Simple Poverty Scorecard® Poverty-Assessment Tool Bangladesh

Mark Schreiner

21 September 2006

This document is at SimplePovertyScorecard.com

Abstract

The Simple Poverty Scorecard®-brand poverty-assessment tool uses ten low-cost indicators from Bangladesh's 2000 Household Income and Expenditure Survey to estimate the likelihood that a household has expenditure below a given poverty line. Field workers can collect responses in about ten minutes. The scorecard's accuracy is reported for a range of poverty lines. The scorecard is a practical way for pro-poor programs in Bangladesh to measure poverty rates, to track changes in poverty rates over time, and to segment clients for targeted services.

Acknowledgements

This work was funded by Grameen Foundation. I am grateful for data from the Bangladesh Bureau of Statistics and for advice from Nurul Alam, Josh Al-Zayed, M. Niaz Asadullah, Nigel Biggar, Nusha Choudhury, Asif Dowla, Anna Fruttero, Varun Gauri, Stephen Haslett, Azizur Khan, Qaiser M. Khan, Jonathan Morduch, Jeff Toohig, Quentin Wodon, and Salman Zaidi. "Simple Poverty Scorecard" is a Registered Trademark of Microfinance Risk Management, L.L.C. for its brand of poverty-assessment tools.

Author

Mark Schreiner directs Microfinance Risk Management, L.L.C. He is also a Senior Scholar at the Center for Social Development at Washington University in Saint Louis.

Simple Poverty Scorecard® Poverty-Assessment Tool

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

Indicator	Response	Points	Score
1. What type of latrine does the household use?	A. Open field	0	
	B. *Kacha* (temporary or permanent) or pit *pacca*	7	
	C. Sanitary or water seal *pacca*	12	
2. How many household members are 11-years-old or younger?	A. Four or more	0	
	B. Three	7	
	C. Two	12	
	D. One	17	
	E. None	26	
3. Does any household member work for a daily wage?	A. Yes	0	
	B. No	7	
4. How many rooms does the house have (excluding ones used for business)?	A. One	0	
	B. Two or three	3	
	C. Four or more	9	
5. Do all children ages 6 to 17 attend school?	A. No	0	
	B. No members ages 6 to 17	4	
	C. Yes	6	
6. Does the household own a television set?	A. No	0	
	B. Yes	11	
7. How many hectares of cultivable land does the household own?	A. Less than 0.34	0	
	B. 0.34 to 0.99	3	
	C. 1.00 to 1.99	4	
	D. 2.00 or more	9	
8. What is the main construction material of the walls of the house?	A. Hemp/hay/bamboo, or mud brick	0	
	B. C.I. sheet/wood	5	
	C. Brick/cement	7	
9. Does the household own drawing-room furniture?	A. No	0	
	B. Yes	9	
10. Does the household have a separate kitchen?	A. No	0	
	B. Yes	4	

Score:

Simple Poverty Scorecard® Poverty-Assessment Tool Bangladesh

1. Introduction

This paper presents the Simple Poverty Scorecard poverty-assessment tool, a low-cost way to help development projects in Bangladesh to target services, track changes in poverty over time, and report on participants' poverty rates.

The low-tech scorecard is derived from Bangladesh's 2000 Household Income and Expenditure Survey (Ahmed, 2004). Indicators were selected to be:

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

All scorecard weights are non-negative integers. Scores range from 0 (most-likely poor) to 100 (least-likely poor). Field workers can compute scores by hand in real time.

A participant's score corresponds to a "poverty likelihood", that is, the probability of being poor. The share of all participants who are poor is the average poverty likelihood. For participants over time, progress is the change in their average poverty likelihood.

While 5-, 10- and 15-indicator scorecards were tested, only the 10-indicator scorecard is presented. Fifteen indicators were about as accurate as 10, and 5 indicators are less sensitive to changes in poverty. Separate rural and urban scorecards were not tested, as Schreiner (2006a) finds that a country-wide scorecard is almost as accurate.

By construction, the scorecard is accurate in that, on average, the estimated poverty likelihoods of individuals and the overall poverty rate of groups are equal to the true values. The scorecard is also accurate for targeting in that poor people are concentrated among low scores and non-poor people are concentrated among high scores.

Precision was measured by bootstrapping a hold-out sample. The 90-percent confidence intervals for estimated poverty likelihoods are about +/–5 percentage points, and the 90-percent interval for estimated overall poverty rates is +/–1.5 percentage points.

The scorecard is an appropriate tool for USAID microenterprise grantees who are required to report (in a proven, objective way) the share of participants who live on less than $1/day.

2. Existing similar tools for Bangladesh

There are already at least six similar tools for Bangaldesh; why one more?[1] The new scorecard is more likely to actually be used than those reviewed below because it is simple, and it is the only tool whose precision is has been measured.

The Grameen Bank's poverty-measurement tool (Dowla and Barua, forthcoming) is based on its in-house expertise and experience, and as such it is well-accepted by staff. Some indicators, however, are subjective (such as "all family members are conscious of their health"). Furthermore, two indicators ("Member's minimum weekly installment is 200 taka" and "Borrower maintains an average annual balance of 5000 taka in her savings accounts") are relevant only for microfinance participants. It is also too stringent, counting a member as non-poor only if all 10 indicators have a non-poor value. But many non-poor households (at least by an expenditure-based poverty line) have some indicators showing a poor value. Most important, Grameen's scorecard is not benchmarked to an expenditure-based poverty line.

The paper by Zeller, Alcaraz V., and Johannsen (2004) is like this paper, except:

- It estimates expenditure directly, rather than the probability that expenditure is below a poverty line
- The sample is 800 rather than 7,440
- Some indicators in some scorecards (such as "clothing expenditures in the past 12 months" and "total value of household assets") are difficult to collect accurately
- Some indicators in some scorecards (such as "household feels that clothing expenditures are below need") are subjective
- The scorecards are more complex mathematically (using ratios, logarithms, squares, and continuous indicators), so scoring in the field requires a portable computer
- Estimates of overall poverty rates are biased and inconsistent[2]
- Weights are not presented, so no program can actually use the scorecard

Wodon (1997) describes a set of poverty-measurement tools derived from the 1991 HES (the predecessor of the HIES). The scorecards estimate poverty likelihoods (not expenditure), but no weights are reported.

---

[1] In general, there is a large literature on "proxy means tests" and "small area poverty mapping", much of which is the same as the scorecard here.

[2] This follows from the fact that the indicator function used to convert estimated expenditure into poor/non-poor poverty status is non-linear and discontinuous.

Kam *et al.* (2004) present two scorecards, one based on a rural survey and one on census data. Both estimate income directly and are simple enough to be practical, although the indicators—due to data limitations—cover fewer aspects of poverty than other scorecards reviewed here.

The scorecard in Bangladesh Bureau of Statistics (2004, hereafter "BBS") uses the 2000 HIES, has indicators similar to those in this paper, and is also simple enough to be practical. It estimates expenditure directly. Like the other scorecards reviewed here, its accuracy was checked with R^2, and precision was not measured.

Finally, Cortez *et al.* (2005, hereafter "Cortez") make an eminently practical 14-indicator scorecard based on the 2000 HIES. They even show how to use the scorecard to estimate expenditure. They do not, however, adjust the poverty line for cost-of-living, they check accuracy with R^2, and they do not measure precision. They find that the scorecard successfully classifies as poor 94 percent of 220 people known to be *extremely* poor, but classifying the extremely poor is less difficult than classifying people near the poverty line.

Considering both accuracy and feasibility, the best scorecard among the six is Cortez. The advantages of the simple scorecard in this paper (some of which—but not all—are shared by some of the other scorecards just reviewed) include:

- Concern for both accuracy and ease-of-use
- Based on 7,440 households from the nationally representative 2000 HIES
- Adjustments to the poverty line for regional cost-of-living
- Benchmarked to an objective, expenditure-based poverty line
- Simple, inexpensive, reliable, objective indicators liable to change with poverty
- Simple weights (all non-negative integers)
- Simple scores, from 0 (least-likely poor) to 100 (most-likely poor)
- Standard measures of accuracy and precision
- Unbiased, consistent estimates of:
    - Individuals' poverty likelihoods
    - Groups' overall poverty rates
- Feasible for use in the field by hand in real time

3. Poverty lines in Bangladesh

The Bangladesh Bureau of Statistics defines a "lower" and an "upper" poverty line with adjustments for cost-of-living in 14 regions (Figure 1).

The "lower" poverty line is the expenditure needed to consume 2,122 kcal of food (and nothing else) per person per day (Ahmed, 2004). By this minimum standard, 33.7 percent of Bangladshis were poor. Of course, people below this "lower" line consumed some non-food items, so they had less than 2,122 kcal per day.

The "upper" poverty line is observed total expenditure in the 2000 HIES among those who actually consumed 2,122 kcal. By this line, 49.8 percent of Bangladeshis were poor.

A third poverty line is the $1/person/day (at purchase-power parity) international benchmark. In 2000, this was 20.20 taka (Sillers, 2006). After adjusting for regional cost-of-living, 43.7 percent of Bangladeshis lived on less than $1/day.

For international comparability and to fit USAID requirements, this paper uses $1/day.

Figure 1: Bangladesh poverty lines (taka/person/day in 2000)

Region	Official		$1/day, adjusted for cost-of-living
	"Lower"	"Upper"	
SMA Dhaka	649	893	832
Other Urban Dhaka	521	629	586
Rural Dhaka	548	659	614
Rural Faridpur, Tangail, and Jamalpur	540	591	551
SMA Chittagong	702	971	905
Other Urban Chittagong	694	818	762
Rural Sylhet and Comilla	572	738	688
Rural Noakhali and Chittagong	582	719	670
Urban Khulna	609	803	748
Rural Barishal and Pathuakali	546	616	574
Rural Khulkna, Jessore, and Kushtia	527	624	581
Urban Rajshahi	557	726	676
Rural Rajshahi and Pabna	586	690	643
Rural Bogta, Rangpur, and Dinajpur	510	582	542
Poverty rate (%), all Bangladesh:	33.7	49.8	43.7

4. Indicator preparation

About 620 potential poverty indicators were prepared for the 5,580 households from the 2000 HIES used to construct the scorecard. Broadly, the indicators cover:

- Household demographics (such as education or number of children)
- Characteristics of the homestead (type of wall, type of toilet, or hectares of land)
- Household consumption (such as liquid milk or beef)
- Household durable goods (such as televisions or wrist watches)

As a first step, the ability of each indicator to predict poverty was tested with the "uncertainty coefficient", an entropy-based measure (Goodman and Kruskal, 1979).

About 110 indicators were selected for further analysis. These are listed in Figure 2 and ranked by the uncertainty coefficient. They are worded as in the HIES, with possible responses ordered starting with those most strongly linked with poverty.

Many indicators in Figure 2 have similar relationships with poverty. For example, households with an electrical connection are also more likely than other households to own a television. If a scorecard includes "owns a television", then also including "electrical connection" adds little. Thus, many indicators with strong links with poverty are not in the scorecards because they are similar to other included indicators.

The scorecard also aims to measure *changes* in poverty through time. Thus, some indicators that are unlikely to change even if poverty changes (such as the highest grade completed by the female spouse/head) are omitted in favor of indicators that are less-powerful but more likely to change.

Some potential indicators did not make it to Figure 2 because they are difficult to collect ("In the past two weeks, did anyone eat any vegetables?"), difficult to compute ("How many square feet does the house have per resident?") or too sensitive ("Did you receive Zakat?"). Finally, all indicators related to past consumption were omitted from the scorecard, as field agents cannot straightforwardly verify whether the respondent's memory or motives are trustworthy when answering questions such as "In the past two weeks, did anyone in the household drink liquid milk" or "In the past year, did anyone in the household acquire leather shoes or sandals".

Figure 2: Poverty indicators ranked by their uncertainty coefficient

Uncertainty coefficient	Indicator (Responses ordered starting from the one most-closely linked with poverty)
140	What is the highest education level of any household member? (None or Class 1–4; Class 5–6; Class 7–8; Class 9 or above)
137	In the past year, did anyone in the household acquire leather shoes or sandals? (No; Yes)
111	What type of latrine does the household use? (Open field; Kacha latrine (temporary or permanent) or pit Pacca; Sanitary or water-seal Pacca latrine)
107	What is the main construction material of the walls of the residence? (Hemp/hay/bamboo or mud brick; C.I. sheet/wood; Brick/cement)
103	Does anyone in the household own a wrist watch? (No; Yes)
101	Does any household member work for a daily wage? (No; Yes)
91	Does anyone in the household own a clock? (No; Yes)
89	Does the residence have an electricity connection? (No; Yes)
87	Does the household own a television? (No; Yes)
86	What daily wage does the male head/spouse earn? (1–70 taka; >70 taka; Does not work for a daily wage)
84	In the past year, did anyone in the household acquire leather shoes? (No; Yes)
84	What is the highest grade completed by the female head/spouse? (None; Class 1–5; Class 6–8; Class 9 or more)
83	In the past year, did the household spend anything on Qurbani? (No; Yes)
82	What is the highest grade completed by the male head/spouse? (None; Class 1–6; Class 7–9 or SSC; HSC or more)
82	Does the household own a fan? (No; Yes)
81	Can the female head/spouse read a letter and write a letter? (No; Yes; No female spouse/head)
77	In the past two weeks, did anyone in the household eat meat, including beef, buffalo, mutton, sheep, hen, or duck? (No; Yes)
75	In the past year, did the household head work for a daily wage? (Yes; No)
75	In the past two weeks, did anyone in the household consume milk or dairy products, including liquid milk, powdered milk, curd, Casein/Ponir/butter, or other milk drinks? (No; Yes)
74	How many household members are 11 years old or younger? (4 or more; 3; 2; 1; 0)
73	How many square feet is the inside of the dwelling? (200 or less; 201-300; 301-600; 601 or more)
71	In the past two weeks, has anyone in the household consumed ginger? (No; Yes)
69	What is the main construction material of the roof of the residence? (Hemp/hay/bamboo or tile/wood; C.I. sheet/wood; Brick/cement)
68	Does the household own a two-in-one cassette player? (No; Yes)
67	In the past two weeks, has anyone in the household consumed clove/black pepper/ Cassia-leaf? (No; Yes)
67	In the past two weeks, has anyone in the household consumed sugar/Misri? (No; Yes)
67	In the past year, did the household have any personal teaching expenses for male or female students? (No; Yes)

Figure 2: Poverty indicators ranked by their uncertainty coefficient (cont.)

Uncertainty coefficient	Indicator (Responses ordered starting from the one most-closely linked with poverty)
67	How many rooms (excluding rooms used for business) does your household occupy? (1; 2 or 3; 4 or more)
65	Can the male head/spouse read a letter and write a letter? (No; Yes; No male spouse/head)
62	How many household members are 14 years old or younger? (4 or more; 3; 2; 1; 0)
60	In the past month, did anyone in the household pay for bus fare, public transport, or rickshaw/van fare? (No; Yes)
60	In the past two weeks, did anyone in the household consume liquid milk? (No; Yes)
59	Does the household own drawing room furniture? (No; Yes)
58	In the past two weeks, did anyone in the household consume beef? (No; Yes)
56	In the past month, did anyone in the household pay rickshaw/van fare? (No; Yes)
55	In the past month, did the household pay salaries or wages for servants, drivers, guards, gardeners, housekeepers, etc.? (No; Yes)
54	What type of employment does the male head/spouse have? (Daily wage; Self-employed in agriculture; No employment; Self-employed in non-agriculture; Salaried wage)
53	In the past two weeks, did anyone in the household consume hens? (No; Yes)
53	In the past month, did the household pay Fitra? (No; Yes)
50	In the past two weeks, did anyone in the household consume lentils/Musur? (No; Yes)
50	In the past month, did the household pay bus fare? (No; Yes)
50	Do all children under age 5 have immunization cards? (No; Yes; No children under age 5)
49	In the past two weeks, did anyone in the household smoke cigarettes? (No; Yes)
48	In the past two weeks, did anyone in the household eat ripe bananas? (No; Yes)
45	How many household members are 17 years old or younger? (5 or more; 4 or 3; 2; 1; 0)
44	In the past two weeks, did anyone in the household consume apples? (No; Yes)
42	In the past two weeks, did anyone in the household consume hen eggs? (No; Yes)
41	How many male household members are 11 years old or younger? (2 or more; 1; 0)
41	How many hectares of cultivable land does the household own? (<0.34; 0.34 to 0.99; 1 to 1.99; 2 or more)
41	How many household members are 5 years old or younger? (2 or more; 1; 0)
41	Does the household own a refrigerator or freezer? (No; Yes)
40	In the past two weeks, did anyone in the household consume Rhui/Katla/Mrigel/Kal Baush? (No; Yes)
39	In the past year, did the household rent land or other property? (No; Yes)
38	Do all children ages 6 to 14 attend school? (No; Yes; No children in ages 6 to 14)
37	In the past month, did the household buy a mosquito coil? (No; Yes)
36	In the past month, did the household buy bath soap, shampoo, toothpaste, etc.? (No; Yes)
36	How many household members are 6 to 11 years old? (3 or more; 2; 1; 0)
35	How many hectares of cultivable land plus homestead land does the household own? (None; 0.01 to 0.99; 1 or more)
35	In the past two weeks, did anyone in the household consume Rasogolla/Chamcham/Shandash? (No; Yes)

Figure 2: Poverty indicators ranked by their uncertainty coefficient (cont.)

Uncertainty coefficient	Indicator (Responses ordered starting from the one most-closely linked with poverty)
35	In the past month, did the household use natural gas or bio-gas? (No; Yes)
34	How many male household members are 14 years old or younger? (3 or more; 2; 1; 0)
33	In the past two weeks, did anyone in the household consume tea/coffee leaf? (No; Yes)
33	Does the house have a separate kitchen? (No; Yes)
32	In the past two weeks, did anyone in the household consume tomatoes? (No; Yes)
31	Does the household own any dining room furniture? (No; Yes)
31	In the past year, did anyone in the household buy gold jewelry? (No; Yes)
31	In the past year, did anyone in the household have medical tests such as x-rays, blood, urine, etc.? (No; Yes)
31	In the past two weeks, did anyone in the household consume bread/Bonroti? (No; Yes)
30	Do all children ages 6 to 17 attend school? (No; Yes; No children ages 6 to 17)
29	In the past two weeks, did the household consume flour? (No; Yes)
29	In the past two weeks, did anyone in the household consume Biri? (No; Yes)
28	How many hectares of cultivable land does the household rent-out, sharecrop, or mortgage out? (None; 0.01 to 0.29; 0.3 or more)
28	In what sector of activity does the male head/spouse work? (Agriculture, wood, forest products, fishing, transport, carpentry, wood-working, metal working, or construction; Food retailing, garments, retail sales, others, or no activity; Public-sector employees, teachers, or health care)
28	In the past two weeks, did anyone in the household consume medium rice? (No; Yes)
28	In the past month, did the household buy Vim/dish-cleaning supplies? (No; Yes)
28	In the past two weeks, did anyone in the household consume Fanta/Coca-Cola/Sherbat, etc.? (No; Yes)
27	In the past two weeks, did anyone in the household consume puffed rice? (No; Yes)
26	In the past two weeks, did anyone in the household consume coarse rice? (No; Yes)
26	In the past two weeks, did anyone in the household consume Hilsa? (No; Yes)
26	How many female household members are 11 years old or younger? (2 or more; 1; 0)
24	In the past two weeks, did anyone in the household consume tea/coffee? (No; Yes)
24	What is the main source of water for non-drinking use? (Pond/river, well, waterfall/spring, or other; Tube well; Supply water)
24	In the past two weeks, did anyone in the household consume grapes? (No; Yes)
24	In the past two weeks, did anyone in the household consume Pangash/Boal/Air? (No; Yes)
24	In the past two weeks, did anyone in the household consume Kai/Magur/Shinghi/Khalisha? (No; Yes)
23	Does any household enterprise have a non-household employee? (No, or no enterprise; Yes)
23	In the past two weeks, did anyone in the household consume wheat? (No; Yes)
23	Does the household own a camera or a camcorder? (No; Yes)
22	Do all children ages 12 to 17 attend school? (No; Yes; No children ages 12 to 17)

Figure 2: Poverty indicators ranked by their uncertainty coefficient (cont.)

Uncertainty coefficient	Indicator (Responses ordered starting from the one most-closely linked with poverty)
22	In the past two weeks, did anyone in the household consume garlic? (No; Yes)
22	What is the household's main source of drinking water? (Not supply water; Supply water)
21	Do all female children ages 6 to 11 attend school? (No; Yes; No female children ages 6 to 11)
21	In the past two weeks, did anyone in the household consume fine rice? (No; Yes)
21	In the past month, did anyone in the household pay taxi/Tempoo/Mishuk fare? (No; Yes)
21	Does the household own a tube well (for drinking water only)? (No; Yes)
20	Does any household member have salaried employment? (No; Yes)
19	Do all female children ages 6 to 14 attend school? (No; Yes; No female children ages 6 to 14)
19	In the past year, did the household buy a wrist watch, clock, etc.? (No; Yes)
17	Do all female children ages 6 to 17 attend school? (No; Yes; No female children ages 6 to 17)
15	Does the household own a telephone/mobile? (No; Yes)
15	Do all children ages 6 to 11 attend school? (No; Yes; No children ages 6 to 11)
13	Where is the residence located? (Rural; Urban)
13	How many household members are involved with self-employment? (0, 1, 2 or more)
12	Does the household own a bicycle? (No; Yes)
11	What is the age of the female head/spouse? (Less than 35; 35 to 44; 45 or more)
9	For how many years has the household's oldest enterprise been operating? (Less than 1 or no enterprise; 1 to 7; 8 or more)
9	What is the age of the male head/spouse? (Less than 30; 30 to 39; 40 to 49; 50 or more)
8	Are any household members involved in non-agricultural self-employment? (No; Yes)
8	Does the household own a motorcycle or scooter? (No; Yes)
8	How many family members work in the household's enterprise? (None; No enterprise; 1 or more)
8	Does the household operate any enterprises? (No; Yes)
8	Does the household own a radio? (No; Yes)
4	Does the household own any cattle? (No; Yes)
3	Are any household members involved in agricultural self-employment? (No; Yes)

5. Selecting indicators for the scorecards

An appropriate statistical approach for classifying people as poor/non-poor is Logit regression. Indicators were selected by combining statistics with the judgment of an analyst with expertise in both scoring and microfinance:

1. Start with a scorecard with no indicators
2. For each candidate indicator not already in the scorecard:
   A. Add the indicator to the scorecard
   B. Derive weights with Logit
   C. Record the improvement in general accuracy measured by the "c" statistic
3. Select an indicator based on (Schreiner *et al.*, 2004; Zeller, 2004):
   A. Likelihood of acceptance by users:
      i. "Face validity" (experience, theory, and common sense)
      ii. Simplicity and cost of collection
   B. Likelihood of changing as poverty status changes
   C. Accuracy
   D. Contrast with indicators already in the scorecard
   E. Verifiability and susceptibility to strategic falsification
4. Add the selected indicator to the scorecard
5. Repeat steps 2–4 until there are 10 indicators
6. Transform the original Logit weights so that:
   A. All weights are non-negative integers
   B. The minimum score is 0 (most likely poor), and the maximum is 100

This "MAXC" algorithm for Logit is analogous to the "MAXR" algorithm for ordinary-least squares in Zeller, Alcaraz V., and Johannsen (2005 and 2004); Zeller and Alcaraz V. (2005a and 2005b); and IRIS (2005a and 2005b). If all classification errors are equally costly, then R^2 and "c" are good general measures of accuracy.

"c" is the area under a Receiver Operator Characteristic curve (Baulch, 2003) that plots the share of poor people (vertical axis) versus the share of all people ranked by score (horizontal axis). It 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. Finally, it is equivalent to the Mann-Whitney U statistic.

Figure 3: Simple Poverty Scorecard tool for Bangladesh

Indicator	Attributes					Points
1. What type of latrine does the household use?			Open field	Kacha (temporary or permanent) or pit Pacca	Sanitary or water-seal Pacca	
			0	7	12	
2. How many household members are 11 years old or younger?	4 or more	3	2	1	0	
	0	7	12	17	26	
3. Does any household member work for a daily wage?				Yes	No	
				0	7	
4. How many rooms does the house have (excluding ones used for business)?			1	2 or 3	4 or more	
			0	3	9	
5. Do all children ages 6 to 17 attend school?			No	No children ages 6 to 17	Yes	
			0	4	6	
6. Does the household own a television set?				No	Yes	
				0	11	
7. How many hectares of cultivable land does the household own?		Less than 0.34	0.34 to 0.99	1 to 1.99	2 or more	
		0	3	4	9	
8. What is the main construction material of the walls of the house?		Hemp/hay/bamboo or mud brick		C.I. sheet/wood	Brick/cement	
		0		5	7	
9. Does the household own drawing room furniture?				No	Yes	
				0	9	
10. Does the house have a separate kitchen?				No	Yes	
				0	4	
0					Total:	

6. Scorecard use

As explained in Schreiner (2005), the main goal is not to maximize accuracy but rather to maximize the likelihood of programs' using scoring. When scoring projects fail, the culprit is usually not inaccuracy but rather the failure of users to accept scorecards and to use them properly (Schreiner, 2002). The roadblocks are less technical than human and organizational, less statistics than change management. "Accuracy" is easier—and matters less—than "practicality".

The simple, low-tech design here is meant to help users understand and trust the scorecard so that they will use it. While accuracy is important, it must be balanced against 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 to them.

This practical focus naturally leads to a one-page scorecard (Figure 3) whose features enable field workers to compute scores by hand in real time:

- Few indicators (10)
- Categorical indicators ("does the household own a TV", not "total value of assets")
- User-friendly weights (non-negative integers, no arithmetic beyond simple addition)

Among other things, this design permits rapid poverty appraisal, for example, determining in a day which village residents qualify for, say, microfinance membership, work-for-food programs, or ration cards.

The scorecard in Figure 3 could be photocopied to take to the field. It could also serve as a template for data-entry screens to record indicators, scores, poverty likelihoods, and changes in poverty likelihood over time.

When using the scorecard, field agents read each question, circle the response and the corresponding points, verify the response, write the points in the right-hand column, add up the points to get the score, and then execute program policy based on the score.

Field agents must be trained how to collect indicators. If they put garbage in, the scorecard will put garbage out. On-going audits of data quality are advisable.

Programs should record in a digital database at least the score, date, and participant identifier, if not also the values of the indicators. This will simplify computation of average poverty likelihoods and other analyses, both at a point in time and for changes through time (Matul and Kline, 2003).

7. Scores and poverty likelihoods

A *score* (sum of scorecard points) is not the same as a *poverty likelihood* (probability of being poor). But each score is associated with a poverty likelihood via a simple table (Figure 4, column "Poverty Likelihood for people with score in range (%)").

For example, scores of 0–4 correspond to a poverty likelihood of 93.0 percent because, in the bootstrapped samples from the hold-out sample from the 2000 HIES (see below), 93.0 percent of people with scores of 0–4 were poor.

In the same way, scores of 85–100 correspond to a poverty likelihood of 0 percent, as no one with scores in this range were poor.

In rough terms, the accuracy of scoring for targeting is the extent of concentration of the poor among low scores and of the non-poor among high scores.

In Figure 4, the column "% of people <=score who are poor" shows the share of all Bangladeshis with a given score or less who are poor. For example, 92.3 percent of those with scores of 15–19 or less are poor.

Likewise, the column "% of people > score who are non-poor" shows the share of all Bangladeshis with scores greater than a given range who are non-poor. For example, 60.8 percent of those with scores of more than 15–19 are poor.

Programs can use Figure 4 to set policy cut-offs for targeting program services. For example, suppose the program decides treat people scoring 19 or less as "poor" and people scoring 20 or more as "non-poor". Then—assuming the program serves a population that mirrors that of Bangladesh as a whole—89.8 percent of those treated by the program as "poor" truly are poor (and 11.2 percent are non-poor), and 60.8 percent treated by the program as "non-poor" truly are non-poor (and 39.2 percent are poor).

Alternatively, the program could aim for a given overall poverty rate (say, 75 percent) and then choose a corresponding cut-off (here, 30–34). Usually, however, program participants will not mirror the population of Bangladesh as a whole, so the two right-hand columns of Figure 4 are not relevant.[3] In this case, a program aiming for a given overall poverty rate would set a policy cut-off and then monitor its overall poverty rate, adjusting the cut-off as required.

[3] Even if participants do not mirror the country as a whole, Figure 4 can still be used with a net-benefit matrix to set policy cut-offs, as discussed later.

Figure 4: Scores and corresponding poverty likelihoods

Score	Poverty likelihood for people with score in range (%)	% of people <=score who are poor
0-4	93.0	93.0
5-9	94.6	94.2
10-14	94.6	94.5
15-19	89.8	92.3
20-24	76.0	85.0
25-29	76.7	82.2
30-34	59.4	75.9
35-39	51.4	70.9
40-44	38.3	65.4
45-49	26.2	60.6
50-54	14.3	56.3
55-59	5.7	53.1
60-64	5.1	50.7
65-69	6.7	48.7
70-74	1.9	47.2
75-79	3.7	46.3
80-84	0.3	45.7
85-89	0.0	45.3
90-94	0.0	45.1
95-100	0.0	45.1
Total:	45.1	
0		

8. Correspondences between scores and poverty likelihoods

Poverty likelihoods in Figure 4 are derived from the 2000 HIES using a bootstrapped hold-out sample.

At the start of the study, a *hold-out sample* of 1,860 households (25 percent of the 2000 HIES) was selected at random, and all steps in scorecard construction were done on the remaining 5,580 households (without "peeking" at the hold-out sample). The hold-out sample was then used to determine poverty likelihoods and to measure precision.

Except Setel *et al.* (2003), all poverty-measurement tools to date (including Schreiner 2006a–2006d) have been built and tested on the same set of households. This overstates accuracy because all scorecards are inevitably overfit to some extent.[4] *Overfit* means that the choice of indicators, the form of indicators, and the weights represent not only universal, permanent patterns present among all households but also random and/or transitory patterns found only among the households used to build the scorecard.

To get measures of precision uncontaminated by overfitting requires testing on households not used to build the scorecard. This also mimics how the scorecard is actually used in practice.

Poverty likelihoods are derived from the hold-out sample as follows:

- Score all households in the hold-out sample
- For a given score, define the *poverty likelihood* as the share of people with that score who are poor

For example, suppose that 85.6 percent of people in the hold-out sample with scores of 15–19 are poor. The poverty likelihood for scores of 15–19 is then 85.6 percent.

Of course, drawing a different hold-out sample would lead to a different poverty likelihood. While estimates vary from sample to sample, a precise estimator is one that usually is close to the true value being estimated.

[4] Bigman *et al.* (2000) show this for a poverty-assessment tool in Burkina Faso.

Bootstrapping is a simple way to measure precision (Efron and Tibshirani, 1993). In this paper, it is also used to produce estimates of poverty likelihoods that are free of overfitting. The algorithm is:

- From the hold-out sample, draw a new sample of 1,860 households *with replacement*[5]
- For each given range of scores, compute the poverty likelihood as described above
- Repeat the previous two steps many times (here, 10,000)
- For a given score range, define the poverty likelihood (first column of Figure 4) as the average of the 10,000 poverty likelihoods in the bootstrapped samples

Precision is expressed as confidence intervals (Figure 5), a standard statistical technique that is well-understood by the general public.[6]

For example, the average across all bootstrap samples of the share of people with scores of 15–19 who were poor was 89.8 percent. In 90 percent of the 10,000 bootstrapped samples, the share was between 84.3–95.0 percent.

The confidence intervals are wider for scores of 0–9 because few people fall is these ranges. They are narrow for scores above 84 because everyone in this range is non-poor. For scores from 10 to 80, the 90-percent interval is about 10 percentage points wide, the 95-percent interval is about 12 percentage points wide, and the 99-percent interval is about 16 percentage points wide.

Narrower confidence intervals mean greater precision. Among other things, the width of confidence intervals depends on the size of the hold-out sample, the number of people with a given score, the accuracy of the scorecard, and the extent of overfitting.

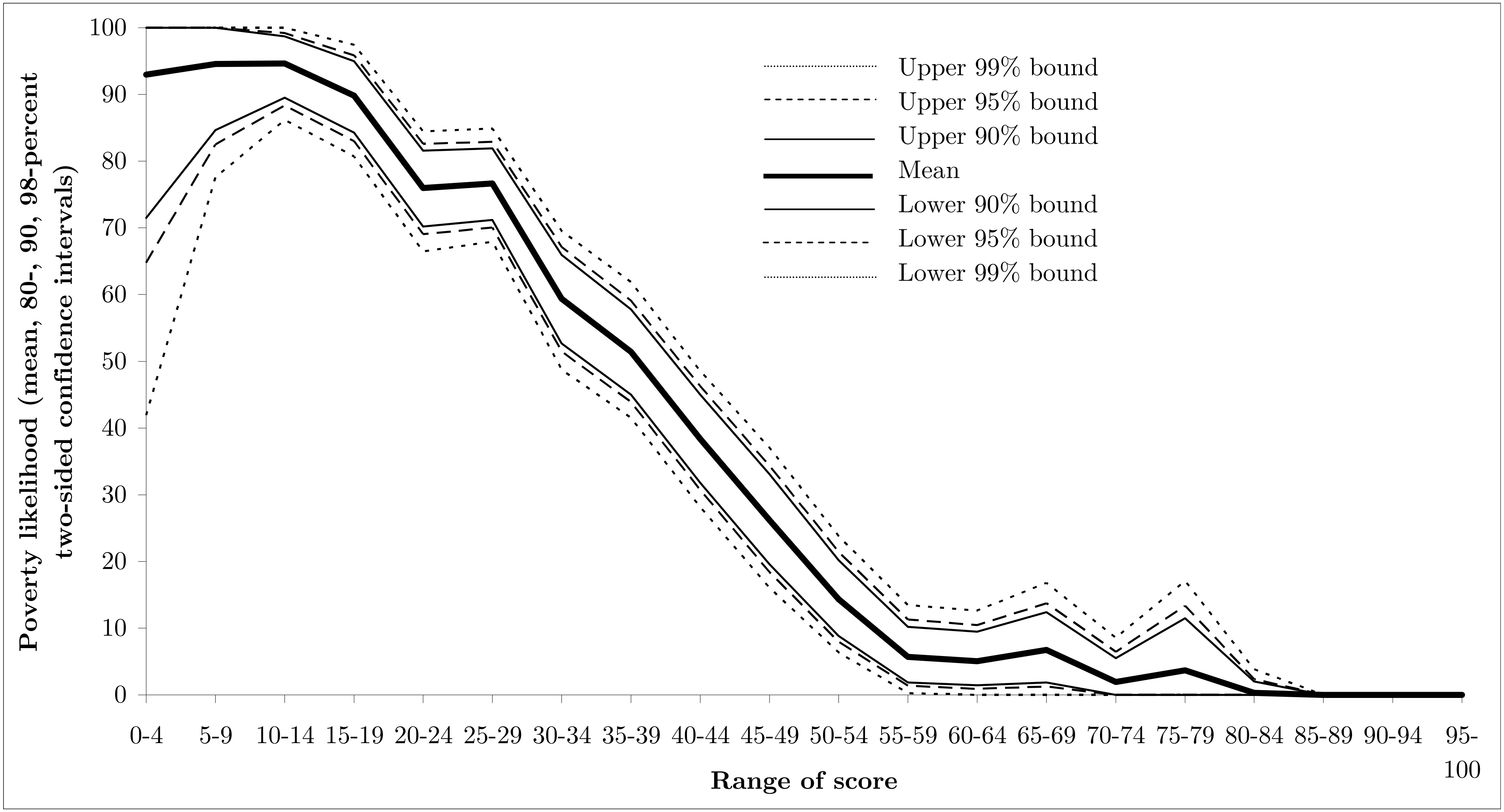
There is no absolute benchmark for what is "precise enough". If other scorecards measured precision with confidence intervals, they could be compared.

Note that this scorecard produces objective (data-based) estimates of poverty likelihood. This holds even though some qualitative judgment is used—along with "MAXC"—to select indicators. In fact, objective scorecards of proven accuracy are often constructed based *only* on qualitative judgment (Caire, 2004; Schreiner *et al.*, 2004; Lovie and Lovie, 1986; Dawes, 1979; Wainer, 1976). What makes for objectivity is not how scorecards are constructed but rather how scores are linked with poverty likelihoods.

---

[5] Because samples are drawn with replacement, the probability that any particular household will appear at least once in a given bootstrap sample is about 63.2 percent.
[6] To my knowledge, this is its first application to scorecards and "proxy means tests". For examples of application in small-area poverty mapping, see Elbers, Lanjouw, and Lanjouw (2003), and Hentschel *et al.* (2000).

Figure 5: Confidence intervals for poverty likelihoods



9. Estimates of poverty rates

The overall poverty rate for all participants—the number that USAID microenterprise grantees must report—is the average of the poverty likelihoods of all participants.

For example, suppose a Bangladeshi development program had three participants on Jan. 1, 2006 with scores of 20, 30, and 40, corresponding to poverty likelihoods of 76.0, 59.4, and 38.3 percent. The overall poverty rate is the participants' average poverty likelihood, that is, $(76.0 + 59.4 + 38.3) \div 3 = 57.9$ percent.

The precision of the estimated poverty rate was measured by drawing 10,000 new bootstrap samples from the hold-out sample.[7] The distribution of the 10,000 differences between the scorecard's estimate of the overall poverty rate (average poverty likelihood) and the true poverty rate is—in line with theory—Normal ($\mu = 0.0049$, $\sigma = 0.0092$).

The scorecard's average estimate is 0.5 percentage points too high. If all the assumptions of Logit regression hold, the estimator should be unbiased, but at least two assumptions do not hold:

- The weights are not exactly optimal due to rounding in the transformation from the original Logit weights into non-negative integers whose sum is between 0 and 100
- Failure of the scorecard to include a complete set of all relevant indicators[8]

This bias has a simple remedy; estimate the overall poverty rate as the average poverty likelihood, less 0.5 percentage points.

How precise are the scorecard estimates? The bootstrap indicates that there is 90-percent confidence that the true poverty rate is within +/–1.5 percentage points of the estimate, 95-percent confidence for +/–1.8 percentage points, and 99-percent confidence for +/–2.4 percentage points. For program purposes, this level of precision is probably more than adequate.

[7] If the 2000 HIES had more cases, it would have been preferable to draw this second set of bootstrap samples from a second hold-out sample. The 2005 HIES—when available—will provide an ideal second "hold-out sample".

[8] Of course, this omitted-variable bias is ubiquitous in any scoring exercise. In contrast, the estimators in IRIS (2005a) and Zeller, Alcaraz V., and Johannsen (2004) are biased even if all their modeling assumptions hold.

10. Progress out of poverty through time

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

Continuing the example from the previous section, suppose that on Jan. 1, 2007, the same three people (some of whom may no longer be participants) have scores of 25, 35, and 60 (poverty likelihoods of 76.7, 51.4, and 5.1 percent). Their average poverty likelihood is now 44.4 percent, an improvement of 57.9 – 44.4 = 13.5 percentage points.

In a large portfolio, this means 13.5 of every 100 participants exited poverty. Given that 57.9 percent of participants were poor in the first place, about one in four (13.5 ÷ 57.9 = 23.3 percent) poor participants left poverty.

Of course, this does not mean that participation in the development program *caused* the progress; the scorecard just measures what happened, regardless of cause.

11. Accuracy in targeting

While accuracy is not the only (nor main) goal, it is important. The individual poverty likelihoods and the overall poverty rate are accurate by their construction from bootstrapping. When using scoring for targeting, greater accuracy means that the poor are more concentrated in low scores and the non-poor in high scores.

At the extreme, a perfect scorecard would assign all the lowest scores to poor people and all the highest scores to non-poor people, for example, if everyone with a score of 49 or less were poor and everyone with a score of 50 or more were non-poor. In reality, no scorecard is perfect. Some non-poor people have low scores, and vice versa.

A person has one of two *poverty statuses*:

- Poor: Expenditure at or below the poverty line
- Non-poor: Expenditure above the poverty line

Poverty status is a fact. If there is data on expenditure (as in the HIES), then poverty status is known.

A person can also be classified into one of two *poverty segments*:

- Poor: Score at or below a poor/non-poor cut-off
- Non-poor: Score above a poor/non-poor cut-off

Poverty segment is program-determined. For example, a program might set a cut-off of 35–39. For program purposes, people with scores at or below 35–39 are treated as if they were poor, and the rest are treated as non-poor.

Because no scorecard is perfect, poverty *status* (expenditure vis-à-vis a poverty line) sometimes differs from poverty *segment* (score vis-à-vis a program's cut-off). That is, some people whose status is truly poor are classified in the non-poor segment, and vice versa. Targeting is accurate to the extent that poverty segment matches poverty status.

Programs use poverty segment as a proxy for (unknown) poverty status. Indeed, the purpose of scoring is to infer poverty status without having to measure expenditure. For people in the 2000 HIES, both poverty status and poverty segment are known, so their coincidence (targeting accuracy) can be measured.

Suppose that a program defines a poor/non-poor cut-off of 35–39. According to the column "% of people <=score who are poor" in Figure 4, 70.9 percent of Bangladeshis with scores at or below 35–39 were truly poor. (The other 29.1 percent were non-poor.)

At the same time, the column “% of people >score who are non-poor” shows that 82.7 percent of those with scores of more than the 35–39 cut-off were truly non-poor (and thus 17.3 percent were poor). In sum, a poor/non-poor cut-off of 35–39 correctly classifies (that is, poverty status matches poverty segment) 70.9 percent of the people classified as poor and 82.7 percent of the people classified as non-poor.

How does classification accuracy depend on the cut-off? Using 40–44 (rather than 35–39) correctly classifies 65.4 percent of those classified as poor and 88.7 percent of those classified as non-poor. This illustrates a general point; better targeting for the poor comes at the cost of worse targeting for the non-poor (and vice versa).

12. Setting the poor/non-poor cut-off

To choose a cut-off, programs need a way to trade off accuracy for the poor versus accuracy for the non-poor. The standard way uses a *classification matrix* and a *net-benefit matrix.*

12.1 Classification matrix

Given a poor/non-poor cut-off, there are four types of classification results:

A. Truly poor correctly classified in poor segment (score at or below the cut-off)
B. Truly poor incorrectly classified in non-poor segment (score above cut-off)
C. Truly non-poor incorrectly classified in poor segment (score at or below cut-off)
D. Truly non-poor correctly classified in non-poor segment (score above cut-off)

These four results can be thought of as a classification matrix:

Figure 6: General classification matrix

		Poverty segment	
		Poor	**Non-poor**
True poverty status	**Poor**	**A.** Truly poor correctly classified in poor segment	**B.** Truly poor incorrectly classified in non-poor segment
	Non-poor	**C.** Truly non-poor incorrectly classified in poor segment	**D.** Truly non-poor correctly classified in non-poor segment

Accuracy improves as greater shares fall in quadrants A and D and fewer in B and C.

Figure 7 is the share of Bangladeshis in each classification for all cut-offs. For 35–39:

A. 36.7% are correctly classified as poor (truly poor in poor segment)
B. 8.3% are incorrectly classified as poor (truly non-poor in poor segment)
C. 15.1% are incorrectly classified as non-poor (truly poor in non-poor segment)
D. 39.9% are correctly classified as non-poor (truly non-poor in non-poor segment)

Figure 7: Share of people by classification

Score	A. % Truly poor correctly classified in poor segment	B. % Truly poor incorrectly classified in non-poor segment	C. % Truly non-poor incorrectly classified in poor segment
0-4	0.3	44.8	0.0
5-9	1.4	43.7	0.1
10-14	5.5	39.5	0.3
15-19	10.2	34.9	0.8
20-24	17.0	28.1	3.0
25-29	24.5	20.5	5.3
30-34	31.3	13.8	9.9
35-39	36.7	8.3	15.1
40-44	40.8	4.2	21.6
45-49	43.1	2.0	28.0
50-54	44.1	0.9	34.2
55-59	44.4	0.6	39.1
60-64	44.6	0.4	43.4
65-69	44.9	0.1	47.2
70-74	45.0	0.1	50.3
75-79	45.1	0.0	52.3
80-84	45.1	0.0	53.5
85-89	45.1	0.0	54.5
90-94	45.1	0.0	54.8
95-100	45.1	0.0	54.9

0

If the cut-off rises to 40–44, more poor (but less non-poor) are correctly classified:

E.	40.8%	are correctly classified	as poor	(truly poor	in poor segment)
F.	4.2%	are incorrectly classified	as poor	(truly non-poor	in poor segment)
G.	21.6%	are incorrectly classified	as non-poor	(truly poor	in non-poor segment)
H.	33.3%	are correctly classified	as non-poor	(truly non-poor	in non-poor segment)

Whether a cut-off of 35–39 is preferred to a cut-off of 40–44 depends on net benefit.

12.2 Net-benefit matrix

Each of the four types of classification results is associated with a net benefit:

α. Benefit of truly poor correctly classified in poor segment
β. Cost (negative net benefit) of truly poor incorrectly classified in non-poor segment
γ. Cost (negative net benefit) of truly non-poor incorrectly classified in poor segment
δ. Benefit of truly non-poor correctly classified in non-poor segment

Figure 8: General net-benefit matrix

		Poverty segment	
		Poor	**Non-poor**
True poverty status	**Poor**	α	β
	Non-poor	γ	δ

Given a net-benefit matrix and a classification matrix, total net benefit is:

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

To choose the optimal cut-off, a program would:

- Define a net-benefit matrix based on the program's values and mission
- Compute total net benefits for each cut-off using Figure 7 and the net-benefit matrix
- Select the cut-off with the highest total net benefit

Most pro-poor development programs care about correctly classifying both the poor and non-poor, even if the poor matter more. Thus, most programs will have non-zero values in at least three of the four quadrants of the net-benefit matrix. The use of a net-benefit matrix allows programs to be explicit and intentional about how they value all the trade-offs inherent when setting cut-offs. This is why the use of classification matrices and net-benefit matrices is standard in scoring (SAS, 2004; SPSS, 2003; Adams and Hand, 2000; Salford Systems, 2000; Hoadley and Oliver, 1998; Greene, 1993).

12.3 "Total Accuracy"

For example, suppose a program selects the net-benefit matrix that corresponds to the "Total Accuracy" criterion (IRIS, 2005b).

Figure 9: "Total Accuracy" net-benefit matrix

Error! Not a valid link.

With "Total Accuracy", total net benefit is the number of people correctly classified:

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

Grootaert and Braithwaite (1998) and Zeller, Alcaraz, and Johannsen (2004) use "Total Accuracy" as a measure of scorecard accuracy.

Figure 10 show "Total Accuracy" for all cut-offs. Total net benefit is highest (76.6) with a cut-off of 35–39; here, poverty segment matches poverty status for about three in four Bangladeshis.

A weakness of "Total Accuracy" is that it weighs correct classifications of the poor and non-poor equally (IRIS, 2005b). If most people are non-poor and/or if a scorecard is more accurate for the non-poor, then "Total Accuracy" might be high even if few poor people are correctly classified. Programs targeting the poor, however, probably value correct classification more for the poor than the non-poor.

Figure 10: Net benefits for common net-benefit matrices

			Non-poverty		
	Total Accuracy	Poverty Accuracy	Accuracy	Undercoverage	Leal
	(A + B)	100*A / (A+B)	100*D / (C+D)	100*B / (A+B)	100*C /
	1 0	1 0	0 0	0 -1	0
Score	0 1	0 0	0 1	0 0	-1
0-4	55.2	0.6	100.0	99.4	7.
5-9	56.2	3.1	99.8	96.9	5.
10-14	60.1	12.2	99.4	87.8	5.
15-19	64.3	22.6	98.5	77.4	7.
20-24	68.9	37.6	94.5	62.4	15
25-29	74.2	54.4	90.4	45.6	17
30-34	76.3	69.4	82.0	30.6	24
35-39	76.6	81.6	72.5	18.4	29
40-44	74.1	90.6	60.6	9.4	34
45-49	70.0	95.6	49.0	4.4	39
50-54	64.8	97.9	37.7	2.1	43
55-59	60.2	98.6	28.7	1.4	46
60-64	56.2	99.1	21.1	0.9	49
65-69	52.6	99.7	14.1	0.3	51
70-74	49.6	99.8	8.5	0.2	52
75-79	47.7	100.0	4.8	0.0	53
80-84	46.5	100.0	2.5	0.0	54
85-89	45.5	100.0	0.9	0.0	54
90-94	45.2	100.0	0.2	0.0	54
95-100	45.1	100.0	0.0	0.0	54

All figures in percentage units.

A simple, transparent way to reflect this is to increase the relative net benefit of correctly classifying the poor. For example, if a program values correctly classifying the poor twice as much as correctly classifying 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 40–44, the point where 2.A + D is highest.

12.4 "Poverty Accuracy"

A criterion that emphasizes the importance of correctly classifying the poor is "Poverty Accuracy" (IRIS, 2005b).

Figure 11: "Poverty Accuracy" net-benefit matrix

True poverty status	Poverty segment: Poor	Poverty segment: Non-poor
Poor	1	0
Non-poor	0	0

"Poverty Accuracy" only counts correct classifications of the poor:

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

The weakness is that correct classification of the poor is rarely the sole criteria. In fact, Figure 10 shows that "Poverty Accuracy" is always maximized with a cut-off of 95–100. While classifying everyone as poor does ensure that all poor people qualify for program services and thus minimizes *undercoverage* of the poor (second-to-last column of Figure 10), it also maximizes *leakage* (the last column), as all non-poor people are also classified as poor.

In short, maximizing "Poverty Accuracy" means universal programs (no targeting). In some contexts, this is appropriate; the point here is to make explicit the implications of "Poverty Accuracy" as a criterion for choosing a poor/non-poor cut-off.

12.5 "Non-poverty Accuracy"

"Non-poverty Accuracy" counts only correct classifications of the non-poor (total net benefit is D). Of course, this is maximized by setting a cut-off of 0–4 so that everyone is classified as non-poor. This is not useful, as it means no one is targeted for program services (leakage is minimized, but undercoverage is maximized).

12.6 "BPAC"

IRIS (2005b) proposes a new measure of scorecard accuracy called the "Balanced Poverty Accuracy Criterion" (BPAC). It attempts to balance two goals:

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

For the first goal, the estimated poverty rate is most accurate when undercoverage B equals leakage C. For the second goal, "Poverty Accuracy" is best when A is maximized. If B > C, then the implicit net-benefit matrix for "BPAC" is:

Figure 12: "BPAC" net-benefit matrix

		Poverty segment	
		Poor	**Non-poor**
True poverty status	**Poor**	1	1
	Non-poor	-1	0

If B>C, then "BPAC" maximizes A while making B as close to C as possible:

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

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

Unfortunately, BPAC is not meaningful for scorecards that estimate poverty likelihoods rather than expenditure (Schreiner, 2005). Instead, this paper takes the standard, well-understood approach to measuring accuracy and precision via the statistical concepts of "bias" (in repeated samples, how close on average the estimate is to the true value) and "confidence intervals" (in repeated samples, how often the estimate falls within a given distance of the true value). Rather than "BPAC is x", the measures are "There is x-percent confidence that the estimated overall poverty rate is within $+/-y$ percentage points of the true value".

12.7 Summary of accuracy discussion

A scorecard is used for:

- Estimating overall poverty rates
- Estimating individual poverty likelihoods
- Classifying people for targeting purposes

Estimates are accurate to the extent that they match the true value being estimated.

For overall poverty rates, estimates from the Bangladesh scorecard have a 90-percent chance of being within 1.5 percentage points of the true poverty rate. For individual poverty likelihoods, estimates have a 90-percent chance of being within 5 percentage points of the true poverty rate.

For targeting with individual classifications, accuracy varies by scorecard and by the poor/non-poor policy cut-offs defined by the program. The most appropriate measure is total net benefit based on a program-specific net-benefit matrix and the classification results in Figure 7. Total net benefit is not an absolute benchmark, but a given program can use it to choose between two different scorecards.

A general, non-program-specific measure of targeting accuracy is "c", the share of all pairs of poor and non-poor households in which the poor household has a lower score. For this scorecard, "c" is 85.8 percent.

13. Comparisons with other Bangladesh tools

How does this scorecard compare with the six other poverty-measurement tools reviewed earlier?

For five of the scorecards, a comparison is not possible. Only 5 of its 10 indicators used by the Grameen Bank are in the 2000 HIES. Likewise, 7 of the 15 indicators used by Zeller, Alcaraz V., and Johansson (2004) in their "LSMS-compatible" scorecard are not in the HIES, and, in any case, weights are not reported. Wodon (1997) lacks weights and a full specification of all the indicators. Kam *et al.* (2004) estimate income rather than expenditure, and they do not use the HIES.

BBS estimates expenditure with the 2000 HIES, but the documentation does not permit reconstructing the scorecard here. Furthermore, the scorecard was part of a local-area poverty-mapping exercise, and as such was not intended for targeting individuals. It has many indicators and would require a computer to calculate its logarithms, proportions, squares, means, and interactions. Its indicators were limited to those that are both in the HIES and the census, limiting accuracy as a targeting tool. Finally, some BBS indicators (marital status, religion, rural/urban location, geographic division) are not sensitive to changes in poverty over time.

Cortez' scorecard uses the 2000 HIES, and it is simple enough to reconstruct from the available documentation (Figure 13).[9] A comparison is possible even though Cortez estimates expenditure directly and the scorecard here estimates poverty likelihoods. The comparison focuses on:

- Overall targeting accuracy as summarizes by ROC curves
- Bias and variance for estimated overall poverty rates

Looking at Cortez's scorecard versus the scorecard here, the main differences are:

- Cortez uses all 7,440 households in the HIES (versus 5,580). Performance for Cortez will therefore be overstated, as the hold-out sample was not held out
- Cortez estimates expenditure (versus poverty likelihood)
- Cortez has 15 indicators (versus 10), two of which are continuous (versus none)
- Cortez has 15 non-zero weights (versus 19)
- Cortez includes household demographics, ownership of durable goods, and quality of residence (versus these plus employment, school attendance, and land ownership)

[9] Cortez (p. 73) has an indicator "Has no private toilet". Whether toilet arrangements are private or shared, however, is not in the HIES. Based on the reported mean, the indicator must be "Does the household use a temporary kacha latrine or open fields?"

Overall, Cortez is slightly simpler, but either scorecard could easily be used in the field.

Figure 13: Bangladesh tool from Cortez *et al.* (2005)

Indicator	Attributes		Points
1. How many household members are there?			
	Number of members multiplied by -1.4		
2. What is the highest grade completed by any household member?			
	Highest year of schooling multiplied by 0.8		
3. Is there any household member aged 16 or older who never attended school?	No	Yes	
	0	-3.3	
4. Does the residence have an electrical connection?	No	Yes	
	0	2.6	
5. Does the household own a telephone/mobile?	No	Yes	
	0	42.5	
6. Does the household own a refrigerator or freezer?	No	Yes	
	0	17.6	
7. Does the household own a television?	No	Yes	
	0	4.2	
8. Does the household own an electric fan?	No	Yes	
	0	3.8	
9. Does the household own dining room furniture?	No	Yes	
	0	4.2	
10. Does the household own drawing room furniture?	No	Yes	
	0	6.2	
11. Is the main construction material of the walls of the residence hemp/hay/bamboo?	No	Yes	
	0	-1.5	
12. Is the main construction material of the roof of the residence hemp/hay/bamboo?	No	Yes	
	0	-1.2	
13. Does the household use a temporary kacha latrine or open fields?	No	Yes	
	0	-4	
14. Is the main source of water for non-drinking use "supply water"?	No	Yes	
	0	10.8	
15. Every household gets 25 points			
			29
Source: Cortez *et al*. (2005)		Total:	

Figure 14 shows two pairs of ROC curves representing overall accuracy for the poor and non-poor for Cortez, BBS, and this paper. As points of reference, there are two additional pairs of ROC curves, corresponding to random targeting (the dotted diagonal line) and perfect targeting (the thick-lined trapezoid).

What do the ROC curves mean? Suppose a program sets a cut-off so as to target the lowest-scoring x percent of potential participants. The ROC curves then show—for a given scorecard—the share of the poor and non-poor who are targeted.

For example, suppose a program sets a cut-off so as to target those 30 percent of potential participants with the lowest scores. With random targeting (the dotted diagonal line), they would pick up 30 percent of all poor people and 30 percent of all non-poor people. With perfect targeting (the thick-lined trapezoid), they would pick up 68.6 percent of the poor and no non-poor.

With Cortez (inner pair of curves), targeting the lowest-scoring 30 percent in the hold-out sample (without bootstrapping) picks up 51.4 percent of the poor and 14.4 percent of the non-poor. The scorecard here (outer pair of curves) is the most accurate, picking up 55.3 percent of the poor and 10.4 percent of the non-poor.

In fact, at all possible cut-offs, the scorecard here has the best targeting accuracy, as its curves are always the closest to the northwest and southeast corners. It encloses 69.2 percent of area within the trapezoid of perfect targeting, versus 61.1 percent for Cortez.

Another general measure of targeting accuracy common in the scoring industry (Hoadley and Oliver, 1998; Wilkie, 1992) is the Kolmogorov-Smirnov (KS) statistic, the maximum vertical distance between a pair of ROC curves. The KS of the scorecard here is 54.0, versus 47.4 for Cortez.

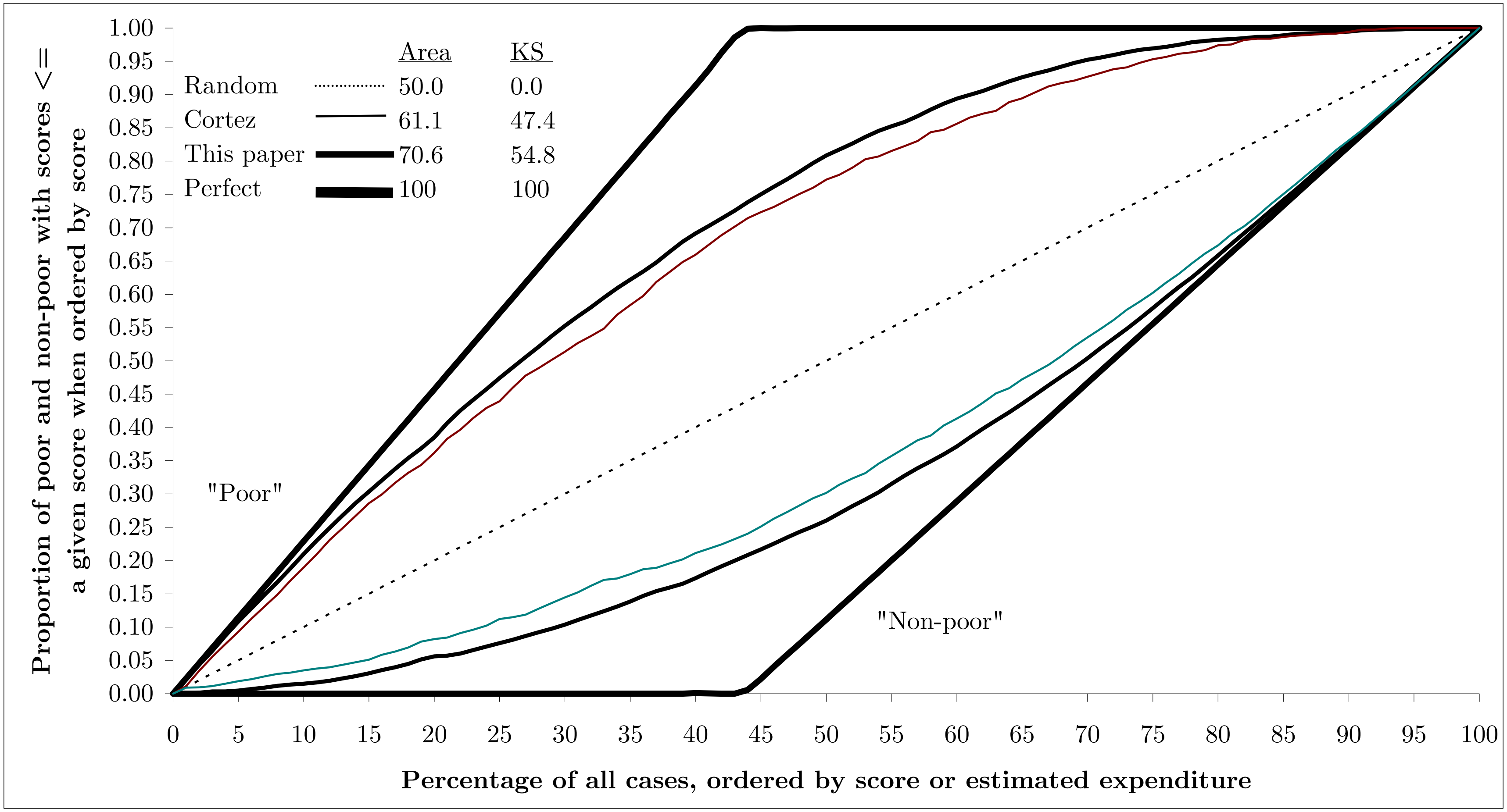
To measure the accuracy (bias) and precision (variance) of estimates of the overall poverty rate, 10,000 bootstrap samples were drawn from the hold-out sample.

For Cortez, the difference between the estimated overall poverty rate and the true value was distributed Normal, with mean 0.0230 and standard deviation 0.0120.[10]

As reported earlier, the distribution for the scorecard here is Normal, with mean 0.0049 and standard deviation 0.0092. Thus, the scorecard here is both more accurate (less bias) and more precise (less variance).

[10] The 90-percent confidence interval is +/–1.9 percentage points, 95 percent is +/–2.3 percentage points, and 99 percent is +/–3.1 percentage points.

Figure 14: Targeting accuracy as measured by ROC curves, Bangladesh Simple Poverty Scorecard from this paper and from tool of Cortez *et al.* (2005)



Why is this? After all, Cortez uses the hold-out sample in construction and directly estimates expenditure, thus using more information than the scorecard here. Both these factors should increase relative accuracy.

Cortez' simplicity is one factor that decreases accuracy. Another factor may be that much of the extra information used by Cortez concerns differences in expenditure far from the poverty line. Also, the Logit approach here is designed for yes/no outcomes; it is—under its assumptions—statistically unbiased and consistent for overall poverty rates. In contrast, Cortez' least-squares approach is—even under its own assumptions—biased and inconsistent for overall poverty rates. Finally, Cortez selected indicators based on R^2, a statistic that does not directly measure accuracy in yes/no predictions.

14. Three classes: very poor, poor, and non-poor

Congress requires USAID microfinance grantees to report the share of participants who are poor versus non-poor. Grameen Foundation U.S.A., however, looks not only at the poor and non-poor but also at the *very poor*, that is, the poorest half of the poor.

For GFUSA, a person has one of three poverty statuses:

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

There are also three poverty 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:

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

Rather than use choosing policy cut-offs using the explicit, rational paradigm of classification matrices and net-benefit matrices, GFUSA has opted instead to leave the choice process and criteria implicit. Thus, a discussion of this paradigm when there are three poverty statuses is omitted. It may be found in Schreiner (2006c and 2006d).

As in those previous documents, it is hereby noted that it is inappropriate to have anyone (such as a consultant) other than the program set their policy cut-offs (of course, experts can assist programs to work through the appropriate process). In particular, it is in general inappropriate to simply set cut-offs so that individuals are assigned to the class for which they have the highest poverty likelihood.

Figure 15: Classification results, 3 classes (very poor, poor, non-poor)

	% people with score in range			% people <= score who are:		
Score	**Very Poor**	**Poor**	**Non-poor**	**Very Poor**	**Poor**	**Non-poor**
0-4	45.5	47.4	7.0	45.5	47.4	7.0
5-9	100.0	0.0	0.0	89.6	9.1	1.3
10-14	57.5	37.2	5.4	65.6	30.1	4.4
15-19	63.1	26.8	10.2	64.4	28.5	7.1
20-24	50.1	25.9	24.0	58.0	27.3	14.7
25-29	43.2	33.5	23.3	53.1	29.4	17.6
30-34	30.4	29.0	40.6	46.8	29.3	23.9
35-39	20.1	31.4	48.6	41.3	29.7	29.0
40-44	7.5	30.8	61.7	35.6	29.9	34.5
45-49	3.0	23.2	73.8	31.6	29.1	39.3
50-54	1.0	13.3	85.7	28.8	27.6	43.6
55-59	0.0	5.7	94.3	27.0	26.2	46.8
60-64	0.0	5.1	94.9	25.6	25.2	49.2
65-69	0.0	6.7	93.3	24.5	24.3	51.2
70-74	2.3	0.0	97.7	23.8	23.5	52.7
75-79	0.0	3.7	96.3	23.2	23.1	53.6
80-84	0.0	0.3	99.7	22.9	22.8	54.2
85-89	0.0	0.0	100.0	22.7	22.6	54.6
90-94	0.0	0.0	100.0	22.7	22.5	54.8
95-100	0.0	0.0	100.0	22.6	22.5	54.9
Total:	22.6	22.4	54.9			

In Figure 15, the three left-side columns under "% people with score in range" link scores with poverty likelihoods. Among people with scores of 15–19, for example, 63.1 percent are very poor, 26.8 percent are poor, and 10.2 percent are non-poor. The scorecard has good accuracy, as the very poor are concentrated in the lowest scores, the poor in "middle" scores, and the non-poor in the highest scores.

The three right-side columns in Figure 15 under "% people <= score who are:" show the poverty composition of a portfolio if a program accepts only participants at or below a given cut-off (and if applicants mirror the population for Bangladesh as a whole). A cut-off of 15–19, for example, leads to a portfolio in which 64.4 percent of participants are very poor, 28.5 percent are poor, and 7.1 percent are non-poor.

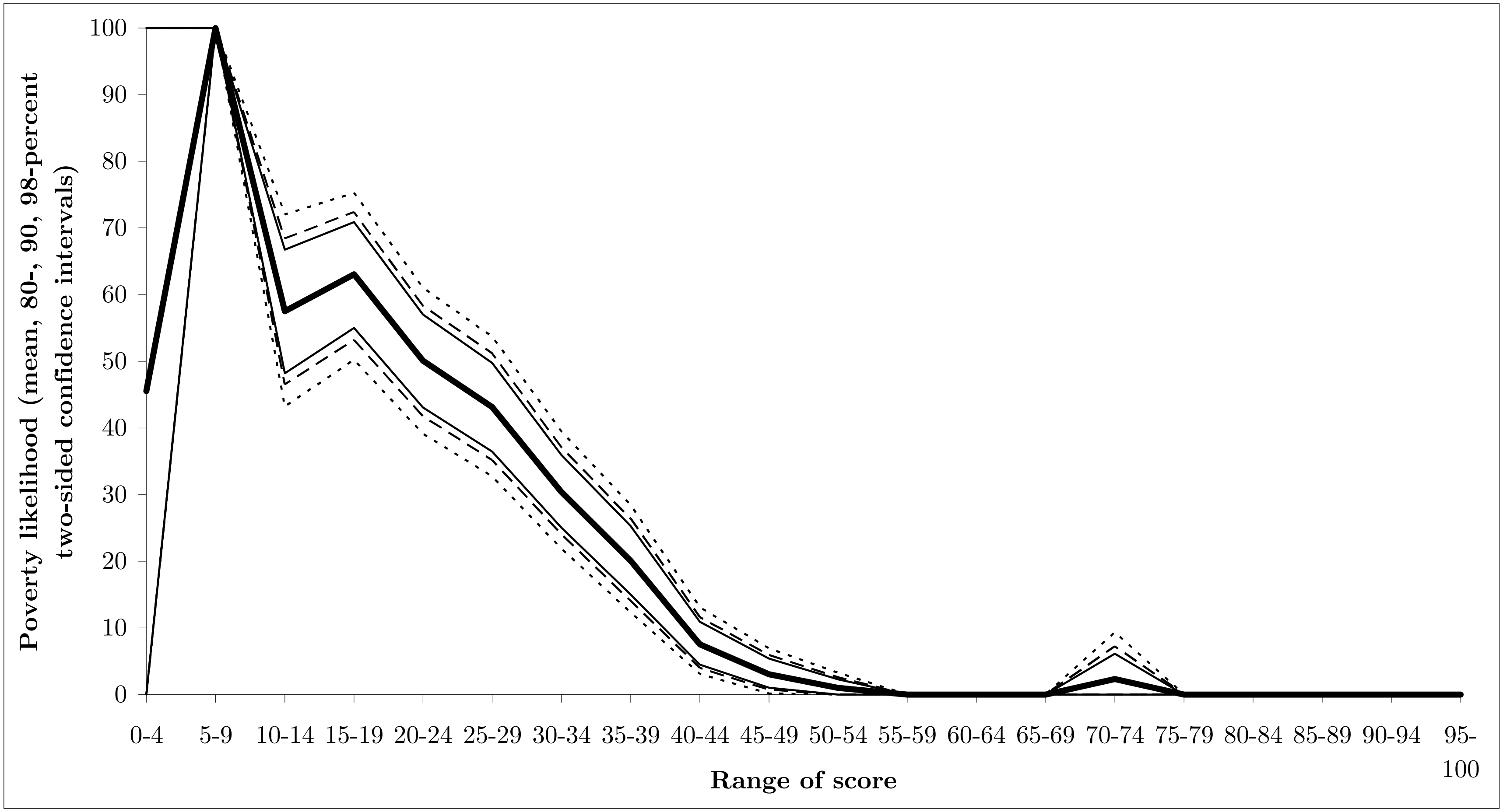
Programs can thus use Figure 15 to set policy cut-offs. Of course, their applicant pool probably does not mirror the population of Bangladesh as a whole, but programs can, as a first cut, assume that it does and then monitor their actual overall poverty rates, adjusting the cut-offs up or down over time as needed.

As in the two-class case, poverty likelihoods are derived from bootstrapping the hold-out sample. Figure 16 depicts 90-, 95-, and 99-percent confidence intervals for the estimated poverty likelihoods. In the range of scores in which there are large numbers of very poor people, the 90-percent confidence interval is abut +/–5 percentage points.

The estimated "very poor" poverty rate is the average "very poor" poverty likelihood among a group of individuals. On average in 10,000 bootstrap samples, the distribution of the difference between the estimated "very poor" poverty rate and the true "very poor" poverty rate was Normal ($\mu = 0.0044$, $\sigma = 0.0082$).

Thus, the best estimate of the overall "very poor" poverty rate is the average "very poor" poverty likelihood, minus 0.4 percentage points. The 90-percent confidence interval is +/–1.4 percentage points, the 95-percent confidence interval is +/–1.6 percentage points, and the 99-percent confidence interval is +/–2.1 percentage points.

Figure 16: Confidence intervals for poverty likelihoods associated with being very poor



15. Summary

- Bangladesh has many poor people. A simple, easy-to-use, inexpensive tool for identifying the poor could improve targeting and help speed progress out of poverty

- The scorecard here estimates the likelihood that a person has expenditure of less than $1/day. It estimates accurately:
  - The likelihood that an individual is poor (within +/–5 percentage points with 90-percent confidence)
  - The overall poverty rate (within +/–1.4 percentage points with 90-percent confidence)

- Accuracy is objectively proven, as scores are related to poverty likelihoods via the 2000 HIES. Precision is measured via bootstrapping on a hold-out sample

- Field workers can compute scores on paper in real time

- The scorecard can be used by any program seeking a quick, easy, inexpensive, and accurate way to identify the poor

- The scorecard can be used to accurately classify people as very poor/poor/non-poor

- Overall, pro-poor development programs in Bangladesh can use the scorecard here to:
  - Target services to the poor
  - Track participants' progress out of poverty through time
  - Report on the share of participants are poor

References

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

Ahmed, Faizuddin. (2004) "Practices of Poverty Measurement and Poverty Profile of Bangladesh", ERD Working Paper No. 54, Asian Development Bank, http://www.adb.org/Documents/ERD/Working_papers/WP054.pdf.

Bangladesh Bureau of Statistics. (2004) "Local Estimation of Poverty and Malnutrition in Bangladesh", United Nations World Food Program, http://www.povertymap.net/publications/doc/SAE%20-%20Final%20Report%20-%20May%202004.pdf.

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

Bigman, David; Dercon, Stefan; Guillaume, Dominique; and Michel Lambotte. (2000) "Community Targeting for Poverty Reduction in Burkina Faso", *World Bank Economic Review*, Vol. 14, No. 1, pp. 167–193.

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

Cortez, Rafeal; Mahmud, Farial; Rabbani, Ghulam; Hossain, Najmul; Faiz, Naushad; Huq, Nazmul; Sultana, Sharmin; and Tarana Tasnim. (2005) "Targeting Resources for the Poor in Bangladesh", Bangladesh Development Studies Paper No. 5, Washington, D.C.: World Bank, http://siteresources.worldbank.org/BANGLADESHEXTN/Resources/FINAL-printversion_Paper-5.pdf.

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

Dowla, Asif; and Barua, Dipal. (forthcoming) *The Poor Always Pay Back: The Grameen II Story*, Bloomfield, CT: Kumarian Press.

Efron, Bradley; and Robert J. Tibshirani. (1993) *An Introduction to the Bootstrap*, New York, NY: Chapman and Hall, ISBN 0–412–04231–2.

Elbers, Chris; Lanjouw, J.O.; and Peter Lanjouw (2003) "Micro-Level Estimation of Poverty and Inequality," *Econometrica*, Vol. 71, No. 1, pp. 355-64.

Goodman, L.A. and Kruskal, W.H. (1979) *Measures of Association for Cross Classification*, New York, NY: Springer-Verlag, ISBN 0–38–790443–3.

Greene, William H. (1993) *Econometric Analysis: Second Edition*, New York, NY: MacMillan, ISBN 0–02–346391–0.

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, Washington, D.C., http://www.worldbank.org/html/dec/Publications/Workpapers/WPS1900series/wps1942/wps1942.pdf.

Hentschel, Jesko; Olsen Lanjouw, Jean; Lanjouw, Peter; and Javier Poggi. (2000) "Combining Census and Survey Data to Trace the Spatial Dimensions of Poverty: A Case Study of Ecuador", *World Bank Economic Review*, Vol. 14, No. 1, pp. 147–165.

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", http://www.povertytools.org/documents/Accuracy%20Results%20for%2012%20Countries.pdf.

IRIS Center. (2005b) "Notes on Assessment and Improvement of Tool Accuracy", http://www.povertytools.org/documents/Assessing%20and%20Improving%20Accuracy.pdf.

Kam, S.P.; Hossain, M.; Bose, M.L.; and L. Villano. (2004) "Mapping Rural Poverty in Bangladesh: How Census Data Systems Can Help", International Rice Research Institute, http://www.nscb.gov.ph/poverty/conference/papers/11_Kam%20Suan%20Pheng.pdf.

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

Matul, Michal; and Sean Kline. (2003) “Scoring Change: Prizma’s Approach to Assessing Poverty”, MFC Spotlight Note No. 4, Warsaw, Poland: Microfinance Centre for Central and Eastern Europe and the New Independent States, http://www.mfc.org.pl/doc/Research/ImpAct/SN/MFC_SN04_eng.pdf.

Salford Systems. (2000) *CART for Windows User’s Guide*, San Diego, CA.

SAS. (2004) *SAS OnlineDoc 9.1.2*, Cary, NC.

Schreiner, Mark. (2006a) “Herramienta Índice de Calificación de Pobreza™: México”, simplepovertyscorecard.com/MEX_2002_SPA.pdf.

Schreiner, Mark. (2006b) “Simple Poverty Scorecard Poverty-Assessment Tool: Philippines”, simplepovertyscorecard.com/PHL_2002_ENG.pdf.

Schreiner, Mark. (2006c) “Simple Poverty Scorecard Poverty-Assessment Tool: India”, simplepovertyscorecard.com/IND_2003_ENG.pdf.

Schreiner, Mark. (2006d) “Poverty Scorecard Poverty-Assessment Tool: Pakistan”, simplepovertyscorecard.com/PAK_2001_ENG.pdf.

Schreiner, Mark. (2005) “IRIS questions on the Simple Poverty Scorecard Poverty-Assessment Tool”, microfinance.com/English/Papers/Scoring_Poverty_Response_to_IRIS.pdf.

Schreiner, Mark. (2002) *Scoring: The Next Breakthrough in Microfinance?* Occasional Paper No. 7, CGAP, cgap.org/docs/OccasionalPaper_07.pdf,

Schreiner, Mark; 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.

Setel, Philip; Abeyasekera, Savitri; Ward, Patrick; Hemed, Yusuf; Whiting, David; Mswia, Robert; Antoninis, Manos; and Henry Kitange. (2003) “Development, Validation, and Performance of a Rapid Consumption Expenditure Proxy for Measuring Income Poverty in Tanzania: Experience from AMMP Demographic Surviellance Sites”, www.rdg.ac.uk/ssc/publications/Inc_PovProx.pdf.

Sillers, Don. (2006) “National and International Poverty Lines: An Overview”, USAID, povertytools.org/Project_Documents/Poverty_lines___An_Overview_1_4_06.pdf

SPSS, Inc. (2003) *Clementine 8.0 User's Guide*.

Wainer, Howard. (1976) "Estimating Coefficients in Linear Models: It Don't Make No Nevermind", *Psychological Bulletin*, Vol. 83, pp. 213–217.

Wilkie, A.D. (1992) "Measuring for Comparing Scoring Systems", pp. 123–138 in L.C. Thomas, J.N. Crook, and D.B. Edelman (eds) *Credit Scoring and Credit Control*.

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

Zeller, Manfred. (2004) "Review of Poverty Assessment Tools", AMAP, `povertytools.org/documents/Review%20of%20Poverty%20Assessment%20Tools.pdf`.

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

_____. (2005b) "Developing and Testing Poverty-Assessment Tools: Results from Accuracy Tests in Uganda", AMAP, `povertytools.org/documents/Uganda%20Accuracy%20Report.pdf`.

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

_____. (2004) "Developing and Testing Poverty-Assessment Tools: Results from Accuracy Tests in Bangladesh", AMAP, `povertytools.org/documents/Bangladesh%20Accuracy%20Report.pdf`.