Certification Criteria: Accuracy
USAID/IRIS Tool for Peru
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 best 15 poverty indicators were determined for each country. If a best 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 Peru, the poverty assessment tool was developed from survey conducted by IRIS in 2004 on a nationally-representative sample of 800 households.[1]

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 the four composite surveys conducted by IRIS, additional categories of variables included food security, subjective poverty, voice and vulnerability, and social capital. For Peru, more than 200 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.[2] This selection of the best model was based on the BPAC and PIE accuracy criteria.

---

[1] Further details on the sample can be found here:
http://www.povertytools.org/Project_Documents/Peru%20Accuracy%20Report.pdf

[2] 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.

Peru **(median)** **Poverty Line:** varies by sampling area **Poverty Rate:** 26.88%	**Total Accuracy**	**Poverty Accuracy**	**Under-coverage**	**Leakage**	**PIE**	**BPAC**
Single-step methods						
OLS	83.75	62.33	37.67	22.79	-4.00	47.44
Quantile regression (estimation point: 41)	**83.38**	**69.30**	**30.69**	**31.16**	**0.13**	**68.84**
Linear Probability	86.38	62.79	37.21	13.49	-6.38	39.07
Probit	86.63	67.91	32.09	17.67	-3.88	53.49
Two-step methods						
OLS – 57 percentile cutoff	85.25	67.91	32.09	22.79	-2.50	58.60
Quantile (estimation points: 41, 23) 57 percentile cutoff	83.50	72.09	27.91	33.49	1.50	66.51
LP – 60 percentile cutoff	87.75	72.56	27.44	18.1395	-2.50	63.26
Probit – 60 percentile cutoff	87.75	73.02	26.98	18.60	-2.25	64.65

	Number of households 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)	149	66
Number of “true” not very-poor households (as determined by benchmark survey)	67	518

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 Peru, 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, 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.[3] 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.[4]

For Peru, the binding poverty line was the 'median' poverty line, which is actually seven poverty lines: one for each of the seven sampling areas in Peru (see table below). Because the tool is based on a Quantile model, those households whose predicted expenditures according to the tool fall below the poverty line in their sampling area will be considered very poor.

Poverty lines by sampling area	annual income equivalent to 50% < nat. pov. line	annual income equivalent to 50% < nat. pov. line
	(soles/ person/ year)	(soles/ person/ year)
	As of July 2004	*As of June 2006*
Lima Metrop.	2182.0	2253.0
Urban Coast	1706.6	1760.4
Rural Coast	1108.0	1145.3
Urban Highland	1474.9	1522.1
Rural Highland	867.4	896.7
Urban Lowland	1397.9	1443.0
Rural Lowland	950.6	979.6

---

[3] 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.

[4] 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.

PERU 1-STEP MAXR/QUANT: variables from MAXR/OLS 100 percentile model
Regression results, estimation point of 41 percentile
Sample: Full

.41 Quantile regression Number of obs = 800
Raw sum of deviations 476.7262 (about 7.5185914)
Min sum of deviations 218.349 Pseudo R2 = 0.5420

Variable	Coefficient	Standard Error	t	P>|t|	[95% Confidence Interval]	
Household size	-0.26234	0.027872	-9.41	0.000	-0.317059	-0.2076299
Household size squared	0.010425	0.002373	4.39	0.000	0.0057668	0.0150838
Household head age	0.00581	0.006	0.97	0.333	-0.00597	0.017587
Household head age squared	-0.00007	5.59E-05	-1.25	0.211	-0.00018	0.0000398
Household lives in Lima Metropolitan	0.391584	0.049782	7.87	0.000	0.29386	0.4893075
Household lives in Rural Coast	0.009713	0.083661	0.12	0.908	-0.154517	0.1739426
Household lives in Urban Highlands	0.142301	0.05745	2.48	0.013	0.0295256	0.2550769
Household lives in Rural Highlands	-0.19801	0.056328	-3.52	0.000	-0.30858	-0.0874362
Household lives in Urban Lowlands	0.075455	0.075416	1	0.317	-0.072588	0.2234988
Household lives in Rural Lowlands	-0.18234	0.08278	-2.2	0.028	-0.34484	-0.0198434
Share of household members excluding household head who have no education	-0.22079	0.111331	-1.98	0.048	-0.43934	-0.0022429
Number of rooms in dwelling	0.02389	0.009792	2.44	0.015	0.00467	0.0431122
Household owns one or more telephones	0.251176	0.044584	5.63	0.000	0.1636571	0.3386948
Household members received in-kind services from food aid programs	-0.14552	0.036862	-3.95	0.000	-0.21788	-0.0731604
Number of household members belong to a water/ waste group	0.153268	0.059601	2.57	0.01	0.0362685	0.2702676
Number of cars household owns	0.31995	0.055177	5.8	0.000	0.2116354	0.4282644
Number of color television household owns	0.096703	0.027964	3.46	0.001	0.0418089	0.1515971
Number of metal pots household owns	0.020322	0.005377	3.78	0	0.0097681	0.0308766
Main entrance lock is key lock or simple padlock	0.132037	0.039427	3.35	0.001	0.0546399	0.2094336
Roof is made of leaves, straw or bamboo/ wood	-0.32458	0.063734	-5.09	0.000	-0.44969	-0.1994677
Exterior walls is made of wood	-0.37464	0.076546	-4.89	0.000	-0.52490	-0.2243791
Primary source of	0.168458	0.057153	2.95	0.003	0.0562646	0.2806512

drinking water is public borehole (open), spring, or public well						
Log of average resale value of tractors and trucks household owns	0.063585	0.015673	4.06	0.000	0.0328182	0.0943518
Log of average resale value of tractors and trucks household owns	0.025854	0.005782	4.47	0.000	0.0145028	0.0372043
Household does not have savings account because of too little income (cannot save)	-0.18303	0.044325	-4.13	0.000	-0.270039	-0.0960176
Intercept	8.397509	0.163469	51.37	0.000	8.076614	8.718403