

**Certification Criteria: Accuracy**  
**USAID/IRIS Tool for Kazakhstan**  
**Submitted: November 8, 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 on 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 Kazakhstan, the poverty assessment tool was developed from survey conducted by IRIS in 2004 on a nationally-representative sample of 817 households.<sup>1</sup>

**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 Kazakhstan, 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.<sup>2</sup> This selection of the best model was based on the BPAC and PIE accuracy criteria.

---

<sup>1</sup> Further details on the sample can be found here:

[http://www.povertytools.org/Project\\_Documents/Kazakhstan%20Accuracy%20Report.pdf](http://www.povertytools.org/Project_Documents/Kazakhstan%20Accuracy%20Report.pdf)

<sup>2</sup> 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.**

<b>Kazakhstan (median)</b> Poverty Line: varies by oblast Poverty Rate: 4.53%	<b>Total Accuracy</b>	<b>Poverty Accuracy</b>	<b>Under-coverage</b>	<b>Leakage</b>	<b>PIE</b>	<b>BPAC</b>
<b>Single-step methods</b>						
OLS	95.47	5.41	94.59	5.41	-4.04	-83.78
Quantile regression (estimation point: 21)	94.00	37.84	62.16	70.27	0.37	29.73
Linear Probability	95.59	5.41	94.59	2.70	-4.16	-86.49
Probit	96.33	32.43	67.57	13.51	-2.45	-21.62
<b>Two-step methods</b>						
OLS – 20 percentile cutoff	96.57	37.84	62.16	13.51	-2.20	-10.81
Quantile (estim. points: 21, 4) 20 percentile cutoff	94.61	43.24	56.76	62.16	0.24	37.84
LP – 24 percentile cutoff	98.04	59.46	40.54	2.70	-1.71	21.62
Probit – 24 percentile cutoff	97.92	59.46	40.54	5.41	-1.59	24.32

	<b>Number of households predicted as very poor by the tool</b>	<b>Number of households predicted as not very-poor by the tool</b>
<b>Number of “true” very poor households (as determined by benchmark survey)</b>	16	21
<b>Number of “true” not very-poor households (as determined by benchmark survey)</b>	23	757

**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 Kazakhstan, the weights are from a 2-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.<sup>3</sup> For Kazakhstan, the binding poverty line used by IRIS was the ‘median’ poverty line, which is actually eight poverty lines according to oblast (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 for the relevant oblast will be considered very poor.

Poverty lines by sampled oblast	per adult equivalent to 50% < nat. pov. line (Tenge/ person/day)	per adult equivalent to 50% < nat. pov. line (Tenge/ person/day)
	<i>As of October 2004</i>	<i>As of June 2006<sup>4</sup></i>
Aktobe	121.47	139.49
Almaty urban	156.48	179.69
Almaty rural	145.04	166.56
Est	162.35	186.43
Jambyl	140.97	161.88
Karaganda	138.88	159.48
Kostanay	184.56	211.94
South Kazakhstan	128.37	147.41

Not all of the oblasts in Kazakhstan were sampled in the IRIS survey. Oblasts that were not sampled were paired with a sampled oblast according to two criteria—Subsistence Minimum and Cultural Zone—as displayed in the table found below.

<sup>3</sup> For a 2-step OLS or Quantile model, the decision rule in the 1<sup>st</sup>-step compares the expenditures predicted for each household to a certain expenditure cutoff.

<sup>4</sup> The poverty lines (and other indicators in local currency) for Kazakhstan were updated for inflation to June 2006—the most recent inflation figure available—for the purposes of regression analysis and prediction. This ensures that prediction by the tool will be in the most current terms possible.

### Matching Sampled and Unsampled Oblasts in Kazakhstan

Region/oblast (bold indicates it was sampled by IRIS in 2004)	Subsistence Minimum per adult equivalent (Tenge per day)	Cultural Zone	Paired with...
Akmola	198	North	Karaganda
<b>Aktobe</b>	<b>214</b>	<b>West</b>	-
<b>Almaty rural</b>	<b>188</b>	<b>South</b>	-
Atyrau	240	West	Aktobe
<b>East Kazakhstan</b>	<b>194</b>	<b>East</b>	-
<b>Jambyl</b>	<b>169</b>	<b>South</b>	-
West Kazakhstan	183	West	Kostanai
<b>Karaganda</b>	<b>193</b>	<b>North</b>	-
<b>Kostanai</b>	<b>186</b>	<b>North</b>	-
Kyzylorda	190	South	Almaty rural
Mangistau	257	West	Aktobe
Pavlodar	185	North	Kostanai
North Kazakhstan	189	North	Kostanai
<b>South Kazakhstan</b>	<b>168</b>	<b>South</b>	-
Astana	199	North	Karaganda
<b>Almaty urban</b>	<b>220</b>	<b>South</b>	-

The following method was used: match an unsampled oblast to the sampled oblast with the most similar Subsistence Minimum (SM) per adult equivalent, but also taking into account Cultural Zone where possible. For example, North Kazakhstan was not sampled; its SM is 189. The closest sampled oblast on SM grounds alone is Almaty rural at 188. However, Kostanai has a SM of 186 and is in the same Cultural Zone (North) and thus would be the match. Therefore, households using the template in North Kazakhstan would have the same oblast dummy and median poverty line as those in Kostanai.

KAZAK 2 STEP MAXR/QUANT: variables from MAXR/OLS 100 percentile model  
 Regression results, estimation point of 21 percentile

.21 Quantile regression

Number of obs = 817

Raw sum of deviations 272.2942 (about 11.223296)

Min sum of deviations 184.8812

Pseudo R2 = 0.3210

Variable	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Intercept	10.287	0.325	31.620	0.000	9.648	10.925
Household size	-0.237	0.047	-5.090	0.000	-0.328	-0.146
Household size squared	0.010	0.004	2.610	0.009	0.002	0.017
Household head age	0.014	0.010	1.350	0.176	-0.006	0.034
Household head age squared	0.000	0.000	-1.570	0.117	0.000	0.000
Household lives in Aktobe Oblast	0.171	0.107	1.600	0.110	-0.039	0.381
Household lives in Almaty Oblast	0.248	0.078	3.170	0.002	0.094	0.402
Household lives in East Kazakhstan Oblast	0.156	0.090	1.740	0.082	-0.020	0.332
Household lives in Jambyl Oblast	-0.054	0.088	-0.610	0.541	-0.227	0.119
Household lives in Karaganda Oblast	0.080	0.091	0.880	0.378	-0.098	0.259
Household lives in Kostanay	0.273	0.112	2.440	0.015	0.053	0.492
Household lives in rural area	0.000	0.053	-0.010	0.994	-0.105	0.104
Household owns one or more cell phones	0.217	0.069	3.150	0.002	0.082	0.353
Household owns one or more sets of dishes	0.175	0.050	3.460	0.001	0.075	0.274
Number of meals served to household members in the last 2 days	0.051	0.013	3.770	0.000	0.024	0.077
Number of days in the last seven that butter was served in a main meal eaten by the household	0.039	0.008	5.050	0.000	0.024	0.054
Number of days out of the last seven that cheese, cooked meats, or sausage was served in a main meal eaten by the household	0.033	0.010	3.230	0.001	0.013	0.053
Number of kilograms of flour household typically buys in a single purchase	0.002	0.001	2.850	0.005	0.001	0.003
Maximum education/highest class passed by any member of household	0.036	0.013	2.820	0.005	0.011	0.061

Head of household is a member of a trade association	-0.548	0.152	-3.600	0.000	-0.846	-0.249
Head of household is unemployed, not seeking work	-0.383	0.192	-2.000	0.046	-0.760	-0.007
Number of adults living in the household who can read and write	0.062	0.029	2.110	0.035	0.004	0.119
Number of carpets and felt mats household owns	0.016	0.008	1.890	0.059	-0.001	0.032
Number of horses household owns	0.136	0.045	3.030	0.003	0.048	0.224
None of the members of the household has a withdrawable savings account, checking account, fixed-term deposit account, or any other type of savings account	-0.193	0.050	-3.870	0.000	-0.291	-0.095
Household owns one or more refrigerators	0.133	0.058	2.280	0.023	0.019	0.248
Head of household is a salaried worker/employee	0.128	0.096	1.330	0.183	-0.061	0.317

KAZAK 2 STEP MAXR/QUANT: variables from MAXR/OLS 20 percentile model  
 Regression results, estimation point of 22 percentile  
 Sample: Predicted expenditures below 20 percentile

.22 Quantile regression

Number of obs = 302

Raw sum of deviations 90.8643 (about 10.89783)

Pseudo R2 = 0.2517

Min sum of deviations 67.99728

Variable	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Intercept	10.420	0.617	16.880	0.000	9.205	11.636
Household size	-0.211	0.080	-2.640	0.009	-0.368	-0.053
Household size squared	0.011	0.007	1.550	0.123	-0.003	0.026
Household head age	0.025	0.021	1.180	0.239	-0.017	0.067
Household head age squared	0.000	0.000	-0.990	0.323	-0.001	0.000
Household lives in Aktobe Oblast	-0.092	0.185	-0.500	0.620	-0.457	0.273
Household lives in Almaty Oblast	-0.126	0.133	-0.950	0.345	-0.389	0.137
Household lives in East Kazakhstan Oblast	-0.079	0.196	-0.400	0.687	-0.466	0.308
Household lives in Jambyl Oblast	-0.252	0.150	-1.670	0.095	-0.548	0.044
Household lives in Karaganda Oblast	-0.169	0.207	-0.820	0.414	-0.577	0.238
Household lives in Kostanay	0.094	0.228	0.410	0.682	-0.356	0.543
Household lives in rural area	0.009	0.098	0.090	0.928	-0.185	0.203
Household head is employed as a craftsman	-0.499	0.254	-1.960	0.050	-1.000	0.001
Floor of dwelling is made of cement or cement with additional covering	-0.011	0.205	-0.060	0.955	-0.415	0.392
Household typically purchases flour monthly	0.127	0.100	1.270	0.205	-0.070	0.324
Number of females in the family with a disability	-0.324	0.158	-2.050	0.041	-0.634	-0.013
Roof of dwelling is made of straw or wood	0.114	0.212	0.540	0.593	-0.304	0.532
Toilet is a flush toilet	0.089	0.133	0.670	0.504	-0.173	0.351
Number of meals served to family members in the last two days	0.079	0.035	2.270	0.024	0.010	0.147
Household owns one or more electric or gas cookers or stoves	0.160	0.109	1.470	0.142	-0.054	0.375
During the last three years residence/household was relocated because of reasons not related to violence, such as natural disasters (drought, flood, etc.)	0.556	0.256	2.170	0.031	0.052	1.061
Household is occasionally denied service or has only limited opportunity to use utility services	-0.137	0.119	-1.150	0.249	-0.370	0.097

Household is occasionally denied service or has only limited opportunity to use social help	0.041	0.118	0.350	0.730	-0.191	0.272
Head of household is unemployed, not seeking work	-0.353	0.263	-1.340	0.181	-0.871	0.165
Household owns one or more musical instruments	0.319	0.192	1.660	0.099	-0.060	0.697
Total number of sick days for female members of household in the last twelve months	0.001	0.000	1.840	0.067	0.000	0.001
Household owns one or more wardrobes	-0.043	0.185	-0.230	0.816	-0.407	0.321