Simple Poverty Scorecard® Poverty-Assessment Toll Kenya

Shiyuan Chen, Mark Schreiner, and Gary Woller

August 27, 2008

This document and related tools are available at `SimplePovertyScorecard.com`.

Abstract

The Simple Poverty Scorecard®-brand poverty-assessment tool uses ten low-cost indicators from Kenya's 1997 Welfare Monitoring 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 Kenya to measure poverty rates, to track changes in poverty rates over time, and to segment clients for targeted services.

Acknowledgements

This paper was funded by the Consultative Group to Assist the Poorest under the CGAP/Ford Social Indicators Project. Data are from Kenya's National Bureau of Statistics. Thanks go to Malika Anand, Nigel Bigger, Mary Bundi, Luc Christiaensen, Frank DeGiovanni, Syed Hashemi, Jeni Klugman, Milu Charles Muyanga, Seeva Ramasawmy, Tony Sheldon, and Don Sillers. This Simple Poverty Scorecard® tool is re-branded by Grameen Foundation (GF) as the Progress out of Poverty Index®. The PPI® is a performance-management tool that GF promotes to help organizations achieve their social objectives more effectively. "Progress out of Poverty Index" and "PPI" are Registered Trademarks of IPA. "Simple Poverty Scorecard" is a Registered Trademark of Microfinance Risk Management, L.L.C. for its brand of poverty-assessment tools.

Authors

Shiyuan Chen is Senior Analyst with Microfinance Risk Management, L.L.C. Mark Schreiner is Director with Microfinance Risk Management, L.L.C. and Senior Scholar, Center for Social Development, Washington University in Saint Louis. Gary Woller is Principal with Woller & Associates.

Simple Poverty Scorecard® Poverty-Assessment Tools

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

Indicator	Value	Points	Score
1. How many household members are aged 25 or younger?	A. Three or more	0	
	B. None, one, or two	8	
2. How many household members aged 6 to 17 are currently attending school?	A. Not all	0	
	B. All	8	
	C. No children aged 6 to 17	21	
3. What is the material of the walls of the house?	A. Mud/cow dung, grass/sticks/makuti, or no data	0	
	B. Other	5	
4. What kind of toilet facility does your household use?	A. Other	0	
	B. Flush to sewer, flush to septic tank, pan/bucket, covered pit latrine, or ventilation improved pit latrine	2	
5. Does the household own a TV?	A. No	0	
	B. Yes	16	
6. Does the household own a couch or sofa?	A. No	0	
	B. Yes	14	
7. Does the household own a gas or electric stove?	A. No	0	
	B. Yes	12	
8. Does the household own a radio?	A. No	0	
	B. Yes	8	
9. Does the household own a bicycle?	A. No	0	
	B. Yes	5	
10. How many head of cattle are owned by the household currently?	A. None or unknown	0	
	B. One or more	9	

SimplePovertyScorecard.com **Score:**

Simple Poverty Scorecard® Poverty-Assessment Tool

Kenya

1. Introduction

This paper presents the Simple Poverty Scorecard® poverty-assessment tool. Pro-poor programs in Kenya can use to monitor groups' poverty rates at a point in time, track changes in groups' poverty rates between two points in time, and target services to individuals.

The direct approach to poverty measurement via expenditure surveys is difficult, lengthy, and costly, asking about a long list of consumption items ("In the past 7 days, did you purchase any bananas? If you purchased bananas, how many did you purchase and what price did you pay? Now then, in the past 7 days, did you purchase any pineapples? . . ."). In contrast, the indirect approach via the scorecard is simple, quick, and inexpensive. It uses 10 verifiable indicators (such as "Does the household own a stove?" or "What is the material of the walls of the house?") to get a score that is highly correlated with poverty status as measured by the exhaustive expenditure survey.

The scorecard here differs from "proxy means tests" (Coady, Grosh, and Hoddinott, 2002) in that it is tailored to the capabilities and purposes not of national governments but rather of local, pro-poor organizations. The feasible poverty-measurement options for these organizations are typically subjective and relative (such as participatory wealth ranking by skilled field workers) or blunt (such as land-ownership cut-offs or housing indices). Results from these approaches are not comparable across organizations nor across countries, they may be costly, and their accuracy is unknown.

If an organization wants to know what share of its participants are below a poverty line (say, \$1/day for the Millennium Development Goals, or the poorest half below the national poverty line as required of USAID microenterprise grantees, see U.S. Congress, 2004), or if it wants to measure movement across a poverty line (for example, movement across \$1/day to report to the Microcredit Summit Campaign), then it needs an expenditure-based, objective tool with known accuracy. While most organizations lack the resources to field expenditure surveys—and even governments cannot survey large shares of all households—many organizations can implement an inexpensive scorecard that can serve for monitoring, management, and targeting.

The statistical approach here aims to be understood by non-specialists. After all, if managers are to adopt the scorecard on their own and apply it to inform their decisions, they must first trust that it works. Transparency and simplicity build trust. Getting "buy-in" matters; proxy means tests and regressions on the "determinants of poverty" have been around for three decades, but they are rarely used to inform decisions, not because they do not work, but because they are presented (when they are presented at all) as tables of regression coefficients incomprehensible to lay people (with cryptic indicator names such as "`HHSIZE_2`", negative values, decimal places, and standard errors). Thanks to the predictive-modeling phenomenon known as the "flat max", the scorecard is almost as accurate as complex tools.

The technical approach here is also innovative in how it associates scores with poverty likelihoods, in the extent of its accuracy tests, and in how it derives sample-size formula. Although these techniques are simple and/or standard, they have rarely or never been applied to proxy means tests.

The scorecard (Figure 1) is based on Kenya's 1997 Welfare Monitoring Survey (WMS). Indicators are selected to be:

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

All points in the scorecard are non-negative integers, and total scores range from 0 (most likely below a poverty line) to 100 (least likely below a poverty line). Non-specialist field workers can collect data and tally scores on paper in about 5 minutes.

The scorecard can be used to estimate three basic quantities. First, it can estimate an individual's "poverty likelihood", that is, the probability that the individual has expenditure below a given poverty line.

Second, the scorecard can estimate a group's poverty rate at a point in time. (The "poverty rate" is also known as the "poverty prevalence", "head-count index", or "share below the poverty line".) This is simply the average poverty likelihood among individuals in the group.

Third, the scorecard can estimate changes in the poverty rate for a group between two points in time. This estimate is the change in the average poverty likelihood of individuals in the group over time.

The scorecard can also be used for targeting. To help managers choose a targeting cut-off, this paper reports the share of Kenya's population who are at or below a given score cut-off and who are also below a given poverty line.

This paper presents a single scorecard (Figure 1) based on Kenya's national poverty line and a random sub-sample of the 1997 WMS. Scores from this single scorecard are calibrated to poverty likelihoods for five poverty lines:

- Kenya's national poverty line
- Kenya's "food" poverty line
- The "extreme" poverty line used by USAID for microenterprise reporting that divides those below the national poverty line into two equal-sized groups
- $1/day
- $2/day

The accuracy of the scorecard is tested against bootstrapped data from a separate sub-sample of the 1997 WMS. All three scoring estimators are unbiased (that is, they match the true value on average in repeated samples) for the population from which the 1997 WMS was drawn (that is, "in-sample"), but they are biased for other populations (that is, "out-of-sample").

Thus, while the indirect scoring approach is less costly than the direct survey approach, it is also biased. (The survey approach is unbiased by assumption.) There is bias because scoring must assume that the relationship between indicators and poverty status in new future samples will be the same as in the sample used to build the scorecard.[1] Of course, this assumption—ubiquitous and inevitable in predictive modeling—does not hold completely.

Still, bias for Kenya is usually less than ±2.3 percentage points. Furthermore, for sample sizes of $n = 8{,}192$, these estimators are precise to ±1.3 percentage points or less with 90-percent confidence. (For samples of $n = 512$, they are precise to about ±5.1 percentage points or less.)

Section 2 below describes data and poverty lines. Section 3 compares the new scorecard here with existing tools for Kenya. Sections 4 and 5 describe scorecard construction and offer practical guidelines for use. Sections 6 and 7 detail the estimation of individuals' poverty likelihoods and of groups' poverty rates at a point in time. Section 8 discusses estimating changes in poverty rates. Section 9 covers targeting. The final section is a summary.

---

[1] Bias may also result from changes in data collection, from imperfect adjustment of poverty lines across time or geographic regions, or from sampling variation across expenditure surveys.

2. Data and poverty lines

The scorecard is based on the 1997 WMS.[2] One-third of WMS households (weighted by household size) are randomly assigned to a "construction" sample used to select indicators and points. Another one-third are randomly selected to form a "calibration" sample used to associate scores with poverty likelihoods. The final one-third of households form a "validation" sample (Figure 2). The "validation" sample is used to test the accuracy of estimates of individuals' poverty likelihoods and groups' poverty rates.

Kenya has two official poverty lines. The "food" line is the expenditure required for 2,250 calories per day per adult-equivalent. The national poverty line is the food line plus expenditure for basic non-food requirements. Both poverty lines are derived from the 1997 WMS by the Kenyan government (World Bank, 2003). For rural areas, the national poverty line is 40.73 Ksh/person/day for a poverty rate of 48.1 percent, and the rural food line is 30.48 Ksh/person/day for a poverty rate of 31.1 percent (Figure 3). For urban areas, the national line is 87.06 Ksh/person/day for a poverty rate of 41.5 percent, and the urban food line is 41.22 Ksh/person/day for a poverty rate of 6.9 percent. This paper focuses on the national line. For Kenya as a whole, 47.1 percent of people are below the national line (Figure 3).

The scorecard is also calibrated to the USAID "extreme" poverty line that divides people below the national line into two equal groups. The "extreme" lines are

[2] The 1997 WMS is the most recent national expenditure survey, as the 2004/5 Kenya Integrated Household Budget Survey is not available.

26.43 Ksh/person/day for rural areas and 61.83 Ksh/person/day for urban areas. The corresponding "extreme" poverty rates are 24.0 percent and 22.8 percent (Figure 3).

Finally, the scorecard is calibrated to the international purchase-power parity benchmarks of \$1/day and \$2/day, using the following data:

- 1993 purchase-power parity exchange rate of 11.77 Ksh per \$1
- Inflation rates of 28.8, 1.6, 9.0, and 11.2 percent for 1994, 1995, 1996 and 1997[3]

The \$1/day line for 1997 (Sillers, 2006) is then:

11.77 x (1+0.288) x (1+0.016) x (1+0.090) x (1+0.112) x 1.08 = Ksh 20.16.

All the poverty lines are adjusted for differences in cost-of-living by rural/urban area within provinces using:

- L, the all-Kenya line in Ksh
- d_i, the provincial deflator (Republic of Kenya, 2000)
- p_i, the provincial population (Republic of Kenya, 2000)
- p_r and p_u, the rural and urban population[4]
- L_u and L_r, the urban and rural national lines (Republic of Kenya, 2000)

L_{ir} is the \$1/day line adjusted for cost-of-living for the rural areas in province i:

$$L_{ir} = L \cdot \frac{d_i}{\sum_{i=1}^{8} p_i \cdot d_i} \cdot \frac{L_r}{(p_r \cdot L_r + p_u \cdot L_u)/(p_r + p_u)}.$$

Switching r and u in the formula gives the urban lines. If the urban and rural national lines are known, then last part in the formula is omitted.

[3] Kenya National Bureau of Statistics, http://www.knbs.go.ke/sectoral/cpi/cpi_inflation_trends.html?SQMSESSID=ad0454c9bf0c426a0f1f83d64fb7ca83.

[4] FAOSTAT, http://faostat.fao.org/site/429/DesktopDefault.aspx?PageID=429

The 1997 average \$1/day line is Ksh20.34, and the \$2/day line is Ksh40.68. The corresponding poverty rates for \$1/day line are 3.4 percent in the rural areas and 3.6 percent in the urban areas. For \$2/day line, the poverty rates are 29.4 percent in rural areas and 23.1 percent in urban areas (Figure 3).

3. Poverty-assessment tools for Kenya

This section discusses some existing poverty studies for Kenya. The main aspects of interest are the purpose of the study, methods, relative/absolute poverty estimation, poverty lines, indicators, accuracy, and sample-size formula.

3.1 Geda, Jong, Kimenyi and Mwabu

Geda *et al.* (2005) use the 1994 WMS to seek causes of poverty in Kenya that might be modified by policy. As here, they use Logit analysis and the official per-adult-equivalent national and "food" poverty lines. Their indicators are:

- Household head:
  - Sex
  - Age (and age squared)
  - Marital status
  - Sector of employment
  - Education
- Whether any household member is literate
- Main occupation of household members
- Area of residence
- Total land holdings (in acres)
- Number of animals owned

Geda *et al.* find that having expenditure below the national line or the food line is associated with education, household size, and being in agriculture. Of course, these associations are not necessarily causes, limiting their policy relevance.

Compare with the indicators in the scorecard here, the indicators in Geda *et al.* have some limitations. First, they omit several plausible indicators of poverty that are in the 1994 WMS. Second, in terms of feasibility, some indicators are difficult to verify (such as area of the residence and total landholdings), are complex (age squared), or are diffusely defined (number of animals, which treats cattle the same as poultry).

3.2 Zeller, Sharma, Henry, and Lapenu

As in this paper, Zeller *et al.* (2006) seek to develop a practical, low-cost, accurate way to assess the poverty of participants in local pro-poor programs. They differ from this paper in that their measure of poverty is relative to other households in the area, rather than linked to an absolute poverty line. Thus, their results are not comparable across countries or even across areas within Kenya. Because they conduct their own local, special-purpose survey, they do not need national expenditure data.

In Kenya, Zeller *et al.* survey a random sample of 200 participants of a microfinance organization and of 300 non-participants from same area and then apply principal component analysis to select indicators (based on statistical significance) for an index (akin to a score) that is assumed to be related to poverty. The loadings of the first principal component are the index's points. The selected indicators are:

- Education of the household head
- Percentage of literate adults
- Whether the walls of the residence are made of permanent material
- Presence of an electrical connection
- Presence of a latrine in the house
- Type of access to water
- Number of televisions
- Value of radio
- Value of electrical devices
- Value of assets per person
- Episodes of hunger in the past 30 days

- Episodes of hunger in the past 12 months
- Number of days in the past week eating "luxury" food 1
- Number of days in the past week eating "luxury" food 2
- Per person expenditure on clothing

These indicators are more costly than those here in that they require more calculation (percentage of literate adults or per-capita value of assets), cannot be verified (such as past episodes of hunger or consumption of "luxury" foods), or require the valuation of assets (such as radios and total household assets).

Other limitations of Zeller *et al.* include not reporting accuracy, small samples (and not reporting how this affects precision), and not testing the identification of poverty with the first principal component.

3.3 Sahn and Stifel

Sahn and Stifel (2000) use factor analysis (akin to principal components) with the 1988, 1993, and 1998 Kenya Demographic and Health Surveys to construct an asset index (akin to a scorecard). They then compare the distribution of the index and poverty rates over time and between countries. (Poverty rates are based on relative lines set at the 25th and 40th percentiles of expenditure.)

Sahn and Stifel use one set of simple, inexpensive, and verifiable indicators for both their Kenya index and for a single cross-country index constructed from the pooled DHS of 16 African countries that permits cross-country analysis:

- Household durables:
  — Radio
  — Television

— Refrigerator
— Bicycle
— Motorized transport
- Household characteristics:
— Source of drinking water (piped or surface)
— Toilet facility (flush or none)
— Floor quality
- Education of the household head

This approach shares many of the strengths of the approach here in that it can be used for targeting and in that it is flexible, low-cost, adaptable to diverse contexts, and comparable within and even across countries. It differs in that it does not require expenditure data (although it still requires national survey data for indicators).

Like Zeller *et al.*, Sahn and Stifel use a relative measure of poverty, do not test accuracy, do not report sample-size formula, and do not test the assumption that their index captures poverty. Unlike Zeller *et al.* (but like this paper), Sahn and Stifel use only simple and verifiable indicators drawn from nationally representative surveys.

3.4 Stifel and Christiaensen

Like Sahn and Stifel, Stifel and Christiaensen (2007) build an asset index to track changes in poverty. They differ in that they use the 1997 WMS and relate indicators to an absolute, expenditure-based poverty line. Other data sources include:

- 1993, 1998, and 2003 DHS
- District-level malaria data from the 1992, 1994, and 1997 WMS
- District-level infrastructure data from the 1999 Census
- District-level rainfall data from the Famine Early Warning System

Using the 1997 WMS, Stifel and Christiaensen's poverty-assessment tool estimates expenditure using only indicators in both the WMS and the DHS. It is then applied to predict household expenditure in the various years of the DHS (which does not collect expenditure).

At the household-level, Stifel and Christiaensen seek indicators that have returns that do not vary much over time (leading to the exclusion of labor and land). At the macro level, they seek indicators that are location-specific, vary annually, and are likely to affect asset returns (such as rainfall and malaria incidence).

Given potential indicators meeting these criteria, Stifel and Christiaensen build three tools (Nairobi, other urban, and rural) with stepwise least-squares regression. The indicators selected are:

- Household demographics:
  - — Dependency ratio
  - — Household size
  - — Share of household members with a given level of education:
    - Secondary
    - Post-secondary
  - — Education of the household head:
    - Secondary
    - Post-secondary
- Housing quality and sanitation:
  - — Floor
  - — Roof
  - — Source of drinking water
  - — Type of toilet
- Ownership of household durables:
  - — Radio
  - — Television
  - — Refrigerator
  - — Bicycle

- Cluster and district characteristics:
  - Share of households with:
    - Low-quality floors
    - Piped water
    - Refrigerator
    - Electricity
  - Cluster-average share of household heads with education:
    - Primary
    - Secondary
    - Post-secondary
  - Cluster-average share of households with someone with post-secondary education
- Rainfall and health:
  - Deviation of early-season rain from long-run average
  - Malaria prevalence
  - Average household height-for-age z-score

Most of Stifel and Christiaensen's indicators are simple, inexpensive, and verifiable. Of course, the district-level indicators are more complex, and the dependency ratio requires division. Also, they do not directly measure accuracy/precision, nor do they report sample-size formula. The use of district-level indicators makes their index more applicable at the national level (as they intend) than at the household level.

In general, Stifel and Christiaensen do well what they intend to do, which is estimate national expenditure-based poverty rates for years without a national expenditure survey. And although they cannot directly test accuracy without a second WMS, they do get a good idea of accuracy via triangulation of other macro indicators.

3.5 A new scorecard for Kenya

The new scorecard for Kenya here has the following strengths. First, it measures accuracy with different data than that used in scorecard construction. Second, it reports indicators and points, so local pro-poor organizations in Kenya can pick up the scorecard and actually use it. Third, the scorecard is based on an absolute poverty line. (Zeller *et al.* and Sahn and Stifel estimate relative poverty, while Geda *et al.* and Stifel and Christiaensen estimate absolute poverty.) Fourth, this study reports sample-size formulas. Fifth, the scorecard here is designed to be practical for local pro-poor organizations. It has 10 indicators, all of them categorical and selected to be not only highly predictive of poverty but also verifiable, quick to answer, and liable to change over time. This facilitates data collection and improves the data quality, which in turn improves accuracy. Only Sahn and Stifel is as practical. The scorecard here also has the most straightforward derivation and the simplest point scheme. Finally, this study adjusts poverty lines for differences in cost-of-living across urban/rural and provinces. It also considers five poverty lines (national, food, USAID "extreme", $1/day, and $2/day), giving users the flexibility to use the line most relevant for their purposes.

4. Scorecard construction

About 100 potential indicators are prepared in the areas of:

- Family composition (such as female headship and number of children)
- Education (such as highest grade completed and school attendance by children)
- Employment (such as job type)
- Housing (such as material of walls, floor, and roof)
- Household services (such as source of cooking fuel, source of drinking water, and type of toilet facility)
- Asset ownership (such as land, cattle, and sofa)

Each indicator is first screened with the entropy-based “uncertainty coefficient” (Goodman and Kruskal, 1979) that measures how well the indicator predicts poverty on its own. Figure 4 lists the top indicators, ranked by their uncertainty coefficients. Responses for each indicator are ordered starting with those most strongly associated with poverty.

Many indicators in Figure 4 are similar to each other in terms of their association with poverty. For example, most households with a wall made of mud, cow dung, grass, sticks, or *makuti*, also have a floor made of these materials. If a scorecard includes an indicator for walls, then data on the floor do not contribute much. For this reason, many indicators strongly associated with poverty are not in the scorecard, as they add little over and above other indicators that are included.

The scorecard also aims to measure *changes* in poverty through time. Thus, some powerful indicators (such as the highest education level completed by a household member) that are relatively insensitive to changes in poverty are omitted in favor of less-powerful indicators (such as ownership of a radio or a stove) that are more sensitive.

Some indicators are not selected because they are awkward to answer or difficult to verify (such as whether a household member looked for work in the past week).

The scorecard itself is built using Logit regression on the construction sample from the 1997 WMS (Figure 2). Indicator selection uses both judgment and statistics (forward stepwise based on "c", see below). The first step is to build one scorecard for each candidate indicator, using Logit to derive points. Each scorecard's accuracy is taken as "c", a measure of ability to rank by poverty status (SAS Institute Inc., 2004).

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

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

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

This algorithm is the Logit analogue to stepwise on R^2. Like R^2 in a least-squares regression on expenditure, "c" in a Logit regression on poverty status is a good measure of global accuracy. The procedure here differs from naïve stepwise in that along with statistical criteria, judgment and non-statistical criteria are also used to select indicators. The use of non-statistical criteria can improve robustness out-of-sample and, more important, helps ensure that indicators are simple and make sense to users.

The single scorecard here applies to all regions of Kenya. Evidence from India and Mexico (Schreiner, 2006a and 2005a), Sri Lanka (Narayan and Yoshida, 2005), and Jamaica (Grosh and Baker, 1995) suggests that segmenting scorecards by rural/urban does not improve accuracy much.

5. Practical guidelines for scorecard use

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

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

To this end, the scorecard here fits on one page (Figure 1). The construction process, indicators, and points are simple and transparent. "Extra" work is minimized; non-specialists can compute scores by hand in the field because the scorecard has:

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

A field worker using the paper scorecard would:

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

Of course, field workers must be trained. Quality results depend on quality inputs. If organizations or field workers gather their own data and have an incentive to exaggerate poverty rates (for example, if they are rewarded for reaching poorer participants), then it is wise to implement on-going quality control via data review and random audits (Matul and Kline, 2003).[5] IRIS Center (2007a) and Toohig (2007) are useful nuts-and-bolts guides for budgeting, training field workers and supervisors, logistics, sampling, interviewing, piloting, recording data, and quality control.

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

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

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

- Employees of the organization

[5] If an organization does not want field workers to know the points associated with indicators, then it is a simple matter to remove the points from the paper scorecard and apply them later in a spreadsheet or database at the central office.

- Third-party contractors

Scores can be recorded:

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

The subjects to be scored can be:

- All participants (or all new participants)
- A representative sample of all participants (or of all new participants)
- All participants (or all new participants) in a representative sample of branches
- A representative sample of all participants (or all new participants) in a representative sample of branches

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

The scorecard's frequency of application can be:

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

When the scorecard is applied more than once so as to measure change in poverty rates, it can be applied:

- With two different representative sets of participants
- Twice with the same set of participants

An example set of choices were made by BRAC and ASA, two microlenders in Bangladesh (each with 7 million participants) who are applying the Simple Poverty Scorecard® tool for Bangladesh (Schreiner, 2006b). Their design is that loan officers in a random sample of branches score all participants each time they visit a homestead as part of their standard due diligence prior to loan disbursement (about once a year). Scores are recorded on paper in the field before being sent to a central office to be entered into a database. ASA's and BRAC's sampling plans cover 50,000–100,000 participants each.

6. Estimates of individual poverty likelihoods

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

To get absolute units, scores must be converted to *poverty likelihoods*, that is, probabilities of being below a poverty line. This is done via simple look-up tables. For the national line, scores of 0–4 correspond to a poverty likelihood of 81.2 percent (Figure 5), and scores of 45–49 correspond to a poverty likelihood of 27.3 percent.

The poverty likelihood associated with a score varies by poverty line. For example, scores of 45–49 are associated with a poverty likelihood of 27.3 percent for the national line but 10.8 percent for the food line.[6]

6.1 Calibrating scores with poverty likelihoods

A given score is associated ("calibrated") with a poverty likelihood non-parametrically by defining the poverty likelihood as the share of people from the 1997 WMS calibration sample who have the score and who are below a given poverty line.

[6] Starting with Figure 5, most figures have five versions, one for each poverty line. To keep them straight, they are grouped by poverty line. Single tables that pertain to all poverty lines are placed with the tables for the national line.

Figure 6 illustrates this. For the example for the national line, there are 4,246 people with a score of 0–4, of whom 3,449 are below the poverty line. The estimated poverty likelihood associated with a score of 0–4 is then 81.2 percent, because 3,449 ÷ 4,246 = 81.2 percent.

As another illustration, with the national line and a score of 45–49, there are 6,042 people in the calibration sample, of whom 1,649 are below the line (Figure 6). Thus, the estimated poverty likelihood for a score of 45–49 is 1,649 ÷ 6,042 = 27.3 percent.

The same method is used to calibrate scores with estimated poverty likelihoods for the other four poverty lines.

Figure 7 shows, for all scores, the likelihood that expenditure falls in a range demarcated by two adjacent poverty lines. For example, the daily expenditure of someone with a score of 35–39 falls in the following ranges with probability:

- 3.0 percent below $1/day
- 15.8 percent between $1/day and the food line
- 0.4 percent between the food line and the USAID "extreme" line
- 2.0 percent between the USAID "extreme" line and $2/day
- 14.3 percent between $2/day and the national line
- 64.7 percent above the national line

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

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

6.2 Accuracy of estimates of poverty likelihoods

As long as the relationship between indicators and poverty does not change, this calibration process produces unbiased estimates of poverty likelihoods. *Unbiased* means that in repeated samples from the same population, the average estimate matches the true poverty likelihood. The scorecard also produces unbiased estimates of poverty rates at a point in time and of changes in poverty rates between two points in time.[7]

Of course, the relationship between indicators and poverty changes over time, so any scorecard applied out-of-sample (as all are in practice) will generally be biased. Still, estimators that are unbiased in-sample should have less bias out-of-sample.

How accurate are estimates of individual poverty likelihoods? To measure this, the scorecard is applied to 1,000 bootstrap samples of size n = 16,384 for households from the validation sample. The bootstrap process entails:

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

For all 20 score ranges, Figure 8 shows the average difference between estimated and true poverty likelihoods and the average confidence intervals around the estimate.

For the example of the $2/day line (*not* the national line), the average poverty likelihood across bootstrap samples for scores of 0–4 was too high by 0.7 percentage points (Figure 8). For a score of 5–9, the estimate is too low by 4.4 percentage points, and for a score of 10–14, the estimate is too high by 2.1 percentage points.

---

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

The 90-percent confidence interval for these estimated diffferences for scores of 0–4 is ±4.7 percentage points (Figure 8).[8] This means that in 900 of 1,000 bootstraps, the difference is between –4.0 and 5.4 percentage points (because 0.7 – 4.7 = –4.0, and 0.7 + 4.7 = 5.4). In 950 of 1,000 bootstraps (95 percent), the difference is 0.7 ± 5.3 percentage points, and in 990 of 1,000 bootstraps (99 percent), the difference is 0.7 ± 6.8 percentage points.

The estimated poverty likelihood for almost every score for every poverty line differs from the true value, sometimes by a lot (Figures 8–10).[9] This is because both calibration and validation are based on a single finite sample. For targeting, however, what matters is less the difference in all score ranges and more the difference in score ranges just above and below the targeting cut-off. This fact reduces the effects of these differences on targeting. Section 9 below looks at targeting accuracy in detail.

Of course, if estimates of groups' poverty rates are to be usefully accurate, then errors for individuals must largely cancel out.

There are three approaches to mitigating these differences. First, poverty likelihoods in Figure 5 could be adjusted to compensate for the biases in Figure 8. For the example of the $2/day line, Figure 5 shows a poverty likelihood for a score of 0–4 of 63.4, but Figure 8 shows that this is too high by 0.7 percentage points. Changing the poverty likelihood associated with scores of 0–4 to 63.4 – 0.7 = 62.7 would remove the difference.

---

[8] Confidence intervals are a standard, widely understood measure of precision.
[9] Figure 9 summarizes Figure 10 for all poverty lines.

A second approach to mitigating differences between estimates and true values is to increase the fineness of the point scale (for example, by allowing points to range from 0 to 200 instead of 0 to 100), to increase the number of ranges into which scores are grouped (for example, from 20 to 40), and/or to increase the number of response categories for indicators. Of course, all of these approaches add complexity.

By construction, the scorecard here is unbiased in-sample. But it may still be *overfit* out-of-sample. That is, it may fit the 1997 construction and calibration data so closely that it captures not only some timeless patterns but also some random patterns that, due to sampling variation, do not show up in the validation data or do not show up in post-1997 data (if such data were available). Or it may be overfit in that it is not robust to changes over time in the relationship between indicators and poverty.

Overfitting can be mitigated by simplifying the scorecard and not relying completely on the data but rather also considering experience, judgment, and theory. Of course, the scorecard here does this. Bootstrapping can also mitigate overfitting by reducing (but not eliminating) dependence on a single sampling instance. For Kenya, scorecard weights in the 10-indicator Logit hardly changed when bootstrapped. Combining scorecards can also mitigate overfitting, at the cost of greater complexity.

Another approach to mitigating these differences is to do nothing. After all, most errors in individual likelihoods cancel out in the estimates of poverty rates for the scorecard applied to the validation sample (see following sections). Furthermore, at least some bias may come from external sources such as changes in the relationship between indicators and poverty, sampling variation, changes in poverty lines, and inconsistencies in cost-of-living adjustments. These factors can be addressed only by improving data quantity and quality (beyond the scope of the scorecard) or by reducing overfitting (which likely has limited returns, given the scorecard's current parsimony).

7. Estimates of group poverty rates at a point in time

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

To illustrate, suppose a program samples three participants on Jan. 1, 2007 and that they have scores of 20, 30, and 40, corresponding to poverty likelihoods of 53.9, 43.7, and 32.8 percent (national line, Figure 5). The group's estimated poverty rate is the participants' average poverty likelihood of (53.9 + 43.7 + 32.8) ÷ 3 = 43.5 percent.[10]

7.1 Accuracy of estimated poverty rates at a point in time

How accurate is this estimate? For a range of sample sizes, Figure 12 reports the average difference between estimated and true poverty rates and precision the average confidence intervals around the estimated bias for the scorecard applied to 1,000 bootstrap samples from the validation sample. For the national poverty line and a sample size of 16,384, the scorecard is too low by about 1.3 percentage points; it estimates a poverty rate of 46.0 percent, but the true value for the validation sample is 47.3 percent (Figure 2). The average difference ranges from –1.3 percentage points (national line) to +2.3 percentage points (USAID "extreme" line, Figure 11).[11]

[10] The group's poverty rate is not the poverty likelihood associated with the average score. Here, the average score is (20 + 30 + 40) ÷ 3 = 30, so the poverty likelihood associated with the average score is 43.7 percent. This is not the 43.5 percent found as the average of the three poverty likelihoods associated with each of the three scores.
[11] Figure 11 summarizes Figure 12 for all poverty lines.

In terms of precision, the 90-percent confidence interval for the estimated poverty rate at a point in time with $n = 16{,}384$ is ± 0.9 percentage points (Figure 12). This means that in 900 of 1,000 bootstrap samples, the estimate is between $46.0 - 0.9 = 45.1$ percent and $46.0 + 0.9 = 46.9$ percent. For $n = 1{,}024$, the 90-percent interval is ± 3.5 percentage points (Figure 12).

7.2 Sample-size formula for estimates of poverty rates at a point in time

How many participants should an organization sample if it wants to estimate their poverty rate at a point in time for a desired confidence interval and desired confidence level? The first paper in the poverty-scoring literature to address practical question is Schreiner (2008a).[12]

[12] IRIS Center (2007a and 2007b) says that $n = 300$ is sufficient to meet the USAID microenterprise reporting requirements. If a scorecard is as precise as direct measurement, if the expected (before measurement) poverty rate is 50 percent, and if the confidence level is 90 percent, then $n = 300$ implies a confidence interval of about $\pm$ 2.2 percentage points. In fact, USAID has not specified confidence levels or confidence intervals. Furthermore, the expected poverty rate may not be 50 percent, and the scorecard could be more or less precise than direct measurement.

With direct measurement, the poverty rate can be estimated as the number of people observed to be below the poverty line, divided by the number of all observed people. The formula for sample size n is then (Cochran, 1977):

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

where

$$z \quad \text{is} \quad \begin{cases} 1.64 \text{ for confidence levels of 90 percent} \\ 1.96 \text{ for confidence levels of 95 percent} \\ 2.58 \text{ for confidence levels of 99 percent} \end{cases},$$

c is the confidence interval as a proportion (for example, 0.02 for an interval of ±2 percentage points), and

$\hat{p}$ is the expected (before measurement) proportion of people below the poverty line.

Poverty-assessment tools, however, do not measure poverty directly, so this formula is not applicable. To derive a similar sample-size formula for the Kenya scorecard, consider the national poverty line and the scorecard applied to the validation sample. Figure 2 shows that the expected (before measurement) poverty rate $\hat{p}$ in the construction sample is 0.470, the weighted average of 0.471 and 0.468. In turn, a sample size n of 16,384 and a 90-percent confidence level correspond to a confidence interval of ±0.86 percentage points (Figure 12).[13] Plugging these into the direct-measurement sample-size formula (1) above gives not n = 16,384 but rather

[13] Due to rounding, Figure 12 displays 0.9, not 0.86.

$n = \left(\frac{1.64}{0.0086} \right)^2 \cdot 0.47 \cdot (1 - 0.47) =$ 9,059. The ratio of the sample size for scoring (derived empirically) to the sample size for direct measurement (derived from theory) is 16,384 ÷ 9,059 = 1.81.

Applying the same method to n = 8,192 gives $n = \left(\frac{1.64}{0.0126} \right)^2 \cdot 0.47 \cdot (1 - 0.47)$ = 4,220. This time, the ratio of the sample size using scoring to the sample size using direct measurement is 8,192 ÷ 4,220 = 1.94. This ratio of 1.94 for n = 8,192 is close to the ratio of 1.81 for n = 16,384. Indeed, applying this same procedure for all $n \geq 256$ in Figure 12 gives ratios that average to 1.90. This can be used to define a sample-size formula for the Kenya scorecard applied to the validation sample:

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

where α = 1.90 and z, c, and $\hat{p}$ are defined as in (1) above.

To illustrate, if c = 0.069 (confidence interval of ± 6.9 percentage points) and z = 1.64 (90-percent confidence), then (2) gives $n = 1.90 \cdot \left(\frac{1.64}{0.069} \right)^2 \cdot 0.47 \cdot (1 - 0.47) =$ 267 which is close to the sample size of 256 for these parameters in Figure 12.

If the sample-size factor α is less than 1.0, it means that the scorecard is more precise than direct measurement. For Kenya, α ranges from 1.8 to 2.7 for the estimates of groups' poverty rates at a point in time (Figure 11).

Of course, the sample-size formula here is specific to Kenya, its poverty lines, its poverty rates, and this scorecard. The derivation method, however, is valid for any poverty-assessment tool following the approach in this paper.

In practice, an organization would select a poverty line (say, Kenya's national poverty line), select a desired confidence level (say, 90 percent, or $z = 1.64$), select a desired confidence interval (say, $\pm$ 2 percentage points, or $c = 0.02$), make an assumption about $\hat{p}$ (perhaps based on a previous measurement or national figures), assume that the scorecard works out-of-sample,[14] and compute the required sample size. In this illustration, $n = 1.90 \cdot \left(\frac{1.64}{0.02}\right)^2 0.47 \cdot (1 - 0.47) = 3{,}182$.

If the scorecard has already been applied to a sample n, then $\hat{p}$ is the scorecard's estimated poverty rate and the confidence interval c is $\pm z \cdot \sqrt{\frac{\alpha \cdot \hat{p} \cdot (1 - \hat{p})}{n}}$.

How does accuracy for indirect measurement via scoring compare with direct measurement via surveys? For Kenya, direct measurement is more accurate because it is unbiased (by definition), and direct measurement is also more accurate because it is more precise (the sample-size factor α exceeds 1.0). The benefit of scoring, therefore, is not in accuracy but rather in cost.

[14] This paper reports accuracy for the scorecard applied to the 1997 validation sample, but it cannot test accuracy for later years. Still, performance after 1997 will most likely resemble performance in 1997, with some deterioration as time passes.

8. Estimates of changes in group poverty rates over time

The change in a group's poverty rate between two points in time is estimated as the change in the average poverty likelihood of the individuals in the group. With data for 1997, this paper cannot estimate changes over time, nor can it present sample-size formula specific to Kenya. Nevertheless, the concepts are presented here because, in practice, pro-poor organizations can generate their own data and measure change through time.

8.1 Warning: Change is not impact

Scoring can estimate change. Of course, poverty could get better or worse, and scoring does not indicate what caused change. This point is often forgotten or confused, so it bears repeating: the scorecard simply estimates change, and it does not, in and of itself, indicate the reason for the change. In particular, estimating the impact of program participation requires knowing what would have happened to participants if they had not been participants (Moffitt, 1991). Knowing this requires either strong assumptions or a control group that resembles participants in all ways except participation. To belabor the point, the scorecard can help estimate program impact only if there is some way to know what would have happened in the absence of the program. And that information must come from somewhere beyond the scorecard. Even measuring simple change usually requires the strong assumptions that the population is constant over time and that program drop-outs do not differ from others.

8.2 Calculating estimated changes in poverty rates over time

Consider the illustration started in the previous section. On Jan. 1, 2007, a program samples three participants who score 20, 30, and 40 and so have poverty likelihoods of 53.9, 43.7, and 32.8 percent (national line, Figure 5). The group's baseline estimated poverty rate is the participants' average poverty likelihood of (53.9 + 43.7 + 32.8) ÷ 3 = 43.5 percent).

After baseline, two sampling approaches are possible:

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

By way of illustration, suppose that on Jan. 1, 2008, the program samples three additional people who are in the same cohort as the three people originally sampled (or scores the same three original people) and finds that their scores are 26, 35, and 45 and so have poverty likelihoods of 51.2, 35.4, and 27.3 percent (national line, Figure 5). Their average poverty likelihood at follow-up is now (51.2 + 35.4 + 27.3) ÷ 3 = 38.0 percent, an improvement of 43.5 – 38.0 = 5.5 percentage points.

This suggests that about 55 of 1000 participants crossed the poverty line in 2007.[15] Among those who started below the line, about one in eight (5.5 ÷ 43.5 = 12.6 percent) ended up above the line.[16]

---

[15] This is a net figure; some people started above the poverty line and ended below it, and vice versa.
[16] The scorecard does not reveal the reasons for this change.

8.3 Accuracy for estimated change for two independent samples

Under direct measurement, the sample-size formula for estimates of changes in poverty rates in two equal-sized independent samples is:

$$n = 2 \cdot \left(\frac{z}{c}\right)^2 \cdot \hat{p} \cdot (1 - \hat{p}) , \tag{3}$$

where z, c, and $\hat{p}$ are defined as in (1). Before measurement, $\hat{p}$ is assumed equal at both baseline and follow-up. n is the sample size at both baseline and follow-up.[17]

The method developed in the previous section can be used again to derive a sample-size formula for indirect measurement via poverty-assessment tools:

$$n = \alpha \cdot 2 \cdot \left(\frac{z}{c}\right)^2 \cdot \hat{p} \cdot (1 - \hat{p}) . \tag{4}$$

As before, α is the average across sample sizes ≥ 256 of the ratio between the empirical sample size required under scoring for a given precision and the theoretical sample size required under direct measurement.

For Peru and India (Schreiner, 2008a and 2008b), the average α across poverty lines is 1.6 and 1.2, so 1.5 may be a reasonably conservative figure for Kenya.

To illustrate how to use (4) to determine sample size for estimating changes in poverty rates across two independent samples, suppose the desired confidence level is 90 percent ($z = 1.64$), the desired confidence interval is 2 percentage points ($c = 0.02$), the poverty line is \$2/day, $\alpha = 1.50$, and $\hat{p} = 0.291$ (from Figure 2). Then baseline sample

[17] This means that, for a given precision, estimating the change in a poverty rate between two points in time requires 4 times as many measurements (not twice as many) as does estimating a poverty rate at a point in time.

size is $n = 1.5 \cdot 2 \cdot \left(\frac{1.64}{0.02}\right)^2 \cdot 0.291 \cdot (1 - 0.291)$ = 4,162, and follow-up sample size is also 4,162.

8.4 Accuracy for estimated change for one sample, scored twice

Under direct measurement, the sample-size formula for estimates of changes in poverty rates in a single sample scored twice is:[18]

$$n = \left(\frac{z}{c}\right)^2 \cdot \left[\hat{p}_{12} \cdot (1 - \hat{p}_{12}) + \hat{p}_{21} \cdot (1 - \hat{p}_{21}) + 2 \cdot \hat{p}_{12} \cdot \hat{p}_{21}\right], \tag{5}$$

where z and c are defined as in (1), $\hat{p}_{12}$ is the expected (before measurement) share of all sampled cases that move from below the poverty line to above it, and $\hat{p}_{21}$ is the expected share of all sampled cases that move from above the line to below it.

How can a user set $\hat{p}_{12}$ and $\hat{p}_{21}$? Before measurement, a reasonable assumption is that the change in poverty rate is zero. Then $\hat{p}_{12} = \hat{p}_{21}$ and (5) becomes:

$$n = 2 \cdot \left(\frac{z}{c}\right)^2 \hat{p}_*, \tag{6}$$

where $\hat{p}_* = \hat{p}_{12} = \hat{p}_{21}$.

Still, $\hat{p}_*$ could take any value between 0 and 1, so (6) cannot determine sample size. The estimate of $\hat{p}_*$ must be based on data available before baseline measurement.

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

Suppose that the observed relationship between $\hat{p}_*$ and the variance of the baseline poverty rate $p_{baseline} \cdot (1 - p_{baseline})$ is—as it was in Peru, see Schreiner, 2008a—close to $\hat{p}_* = 0.0085 + 0.206 \cdot [p_{baseline} \cdot (1 - p_{baseline})]$. Of course, $p_{baseline}$ is not known before the measurement, but it is reasonable to use as its expected value the observed poverty rate from the previous year. Given this and a poverty line, a sample-size formula for a single sample directly measured twice for Kenya after 1997 is:

$$n = 2 \cdot \left(\frac{z}{c}\right)^2 \cdot \{0.0085 + 0.206 \cdot [p_{1997} \cdot (1 - p_{1997})]\}. \quad (7)$$

As usual, (7) is modified with α to get the scorecard sample-size formula:

$$n = \alpha \cdot 2 \cdot \left(\frac{z}{c}\right)^2 \cdot \{0.0085 + 0.206 \cdot [p_{1997} \cdot (1 - p_{1997})]\}. \quad (8)$$

In Peru (the only other country for which there is an estimate), the average α across years and poverty lines is about 1.8 (Schreiner, 2008a).

To illustrate the use of (8), suppose the desired confidence level is 90 percent (z = 1.64), the desired confidence interval is 2 percentage points (c = 0.02), the poverty line is \$2/day, and the panel will be scoredin 1997. The before-baseline poverty rate is taken as 29.1 percent (p_{1997} =0.291, Figure 2), and suppose α = 1.8. Then baseline sample size is $n = 1.8 \cdot 2 \cdot \left(\frac{1.64}{0.02}\right)^2 \cdot \{0.0085 + 0.206 \cdot [0.291 \cdot (1 - 0.291)]\}$ = 1,235. Of course, n = 1,235 for the follow-up sample as well.

9. Targeting

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

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

Targeting is successful when people truly below a poverty line are targeted (*inclusion*) and people truly above a poverty line are not targeted (*exclusion*). Of course, no scorecard is perfect, and targeting is unsuccessful when people truly below a poverty line are not targeted (*undercoverage*) or people truly above a poverty line are targeted (*leakage*). Figure 13 illustrates these four possible targeting outcomes. Targeting accuracy varies by cut-off; a higher cut-off has better inclusion (but worse leakage), while a lower cut-off has better exclusion (but worse undercoverage).

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

Figure 14 shows the percentage of people by targeting outcome for the scorecard applied to the validation sample. Given an example cut-off of 15–19, outcomes for the national poverty line applied to the validation sample are:

- Inclusion: 21.1 percent are below the line and correctly targeted
- Undercoverage: 26.1 percent are below the line and mistakenly not targeted
- Leakage: 10.1 percent are above the line and mistakenly targeted
- Exclusion: 42.7 percent are above the line and correctly not targeted

Increasing the cut-off to 20–24 improves inclusion and undercoverage but worsens leakage and exclusion:

- Inclusion: 27.6 percent are below the line and correctly targeted
- Undercoverage: 19.6 percent are below the line and mistakenly not targeted
- Leakage: 14.4 percent are above the line and mistakenly targeted
- Exclusion: 38.4 percent are above the line and correctly not targeted

Which cut-off is preferred depends on total net benefit. Suppose each targeting outcome has a per-person benefit or cost. Then total net benefit for a given cut-off is:

Benefit per person correctly included	x	People correctly included	+
Cost per person mistakenly not covered	x	People mistakenly not covered	+
Cost per person mistakenly leaked	x	People mistakenly leaked	+
Benefit per person correctly excluded	x	People correctly excluded.	

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

- Assign benefits and costs to possible outcomes, based on its values and mission
- Tally total net benefits for each cut-off using Figure 14 for a poverty line
- Select the cut-off with the highest total net benefit

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

A common choice of benefits and costs is "Total Accuracy" (IRIS, 2005).[19] With this, total net benefit is the number of people correctly included or excluded:

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

Figures 14 shows "Total Accuracy" for all cut-offs for the Kenya scorecard applied to the validation sample. For the national line, total net benefit is greatest (66.0) for a cut-off of 20-24; that cut-off would correctly classify about two in three Kenyans.[20]

"Total Accuracy" weighs successful inclusion of those below the poverty line equally with successful exclusion of those above the poverty line. If a program valued inclusion more (say, twice as much) than exclusion, it could reflect this by setting the

---

[19] Grootaert and Braithwaite (1998) also use this criterion for poverty-assessment tools.
[20] Beyond "Total Accuracy", IRIS (2005) proposes a new yardstick called the "Balanced Poverty Accuracy Criterion" that is meant to account for inclusion. USAID uses BPAC as its criteria for certifying poverty-assessment tools. After normalizing by the number of people below the poverty line, the BPAC formula is:

BPAC = (Inclusion + |Undercoverage – Leakage|) x [100 ÷ (Inclusion+Undercoverage)].

Although inclusion (and therefore targeting accuracy) is in the BPAC formula, BPAC is in fact maximized by minimizing the difference between undercoverage and leakage, regardless of inclusion. But the difference between undercoverage and leakage is the same as the difference between the estimated poverty rate and the true poverty rate. Thus, it would be less obscure to discard the BPAC nomenclature and speak directly in terms of the accuracy of the estimated poverty rate.

benefit for inclusion to 2 and the benefit for exclusion to 1. Then the chosen cut-off would maximize (2 x People correctly included) + (1 x People correctly excluded).

As an alternative to assigning benefits and costs to targeting outcomes and then choosing a cut-off to maximize total net benefit, a program could set a cut-off to achieve a desired poverty rate. Figure 15 shows, for the Kenya scorecard applied to the validation sample, the expected poverty rate among people who score at or below all possible cut-offs. For the example of the national poverty line in 1997, targeting people who score 25–29 or less would lead to a poverty rate among those targeted of 62.5 percent and would mean targeting 52.8 percent of all Kenyans.

10. Conclusion

This paper presents the Simple Poverty Scorecard® tool. Pro-poor organizations in Kenya can use it to estimate the likelihood that an individual has expenditure below a given poverty line, to estimate a group's poverty rate at a point in time, and to estimate changes in a group's poverty rate between two points in time. The scorecard can also be used for targeting.

The scorecard is inexpensive to implement and can be understood by non-specialists. It is designed to be practical for local pro-poor organizations who want to improve how they monitor and manage their social performance so as to speed up their participants' progress out of poverty.

The scorecard is built with data from the 1997 WMS, calibrated to five poverty lines (national, food, USAID "extreme", $1/day, and $2/day), and tested on data different from that used to build the scorecard. Accuracy/precision and sample-size formula are reported for estimates of individuals' poverty likelihoods, groups' poverty rates at a point in time, and changes in groups' poverty rates over time.

When the scorecard is applied to the validation sample, the difference between estimates of group poverty rates at a point in time and their true values ranges from –1.3 to +2.3 percentage points. For n = 8,192 and 90-percent confidence, precision is usually better than ±1.3 percentage points, and for n = 512, precision is usually better than ± 5.1 percentage points. Compared with direct measurement, scoring is less precise, but also less expensive.

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

Although the statistical technique is innovative, and although technical accuracy is important, the design of the scorecard here focuses on ease-of-use. After all, a perfectly accurate scorecard is worthless if programs feel so daunted by its complexity or its cost that they do not even try to use it. For this reason, the scorecard is kept simple, using 10 indicators that are inexpensive to collect and that are straightforward to observe and verify. Indicator weights are all zeros or positive integers, and scores range from 0 (most likely to be below a poverty line) to 100 (least likely to be below a poverty line). Scores are related to poverty likelihoods via simple look-up tables, and targeting cut-offs are likewise simple to apply. The design attempts to facilitate adoption by helping managers understand and trust scoring and by allowing non-specialists can compute scores in the field.

In sum, the scorecard is a practical, objective way for pro-poor programs in Kenya to monitor poverty rates, track changes in poverty rates over time, and target services. The approach used here for Kenya can be applied to any country with similar data from a national expenditure survey.

References

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

Baesens, B.; Van Gestel, T.; Viaene, S.; Stepanova, M.; Suykens, J.; and J. Vanthienen. (2003) "Benchmarking State-of-the-Art Classification Algorithms for Credit Scoring", *Journal of the Operational Research Society*, Vol. 54, pp. 627–635.

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

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

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

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

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

Geda, Alemayehu; Jong, Niek de; Kimenyi, Mwangi S.; and Germano Mwabu. (2005) "Determinants of Poverty in Kenya: A Household Level Analysis", University of Connecticut Storrs Department of Economics Working Paper Series No. 2005–44, `www.econ.uconn.edu/working/2005-44.pdf`.

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

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

Grosh, Margaret; and Judy L. Baker. (1995) "Proxy-Means Tests for Targeting Social Programs: Simulations and Speculation", World Bank LSMS Working Paper No. 118, `go.worldbank.org/W9OWN57PD0`, retrieved 13 May 2016.

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

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. (2007a) "Manual for the Implementation of USAID Poverty Assessment Tools", `povertytools.org/USAID_Tools/USAID_Manual_4-2007.pdf`. See also `povertytools.org/Tool_Implementation/Tool_Implementation.html`.

_____. (2007b) "Introduction to Sampling for the Implementation of PATs", `povertytools.org/USAID_Tools/Training/Sampling/Introduction_Sampling.ppt`.

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

Johnson, Glenn. (2007) "Lesson 3: Two-Way Tables—Dependent Samples", `http://www.stat.psu.edu/~jglenn/stat504/03_2way/53_2way_compare.htm`.

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

Lovie, 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", Microfinance Centre for Central and Eastern Europe and the New Independent States Spotlight Note No. 4, `www.mfc.org.pl/doc/Research/ImpAct/SN/MFC_SN04_eng.pdf`.

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

Moffitt, Robert. (1991) "Program Evaluation With Non-experimental Data", *Evaluation Review*, Vol. 15, No. 3, pp. 291–314.

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

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

Republic of Kenya. (2000) *Second Report on Poverty in Kenya, Volume I: Incidence and Depth of Poverty.*

Sahn, David E.; and David C. Stifel. (2000) "Poverty Comparisons Over Time and Across Countries in Africa", *World Development*, Vol. 28, No. 12, pp. 2123–2155.

SAS Institute Inc. (2004) "The LOGISTIC Procedure: Rank Correlation of Observed Responses and Predicted Probabilities", in *SAS/STAT User's Guide, Version 9*, Cary, NC.

Schreiner, Mark. (2008a) "Simple Poverty Scorecard Poverty-Assessment Tool: Peru", , SimplePovertyScorecard.com/PER_2003_ENG.pdf, retrieved 23 September 2016.

_____. (2008b) "Simple Poverty Scorecard Poverty-Assessment Tool: India", SimplePovertyScorecard.com/IND_2005_ENG.pdf, retrieved 23 September 2016.

_____. (2006a) "Is One Simple Poverty Scorecard Poverty-Assessment Tool Enough for India?", microfinance.com/English/Papers/Scoring_Poverty_India_Segments.pdf, retrieved 23 September 2016.

_____. (2006b) "Simple Poverty Scorecard Poverty-Assessment Tool: Bangladesh", SimplePovertyScorecard.com/BGD_2000_ENG.pdf, retrieved 23 September 2016.

_____. (2005a) "Índice de Calificación de Pobreza™: México", SimplePovertyScorecard.com/MEX_2002_SPA.pdf, retrieved 23 September 2016.

_____. (2005b) "IRIS Questions on the Simple Poverty Scorecard Poverty-Assessment Tool", microfinance.com/English/Papers/Scoring_Poverty_Response_to_IRIS.pdf, retrieved 23 September 2016.

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

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

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

Stifel, David; and Luc Christiaensen. (2007) "Tracking Poverty Over Time in the Absence of Comparable Consumption Data", *World Bank Economic Review*, Vol. 21, pp. 317–341.

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

Toohig, Jeff. (2007) "Progress Out of Poverty Index: Training Guide".

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

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

World Bank. (2003) "Kenya: A Policy Agenda to Restore Growth", Report No. 25840–KE.

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

_____; Sharma, Manohar; Henry, Carla; and Cecile Lapenu. (2006) "An Operational Method for Assessing the Poverty Outreach Performance of Development Policies and Projects: Results of Case Studies in Africa, Asia, and Latin America", *World Development*, Vol. 34, No. 3, pp. 446–464.

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

Sub-sample	Year of Survey	Households	% with expenditure below a poverty line: National	National food	USAID 'Extreme'	International $1/day	International $2/day
Construction							
Selecting indicators and weights	'97	3,661	47.1	27.5	24.6	3.3	29.1
Calibration							
Associating scores with likelihoods	'97	3,567	46.8	27.6	24.9	3.2	29.0
Validation							
Applying scorecard	'97	3,646	47.3	26.8	22.1	3.6	27.7
Difference in poverty rate between constructing and validation samples (percentage points)							
Between samples	'97 to '97	7,228 and 3,646	-0.3	0.8	2.6	-0.3	1.4

Figure 3: Average poverty lines and poverty rates by region

		Poverty line (Ksh/person/day) and poverty rate (%)														
	Line				National			USAID			International					
	or	National			food			'Extreme'			$1/day			$2/day		
Province	rate	Rural	Urban	All	Rural	Urban	All	Rural	Urban	All	Rural	Urban	All	Rural	Urban	All
Coast	Line	41.10	87.85		30.76	41.60		26.67	62.39		14.88	31.80		29.76	63.60	
	Rate	59.0	39.6		41.7	9.1		33.4	24.0		4.6	4.7		39.5	24.6	
Eastern	Line	41.10	87.85		30.76	41.60		26.67	62.39		14.88	31.80		29.76	63.60	
	Rate	53.6	44.5		35.9	14.8		28.8	27.4		2.9	4.0		34.0	28.2	
Central	Line	38.96	83.27		29.15	39.43		25.28	59.14		14.10	30.14		28.21	60.29	
	Rate	26.8	40.2		13.1	9.8		9.6	27.1		0.6	4.7		12.4	27.3	
Nyanza	Line	39.82	85.10		29.79	40.30		25.84	60.44		14.41	30.81		28.83	61.61	
	Rate	59.2	56.2		39.0	21.2		31.0	35.6		5.9	10.6		37.2	35.6	
North Eastern	Line	41.53	88.76		31.08	42.03		26.95	63.04		15.03	32.13		30.07	64.26	
	Rate	—	57.9		—	0.6		—	28.6		—	—		—	28.6	
Rift Valley	Line	41.10	87.85		30.76	41.60		26.67	62.39		14.88	31.80		29.76	63.60	
	Rate	43.0	38.0		27.2	4.6		20.5	19.2		2.6	0.7		25.8	19.5	
Western	Line	41.96	89.68		31.40	42.46		27.22	63.69		15.19	32.46		30.38	64.93	
	Rate	54.8	48.7		37.2	17.1		27.2	34.4		4.5	9.5		34.6	34.4	
Nairobi	Line	—	91.51		—	43.33		—	64.99		—	33.13		—	66.25	
	Rate	—	39.7		—	3.2		—	19.9		—	2.6		—	20.1	
All Kenya	Line	40.73	87.06		30.48	41.22		26.43	61.83		14.75	31.52		29.49	63.03	
	Rate	48.1	41.5	47.1	31.1	6.9	27.3	24.0	22.8	23.8	3.4	3.6	3.4	29.4	23.1	28.4

In North Eastern province, 140 urban households were surveyed in the 1997 WMS. None of them were rural.

Figure 4: Poverty indicators by uncertainty coefficient

Uncertainty coefficient	Indicator (Possible responses ordered starting with those most strongly indicative of poverty)
476	Does the household own a couch or sofa? (No; Yes)
388	What is the highest level completed by the female head/spouse? (STD7 or lower, or no data; STD7 and CPE to Form2; Form2 and KJSE or higher)
275	What is the material of the walls of the house? (Mud/cow dung, grass/sticks/makuti, or no data; Stone, cement/bricks, roasted bricks, iron sheets, or other)
273	Does the household own a TV? (No; Yes)
266	What is the material of the floor of the house? (Mud/cow dung, grass/sticks/makuti, or no data; Stone,cement/bricks,wood, Roasted bricks, or other)
260	How many household members aged 6 to 17 are currently attending school?(Not all; All; No children aged 6 to 17)
246	Does the household own a radio? (No; Yes)
222	What is the highest level completed by the male head/spouse? (STD8 or lower, or no data; STD8 and CPE or higher)
217	Does the household own a gas or electric stove? (No; Yes)
177	Does the household own a car? (No; Yes)
164	How many household members work as regular employees (skilled or unskilled)? (0; 1 or more)
164	What is the highest level completed by any family member? (Form 2 or lower, or no data; Form 2 and KJSE or higher)
148	How many household members work in public sector? (0; 1 or more)
145	What is the main source of lighting fuel? (Firewood or no data; Kerosene/oil or electricity; Other)
134	What is the material of the roof of the house? (Wood, grass/sticks/makuti, or no data; Other)
125	Does the household own a phone? (No; Yes)
120	How many household members are aged 25 or younger? (3 or more; 0, 1 or 2)
105	Does the female head/spouse works in the public sector? (No; Yes)
102	What is the main source of drinking water during the rainy season? (Unprotected well, rain water, river, lake, pond, or no data, public outdoor tap, borehole, or protected well; Piped into dwelling or compound, vendor, truck, or other)
102	What is the main source of cooking fuel? (Electricity or other; Kerosene/paraffin, gas, candles, firewood)

Source: 1997 WMS, national poverty line.

Figure 4 (continued): Poverty indicators by uncertainty coefficient

Uncertainty coefficient	Indicator (Possible responses ordered starting with those most strongly indicative of poverty)
102	How old is the female head/spouse? (35 or older; 34 or younger)
100	Does the household own a fan? (No; Yes)
99	Does the male head/spouse works in sales, manufacturing, mining, transportation, or construction? (No; Yes)
93	Does the household own a refrigerator? (No; Yes)
88	What kind of toilet facility does your household use? (None, uncovered pit latrine, other or no data; Flush to sewer, flush to septic tank, pan/bucket, covered pit latrine, or ventilation improved pit latrine)
76	Does the household own a stereo? (No; Yes)
76	Does the household own a sewing machine? (No; Yes)
73	What is the main source of drinking water during the dry season? (Unprotected well, rain water, river, lake, pond, or no data, public outdoor tap, borehole, or protected well; Piped into dwelling or compound, vendor, truck, or other)
69	Is any household member attending a private school? (No; Yes)
69	Does the household own a bicycle? (No; Yes)
40	Does the male head/spouse work in the private sector? (No; Yes)
35	Does the female head/spouse work in sales, manufacturing, mining, transportation, or construction? (No; Yes)
32	How many head of cattle are owned currently by the household? (0 or no data; 1 or more)
32	Does the female head/spouse work in the private sector? (No; Yes)
28	Does the household own a motorcycle? (No; Yes)
23	How many household members worked in the past 12 months? (None, or 3 or more; 1 or 2)
20	What is the structure of household headship? (Female head/spouse only; Male head/spouse only; Male and female spouses)
14	Does the household own an animal cart? (No; Yes)
13	Does the female head/spouse work in the semi-public sector? (No; Yes)
13	How old is the male head/spouse? (35 or older; 34 or younger)

Source: 1997 WMS, national poverty line.

National Poverty Line Figures

(and figures pertaining to all five poverty lines)

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

If an individual's score is . . .	. . . then the likelihood (%) of being below the poverty line is:
0–4	81.2
5–9	71.6
10–14	66.1
15–19	58.1
20–24	53.9
25–29	51.2
30–34	43.7
35–39	35.4
40–44	32.8
45–49	27.3
50–54	26.8
55–59	21.4
60–64	19.6
65–69	9.6
70–74	2.9
75–79	6.7
80–84	0.5
85–89	1.5
90–94	18.4
95–100	0.0

Surveyed cases weighted to represent Kenya's population.
Based on the 1997 WMS

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

Score	People below poverty line		All people at score		Poverty likelihood (estimated, %)
0–4	3,449	÷	4,246	=	81.2
5–9	4,386	÷	6,126	=	71.6
10–14	5,063	÷	7,657	=	66.1
15–19	7,643	÷	13,155	=	58.1
20–24	5,815	÷	10,799	=	53.9
25–29	5,557	÷	10,857	=	51.2
30–34	4,632	÷	10,601	=	43.7
35–39	3,185	÷	9,010	=	35.4
40–44	1,982	÷	6,046	=	32.8
45–49	1,649	÷	6,042	=	27.3
50–54	1,433	÷	5,354	=	26.8
55–59	463	÷	2,163	=	21.4
60–64	300	÷	1,531	=	19.6
65–69	250	÷	2,621	=	9.6
70–74	45	÷	1,520	=	2.9
75–79	71	÷	1,058	=	6.7
80–84	3	÷	523	=	0.5
85–89	4	÷	258	=	1.5
90–94	26	÷	140	=	18.4
95–100	0	÷	292	=	0.0

Number of all people normalized to sum to 100,000.
Based on the WMS 1997

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

	Likelihood of having expenditure in range demarcated by daily per capita poverty lines					
	<$1	>=$1 and <National food	>=National food and <USAID 'Extreme'	>=USAID 'Extreme' and <$2	>=$2 and <National	>=National
Score	<KSH20.16	>=KSH20.16 and <KSH33.95	>=KSH33.95 and <KSH37.86	>=KSH37.86 and <KSH40.32	>=KSH40.32 and <KSH55.7	>=KSH55.7
0–4	10.0	53.2	0.0	0.3	17.8	18.8
5–9	5.1	35.0	1.8	3.9	25.8	28.4
10–14	6.3	44.8	0.0	0.0	15.0	33.9
15–19	5.7	32.0	0.0	1.7	18.7	41.9
20–24	2.8	25.1	0.0	0.0	26.0	46.2
25–29	1.1	30.3	0.0	0.0	19.7	48.8
30–34	3.2	20.2	0.0	0.7	19.7	56.3
35–39	3.0	15.8	0.4	2.0	14.3	64.7
40–44	0.9	10.5	0.0	3.5	18.0	67.2
45–49	1.0	9.8	0.7	1.2	14.7	72.7
50–54	0.0	12.0	4.1	1.8	8.9	73.2
55–59	0.0	9.3	3.2	0.0	8.9	78.6
60–64	0.0	10.6	0.0	0.8	8.3	80.4
65–69	0.0	0.6	0.0	0.0	8.9	90.5
70–74	0.0	1.1	0.0	0.0	1.8	97.1
75–79	0.0	2.7	0.0	0.0	4.0	93.3
80–84	0.0	0.0	0.0	0.0	0.5	99.5
85–89	0.0	0.0	0.0	0.0	1.5	98.5
90–94	0.0	0.0	0.0	0.0	18.4	81.6
95–100	0.0	0.0	0.0	0.0	0.0	100.0

All poverty likelihoods in percentage units.

Figure 8 (National poverty line): Bootstrapped differences between estimated and true poverty likelihoods for individuals in a large sample (n=16,384) from the validation sample, with confidence intervals

	Scorecard applied to validation sample			
		Confidence interval (+/- percentage points)		
Score	**Bias**	**90-percent**	**95-percent**	**99-percent**
0–4	5.1	3.6	4.2	5.5
5–9	-3.7	3.3	3.5	4.5
10–14	-4.6	3.9	4.1	5.0
15–19	-1.8	2.5	2.9	3.6
20–24	-6.3	4.8	5.1	5.5
25–29	1.7	2.6	3.1	3.9
30–34	-2.2	3.2	3.7	5.0
35–39	-0.4	2.9	3.5	4.8
40–44	1.2	3.0	3.6	4.7
45–49	-1.6	3.6	4.5	5.8
50–54	-1.1	5.0	5.9	7.9
55–59	4.5	4.6	5.4	7.1
60–64	5.1	4.2	4.9	6.7
65–69	-0.8	4.4	5.4	7.2
70–74	-6.0	4.5	4.8	5.2
75–79	6.7	0.0	0.0	0.0
80–84	-3.9	4.0	4.5	5.7
85–89	0.3	1.0	1.2	1.5
90–94	18.4	0.0	0.0	0.0
95–100	0.0	0.0	0.0	0.0

Figure 9 (All poverty lines): Bias and precision for bootstrapped estimates of individuals' poverty likelihoods for the validation sample

	Poverty line				
		National	USAID		
Year scorecard applied	National	food	'Extreme'	$1/day	$2/day
Bias	0.7	0.4	1.5	-0.4	0.8
Precision	0.8	0.5	0.6	0.2	0.6

Precision is measured as 90-percent confidence intervals in units of +/- percentage points.
Scorecard based on 1997 WMS; scorecard is applied to validation sample.
Bias and precision estimated from 1,000 bootstraps of size n=16,384.

Figure 10 (National poverty line): Bias and precision for bootstrapped estimates of individuals' poverty likelihoods, by sample size, scorecard applied to validation sample

	Scorecard applied to validation sample			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	4.0	56.1	62.6	72.9
4	3.0	44.7	49.6	61.0
8	1.5	35.1	41.3	52.5
16	0.5	26.6	30.8	41.4
32	-0.9	21.4	24.9	32.0
64	-1.2	16.1	19.5	27.4
128	-1.4	12.7	14.7	20.4
256	-1.1	8.3	10.1	13.2
512	-0.5	5.7	7.0	9.7
1,024	-0.1	3.7	4.3	5.9
2,048	0.3	2.6	3.0	4.0
4,096	0.5	1.8	2.1	2.9
8,192	0.7	1.2	1.4	1.9
16,384	0.7	0.8	0.9	1.3

Figure 11 (All poverty lines): Bias, precision, sample-size α, and mean-squared error for bootstrapped estimates of groups' poverty rates at a point in time for the scorecard applied to the validation sample

	Poverty line				
Year scorecard applied	**National**	**National food**	**USAID 'Extreme'**	**$1/day**	**$2/day**
Bias	-1.3	0.2	2.3	-0.5	1.2
Precision	0.9	0.8	0.8	0.4	0.8
α for sample size	1.9	1.8	1.9	2.7	1.8

Precision is measured as 90-percent confidence intervals in units of +/- percentage points.
Scorecard based on 1997 WMS; scorecard is applied to validation sample.
Bias and precision estimated from 1,000 bootstraps of size n=16,384.
α is estimated from 1,000 bootstrap samples of n=256, 512, 1,024, 2,048, 4,096, 8,192, and 16,384.

Figure 12 (National poverty line): Bias and precision for bootstrapped estimates of groups' poverty rates at a point in time, by sample size, scorecard applied to validation sample

	Scorecard applied to validation sample			
		Confidence interval (+/- percentage points)		
Sample size (n)	**Bias**	**90-percent**	**95-percent**	**99-percent**
2	4.0	56.1	62.6	72.9
4	2.8	44.8	48.8	60.8
8	1.6	33.9	40.2	50.3
16	0.7	25.2	29.5	38.7
32	-0.3	18.9	22.6	29.3
64	-0.8	13.8	16.9	21.5
128	-1.1	9.7	12.2	15.6
256	-1.2	6.9	8.0	10.7
512	-1.1	5.1	5.9	7.8
1,024	-1.2	3.5	4.2	5.7
2,048	-1.2	2.5	3.0	4.4
4,096	-1.3	1.8	2.2	2.8
8,192	-1.3	1.3	1.5	1.9
16,384	-1.3	0.9	1.0	1.4

Figure 13 (All poverty lines): Possible types of outcomes from targeting by poverty score

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

Figure 14 (National poverty line): People by targeting classification and score, along with "Total Accuracy" and BPAC, scorecard applied to validation sample

Score	<u>Inclusion:</u> < poverty line correctly targeted	<u>Undercoverage:</u> < poverty line mistakenly non-targeted	<u>Leakage:</u> ≥ poverty line mistakenly targeted	<u>Exclusion:</u> ≥ poverty line correctly non-targeted	<u>Poverty Accuracy</u> Inclusion + Exclusion	<u>BPAC</u> See text
0–4	3.2	44.0	1.0	51.7	55.0	-84.2
5–9	7.8	39.4	2.5	50.2	58.1	-61.5
10–14	13.3	34.0	4.8	48.0	61.2	-33.8
15–19	21.1	26.1	10.1	42.7	63.8	10.7
20–24	27.6	19.6	14.4	38.4	66.0	47.3
25–29	33.0	14.2	19.8	32.9	65.9	58.0
30–34	37.9	9.4	25.6	27.2	65.0	45.9
35–39	41.1	6.2	31.4	21.4	62.5	33.6
40–44	43.0	4.3	35.5	17.3	60.3	24.9
45–49	44.7	2.5	39.8	13.0	57.7	15.8
50–54	46.2	1.0	43.7	9.1	55.3	7.6
55–59	46.6	0.7	45.5	7.3	53.9	3.8
60–64	46.8	0.4	46.8	6.0	52.8	1.0
65–69	47.1	0.2	49.1	3.6	50.7	-4.0
70–74	47.2	0.0	50.5	2.2	49.5	-6.9
75–79	47.2	0.0	51.6	1.2	48.4	-9.1
80–84	47.2	0.0	52.1	0.7	47.9	-10.2
85–89	47.3	0.0	52.3	0.4	47.7	-10.7
90–94	47.3	0.0	52.5	0.3	47.5	-11.0
95–100	47.3	0.0	52.7	0.0	47.3	-11.6

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

Figure 15 (National poverty line): People below the poverty line and all people, at a given score or at or below a given score cut-off, scorecard applied to validation sample

	People below poverty line (%)		All people (%)	
Score	At score	At or below score	At score	At or below score
0–4	76.1	76.1	4.2	4.2
5–9	75.3	75.6	6.1	10.4
10–14	70.7	73.5	7.7	18.0
15–19	59.9	67.8	13.2	31.2
20–24	60.2	65.8	10.8	42.0
25–29	49.5	62.5	10.9	52.8
30–34	45.8	59.7	10.6	63.4
35–39	35.8	56.7	9.0	72.5
40–44	31.6	54.8	6.0	78.5
45–49	28.9	52.9	6.0	84.5
50–54	27.6	51.4	5.4	89.9
55–59	16.8	50.6	2.2	92.1
60–64	14.7	50.0	1.5	93.6
65–69	10.5	48.9	2.6	96.2
70–74	9.1	48.3	1.5	97.7
75–79	0.0	47.8	1.1	98.8
80–84	4.4	47.6	0.5	99.3
85–89	1.2	47.5	0.3	99.6
90–94	0.0	47.4	0.1	99.7
95–100	0.0	47.3	0.3	100.0

National Food Poverty Line Figures

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

If an individual's score is . . .	. . . then the likelihood (%) of being below the poverty line is:
0–4	63.1
5–9	40.1
10–14	51.1
15–19	37.7
20–24	27.9
25–29	31.5
30–34	23.4
35–39	18.8
40–44	11.3
45–49	10.8
50–54	12.0
55–59	9.3
60–64	10.6
65–69	0.6
70–74	1.1
75–79	2.7
80–84	0.0
85–89	0.0
90–94	0.0
95–100	0.0

Surveyed cases weighted to represent Kenya's population.
Based on the WMS 1997

Figure 6 (National food line): Illustration of derivation of estimated poverty likelihoods associated with scores

Score	People below poverty line		All people at score		Poverty likelihood (estimated, %)
0–4	2,680	÷	4,246	=	63.1
5–9	2,454	÷	6,126	=	40.1
10–14	3,914	÷	7,657	=	51.1
15–19	4,962	÷	13,155	=	37.7
20–24	3,012	÷	10,799	=	27.9
25–29	3,416	÷	10,857	=	31.5
30–34	2,476	÷	10,601	=	23.4
35–39	1,689	÷	9,010	=	18.8
40–44	685	÷	6,046	=	11.3
45–49	653	÷	6,042	=	10.8
50–54	644	÷	5,354	=	12.0
55–59	201	÷	2,163	=	9.3
60–64	162	÷	1,531	=	10.6
65–69	16	÷	2,621	=	0.6
70–74	17	÷	1,520	=	1.1
75–79	28	÷	1,058	=	2.7
80–84	0	÷	523	=	0.0
85–89	0	÷	258	=	0.0
90–94	0	÷	140	=	0.0
95–100	0	÷	292	=	0.0

Surveyed cases weighted to represent Kenya's population.
Based on the WMS 1997

Figure 8 (National food line): Bootstrapped differences between estimated and true poverty likelihoods for individuals in a large sample (n=16,384) from the validation sample, with confidence intervals

	Scorecard applied to validation sample			
		Confidence interval (+/- percentage points)		
Score	**Bias**	**90-percent**	**95-percent**	**99-percent**
0–4	-0.1	4.7	5.3	6.8
5–9	-10.2	6.9	7.2	7.9
10–14	2.9	3.6	4.3	5.6
15–19	1.2	2.5	2.9	3.7
20–24	-4.4	3.7	4.0	4.9
25–29	4.8	2.2	2.5	3.2
30–34	2.6	2.1	2.5	3.4
35–39	0.1	2.3	2.7	3.6
40–44	-5.1	3.8	4.1	4.6
45–49	1.3	2.6	3.2	4.3
50–54	3.8	2.2	2.7	3.5
55–59	2.0	3.9	4.6	6.0
60–64	3.3	3.3	4.0	5.1
65–69	0.6	0.0	0.0	0.0
70–74	1.1	0.0	0.0	0.0
75–79	2.7	0.0	0.0	0.0
80–84	0.0	0.0	0.0	0.0
85–89	0.0	0.0	0.0	0.0
90–94	0.0	0.0	0.0	0.0
95–100	0.0	0.0	0.0	0.0

Based on 1997 scorecard applied to households from 1997.

Figure 10 (National food line): Bias and precision for bootstrapped estimates of individuals' poverty likelihoods, by sample size, scorecard applied to validation sample

	Scorecard applied to validation sample			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	3.0	50.9	57.8	67.8
4	2.0	39.1	46.2	56.1
8	1.3	31.6	36.9	46.6
16	0.4	24.4	27.5	37.7
32	0.0	19.0	22.5	27.7
64	-0.2	13.8	16.3	21.4
128	0.1	9.1	10.8	14.8
256	0.2	6.3	7.6	10.1
512	0.3	4.1	5.0	6.3
1,024	0.4	2.5	3.1	4.3
2,048	0.4	1.7	2.0	2.8
4,096	0.4	1.2	1.4	1.7
8,192	0.4	0.8	0.9	1.3
16,384	0.4	0.5	0.7	0.9

Figure 12 (National food line): Bias and precision for bootstrapped estimates of groups' poverty rates at a point in time, by sample size, scorecard applied to validation sample

		Scorecard applied to validation sample		
		Confidence interval (+/- percentage points)		
Sample size (n)	**Bias**	**90-percent**	**95-percent**	**99-percent**
2	3.0	50.9	57.8	67.8
4	1.8	38.6	46.2	56.8
8	1.3	31.1	36.2	45.4
16	0.4	23.3	26.8	34.0
32	0.2	16.6	20.5	25.3
64	-0.2	12.5	14.5	19.0
128	0.0	8.6	10.2	12.7
256	0.1	6.0	7.2	9.7
512	0.2	4.2	5.1	6.7
1,024	0.2	3.0	3.6	4.6
2,048	0.3	2.1	2.6	3.3
4,096	0.2	1.6	1.8	2.5
8,192	0.2	1.1	1.3	1.7
16,384	0.2	0.8	0.9	1.3

Figure 14 (National food line): People by targeting classification and score, along with "Total Accuracy" and BPAC, scorecard applied to validation sample

Score	Inclusion: < poverty line correctly targeted	Undercoverage: < poverty line mistakenly non-targeted	Leakage: ≥ poverty line mistakenly targeted	Exclusion: ≥ poverty line correctly non-targeted	Poverty Accuracy Inclusion + Exclusion	BPAC See text
0–4	2.7	24.1	1.6	71.6	74.3	-74.2
5–9	5.7	21.1	4.6	68.6	74.3	-39.9
10–14	9.4	17.4	8.6	64.6	74.0	2.5
15–19	14.2	12.6	16.9	56.3	70.5	36.8
20–24	17.7	9.1	24.2	48.9	66.7	9.6
25–29	20.6	6.2	32.2	41.0	61.6	-20.1
30–34	22.8	4.0	40.6	32.6	55.4	-51.4
35–39	24.5	2.3	47.9	25.3	49.8	-78.7
40–44	25.5	1.3	53.0	20.2	45.7	-97.6
45–49	26.1	0.7	58.4	14.8	40.9	-118.0
50–54	26.5	0.3	63.4	9.8	36.4	-136.3
55–59	26.7	0.1	65.4	7.8	34.5	-143.8
60–64	26.8	0.0	66.8	6.4	33.2	-149.1
65–69	26.8	0.0	69.4	3.8	30.6	-158.9
70–74	26.8	0.0	70.9	2.3	29.1	-164.5
75–79	26.8	0.0	72.0	1.2	28.0	-168.5
80–84	26.8	0.0	72.5	0.7	27.5	-170.4
85–89	26.8	0.0	72.8	0.4	27.2	-171.4
90–94	26.8	0.0	72.9	0.3	27.1	-171.9
95–100	26.8	0.0	73.2	0.0	26.8	-173.0

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

Figure 15 (National food line): People below the poverty line and all people, at a given score or at or below a given score cut-off, scorecard applied to validation sample

	People below poverty line (%)		All people (%)	
Score	**At score**	**At or below score**	**At score**	**At or below score**
0–4	63.0	63.0	4.2	4.2
5–9	50.2	55.4	6.1	10.4
10–14	48.2	52.4	7.7	18.0
15–19	36.6	45.7	13.2	31.2
20–24	32.3	42.2	10.8	42.0
25–29	26.7	39.0	10.9	52.8
30–34	20.9	36.0	10.6	63.4
35–39	18.7	33.9	9.0	72.5
40–44	16.4	32.5	6.0	78.5
45–49	9.5	30.9	6.0	84.5
50–54	8.2	29.5	5.4	89.9
55–59	7.3	29.0	2.2	92.1
60–64	7.4	28.6	1.5	93.6
65–69	0.0	27.9	2.6	96.2
70–74	0.0	27.4	1.5	97.7
75–79	0.0	27.1	1.1	98.8
80–84	0.0	27.0	0.5	99.3
85–89	0.0	26.9	0.3	99.6
90–94	0.0	26.9	0.1	99.7
95–100	0.0	26.8	0.3	100.0

USAID “Extreme” Poverty Line Figures

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

If an individual's score is . . .	. . . then the likelihood (%) of being below the poverty line is:
0–4	55.8
5–9	41.9
10–14	43.6
15–19	35.8
20–24	22.6
25–29	24.5
30–34	18.6
35–39	19.1
40–44	11.3
45–49	11.5
50–54	16.2
55–59	12.5
60–64	2.3
65–69	0.6
70–74	1.1
75–79	0.0
80–84	0.0
85–89	0.0
90–94	0.0
95–100	0.0

Surveyed cases weighted to represent Kenya's population.
Based on the WMS 1997

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

Score	People below poverty line		All people at score		Poverty likelihood (estimated, %)
0–4	2,369	÷	4,246	=	55.8
5–9	2,566	÷	6,126	=	41.9
10–14	3,337	÷	7,657	=	43.6
15–19	4,703	÷	13,155	=	35.8
20–24	2,435	÷	10,799	=	22.6
25–29	2,656	÷	10,857	=	24.5
30–34	1,969	÷	10,601	=	18.6
35–39	1,725	÷	9,010	=	19.1
40–44	685	÷	6,046	=	11.3
45–49	694	÷	6,042	=	11.5
50–54	865	÷	5,354	=	16.2
55–59	271	÷	2,163	=	12.5
60–64	36	÷	1,531	=	2.3
65–69	16	÷	2,621	=	0.6
70–74	17	÷	1,520	=	1.1
75–79	0	÷	1,058	=	0.0
80–84	0	÷	523	=	0.0
85–89	0	÷	258	=	0.0
90–94	0	÷	140	=	0.0
95–100	0	÷	292	=	0.0

Surveyed cases weighted to represent Kenya's population.
Based on the 1997 WMS

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

	Scorecard applied to validation sample			
		Confidence interval (+/- percentage points)		
Score	**Bias**	**90-percent**	**95-percent**	**99-percent**
0–4	6.8	5.2	6.1	8.2
5–9	-3.7	3.8	4.6	6.0
10–14	7.0	3.5	4.2	5.4
15–19	7.3	2.3	2.8	3.6
20–24	-4.5	3.9	4.1	4.7
25–29	4.3	2.0	2.4	3.2
30–34	1.6	2.0	2.4	3.0
35–39	1.5	2.3	2.9	3.9
40–44	-4.4	3.3	3.6	4.0
45–49	0.5	3.0	3.6	4.6
50–54	10.3	1.9	2.4	3.3
55–59	8.8	1.2	1.4	2.1
60–64	-3.5	3.3	3.7	4.7
65–69	-0.6	1.0	1.1	1.4
70–74	0.0	0.6	0.7	1.0
75–79	0.0	0.0	0.0	0.0
80–84	0.0	0.0	0.0	0.0
85–89	0.0	0.0	0.0	0.0
90–94	0.0	0.0	0.0	0.0
95–100	0.0	0.0	0.0	0.0

Based on 1997 scorecard applied to households from 1997.

Figure 10 (USAID "extreme" line): Bias and precision for bootstrapped estimates of individuals' poverty likelihoods, by sample size, scorecard applied to validation sample

	Scorecard applied to validation sample			
		Confidence interval (+/- percentage points)		
Sample size (n)	**Bias**	**90-percent**	**95-percent**	**99-percent**
2	3.2	50.7	57.5	66.4
4	3.0	37.6	45.3	56.3
8	2.9	30.6	36.5	46.0
16	2.8	22.9	27.7	36.5
32	2.5	17.9	21.3	29.3
64	2.0	13.0	15.8	21.9
128	1.9	9.5	11.7	17.0
256	2.1	6.6	8.4	11.4
512	2.0	4.3	5.3	7.6
1,024	1.8	2.8	3.3	4.5
2,048	1.7	1.9	2.2	2.9
4,096	1.6	1.3	1.5	2.3
8,192	1.5	0.9	1.1	1.4
16,384	1.5	0.6	0.7	0.9

Figure 12 (USAID "extreme" line): Bias and precision for bootstrapped estimates of groups' poverty rates at a point in time, by sample size, scorecard applied to validation sample

	Scorecard applied to validation sample			
		Confidence interval (+/- percentage points)		
Sample size (n)	**Bias**	**90-percent**	**95-percent**	**99-percent**
2	3.2	50.7	57.5	66.4
4	2.8	37.5	45.2	55.5
8	2.9	30.2	35.2	44.8
16	2.7	22.4	25.7	32.7
32	2.7	16.7	19.2	24.7
64	2.2	11.8	14.2	17.9
128	2.1	8.4	10.3	13.2
256	2.2	6.1	7.4	10.3
512	2.2	4.3	5.1	6.7
1,024	2.2	3.0	3.6	5.0
2,048	2.3	2.2	2.6	3.3
4,096	2.3	1.6	1.8	2.4
8,192	2.3	1.1	1.3	1.7
16,384	2.3	0.8	0.9	1.2

Figure 14 (USAID "extreme" line): People by targeting classification and score, along with "Total Accuracy" and BPAC, scorecard applied to validation sample

Score	Inclusion: < poverty line correctly targeted	Undercoverage: < poverty line mistakenly non-targeted	Leakage: ≥ poverty line mistakenly targeted	Exclusion: ≥ poverty line correctly non-targeted	Poverty Accuracy Inclusion + Exclusion	BPAC See text
0–4	2.1	20.0	2.2	75.8	77.8	-71.3
5–9	4.9	17.2	5.5	72.4	77.3	-31.0
10–14	7.7	14.4	10.3	67.6	75.2	16.4
15–19	11.4	10.7	19.7	58.2	69.6	10.6
20–24	14.4	7.7	27.6	50.3	64.6	-25.1
25–29	16.5	5.5	36.3	41.6	58.2	-64.3
30–34	18.4	3.7	45.1	32.8	51.2	-104.1
35–39	19.9	2.1	52.5	25.4	45.3	-137.7
40–44	20.9	1.2	57.6	20.3	41.2	-160.7
45–49	21.6	0.5	63.0	14.9	36.5	-185.1
50–54	21.9	0.2	68.0	9.9	31.8	-207.9
55–59	22.0	0.1	70.1	7.8	29.8	-217.3
60–64	22.0	0.0	71.5	6.4	28.4	-223.8
65–69	22.1	0.0	74.1	3.8	25.9	-235.6
70–74	22.1	0.0	75.6	2.3	24.4	-242.4
75–79	22.1	0.0	76.7	1.2	23.3	-247.2
80–84	22.1	0.0	77.2	0.7	22.8	-249.5
85–89	22.1	0.0	77.5	0.4	22.5	-250.7
90–94	22.1	0.0	77.6	0.3	22.4	-251.3
95–100	22.1	0.0	77.9	0.0	22.1	-252.7

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

Figure 15 (USAID "extreme" line): People below the poverty line and all people, at a given score or at or below a given score cut-off, scorecard applied to validation sample

	People below poverty line (%)		All people (%)	
Score	**At score**	**At or below score**	**At score**	**At or below score**
0–4	49.2	49.2	4.2	4.2
5–9	45.6	47.1	6.1	10.4
10–14	36.6	42.6	7.7	18.0
15–19	28.5	36.7	13.2	31.2
20–24	27.0	34.2	10.8	42.0
25–29	20.2	31.3	10.9	52.8
30–34	17.1	28.9	10.6	63.4
35–39	17.6	27.5	9.0	72.5
40–44	15.7	26.6	6.0	78.5
45–49	10.9	25.5	6.0	84.5
50–54	5.9	24.3	5.4	89.9
55–59	3.7	23.8	2.2	92.1
60–64	5.9	23.6	1.5	93.6
65–69	1.2	22.9	2.6	96.2
70–74	1.1	22.6	1.5	97.7
75–79	0.0	22.4	1.1	98.8
80–84	0.0	22.2	0.5	99.3
85–89	0.0	22.2	0.3	99.6
90–94	0.0	22.2	0.1	99.7
95–100	0.0	22.1	0.3	100.0

$1/Day Poverty Line Figures

Figure 5 ($1/day line): Estimated poverty likelihoods associated with scores

If an individual's score is . . .	. . . then the likelihood (%) of being below the poverty line is:
0–4	10.0
5–9	5.1
10–14	6.3
15–19	5.7
20–24	2.8
25–29	1.1
30–34	3.2
35–39	3.0
40–44	0.9
45–49	1.0
50–54	0.0
55–59	0.0
60–64	0.0
65–69	0.0
70–74	0.0
75–79	0.0
80–84	0.0
85–89	0.0
90–94	0.0
95–100	0.0

Surveyed cases weighted to represent Kenya's population.
Based on the WMS 1997

Figure 6 ($1/day line): Illustration of derivation of estimated poverty likelihoods associated with scores

Score	People below poverty line		All people at score		Poverty likelihood (estimated, %)
0–4	423	÷	4,246	=	10.0
5–9	312	÷	6,126	=	5.1
10–14	482	÷	7,657	=	6.3
15–19	750	÷	13,155	=	5.7
20–24	298	÷	10,799	=	2.8
25–29	124	÷	10,857	=	1.1
30–34	336	÷	10,601	=	3.2
35–39	267	÷	9,010	=	3.0
40–44	52	÷	6,046	=	0.9
45–49	58	÷	6,042	=	1.0
50–54	0	÷	5,354	=	0.0
55–59	0	÷	2,163	=	0.0
60–64	0	÷	1,531	=	0.0
65–69	0	÷	2,621	=	0.0
70–74	0	÷	1,520	=	0.0
75–79	0	÷	1,058	=	0.0
80–84	0	÷	523	=	0.0
85–89	0	÷	258	=	0.0
90–94	0	÷	140	=	0.0
95–100	0	÷	292	=	0.0

Surveyed cases weighted to represent Kenya's population.
Based on the 1997 WMS

Figure 8 ($1/day line): Bootstrapped differences between estimated and true poverty likelihoods for individuals in a large sample (n=16,384) from the validation sample, with confidence intervals

	Scorecard applied to validation sample			
		Confidence interval (+/- percentage points)		
Score	**Bias**	**90-percent**	**95-percent**	**99-percent**
0–4	0.7	2.0	2.4	3.1
5–9	-6.3	4.5	4.7	5.2
10–14	-0.9	2.5	3.0	3.8
15–19	1.9	1.3	1.6	2.0
20–24	-3.1	2.4	2.6	2.9
25–29	-0.7	0.6	0.7	0.9
30–34	1.2	0.6	0.7	0.9
35–39	1.2	0.6	0.7	0.9
40–44	-3.4	2.6	2.7	3.1
45–49	1.0	0.0	0.0	0.0
50–54	-0.2	0.2	0.2	0.2
55–59	0.0	0.0	0.0	0.0
60–64	0.0	0.0	0.0	0.0
65–69	0.0	0.0	0.0	0.0
70–74	0.0	0.0	0.0	0.0
75–79	0.0	0.0	0.0	0.0
80–84	0.0	0.0	0.0	0.0
85–89	0.0	0.0	0.0	0.0
90–94	0.0	0.0	0.0	0.0
95–100	0.0	0.0	0.0	0.0

Based on 1997 scorecard applied to households from 1997.

Figure 10 ($1/day line): Bias and precision for bootstrapped estimates of individuals' poverty likelihoods, by sample size, scorecard applied to validation sample

	Scorecard applied to validation sample			
		Confidence interval (+/- percentage points)		
Sample size (n)	Bias	90-percent	95-percent	99-percent
2	0.3	4.2	31.2	44.3
4	0.2	12.5	20.6	36.8
8	-0.1	11.5	17.5	27.0
16	-0.2	9.4	14.2	20.9
32	-0.3	8.2	11.6	16.6
64	-0.6	6.1	8.3	12.2
128	-0.6	4.4	5.4	9.0
256	-0.5	2.7	3.6	5.9
512	-0.5	1.8	2.3	3.1
1,024	-0.4	1.2	1.4	1.9
2,048	-0.4	0.7	0.9	1.2
4,096	-0.4	0.5	0.6	0.8
8,192	-0.4	0.4	0.4	0.6
16,384	-0.4	0.2	0.3	0.4

Figure 12 ($1/day line): Bias and precision for bootstrapped estimates of groups' poverty rates at a point in time, by sample size, scorecard applied to validation sample

	Scorecard applied to validation sample			
		Confidence interval (+/- percentage points)		
Sample size (n)	**Bias**	**90-percent**	**95-percent**	**99-percent**
2	0.3	4.2	31.2	44.3
4	0.2	12.0	20.7	37.1
8	-0.1	11.0	16.9	26.9
16	-0.1	9.4	12.7	18.7
32	-0.1	7.1	9.2	12.7
64	-0.4	5.7	6.7	9.7
128	-0.5	4.1	5.0	6.8
256	-0.5	2.9	3.6	4.8
512	-0.5	2.1	2.4	3.1
1,024	-0.5	1.6	1.9	2.5
2,048	-0.5	1.0	1.2	1.5
4,096	-0.5	0.7	0.9	1.1
8,192	-0.5	0.5	0.6	0.8
16,384	-0.5	0.4	0.4	0.6

Figure 14 ($1/day line): People by targeting classification and score, along with "Total Accuracy" and BPAC, scorecard applied to validation sample

Score	Inclusion: < poverty line correctly targeted	Undercoverage: < poverty line mistakenly non-targeted	Leakage: ≥ poverty line mistakenly targeted	Exclusion: ≥ poverty line correctly non-targeted	Poverty Accuracy Inclusion + Exclusion	BPAC See text
0–4	0.4	3.2	3.9	92.5	92.9	-6.8
5–9	1.1	2.5	9.3	87.1	88.2	-157.5
10–14	1.6	2.0	16.4	80.0	81.6	-354.7
15–19	2.1	1.5	29.0	67.3	69.5	-705.5
20–24	2.8	0.8	39.2	57.2	60.0	-987.4
25–29	3.0	0.6	49.9	46.5	49.5	-1,283.0
30–34	3.2	0.4	60.3	36.1	39.3	-1,571.3
35–39	3.3	0.3	69.1	27.3	30.6	-1,816.7
40–44	3.6	0.0	74.9	21.5	25.1	-1,977.3
45–49	3.6	0.0	80.9	15.5	19.0	-2,144.8
50–54	3.6	0.0	86.3	10.1	13.7	-2,293.1
55–59	3.6	0.0	88.5	7.9	11.5	-2,353.1
60–64	3.6	0.0	90.0	6.4	10.0	-2,395.5
65–69	3.6	0.0	92.6	3.8	7.4	-2,468.2
70–74	3.6	0.0	94.1	2.3	5.9	-2,510.4
75–79	3.6	0.0	95.2	1.2	4.8	-2,539.7
80–84	3.6	0.0	95.7	0.7	4.3	-2,554.2
85–89	3.6	0.0	96.0	0.4	4.0	-2,561.4
90–94	3.6	0.0	96.1	0.3	3.9	-2,565.3
95–100	3.6	0.0	96.4	0.0	3.6	-2,573.4

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

Figure 15 ($1/day line): People below the poverty line and all people, at a given score or at or below a given score cut-off, scorecard applied to validation sample

	People below poverty line (%)		All people (%)	
Score	At score	At or below score	At score	At or below score
0–4	9.3	9.3	4.2	4.2
5–9	11.3	10.5	6.1	10.4
10–14	7.1	9.1	7.7	18.0
15–19	3.8	6.9	13.2	31.2
20–24	5.9	6.6	10.8	42.0
25–29	1.8	5.6	10.9	52.8
30–34	2.0	5.0	10.6	63.4
35–39	1.8	4.6	9.0	72.5
40–44	4.3	4.6	6.0	78.5
45–49	0.0	4.3	6.0	84.5
50–54	0.2	4.0	5.4	89.9
55–59	0.0	3.9	2.2	92.1
60–64	0.0	3.9	1.5	93.6
65–69	0.0	3.7	2.6	96.2
70–74	0.0	3.7	1.5	97.7
75–79	0.0	3.7	1.1	98.8
80–84	0.0	3.6	0.5	99.3
85–89	0.0	3.6	0.3	99.6
90–94	0.0	3.6	0.1	99.7
95–100	0.0	3.6	0.3	100.0

$2/Day Poverty Line Figures

Figure 5 ($2/day line): Estimated poverty likelihoods associated with scores

If an individual's score is . . .	. . . then the likelihood (%) of being below the poverty line is:
0–4	63.4
5–9	45.8
10–14	49.7
15–19	39.4
20–24	27.7
25–29	30.8
30–34	24.0
35–39	21.1
40–44	14.8
45–49	12.6
50–54	17.9
55–59	12.5
60–64	11.3
65–69	0.6
70–74	1.1
75–79	2.7
80–84	0.0
85–89	0.0
90–94	0.0
95–100	0.0

Surveyed cases weighted to represent Kenya's population.
Based on the WMS 1997

Figure 6 ($2/day line): Illustration of derivation of estimated poverty likelihoods associated with scores

Score	People below poverty line		All people at score		Poverty likelihood (estimated, %)
0–4	2,692	÷	4,246	=	63.4
5–9	2,804	÷	6,126	=	45.8
10–14	3,808	÷	7,657	=	49.7
15–19	5,183	÷	13,155	=	39.4
20–24	2,993	÷	10,799	=	27.7
25–29	3,343	÷	10,857	=	30.8
30–34	2,547	÷	10,601	=	24.0
35–39	1,901	÷	9,010	=	21.1
40–44	896	÷	6,046	=	14.8
45–49	764	÷	6,042	=	12.6
50–54	958	÷	5,354	=	17.9
55–59	271	÷	2,163	=	12.5
60–64	173	÷	1,531	=	11.3
65–69	16	÷	2,621	=	0.6
70–74	17	÷	1,520	=	1.1
75–79	28	÷	1,058	=	2.7
80–84	0	÷	523	=	0.0
85–89	0	÷	258	=	0.0
90–94	0	÷	140	=	0.0
95–100	0	÷	292	=	0.0

Surveyed cases weighted to represent Kenya's population.
Based on the 1997 WMS

Figure 8 ($2/day line): Bootstrapped differences between estimated and true poverty likelihoods for individuals in a large sample (n=16,384) from the validation sample, with confidence intervals

	Scorecard applied to validation sample			
		Confidence interval (+/- percentage points)		
Score	**Bias**	**90-percent**	**95-percent**	**99-percent**
0–4	0.7	4.7	5.3	6.8
5–9	-4.4	4.0	4.4	5.8
10–14	2.1	3.5	4.2	5.5
15–19	4.4	2.4	2.9	3.7
20–24	-3.6	3.4	3.8	5.3
25–29	4.7	2.2	2.5	3.2
30–34	2.7	2.1	2.5	3.4
35–39	0.6	2.4	2.8	3.9
40–44	-4.7	3.7	4.0	4.4
45–49	0.7	2.9	3.5	4.6
50–54	7.3	2.6	3.1	4.3
55–59	3.5	4.0	4.7	5.9
60–64	2.9	3.5	4.2	5.7
65–69	-0.6	1.0	1.1	1.4
70–74	0.0	0.6	0.7	1.0
75–79	2.7	0.0	0.0	0.0
80–84	-4.4	4.3	4.7	5.7
85–89	0.0	0.0	0.0	0.0
90–94	0.0	0.0	0.0	0.0
95–100	0.0	0.0	0.0	0.0

Based on 1997 scorecard applied to households from 1997.

Figure 10 ($2/day line): Bias and precision for bootstrapped estimates of individuals' poverty likelihoods, by sample size, scorecard applied to validation sample

	Scorecard applied to validation sample			
		Confidence interval (+/- percentage points)		
Sample size (n)	**Bias**	**90-percent**	**95-percent**	**99-percent**
2	3.4	52.4	59.5	70.5
4	3.0	39.8	47.2	59.0
8	2.3	32.9	39.3	47.9
16	1.6	24.4	28.7	36.7
32	1.0	19.2	22.9	27.9
64	0.8	14.2	16.4	22.0
128	0.8	9.8	11.8	15.6
256	0.8	6.5	7.9	10.3
512	0.9	4.3	5.1	6.7
1,024	0.8	2.7	3.2	4.4
2,048	0.9	1.9	2.2	2.9
4,096	0.8	1.3	1.6	2.0
8,192	0.8	0.9	1.1	1.4
16,384	0.8	0.6	0.7	1.0

Figure 12 ($2/day line): Bias and precision for bootstrapped estimates of groups' poverty rates at a point in time, by sample size, scorecard applied to validation sample

	Scorecard applied to validation sample			
		Confidence interval (+/- percentage points)		
Sample size (n)	**Bias**	**90-percent**	**95-percent**	**99-percent**
2	3.4	52.4	59.5	70.5
4	2.8	39.8	47.5	58.2
8	2.4	32.2	38.7	45.4
16	1.6	23.6	27.0	33.5
32	1.4	17.0	21.1	25.8
64	1.1	12.6	14.9	18.3
128	1.2	8.7	10.5	13.5
256	1.2	6.0	7.4	10.7
512	1.2	4.3	5.3	7.0
1,024	1.2	3.2	3.7	4.8
2,048	1.3	2.2	2.6	3.5
4,096	1.2	1.6	1.9	2.6
8,192	1.2	1.1	1.3	1.8
16,384	1.2	0.8	0.9	1.2

Figure 14 ($2/day line): People by targeting classification and score, along with "Total Accuracy" and BPAC, scorecard applied to validation sample

Score	Inclusion: < poverty line correctly targeted	Undercoverage: < poverty line mistakenly non-targeted	Leakage: ≥ poverty line mistakenly targeted	Exclusion: ≥ poverty line correctly non-targeted	Poverty Accuracy Inclusion + Exclusion	BPAC See text
0–4	2.7	24.5	1.6	71.2	73.9	-74.6
5–9	5.7	21.5	4.6	68.2	73.9	-40.8
10–14	9.4	17.8	8.7	64.2	73.5	0.8
15–19	14.0	13.2	17.2	55.6	69.6	36.7
20–24	17.4	9.8	24.6	48.2	65.6	9.5
25–29	20.2	7.0	32.6	40.2	60.4	-20.0
30–34	22.5	4.7	41.0	31.8	54.3	-50.7
35–39	24.3	2.9	48.1	24.7	49.0	-77.0
40–44	25.5	1.7	53.0	19.8	45.3	-94.9
45–49	26.2	1.0	58.3	14.5	40.7	-114.5
50–54	26.8	0.4	63.1	9.7	36.5	-132.1
55–59	27.0	0.2	65.1	7.7	34.7	-139.3
60–64	27.1	0.1	66.5	6.3	33.5	-144.5
65–69	27.2	0.0	69.1	3.8	30.9	-154.0
70–74	27.2	0.0	70.6	2.2	29.4	-159.5
75–79	27.2	0.0	71.6	1.2	28.4	-163.4
80–84	27.2	0.0	72.1	0.7	27.9	-165.2
85–89	27.2	0.0	72.4	0.4	27.6	-166.2
90–94	27.2	0.0	72.5	0.3	27.5	-166.7
95–100	27.2	0.0	72.8	0.0	27.2	-167.8

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

Figure 15 ($2/day line): People below the poverty line and all people, at a given score or at or below a given score cut-off, scorecard applied to validation sample

	People below poverty line (%)		All people (%)	
Score	At score	At or below score	At score	At or below score
0–4	62.5	62.5	4.2	4.2
5–9	50.1	55.2	6.1	10.4
10–14	47.7	52.0	7.7	18.0
15–19	35.0	44.8	13.2	31.2
20–24	31.4	41.4	10.8	42.0
25–29	26.1	38.2	10.9	52.8
30–34	21.5	35.4	10.6	63.4
35–39	20.5	33.6	9.0	72.5
40–44	19.6	32.5	6.0	78.5
45–49	11.9	31.0	6.0	84.5
50–54	10.6	29.8	5.4	89.9
55–59	8.9	29.3	2.2	92.1
60–64	8.5	29.0	1.5	93.6
65–69	1.2	28.2	2.6	96.2
70–74	1.1	27.8	1.5	97.7
75–79	0.0	27.5	1.1	98.8
80–84	4.4	27.4	0.5	99.3
85–89	0.0	27.3	0.3	99.6
90–94	0.0	27.3	0.1	99.7
95–100	0.0	27.2	0.3	100.0