

GALLUP®

**The World Bank Listening to LAC (L2L) Pilot Project
Baseline Face-to-Face Surveys in Honduras and Peru
Methodological Report**

June 26 2012

Contents

Background	2
World Bank L2L Survey of Honduras	3
Honduras L2L Sample Design	3
Honduras L2L Sample Weighting	4
Demographic Comparison Between Honduras L2L and GWP Surveys	5
Honduras L2L Sample Geographic Dispersion	6
Map Secondary Sampling Units selected in Honduras as part of L2L Survey.....	7
World Bank L2L Survey of Peru.....	8
Peru L2L Sample Design.....	8
Drawing the L2L sample	9
L2L Strata oversample correction (Weighting)	10
Post-Stratification Weighting.....	10
Demographic Comparison Between Peru L2L and ENAHO.....	11
Peru L2L Sample Geographic Dispersion	12
Map of Secondary Sampling Units sampled for the L2L F2F survey in Peru	13
Conclusion	14

Background

The rapid and massive dissemination of mobile phones in the developing world is creating new opportunities for the discipline of survey research. Private sector organizations and academic institutions concerned with the study of public opinions have embarked themselves in intensive experimentation in an attempt to reap the benefits of the faster and more convenient ways to engage survey respondents afforded by mobile technologies. Mobile phones allow researchers to survey respondents in “real time”, as relevant historical events are occurring, and to simultaneously capture responses in digital formats that can be seamlessly and readily integrated into data processing, visualization and analysis software, none of which could be easily accomplished by means of more traditional survey methods.

However, the survey research community still debates the methodological implications of using mobile phones for survey research. Among the focal points of such debate, has been the ability of mobile surveys to represent all geographic and demographic segments within a given country. While mobile phone penetration has increased dramatically in the developing world, researchers are still concerned with the ability of mobile sampling frames to represent rural areas, and questions still remain with regards to the quality of mobile phone signals outside large urban areas.

The World Bank is interested in leveraging the mobile “Short Messaging Service” (SMS) technology as a means of direct communication with poor households in the developing world in order to gather rapid feedback on the impact of economic crises and other historical events on the economy of such households. With this objective in mind, the World Bank has launched the “Listening to LAC” (L2L) pilot program. A research exercise aimed at testing the feasibility of the SMS technology as a data collection method for conducting quick turnaround, self-administered, longitudinal surveys among households in Peru and Honduras. In order to overcome the above mentioned problems with mobile phone sampling frames, the L2L study relies on probability-based household panels, recruited by means of Door to Door, Face to Face contacts, as part of a baseline survey conducted by trained interviewers using paper and pencil questionnaires.

The following report examines the design of the L2L baseline surveys in Peru and Honduras, with an emphasis on the evaluation of their effectiveness for overcoming the coverage problems commonly attributed to surveys conducted via mobile phones. More specifically, the report compares the geographic and demographic distributions achieved with the L2L Face to Face surveys with those obtained by criterion surveys (National Household “ENAHO” Survey in Peru and Gallup World Poll survey in Honduras), and the official census of each country. The report starts with a description of the surveys’ sample design, then offers comparative data for relevant variables, as well as maps illustrating the dispersion of the sampling units selected in each country.

World Bank L2L Survey of Honduras

The L2L Face to Face survey of Honduras was based on a nationally representative sample of 1,500 households. At each selected household, interviewers asked to speak to the person recognized by the family group as the head of the household. If the head of the households could not be located after two attempts, interviewers proceeded to interview informants aged 15 years or older who were permanent residents of the household.

Honduras L2L Sample Design

The L2L sample was designed under the following premises:

1. The sample had to be nationally representative including both, urban and rural areas;
2. The L2L survey had to adopt the sampling frame used by the Gallup World Poll (GWP) because the project's objectives called for a comparison of data collected by both surveys, at the level of Secondary Sampling Units (SSU's).
3. The World Bank had originally requested a disproportionate design that over-represented poor households. However, after examining the proportion of the country's population around the poverty line (60%, per INEI 2010)¹, such disproportionate design was deemed unnecessary. Therefore, Gallup recommended a stratification based on population size, per specifications provided hereafter.

Similarly to the GWP, the L2L survey used the Honduras Census as its sampling frame. The least disaggregated territorial unit reported by the country's National Institute of Statistics (INEI), are geographic conglomerates known as "barrios"/ "aldeas"/ "caseríos" and "colonias". Therefore, these conglomerates became the Primary Sampling Units (PSU's) in the L2L sample design. Within each PSU, Gallup selected "random origins" (starting points also known as Secondary Sampling Units, or SSU's) from which interviewers followed a systematic random route for selecting households. For the L2L survey Gallup selected the same 125 SSU's that were used in the GWP sample in 2010.

The Honduras L2L sample was drawn by means of a random, stratified, multistage design. The following is a description of such design, as well as the procedures involved in selecting the sampling units:

1. Census-defined municipalities were classified into five strata according to population size, as follows:
 - I.- Municipalities with 500.000 to 999.000 inhabitants
 - II.- Municipalities with 100.000 to 499.000 inhabitants.
 - III. Municipalities with 50.000 to 99.000 inhabitants.
 - IV. Municipalities with 10.000 and 49.000 inhabitants.
 - V. Municipalities with less than 10.000 inhabitants.
2. Interviews were then proportionally allocated to these five strata according to their share of the country's population.
3. The **First Stage** of the design consisted of a random selection of PSU's within each of the five strata previously defined. This procedure was performed by assigning each PSU a probability of selection

proportional to the size of its population. As a result, larger PSU's were not only more likely to be selected than smaller ones, but the number of interviews allocated to them was also greater.

4. In each PSU, one or more SSU's were then selected. The number of SSU's to be selected was determined based on the total number of interviews allocated to the PSU, and the number of interviews to be conducted in each SSU, as mandated by the design (a maximum of 12 in the case of the L2L survey). The selection of SSU's was the **Second Stage** of the sample design.
5. Once SSU's were selected, interviewers were sent to the field to proceed with the **Third Stage** of the sample design, which consisted of selecting households by means of a systematic "random route" procedure. Interviewers started from the previously selected "random origin" and walked around the block in clockwise direction, selecting every third household on their right hand side. They were also trained to handle vacant, non-responsive, non-cooperative households, as well as other failed attempts, in a systematic manner.

The following table offers further details about the sample, as designed for the L2L survey, based on the 2010 administration of the GWP in Honduras.

L2L HONDURAS SAMPLE	
1. Universe	All the households that exist in the neighborhoods of Honduras, as reported by the INE 2001 Census. Institutions such as military, religious or educational living quarters are not included in the universe.
2. Geographic Coverage	Includes the entire national territory, with the exception of neighborhoods where access of interviewers is extremely difficult, due to lack of transportation infrastructure or for situations that threaten the physical integrity of the interviewers and supervisors (i.e. extremely high crime rate, warfare, etc.)
3. Number of Cases	1,500 cases.
4. Sampling Error	±2.6 percent points for results based on the total sample
5. Design effect due to sample clustering	1.06%
6. Sample type	Random multi-stage stratified sample, based on SSU's from the most recent census conducted in Honduras (2001).

Honduras L2L Sample Weighting

Given the socio-economic focus of the L2L survey, and the stark socio-economic differences that exist in Honduras, ensuring proportionality of socio-economic variables was in order.

Age and educational attainment of the head of the household are typically used as proxy variables to a household's socio-economic status due to their high correlation with income and other economic indicators. In the L2L survey of Honduras weighting was performed to correct for observed disproportions in these variables, relative to census data.

Tables A and B below, show frequency distributions for age and educational attainment of the head of the household, respectively. They also show weighted and unweighted estimators for the same variables, as well as census parameters and expansion factors.

Table A
Age of Head of Household
-Honduras-

Age group (HHH)	L2L Frequency	L2L Unweighted Percent	L2L Weighted Percent	Target Parameter (Census)	Weights
15-24	101	6.7%	9.3%	9.3273	1.388059701
25-34	324	21.5%	23.4%	23.4421	1.088372093
35-44	411	27.2%	24.2%	24.1672	0.889705882
45-54	330	21.9%	18.8%	18.8082	0.858447489
55+	344	22.8%	24.3%	24.2552	1.065789474

Table B
Educational Attainment of Head of Household
-Honduras-

Education group (HHH)	L2L Frequency	L2L Unweighted Percent	L2L Weighted Percent	Target percent (from Census)	Weights
Less than Primary	310	20.53	27.6525	27.6525	1.346931
Primary	818	54.17	51.9965	51.9965	0.959876
Secondary +	382	25.3	20.351	20.351	0.804387

Demographic Comparison Between Honduras L2L and GWP Surveys

In order to evaluate the L2L's sample performance relative to the GWP and the Honduras Census, a comparative analysis was conducted for the following demographic variables: a) household age composition and b) PSU's urbanity.

The L2L estimators for household age composition mirror the census parameters quite closely. The GWP overestimated the average number of 15+ year old adults in the household (3.82 vs. 2.91 reported by the Census). Such disproportion was not observed for the L2L weighted or unweighted estimators.

Both, the L2L and the GWP surveys overestimated the proportion of households in rural areas relative to the census. However, since the objective of this exercise is to evaluate the performance of the L2L sample at overcoming the coverage issues commonly attributed to mobile phone sampling frames (particularly the coverage of rural areas), the observed disproportions do not disqualify the L2L sample in any way. On the contrary, the L2L sample design over represents households of interest for the Word Bank and allows for a stringent test of the SMS surveys.

Table C below, shows the performance of the L2L sample on the above mentioned variables, relative to the GWP and the Honduras Census.

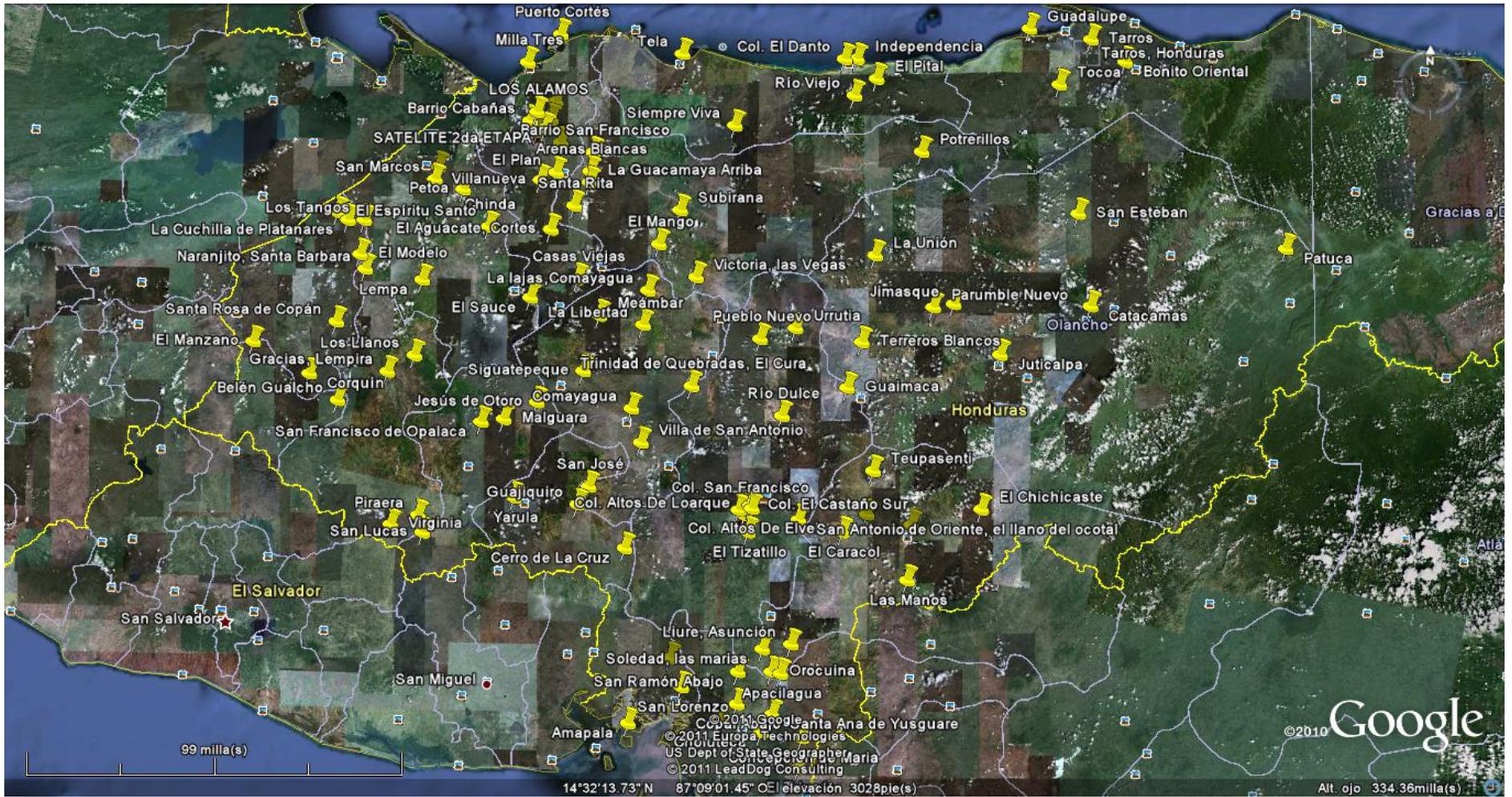
Table C
Average Number of Children and Adults per Household, Level of Urbanization
-Honduras-

Measure	L2L(unweighted)	L2L(weighted)	Census (2001)	Gallup World Poll (unweighted)	Gallup World Poll (weighted)
Number of children under 15 (mean)	2.16	2.16	2,11	1.87	1.96
Number of adults 15 plus (mean)	2.88	2.85	2.91	3.34	3.82
Urban	36.7%	33.3%	46.0%	36.8%	36.8%
Rural	63.6%	66.7%	54.0%	63.2%	63.2%

Honduras L2L Sample Geographic Dispersion

Another criterion for evaluating the performance of the Honduras L2L baseline sample is its geographic coverage. The map on the next page depicts the dispersion of the Honduras L2L Secondary Sampling Units (SSU's), showing the sample covered most parts of Honduras' territory, with an evident concentration of sampling units in the northwestern and southern areas of the countries, consistent with the country's population density.

Map Secondary Sampling Units selected in Honduras as part of L2L Survey



World Bank L2L Survey of Peru

Similarly to the Honduras survey, the L2L Face to Face survey of Peru was based on a nationally representative sample of 1,500 households. At each selected household, interviewers asked to speak to the person recognized by the family as the head of the household. If the head of the households could not be located after two attempts, interviewers proceeded to interview informants aged 15 years or older who were permanent residents of the household.

Peru L2L Sample Design

The sampling design for the L2L survey of Peru was guided by the following criteria:

1. Nationally representative, including both urban and rural areas;
2. Use of the most recent National Household Survey's (ENAHO) sampling frame available (fourth quarter 2010), down to the Secondary Sampling Unit (SSU) level;
3. Oversample households close to the poverty line. For the purposes of this project, "close to poverty line" was defined as 40 percent of income distribution that symmetrically bands the national poverty line: 20% above and 20% below. In 2010 in 27% of Peruvian households monthly per capita consumption was below the moderate poverty line (ENAHO 2010)². Consequently, those households with monthly per capita consumption between 7 and 47 percent of the national distribution, were oversampled.

Since the L2L sample was based on the ENAHO sampling frame, a description of the ENAHO sample design and procedure is in order.

The ENAHO sample is selected in three stages in urban and in more densely populated rural areas, and in two stages in less densely populated areas. At the first stage, the selection of the Primary Sampling Units (PSUs) is performed. All PSUs are grouped in 8 strata, defined by size (number of residents). Strata 1 through 5 correspond to urban areas, strata 6 through 8 correspond to rural areas. The selection occurs within department-specific strata, and the probability of selection of PSUs is proportional to the number of households belonging to them.

The second selection stage is performed based on which of the two large categories a PSU belongs to. The first category is comprised of urban PSUs and more populous rural PSUs. The second category consists of less populous rural PSUs. For simplicity, this document refers to the first category as "urban" and the second as "rural". For the urban category, Secondary Sampling Units – conglomerates- are selected from each PSU with the probability proportional to their size, and with implicit stratification. Implicit stratification is based on a number of socio-economic variables. In the rural category PSUs consist of one conglomerate (SSU) only. For these PSUs, the second stage is omitted.

At the third stage, households are randomly selected from SSUs. In rural areas 8 households are randomly drawn from each SSU; in urban areas 6 households per SSU are selected. The ENAHO survey includes a panel component. The SSUs (conglomerates) from which panel households are drawn are randomly selected. In order to be able to compare L2L data to the future rounds of ENAHO at the SSU level, the L2L sample was drawn from panel conglomerates only.

Drawing the L2L sample

There are multiple ways to oversample households 20 percent above and 20 percent below the poverty line. This report describes the implementation of an oversample of PSUs with higher proportions of households of interest. The number of households drawn from each SSU were the same as in ENAHO (6 households per urban SSU and 8 per rural SSU).

(1) PSU selection. The ENAHO 2010 4th trimester panel sample includes 3,782 households. The 1,500 households of the L2L sample correspond to 40 percent of it. Additional stratification (based on the proportion of households from 7th to 47th income percentile). All PSUs were divided into two strata: PSUs where the proportion of households from 7th to 47th percentiles of the income distribution constitute the majority – stratum I, and PSUs where this proportion was less than 50 percent – stratum II. These two strata were treated as separate samples; subsequent steps (2) through (X) are applied to each stratum. There were 181 and 391 PSUs in stratum I and stratum II, respectively.

(2) PSU selection. 60 percent of the L2L sample (900 households) were drawn from stratum I and 40 percent (600 households) from stratum II. As the table below shows, there were approximately 1,346 households in stratum 1 and 2,436 households in stratum II. Consequently, 67%³ of households from stratum I and 25% of households from stratum II were selected. PSUs were drawn with probability proportional to size from each of 8 ENAHO strata.

Table: L2L strata (based on fraction of hh in 7-47 income percentile band)

ENAHO strata	L2L stratum II		L2L stratum I	
	number of conglomerados	number of hh	number of conglomerados	number of hh
1	109	654	0	0
2	115	690	4	24
3	46	276	3	18
4	36	216	18	108
5	40	240	26	156
6	9	72	15	120
7	19	152	79	632
8	17	136	36	288
Total		2,436		1346

(3) SSU selection – urban areas. The number of SSUs (conglomerates) per PSU in urban strata varied highly: from 1 to 38. We drew 940 households⁴, treating all urban PSUs as a single stratum, according to

³ $0.67 = 900/1346$; $0.25 = 600/2,436$

⁴ $940 = 1500 - 560$; 1,500 – L2L sample size, 616 –number of rural households selected in (2)

probability proportional to the size of SSU. Before drawing, we sorted all SSUs according to socio-economic index, as in the ENAHO sampling, to ensure implicit stratification along the socio-economic dimension.

(4) Household selection: 6 households were randomly selected from each urban SSU, and 8 households from each rural SSU.

L2L Strata oversample correction (Weighting)

As stated before, the sample design overrepresented conglomerates where household income was close to the poverty line (20% above or 20% below it) versus all other conglomerates. The oversample needed to be corrected by accounting for the sample selection probability by creating a base weight. The ENAHO data on the size of the conglomerates that made up the 2 L2L strata was used to create targets for the size of the L2L strata.

Post-Stratification Weighting

Using the ENAHO data, Gallup created targets for the distribution of the age of the head of household, and the highest level of education of the head of household. Tables D and E below, show the L2L unweighted and weighted results for these variables, as well as the ENAHO targets and weights.

Table D
Average Age Head of Household
(Peru)

Variable	L2L(unweighted)	L2L(weighted)	ENAHO
Age head of household (mean)	47	50	50
Age head of household (median)	45	49	49

Table E
Education head of household
(Peru)

Variable	L2L (unweighted)	L2L (weighted)	ENAHO	Weights
1 No formal Education	4.3%	6.70%	8.0%	1.55814
2 Initial Education	.8%	1.20%	.0%	1.5
3 Primary/Elementary school incomplete	12.8%	19.30%	21.0%	1.507813
4 Primary/Elementary school complete	15.3%	17.00%	17.2%	1.111111
5 Secondary school incomplete	13.6%	13.50%	13.3%	0.992647
6 Secondary school complete	30.0%	20.10%	20.0%	0.67
7 Superior no university incomplete	3.3%	3.00%	2.9%	0.909091
8 Superior no university complete	8.8%	8.50%	7.3%	0.965909
9 Superior university incomplete	2.6%	2.40%	3.1%	0.923077

10 Superior university complete	6.2%	6.10%	5.8%	0.983871
11 Post grade/ Master	.5%	0.50%	1.5%	1

The L2L data were trimmed so as to balance bias (how close the demographics of the weighted data align to the targets) and efficiency (the size of the design effect from weighting). Then the data were renormalized so that the sum of the weights equals the number of observations.

Demographic Comparison Between Peru L2L and ENAHO

In order to further test the performance of the Peru L2L sample versus the criterion survey (ENAHO), the following demographic variables were compared: a) Gender of the head of the household, b) household age composition, and c) household possession of certain durable goods.

According to ENAHO 77% of the heads of a household in Peru were men. L2L's result for the same variable is quite comparable (81%), as is shown in Table F below.

Household age composition, on the other hand was compared to the GWP results for Peru. Even though L2L did not utilize the same sampling frame as the GWP, the data for this variable came out quite close, as noted in Table F.

Furthermore, as shown in Table G possession of durable goods - a variable often used as a proxy for socio-economic status - shows a remarkable comparability between L2L and ENAHO. Furthermore, the results for these variables suggest that the weighting scheme worked quite well, as weighted L2L results correct the deviations observed in the unweighted results, closely mirroring the ENAHO survey data on the possession of durable goods.

Table F
Gender of Head of Household, Average Number of Children and Adults per Household (Peru)

Variable	L2L(unweighted)	L2L(weighted)	ENAHO	Gallup World Poll (unweighted)	Gallup World Poll (weighted)
Gender head of household	80.5% male	80.6% male	76.6% male	Not available	Not available
Number of children under 15 (mean)	1.35	1.29	Not available	1.39	1.63
Number of adults 15 plus (mean)	3.15	3.12	Not available	3.15	3.70

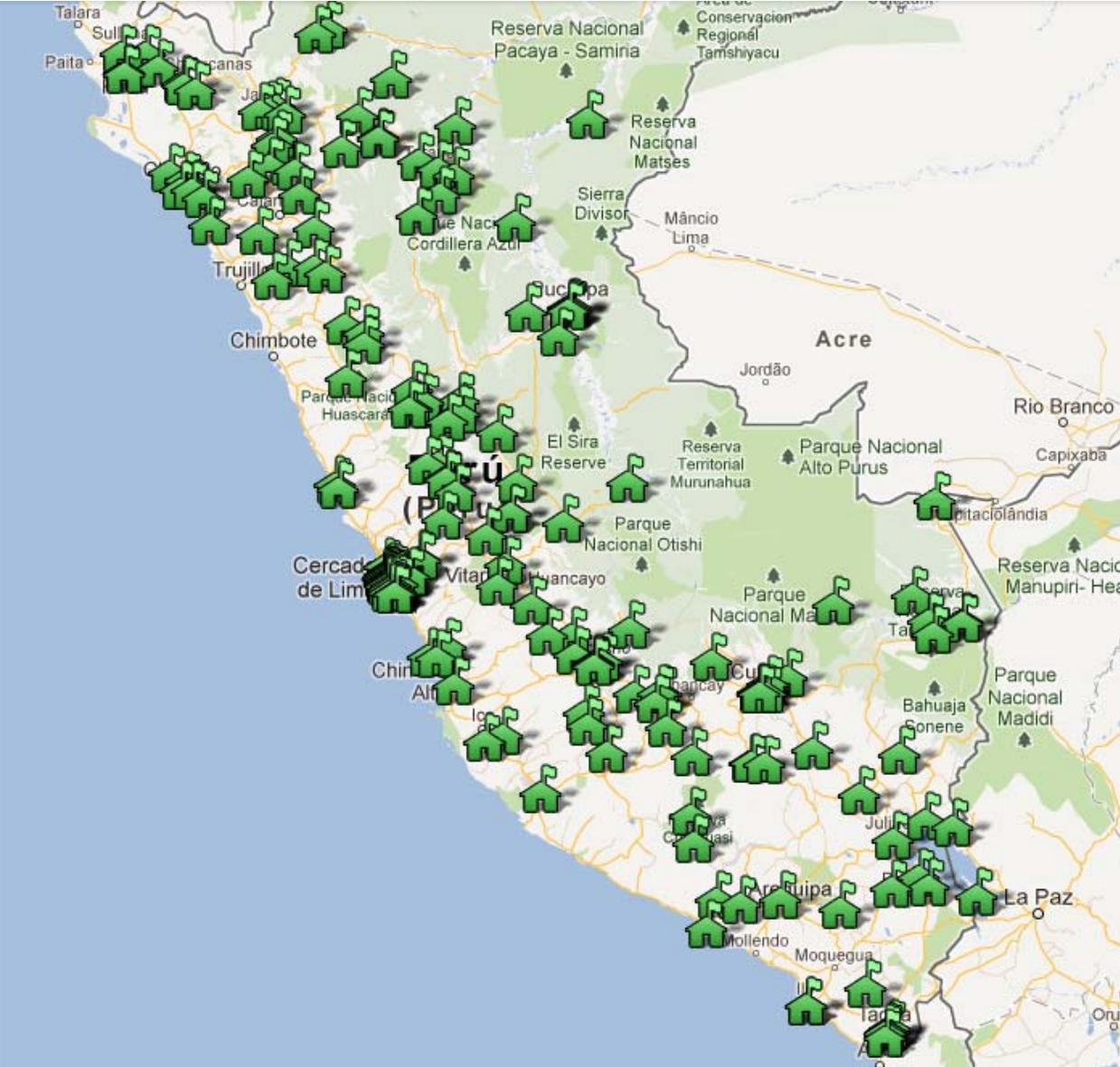
Table G
Possession of Durable Goods
(Peru)

Color TV	L2L (unweighted)	L2L (weighted)	ENAHO
Yes	70.3%	64.8%	66.0%
No	29.7%	35.2%	34.0%
Refrigerator/ freezer	L2L (unweighted)	L2L (weighted)	ENAHO
Yes	41.1%	36.2%	35.4%
No	58.9%	63.8%	64.6%
Gas stove	L2L (unweighted)	L2L (weighted)	ENAHO
Yes	66.5%	60%	64.1%
No	33.5%	40%	35.9%
Car, Van	L2L (unweighted)	L2L (weighted)	ENAHO
Yes	8.1%	7.4%	8.1%
No	91.9%	92.6%	91.9%

Peru L2L Sample Geographic Dispersion

Another criterion for evaluating the performance of the Peru L2L baseline sample is its geographic coverage. The map on the next page depicts the dispersion of the Peru L2L Secondary Sampling Units (SSU's), showing these are scattered throughout Peru's territory, with an evident concentration of sampling units in the central-coastal region, consistent with the country's population density.

Map of Secondary Sampling Units sampled for the L2L F2F survey in Peru



Conclusion

The L2L baseline survey served as the platform for building probability-based household panels for testing the feasibility of the SMS technology as a method for survey research in Peru and Honduras. As such the L2L baseline survey samples had to help overcome the coverage bias commonly encountered in mobile phone sampling frames. In order to accomplish this goal, the L2L surveys leveraged the household based sampling frames utilized by established criterion surveys (ENAH0 in Peru and GWP in Honduras), both of which rely on census information.

The data shown in this report, demonstrates that the L2L baseline surveys performed adequately at generating geographic and demographic estimators comparable to those produced by their respective criterion surveys, as well as census parameters. That is, they generated demographic estimators whose deviations from comparable estimators or parameters could be corrected by means of conventionally accepted post-stratification weighting schemes.

Therefore, the L2L baseline surveys prove to be an ideal platform for building a probability-based panel that overcomes the coverage limitations of the existing mobile phone sampling frames.