0	48.3	17.1
35-39	80.9	62.3	89.7	37.7	26.0
40-44	80.7	72.8	84.3	27.2	31.3
45-49	78.5	79.7	77.9	20.3	37.0
50-54	76.0	88.2	70.3	11.8	41.6
55-59	73.7	91.1	65.5	8.9	44.5
60-64	66.8	93.9	54.0	6.1	51.0
65-69	58.9	97.0	41.0	3.0	56.3
70-74	52.3	98.6	30.5	1.4	59.9
75-79	45.9	99.5	20.5	0.5	62.9
80-84	38.0	100.0	8.7	0.0	65.9
85-89	34.5	100.0	3.6	0.0	67.1
90-94	33.4	100.0	1.9	0.0	67.5
95-100	32.1	100.0	0.0	0.0	67.9
90-94	33.4	100.0	1.9	0.0	67.5
95-100	32.1	100.0	0.0	0.0	67.9

All figures in percentage units.

Figure 14. “Povery Accuracy” net-benefit matrix

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

Figure 15. Net-benefit matrix for BPAC

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

For scorecards similar to the one presented here that estimate 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.

CONCLUSION

One in three Filipinos is poor. An easy-to-use, inexpensive tool for identifying the poor could improve targeting and speed progress out of poverty. This paper presents a simple scorecard that estimates the likelihood that a person is poor.

The scorecard is built and tested using data on 38,014 households from the 2002 APIS. The scorecard is calibrated to estimate the likelihood of being poor, that is, having income below the official poverty line.

Out-of-sample bootstrap tests show that the estimates are both accurate and precise. For individual poverty likelihoods, estimates are within six percentage points of the true value with 90 percent confidence. For a group's overall poverty rate (again, whether poor or very poor), estimates are within one percentage point of the true value with 99 percent confidence.

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 program managers 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. Indicator weights are either zeros or positive integers, and scores range from zero (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.

Overall, the poverty scorecard can help development programs to target services to the poor, track participants' progress out of poverty through time, and report on participants' overall poverty rate.

REFERENCES

Adams, N.M. and D.J. Hand. 2000. Improving the practice of classifier performance assessment. Neural Computation 12:305-311.

Baulch, B. 2003. Poverty monitoring and targeting using ROC curves: examples from Vietnam. IDS Working Paper No. 161 [online]. http://www.ids.ac.uk/ids/bookshop/wp/wp161.pdf [accessed October 28, 2006].

Caire, D. 2004. Building credit scorecards for small business lending in developing markets. Bannock Consulting [online]. http://www.microfinance.com/English/Papers/Scoring_SMEs_Hybrid.pdf [accessed October 28, 2006].

Efron, B. and R.J. Tibshirani. 1993. An Introduction to the bootstrap. New York: Chapman and Hall.

Ericta, C.N. 2005. 2004 Annual Poverty Indicators Survey (APIS): preliminary results [online]. http://www.census.gov.ph/data/pressrelease/2005/ap2004ptx.html [accessed October 28, 2006].

Fuller, R. 2006. Measuring poverty of microfinance clients in Haiti [online]. http://www.microfinance.com/English/Papers/Scoring_Poverty_Haiti_Fuller.pdf [accessed October 28, 2006].

Goodman, L.A. and W.H. Kruskal. 1979. Measures of association for cross classification. New York: Springer-Verlag.

Greene, W.H. 1993. Econometric analysis: second edition. New York: MacMillan.

Grootaert, C. and J. Braithwaite. 1998. Poverty correlates and indicator-based targeting in Eastern Europe and the former Soviet Union. World Bank Policy Research Working Paper No. 1942. Washington, D.C.:WB [online]. http://www.worldbank.org/html/dec/Publications/Workpapers/WPS1900series/wps1942/wps1942.pdf [accessed October 28, 2006].

Hoadley, B. and R.M .Oliver. 1998. Business measures of scorecard benefit. IMA Journal of mathematics applied in business and industry 9:55-64.

IRIS Center. 2005a. Accuracy results for 12 poverty assessment tool countries [online]. http://www.povertytools.org/documents/Accuracy%20Results%20for%2012%20Countries.pdf.

———. 2005b Notes on assessment and improvement of tool accuracy [online]. http://www.povertytools.org/documents/Assessing%20and%20 Improving% 20Accuracy.pdf.

National Statistic Coordination Board. 2004. Poverty threshold: P11,906 in 2002 [online]. http://www.nscb.gov.ph/headlines/2004/30Jan04_povtresh.asp [accessed September 3, 2005].

Schreiner, M. 2002. Scoring: the next breakthrough in microfinance? Occasional Paper No. 7, Consultative Group to Assist the Poorest, Washington, D.C. [online]. http://www.cgap.org/docs/OccasionalPaper_07.pdf [accessed October 28, 2006].

———. 2005a Un indice de pobreza para México. Memo for Grameen Foundation USA [online]. http://www.microfinance.com/Castellano/Documentos/Scoring_Pobreza_Mexico.pdf [accessed October 28, 2006].

———. 2005b. IRIS questions on poverty scorecards. Memo for Grameen Foundation USA [online]. http://www.microfinance.com/English/Papers/Scoring_Poverty_Response_to_IRIS.pdf [accessed October 28, 2006].

———. 2006. Is one simple poverty scorecard enough for India? Memo for Grameen Foundation USA [online]. http://www.microfinance.com/English/Papers/Scoring_Poverty_India_Segments.pdf [accessed October 28, 2006].

Schreiner, M, M. Matul, E. Pawlak, and S. Kline. 2004. Poverty scorecards: lessons from a microlender in Bosnia-Herzegovina. Microfinance Risk Management [online]. http://www.microfinance.com/English/Papers/Scoring_Poverty_in_BiH_Short.pdf [accessed October 28, 2006].

Wodon, Q.T. 1997. Targeting the poor using ROC curves. World Development 25(12):2083-2092.

Zeller, M. 2004. Review of poverty assessment tools. Accelerated Microenterprise Advancement Project [online]. http://www.povertytools.org/documents/Review%20of%20Poverty%20Assessment%20Tools.pdf.

Zeller, M., G. Alcaraz .and J. Johannsen. 2005. Developing and testing poverty-assessment tools: results from accuracy tests in Peru. Accelerated Microenterprise Advancement Project [online]. http:/www.poverty tools.org/documents/Peru%20Accuracy%20Report.pdf