Developing Poverty Assessment Tools Project

Note on Assessment and Improvement of Tool Accuracy

The IRIS Center

June 2, 2005

At the workshop organized by the project on January 30, 2004, practitioners identified a number of standards for evaluating the performance of poverty assessment tools.[1] This note summarizes key issues in refining and evaluating the accuracy of these tools, beginning with the conceptual problems with using a "bright line" to categorize households' level of well-being.

1. Alternative Criteria for Assessing Tool Accuracy[2]

Background

The USAID/IRIS project on Developing Poverty Assessment Tools is collecting new data in four countries to assess a selected set of indicators against the task of identifying "very poor" households (according to the statutory definition of extreme poverty, discussed below). A composite survey questionnaire, compiled from several practitioner tools, is administered to a sample of 800 households. A benchmark for assessing measurement accuracy is developed using the expenditure module of the World Bank's *Living Standards Measurement Survey* (LSMS). Administered to the same set of households exactly 14 days later, this benchmark provides the best available quantitative information on the "true" poverty status of each sample household.[3] Statistical methods are then used to identify the 5, 10, or 15 indicators within this composite survey that most accurately reflect the "true" poverty status of each household – that is, that most closely track the benchmark results.

In addition, a comparative analysis draws on existing LSMS data sets from an additional eight countries to identify the 5, 10, or 15 best poverty predictors (using a similar methodology and set of variables), to facilitate generalization of findings over a larger number of countries.

Any effort to assess the poverty status of a set of households – to classify each household as either very poor or not – must start with the choice of an appropriate poverty line. This project is tasked with finding tools to identify households living in *extreme poverty* – the very poor,

---

[1] Notes from this session can be found at http://www.povertytools.org/documents/accuracy.pdf. For the full report from the Certification Criteria workshop, visit http://www.povertytools.org/documents/Criteria%20Workshop%20Report.pdf

[2] "Tool" in the context of this paper refers only to the set of indicators used to assess poverty. A "poverty assessment tool" for the purpose of this project encompasses the range of issues involved in collecting and analyzing data, as well as the indicators used.

[3] Most development specialists agree that poverty is a multi-dimensional problem, of which an inadequate level of income or expenditures is but one facet. Vulnerability to various kinds of risk, political and social disempowerment, and lack of access to social services and assets are equally important dimensions of the reality of poverty. However, because the language of the Congressional legislation that underlies this project defines poverty only in monetary terms, the project focuses exclusively on one dimension of poverty: measuring household incomes or expenditures.

defined as all households living below the "extreme poverty" line established in the Amendment to the *Microenterprise for Self-Reliance and International Anti-Corruption Act* of 2000.

According to that legislation, a household is classified as "very poor" if *either*

(1) the household is "living on less than the equivalent of a dollar a day" ($1.08 per day at 1993 Purchasing Power Parity) — the definition of "extreme poverty" under the Millennium Development Goals;

or

(2) the household is among the poorest 50 percent of households below the country's own national poverty line.

The wording of the legislation suggests that Congress intends for the *higher* of these two alternative criteria to provide the applicable extreme poverty line for a given country.

Seven Key Concepts

As a convenient shorthand, all households living below the extreme poverty line are referred to as "very-poor" or "VP"; all households living above the extreme poverty line are referred to as "not very-poor" or "NVP." The technical term "not very-poor" may tend to obscure the fact that many of these households would be considered poor or very poor even by the standards of many developing countries, and desperately poor by the standards of developed countries like the United States. In no case does the term "not very-poor" signify that such a household might be considered "comfortable" or "well-off."

Moreover, applying a poverty line to divide a population into two groups can create the misleading impression that all households in the resulting group are relatively similar to one another and very different from households in the other group. In reality, living standards within each group vary widely, and the living standards of households just above the selected poverty line may be virtually indistinguishable from those just below it.

As indicated above, the accuracy of a particular set of indicators is assessed by comparing the poverty status predicted by each potential tool with the "true" poverty status as established by the benchmark (LSMS) data. Four situations are possible, as summarized in the following table.

	Predicted as **VP** by the tool	Predicted as **NVP** by the tool
"True" **VP** (as determined by benchmark survey)	A	B
"True" **NVP** (as determined by benchmark survey)	C	D

Seven key concepts can be derived from this matrix.

Three accuracy criteria

1. *Total Accuracy* = sum of correctly predicted VP plus correctly predicted NVP, expressed as a percentage of the total sample.
   From the matrix, Total Accuracy = 100 * (A + D) / (A + B + C + D)

2. *Poverty Accuracy* = correctly predicted VP as a percentage of total "true" VP.
   From the matrix, Poverty Accuracy = 100 * A / (A + B)

3. *Non-poverty Accuracy* = correctly predicted NVP as a percentage of total "true" NVP.[4]
   From the matrix, Non-poverty Accuracy = 100 * D / (C + D)

Two incidence figures

4. *Actual Poverty Incidence* = respondents who are "true" VP, regardless of whether or not they are correctly predicted, expressed as percentage of the total sample.
   From the matrix, Actual Poverty Incidence = 100 * (A + B) / (A + B + C + D)

5. *Predicted Poverty Incidence* = respondents who are predicted as VP, regardless of their actual poverty status, expressed as percentage of the total sample.
   From the matrix, Predicted Poverty Incidence = 100 * (A + C) / (A + B + C + D).

Three errors

6. *Undercoverage* = "true" VP incorrectly predicted as NVP, expressed as a percentage of total "true" VP.
   From the matrix, Undercoverage =100 * B / (A + B). (By definition, this ratio is equal to (100 – Poverty Accuracy.)

7. *Leakage* = "true" NVP incorrectly predicted as VP, expressed as a percentage of total "true" VP.
   From the matrix, Leakage = 100 * C / (A + B).

[4] "Poverty accuracy" might be more precisely expressed as "accuracy among the very-poor," while "non-poverty accuracy" could be more precisely worded as "accuracy among the not very-poor." Unfortunately, these more precise phrases are also quite cumbersome, tending to make the discussion harder to follow. For this reason, the discussion uses the former terms as shorthand for their longer and more precise equivalents.

8. *Poverty Incidence Error ("PIE")* = difference between Predicted Poverty Incidence and Actual Poverty Incidence, expressed in percentage points.
   From the matrix, Poverty Incidence Error = 100 * ((A + C) – (A + B)) / (A + B + C + D), or, simplifying, 100 * (C – B) / (A + B + C + D).

Example 1 presents a fictitious sample of 200 respondents (60 "true" VP and 140 "true" NVP).

Example 1	Predicted as VP	Predicted as NVP	*Total*
"True" VP	**40**	20	*60*
"True" NVP	80	**60**	*140*
Total	*120*	*80*	***200***

Example 1 would produce the following percentages:

1. Total Accuracy = 100 * (40 + 60) / 200 = 50%. Out of a total sample of 200, 40 VP respondents are correctly predicted as VP, and 60 NVP respondents are correctly predicted as NVP.

2. Poverty Accuracy = 100 * 40 / 60 = 66.67%. 40 out of 60 VP respondents are accurately predicted.

3. Non-poverty Accuracy = 100 * 60 / 140 = 42.86%. 60 out of 140 NVP respondents are correctly predicted.

4. Actual Poverty Incidence = 100 * 60 / 200 = 30%. There are 60 "true" VP respondents in the sample.

5. Predicted Poverty Incidence = 100 * 120 / 200 = 60%. The tool predicts 120 respondents in the sample as VP.

6. Undercoverage error = 100 * 20 / 60 = 33.33%. 20 VP respondents (out of 60) are incorrectly predicted as NVP.

7. Leakage error = 100 * 80 / 60 = 133.33%. 80 NVP respondents (out of 140) are incorrectly predicted as VP.

The remainder of this note discusses the merits and drawbacks of different accuracy measures, presents two new alternative measures, and discusses four analytic approaches developed by the team to improve accuracy results.

The Case for and against "Total Accuracy"

Total Accuracy is a relatively intuitive measure of accuracy; in the above example, the tool identifies half the respondents correctly (100 out of a sample of 200).

However, because Total Accuracy combines accurate identification of both types of household – very-poor and not very-poor – this measure is only useful if we are interested in an aggregate assessment of poverty status without wanting to target funding specifically to the very-poor population. In some cases, a tool with high Total Accuracy might give a substantially inaccurate identification of very-poor households.

For example, Example 2 would yield the same Total Accuracy (50%) as Example 1, but in this case only one out of six VP respondents is correctly predicted (i.e., Poverty Accuracy = 100 * 10 / 60 = 16.67%).

Example 2	Predicted as VP	Predicted as NVP	*Total*
"True" VP	**10**	50	*60*
"True" NVP	50	**90**	*140*
Total	*60*	*140*	***200***

Moreover, as Example 3 demonstrates, a tool might in fact fail to identify *any* of the 60 "true" VP respondent as VP (Poverty Accuracy = 0), and still yield a Total Accuracy of 50%.

Example 3	Predicted as VP	Predicted as NVP	*Total*
"True" VP	**0**	60	*60*
"True" NVP	40	**100**	*140*
Total	*40*	*160*	***200***

The Case for and against "Poverty Accuracy"

Examples 1 through 3 suggest that Poverty Accuracy may be a more relevant criterion than Total Accuracy to satisfy the Congressional Mandate requiring tools that assess *poverty incidence* rather than the poverty status of the population at large.

However, a tool with high Poverty Accuracy may also make significant errors, as Example 4 suggests.

Example 4	Predicted as VP	Predicted as NVP	*Total*
"True" VP	**50**	10	*60*

"True" NVP	40	**100**	*140*
Total	*90*	*110*	***200***

In this example, the tool correctly classifies 50 out of the 60 "true" VP (hence Poverty Accuracy is a respectable 83.33%). However, it misclassifies 40 out of 140 "true" NVP, including them in the category VP (Leakage error of 66.67%). Thus, of the 90 respondents predicted as VP, 40 are in fact NVP. It also misclassifies 10 "true" VP as NVP (Undercoverage of 16.67%).

In an extreme case, the tool could identify all 60 "true" VP respondents as VP (Poverty Accuracy of 100%) and still produce a large Leakage error, as indicated in Example 5.

Example 5	Predicted as VP	Predicted as NVP	*Total*
"True" VP	**60**	0	*60*
"True" NVP	40	**100**	*140*
Total	*100*	*100*	***200***

The possibility that high Poverty Accuracy can be combined with significant overestimation of the number of VP (high Leakage error) is a serious concern, if USAID is committed to targeting its funding to the VP. The tool illustrated in Example 5, for example, would allow USAID to develop assistance programs intended to benefit all 100 microentrepreneurs classified as VP, of whom only 60 are "truly" VP.

Poverty Accuracy, considered alone, cannot therefore be a sufficient accuracy criterion to develop targeted programs of microenterprise support.

The Need for New Accuracy Criteria

If the sole objective of the Congressional Mandate is to develop tools to evaluate the *aggregate* poverty level of populations served by USAID's microenterprise programs, then the most relevant criterion would be one that minimizes the difference between Predicted Poverty Incidence and Actual Poverty Incidence (Example 6).

Example 6	Predicted as VP	Predicted as NVP	*Total*
"True" VP	**40**	20	*60*
"True" NVP	20	**120**	*140*
Total	*60*	*140*	*200*

In this example, Predicted Poverty Incidence (total "true" VP) is 60, and Actual Poverty Incidence also happens to be 60. To derive the Poverty Incidence Error, we subtract Actual Poverty Incidence from Predicted Poverty Incidence: 60 - 60 = 0. This tool thus completely satisfies the language of the Congressional Mandate – although it does raise two potential objections.

The first objection is that minimizing the Poverty Incidence Error may mask high Undercoverage and Leakage errors, as shown in Example 7.[5]

Example 7	Predicted as VP	Predicted as NVP	*Total*
"True" VP	**10**	50	*60*
"True" NVP	50	**90**	*140*
Total	*60*	*140*	***200***

In this example, the Undercoverage and Leakage errors are both 83%, but since the absolute size of the two errors is the same, they fully offset each other: the tool (in this case) provides an accurate measure of the number of VP, at 60 households, but its potential for inaccuracy remains high.

The second objection is that minimizing the Poverty Incidence Error does not necessarily imply high rates of Total Accuracy or Poverty Accuracy. These rates are robust for Example 6, at 80% and 67% respectively, while for Example 7 they are a disappointing 50% and 17%.

A tool that produces such low Total and Poverty Accuracy rates, while it satisfies the letter of the Congressional Mandate, would seem to fall short of the intention of the law. The IRIS team proposes a new potential accuracy criterion that combines Poverty Accuracy and the Poverty Incidence Error. This new measure, which we call the *Balanced Poverty Accuracy Criterion* ("**BPAC**"), is defined as follows:

> ***BPAC*** *= Poverty Accuracy minus the absolute difference between Undercoverage and Leakage, each expressed in absolute numbers or in ratios with the same denominator.*

When Undercoverage and Leakage are equal, as in Examples 6 and 7, the BPAC is equal to Poverty Accuracy (at 67% in Example 6, and 17% in Example 7). In the situation depicted under Example 1, this criterion has a value of -33.33%, derived as follows:

BPAC = 100 * A/(A+B) – |100 * B/(A + B) – 100 * C/(A + B) |
= 66.7 – |33.33 – 133.33| (using the number of "true" VP as common denominator)

The application of the BPAC is based on the following assumptions:

[5] By definition, minimizing the Poverty Incidence Error (defined on page 3) is equivalent to minimizing the absolute value of the difference between Undercoverage and Leakage. Where these two errors are equal (i.e., they cancel each other out), Poverty Incidence Error is equal to zero, as in Examples 6 and 7. (Actual Poverty Incidence and Predicted Poverty Incidence are both at 60, so Undercoverage and Leakage are equal (33% in Example 6, and 83% in Example 7).)

1. Undercoverage and Leakage are considered equally problematic (i.e., it is equally bad to misclassify a VP person as to misclassify a NVP person).
2. When predicting the poverty rate, erring above or below the poverty line is considered equally problematic.
3. USAID is indifferent between a one unit gain in accuracy and a one unit decrease in net error.

While there may be other possible criteria to measure accuracy (for example, a modified BPAC that would use Total Accuracy rather than Poverty Accuracy as its starting point), the choice of the accuracy criterion to be used by USAID as part of the tool certification process will require balancing the stipulations of the Congressional Mandate against the practical implications of the assessment tools.

2. Alternative Estimation Techniques to Improve Tool Accuracy

So far we have discussed differing approaches, or criteria, for measuring the performance of selected measurement tools. Once a decision has been taken on the most appropriate criterion(a) to assess tool accuracy, researchers will need to develop techniques to increase their degree of accuracy, taking into account the very different conditions—in particular poverty incidence—that characterize the USAID partner countries. We now address the various approaches being considered to increase accuracy of the tools, regardless of the criterion or standard of measurement that may be adopted.

These approaches do a better job than the currently utilized method for identifying indicators that correctly predict the very-poor. The current method selects indicators based on their contribution to Total Accuracy. The results from the Bangladesh data (as well as from the LSMS data sets) reveal that, in practice, too many of the indicators identified as "accurate" actually perform best at the *higher reaches* of the income distribution. This finding, which could not have been anticipated, explains why Total Accuracy is high at the same time that Poverty Accuracy is low.

The newly developed methods focus on finding indicators that correctly identify people at the *low end* of the income distribution. The most promising of these methods are the following four.

1. *Two-step method*. This approach (a) predicts who the "non very-poor" are and (b) then eliminates them from analysis. In step (b), the model with the best 5, 10 or 15 predictors is applied to the remaining part of the sample.
2. *Quantile regression method*. Regressions are estimated through different points of the distribution, allowing us to assess the relative importance of different variables as we move along the distribution.
3. *Linear probability*. This method selects variables based on a linear model with a dependent variable with a binary value—"truly" very-poor or not very-poor —rather than the log of household consumption expenditures.
4. *Variance ratio method*. The ideal predictor has zero variance within the very-poor and within the not very-poor but maximum variance between the two groups (for example, *all* NVP people own a car, and *no* VP people own a car). Variables are selected that maximize the ratio of between-variance over within-variance.

While all of these methods are expected to improve the ability to correctly identify the very-poor, this is likely to be at the cost of lower total accuracy. Therefore, these methods will require pre-testing for each country studied.[6] It is impossible to determine on a theoretical basis which method will most increase the ability to correctly identify the VP in a given country, in relation to the decrease in Total Accuracy for all 12 countries in the sample. Hence the only way to determine which method is more promising overall is to try all four methods on all 12 countries.[7]

[6] It is important to note that the analysis techniques described here only relate to the choice of indicators for the practicality tests. Once these have been selected, they will be incorporated in the design of the tools to be tested—and ultimately certified by USAID. In other words, practitioners using the tools will not be required to be familiar with these techniques, since the data entry shells for the tools will automatically incorporate the value of the coefficients resulting from the analysis described here.

[7] It is possible that convergence will appear before the four methods are applied to all countries, but a concern for sufficient confidence would suggest that the tests be run on at least eight countries (all four field countries and four LSMS countries) before a convergence can be confirmed.