Certification Criteria: Accuracy
USAID/IRIS Tool for India (Bihar and Uttar Pradesh)
Submitted: October 31, 2006
Revised: November 9, 2006

1. Please describe the overall approach to the tool development

The USAID/IRIS project on Developing Poverty Assessment Tools collected new data in four countries—Bangladesh, Peru, Kazakhstan, and Peru—to assess a selected set of indicators against the task of identifying "very poor" households based on household per capita expenditures. A composite survey questionnaire, compiled from several practitioner tools, was administered to a sample of 800 households. A benchmark for assessing measurement accuracy was developed using the expenditure module of the World Bank's *Living Standards Measurement Survey* (LSMS). Administered to the same set of households exactly fourteen days later, this benchmark provided the best available quantitative information on the "true" poverty status of each sample household. Multiple statistical methods were then used to identify the 15 indicators (for each "step" if a 2-step model) within this composite survey that most accurately reflect the "true" poverty status of each household – that is, that most closely track the benchmark expenditure results. Identification of the most accurate set of indicator and the weights attached to them were done of the basis of criteria developed especially for this project. In addition to the four countries already listed, a comparative analysis drew on existing LSMS data sets from an additional eight countries to identify the 15 best poverty predictors (using a similar methodology and set of variables), to facilitate generalization of findings over a larger number of countries. The eight LSMS countries are: Albania, Ghana, Guatemala, India (Bihar and Uttar Pradesh), Jamaica, Madagascar, Tajikistan, and Vietnam. Thus, statistical testing for accuracy was carried out for twelve countries in total. The 110 indicators that appeared in the 'best 15' from at least one of the twelve countries were included in the next part of the project: testing for practicality.

The 110 indicators were divided into six surveys to be tested for practicality. Seventeen microenterprise organizations were selected by USAID to conduct the field tests of practicality. Each question was rated as to whether the respondent found it to be sensitive, difficult, or that it was perceived that she falsified her answer. The lessons learned from the practicality testing were brought in after the top 15 poverty indicators were determined for each country. If a top 15 indicator caused difficulties in testing, the indicator was dropped for the list and the next best indicator replaced it. The end result of this development process was a country-specific poverty assessment tool for each of the twelve countries that predicts—rather than directly measures—household per capita expenditure based on a short set of indicators. Each country tool is incorporated into a data entry template that allows microenterprise practitioner to easily enter and store the responses of its sampled clients to indicator questions and will also calculate the percentage of that practitioner's clients that are predicted to be very poor.

2. Please describe the data source used to develop or calibrate the tool.

Eight of the twelve country tools for this project were developed from existing LSMS data. For the other four countries, original survey data was collected, using both a

composite survey consisting of poverty indicators from multiple sources and a benchmark expenditure survey based on the LSMS expenditure module. The sample was selected to be nationally representative.

For India, the poverty assessment tool was developed from a 1998 LSMS survey on a representative sample for the states of Bihar and Uttar Pradesh.[1] Economic growth (GDP per capita) in India from 1998 to present has averaged 4.6 percent per year, while inflation has averaged 4.2 percent per year.[2] Bihar and Uttar Pradesh likely lag the rest of India in economic performance, but both are growing nonetheless.[3] While the currency values in the data set have been adjusted for inflation, the pace of economic change in India raises concerns about the suitability of the IRIS poverty assessment tool. The choice is between a somewhat-outdated tool and a loan size tool with (most likely) strongly inaccurate predictions.[4] Although IRIS cannot definitively prove this without testing the two alternative tools on new data, we would expect that the accuracy performance of a loan size tool is so poor that it would be worth certifying the India (Bihar and Uttar Pradesh) tool. The sample size used for the statistical testing of the tools was 2133 households.

3. Please describe the process used to select the indicators included in the tool.

LSMS data sets typically include variables related to education, housing characteristics, consumer durables, agricultural assets, financial assets, illness and disability, and employment. For India, roughly seventy indicators from all categories were considered.

The MAXR procedure in SAS was used to select the best poverty indicators (for variables found to be practical) from the pool of potential indicators in an automated manner. MAXR is commonly used to narrow a large pool of possible indicators into a more limited, yet statistically powerful set of indicators. The MAXR technique seeks to maximize explained variance (i.e., R^2) by adding one variable at a time (per step) to the regression model, and then considering all combinations among pairs of regressors to move from one step to the next. Thus, the MAXR technique allows us to identify the best model containing 15 variables (not including control variables for household size, age of the household head, and location).

The MAXR procedure yielded the best 15 variables for the OLS model (also used for the Quantile model) and another set of best 15 variables for the Linear Probability model (also used for the Probit model). The final set of indicators and their weights, therefore depended on selecting one of these four statistical models—OLS, Quantile, Linear Probability, or Probit—as the best model.[5] This selection of the best model was based on the BPAC and PIE accuracy criteria.

---

[1] Additional information on the data set is available here: http://www.worldbank.org/lsms/country/india/upbhdocs.html

[2] Data available from: http://www.imf.org/external/pubs/ft/weo/2006/01/data/dbginim.cfm

[3] http://www.thehindubusinessline.com/2006/09/17/stories/2006091702650500.htm

[4] When a loan size tool was tested in Bangladesh, its BPAC was strongly negative (-49.5). Please see: http://www.povertytools.org/Project_Documents/Bangladesh%20Accuracy%20Report%20Final.pdf

[5] The set of indicators and their weights also depended on the selection of a 1-step or 2-step statistical model.

4. Please describe the estimation methods used to identify final indicators and their weights/coefficients.

India (PPP) **Poverty Line:** 3,970 rupees per person per year[6] **Poverty Rate:** 77.87%	**Total Accuracy**	**Poverty Accuracy**	**Under-coverage**	**Leakage**	**PIE**	**BPAC**
Single-step methods						
OLS	83.63	94.39	5.60	15.24	7.57	84.76
Quantile regression (estimation point: 66)	**83.98**	**89.77**	**10.23**	**10.16**	**-0.06**	**89.69**
Linear Probability	83.83	95.94	4.06	16.52	9.79	83.48
Probit	84.44	94.69	5.311	14.50	7.22	85.50
Two-step methods						
OLS – 78 percentile cutoff	84.18	93.28	6.72	13.43	5.27	86.57
Quantile (estimation points: 66, 29.64) 78 percentile cutoff	84.23	89.29	10.71	9.36	-1.06	87.93
LP – 78 percentile cutoff	85.01	94.96	5.04	14.04	7.07	85.96
Probit – 78 percentile cutoff	85.43	92.51	7.49	11.05	2.79	88.95

	Number of households[7] predicted as very poor by the tool	**Number of households predicted as not very-poor by the tool**
Number of "true" very poor households (as determined by benchmark survey)	1504.139	171.4451
Number of "true" not very-poor households (as determined by benchmark survey)	170.217	287.1991

[6] The poverty line reported above is expressed in the prevailing prices at the time of data collection (March 1998). The poverty line and other monetary variables used in the accuracy results, however, are inflation-adjusted through the time of tool construction (May 2006), to ensure that the resulting poverty predictions are as current as possible. The poverty line for India, expressed in the most current available prices, 5,783.79 rupees per person per year.

[7] Values in the matrix are weighted for national representation.

5. Please describe how coefficients and weights are used to compute prediction of poverty status or estimate of household expenditures.

The weights attached to the indicators in the tool in each country are simply the regression coefficients for the statistical model exhibiting the highest statistical accuracy (according to the BPAC criterion). For India, the weights are from a 1-step Quantile model.

The weights are located in the "backend" analysis program of the EPI template as part of the extreme poverty rate calculation. While a skilled EPI user would be able to locate these values and potentially manipulate them, they would not be seen by the client or the interviewer during the normal course of interviewing, entering the data, or in calculating the extreme poverty rate.

6. Please describe the decision rule used to classify households as very poor or not very-poor.

The extreme poverty line for each of the twelve countries in the project was the higher of the two potential poverty lines specified in the legislation: $1.08 a day (in PPP terms) OR the bottom half of households living below the poverty line (termed the 'median' poverty line). Two of statistical models (OLS and Quantile) used by the IRIS team predict the per capita consumption expenditures for each household, which is then compared to the binding poverty line to decide whether the household is very poor.[8] The other two statistical models (Linear Probability and Probit) predict the probability that a household is very poor (according to the binding, absolute poverty line). If this probability exceeds 0.5, the household is predicted to be very poor.[9]

For India, the binding poverty line is the $1.08-a-day line of 3,970 rupees per person per year. Because the tool is based on a Quantile model, those households whose predicted expenditures according to the tool fall below this line will be considered very poor.

[8] For a 2-step OLS or Quantile model, the decision rule in the 1st-step compares the expenditures predicted for each household to a certain expenditure cutoff.

[9] For a 2-step Linear Probability or Probit model, the decision rule in the 1st-step compares the predicted probability that the households' expenditures exceed a certain cutoff to the 0.5 value.

INDIA 1 STEP MAXR/QUANT: Regression results, estimation point of 66 percentile
Sample: Full
(sum of wgt is 2.1330e+03)

.66 Quantile regression — Number of obs = 2133
Raw sum of deviations 824.0891 (about 8.427021)
Min sum of deviations 567.293 — Pseudo R2 = 0.3116

Variable	Coefficient	Standard Error	t	P>|t|	[95% Confidence Interval]	
Household size	-0.12292	0.006741	-18.23	0.000	-0.1361388	-0.1096994
Household size squared	0.003175	0.0002988	10.63	0.000	0.0025894	0.0037615
Household head age	0.008291	0.0034877	2.38	0.018	0.0014507	0.0151302
Household head age squared	-6.2E-05	0.0000354	-1.74	0.082	-0.0001311	7.78E-06
Household lives in Bihar North	0.236544	0.0184603	12.81	0.000	0.2003421	0.2727465
Household lives in Bihar Central	0.169266	0.0216236	7.83	0.000	0.1268597	0.2116713
Household lives in UP South	0.204182	0.0295427	6.91	0.000	0.1462459	0.2621177
Household head worked as a casual laborer	-0.10484	0.0190626	-5.5	0.000	-0.1422238	-0.0674568
Household head is never married, widowed, or divorced/ separated	-0.13958	0.0248199	-5.62	0.000	-0.1882534	-0.0909053
Share of household members excluding household head who are illiterate or literate without formal schooling	-0.34889	0.0366317	-9.52	0.000	-0.4207302	-0.2770541
Share of household members excluding household head whose educational attainment is matriculate or intermediate	0.240714	0.0861637	2.79	0.005	0.0717388	0.4096883
Number of rooms in dwelling	0.042748	0.0045504	9.39	0.000	0.0338239	0.0516716
Structure is semi-pucca, pucca through weaker sector housing schemes, or pucca	0.136993	0.0175191	7.82	0.000	0.1026362	0.1713493
Toilet is flush system or septic tank	0.272194	0.0400875	6.79	0.000	0.1935785	0.3508088
Household owns one or more cameras	0.172428	0.0560427	3.08	0.002	0.0625236	0.282333
Household owns one or more pressure lamps	0.125405	0.0490166	2.56	0.011	0.0292796	0.2215313
Household owns one or more watches	0.127357	0.0182188	6.99	0.000	0.0916282	0.1630857
Household owns one or more radios	0.099518	0.0183061	5.44	0.000	0.0636178	0.1354175
Household owns one or more televisions	0.082718	0.0304114	2.72	0.007	0.0230783	0.1423574
Household owns one or more threshers	0.149169	0.0301177	4.95	0.000	0.0901052	0.2082322

Number of cows household owns	0.053505	0.0055865	9.58	0.000	0.0425496	0.064461
Number of buffalo household owns	0.099132	0.0080806	12.27	0.000	0.0832851	0.1149786
Intercept	8.416994	0.0840719	100.12	0.000	8.252121	8.581866