

Research Design for the Aceh Reintegration and Livelihood Surveys

Patrick Barron
World Bank

Macartan Humphreys
Columbia University

Laura Paler
Columbia University

Yuhki Tajima
Harvard University

Jeremy Weinstein
Stanford University

1 OVERVIEW

This document describes the research and sampling design for the Aceh Reintegration and Livelihood Surveys (ARLS). This study was funded by the World Bank and designed by a team of researchers from Columbia, Harvard and Stanford Universities. The surveys were implemented in Aceh by the research firm A.C. Nielsen from July-September 2008.

The over-arching goal of these surveys is to assess prospects for peace and reintegration among both civilian and former-combatant populations in Aceh. The immediate goals of the ARLS were twofold. One was to collect individual-level data for an impact evaluation of the World Bank's BRA-KDP project, a post-conflict community-drive development program. The second was to collect livelihood and reintegration data on a representative sample of ex-combatants, and a control group of civilian males. Surveys were conducted in a representative sample of 754 villages throughout Aceh. In sampled villages, four over-lapping surveys were implemented:

- **Long Household Survey (LHS):** Conducted in a representative sample of villages from the 67 rural subdistricts that received BRA-KDP, as well as in a representative sample of villages in 67 matched subdistricts (described in Section 2). Five households were randomly sampled in sampled villages, and main respondents were selected randomly from all males and females between the ages of 18-65 who had lived in the household for at least one month. The LHS is a representative sample of men and women from BRA-KDP treatment and control subdistricts. It is not representative of other subdistricts and is not representative at the district level. (Contained in R_DATA).
- **Short Household Survey (SHS):** Implemented in a representative sample of all subdistricts not included in the LHS. The main goal of this survey is to provide, in conjunction with the LHS, an Aceh-wide representative control group of adult males between the ages of 18-65 for the ex-combatant survey. Male respondents were sampled from two randomly sampled households in selected rural villages and eight randomly sampled households in urban villages. A representative sample of male respondents can be achieved by combining the SHS and the male sub-population of the LHS. (Contained in R_DATA).
- **Ex-TNA Survey:** An Aceh-wide representative sample of ex-combatants. Eligible respondents included anyone who fought with GAM-TNA, or was in the GAM-TNA command structure, for at least one month since 1998. A full list of ex-TNA was enumerated in each of the 754 villages and ex-TNA were sampled with a 6 in 10 probability (Contained in R_DATA).
- **Village Head Survey (VHS):** A survey of village-head characteristics, as well as village-level characteristics in all sampled villages. (Contained in VH_DATA).

- **Household Rosters:** In the LHS, SHS and Ex-TNA surveys, data was collected on every member of the respondent's 1998 and 2008 households. The roster includes demographic, welfare, recruitment and conflict data on all members in the household at those times. (Contained in H_Data).

Table 1: Summary of Sample

	Long Household		Short Household		Ex-TNA	Village Head*
	Treated	Controls	Rural	Urban		
Number of villages	218	245	265	25	320	756
Number of subdistricts	67	69	99	15	153	247
Number of districts/kota	17	17	17	3	19**	20
Number of respondents	1090	1225	531	200	1075	756
Total respondents	2315		731		1075	756

* Two additional VH surveys were conducted

** There were no ex-TNA surveys conducted in Aceh Tenggara

2 SELECTING SUBDISTRICTS FOR LONG AND SHORT SURVEYS

The LHS was designed for an impact evaluation of the BRA-KDP program. It was conducted in 67 BRA-KDP treatment subdistricts and 69 control subdistricts. The World Bank did not randomize assignment to subdistricts; they did, however, use a clear assignment rule. A suitable control group was identified based on the known assignment rule. All villages in treatment and control groups received the long survey, with remaining sampled villages receiving the short survey.

For all rural subdistricts, assignment to BRA-KDP was done on the basis of two variables: *conflict intensity* and *spending capacity*. The World Bank produced a measure of conflict intensity through a factor analysis of nine conflict-related indicators.¹ This produced a continuous measure of conflict intensity at the subdistrict level. The World Bank divided this measure into high, medium and low conflict intensity levels so that each group had approximately the same number of subdistricts.² The second assignment variable, spending capacity, was meant to render ineligible those subdistricts that were not able to handle effectively the inflow of funds from BRA-KDP. Only those subdistricts that had spent at least 60 percent of their 2005 KDP funds at the time of treatment assignment were deemed eligible to participate in the program.

The assignment rule used by the World Bank aimed to treat the most conflict-affected subdistricts in a district by category (with a minimum of one subdistrict per district), *conditional* on a subdistrict meeting the spending capacity criterion. An assignment algorithm can be produced as follows:

f(conflict intensity, spending capacity):

1. Select a target number of subdistricts to treat in each district. The target is determined by the number of high conflict-affected subdistricts in a district.
 - a. If there are no high conflict-affected subdistricts, then set the target equal to the number of medium conflict-affected subdistricts.
 - b. If there are no medium conflict-affected subdistricts, then set target equal to one (low conflict-affected subdistrict).

¹ Indicators included data on the number of conflict victims in each of the three years preceding the end of hostilities; the number of reported clashes between GAM and GoI forces; and perceptions of conflict intensity from survey data.

² This process also created clear cutoffs scores between low and medium conflict intensity at 1.9 and between medium and high conflict intensity subdistricts at 2.5.

2. Rank subdistricts in each district by order of conflict intensity, from high to low. Select subdistricts up to the target, conditional on their meeting the 60 percent spending criterion.
3. If no subdistrict in a district meets the spending criterion, then select the subdistrict with the highest spending capacity.

Implementing this algorithm on the original assignment data correctly classifies 207 of 225 rural subdistricts, or 92 percent of all eligible subdistricts.

The above procedure creates a bivariate indicator of whether a subdistrict should have received treatment (according to the rule) or not. In truth, both the measures of conflict intensity and spending capacity contain noise. The approach used to extract a continuous measure of treatment propensity from the binary indicator of treatment assignment essentially selects controls by ascertaining who would have been selected to receive treatment had the noise resulted in slightly different values on the assignment variables. First, a small shock was administered to each assignment variable, distributed with mean zero and variance equal to $\frac{1}{2}$ that variable's standard deviation. The assignment algorithm was then run 10,000 times, with each run administering a new shock to the assignment variables. This produced 10,000 'treatment assignment' datasets, where, in each dataset, a subdistrict was coded '1' if assigned to treatment given that particular shock to the assignment variables, and '0' otherwise. The number of times a subdistrict was selected was then averaged over the 10,000 datasets to get a continuous measure of 'treatment propensity'. We then calculated for each kecamatan the probability of selection under perturbations and selected the 67 non-treated kecamatan that were 'most likely' to have been treated. This produced a total of 134 subdistricts for the LHS survey.

In sum, the LHS is representative of males and females living in BRA-KDP treatment and control subdistricts. BRA-KDP treatment and control subdistricts are predominantly rural and high capacity. They do represent a range of levels of conflict-affectedness. This is not an easy population to generalize to. The LHS data on its own should not be used to make inferences about other subdistricts or be considered representative at the district level. The SHS survey was conducted in the remainder of rural and urban villages in Aceh. The main goal of the SRS was to complete an Aceh-wide representative sample of civilian males. Using the SHS in conjunction with the male subpopulation of the LHS provides an Aceh-wide representative sample of civilian males to be used as a control group for the ex-TNA survey.

3 SAMPLING

3.1 Strata and Clusters/PSUs

To achieve geographic and population representativeness, both subdistrict and population categories were used as strata for all surveys. Within strata, villages are the primary sampling unit (psu) and were sampled with equal probability.

Three sources of data were consulted to achieve the most current population and subdistrict information for Aceh. The main data source comes from the World Bank Conflict and Development Team in Aceh, and was used in assigning the BRA-KDP program.³ Since this data only covered rural areas, this data was merged with 2005 BPS census data on urban populations. These figures were then checked against a third data source, the 2007 RMU data. When comparing the data sources, significant inconsistencies in several districts were found (see Table 3).

³ As an island that largely escaped the conflict, Sabang was not included in the national sample.

To determine which data source was most accurate, villages with identical population data in both WB/BPS and RMU datasets were identified. There were two types of inconsistencies in the remaining data: 1) differences in population numbers for the same village, and 2) villages and/or subdistricts that appeared in one dataset but not the other. We identified all villages with population ≥ 700 in either the WB/BPS or RMU datasets for which there were major inconsistencies between the data sources.⁴ Focusing on large villages was especially important because the potential for sampling error is greater in large population centers. For each village, the World Bank checked data by consulting the appropriate data collection institution or local officials. In most cases, the WB/BPS data was deemed more accurate than the RMU data in rural areas, while the RMU data was deemed more accurate in urban areas.⁵

Given these checks:

- WB/BPS data is used as the master population data.
- RMU data is used for all rural villages not in the WB/BPS data (174 villages) and for the few instances in which consistency checks proved the RMU data was more accurate than World Bank data (e.g. for several kecamatan in Bireuen).⁶
- RMU data is used for all urban areas (kotas).

The final dataset contains 6,202 villages in 248 kecamatan.

A target of 750 villages (that is, slightly less than one in eight or $6202/8$) was set for sampling. In addition, a floor of at least one village per kecamatan was set, bringing the total target to 754. Sub-districts and population categories were used as strata with 1 in 8 villages from each strata randomly selected with equal probability (independent of village size) for enumeration. The size stratum was set in accordance with World Bank population categories: villages with fewer than 300 people were designated as *small*, villages with 300-700 people as *medium*, and villages with more than 700 as *large*.

Since strata were small, integer issues were important for sampling. Allocations were made as follows: Let n_{ij} denote the number of villages in kecamatan i of population category j . The target number of villages to be selected is given by $m_{ij}=n_{ij}/8$. Let $n_{ij}^*=\text{floor}(m)$. Then with probability $m_{ij} - n_{ij}^*$, n_{ij}^*+1 villages were selected. With probability $n_{ij}^*+1 - m_{ij}$, we selected n_{ij}^* villages. The expected number of villages selected is thus $(m_{ij} - n_{ij}^*) \times (n_{ij}^* + 1) + (n_{ij}^* + 1 - m_{ij}) \times n_{ij}^* = m_{ij}$. In each subdistrict the actual target differs from the expected number by no more than one.

3.2 Sampling Households for the LHS and SHS

Following the selection of PSUs, households were treated as SSUs and were sampled with equal probability in selected villages. For the LHS, five households were sampled using a simple SRS. For the SHS, two households were sampled in rural areas and eight households in urban areas. Three methods were used for selecting households, based on the availability of information on households and the village population.

(1) For all villages where there was a complete list of all households updated within the past year.

Systematic random sampling was used to select households. All households in the village were counted (n) and assigned a number. Let m equal the target number of households in the village with $m=5$ for LHS, $m=2$ for SHS rural and $m=8$ for SHS urban. The sampling interval is then given by $k=n/m$.

⁴ Specifically, we queried the 94 villages outside of the 95 percent confidence interval of the log of the WB population data and the RMU population data.

⁵ Sabang was not included in this study.

⁶ Changes were made at the kecamatan level, rather than the village level. For instance, if we had asked for clarification on two villages in a kecamatan, and the WB confirmed the RMU data was more accurate, we used the RMU data for all villages in that kecamatan.

A starting household, h , was selected randomly choosing a number between 1 and k . Within each village, $h, h+k, \dots, h+(m-1)k$ were selected.

(2) Villages where a) there is no complete household list and b) village population < 300. Systematic random sampling was used to select households. Unlike in (1), enumerators did their own full enumeration of all households in the village. They followed the steps in (1) for sampling households.

(3) Villages where a) there is no complete household list updated within the last year and b) village population ≥ 300 . Because of the size of these villages, enumerators first randomly sampled with equal probability two dusun, which are formal subunits within villages. Enumerators then obtained or made a complete list of all households in each selected dusun. If no list was available, the teams used a variant of the compass method. This involved selecting a geographic central point in the dusun and randomly selecting a direction. Enumerators then walked in that direction to the edge of the dusun, making a complete list of households along the way. For LHS and urban SHS, the household lists were combined and households were sampled using the method described in (1) above. For rural SHS, one household in each dusun was randomly sampled.

If enumerators failed to make contact with the sampled households, then they attempted to survey these households two more times. If they failed to contact the household, then they selected the neighboring household on the right.

3.3 Respondent Selection

Respondents for household surveys were randomly sampled from all eligible household members as follows.

- **LHS:** All male and female household members between the ages of 18-65 who had lived in the household for at least one month.
- **SHS:** All male household members between the ages of 18-65 who had lived in the household for at least one month.

Upon arriving at the household, enumerators first made a complete list of all eligible respondents. They then used a random number and a kish grid to select the respondent within the household. Two kish grids were used and randomized across surveys to balance the probability of selection across all surveys. If the first selected respondent was not present, enumerators made three attempts to return. If that individual was not present on the third attempt, the next person on the list was selected as a replacement respondent.

3.4 Ex-Combatant Survey

Ex-GAM (TNA) combatants were interviewed in all 754 sampled villages, as long as former combatants were present in that village. For the purpose of the study, a former-combatant was defined as anyone who fought with GAM-TNA or was in the military command structure for at least one month since 1998. Across all sampled villages with ex-combatants, six out of every ten (6/10) ex-combatants would be surveyed in each village, with a floor of one per village.

Within each selected village an **exhaustive** list of all ex-combatants was compiled through consultations with village leaders and GAM/KPA representatives. The total number of ex-combatants in every village was recorded. Each ex-combatant on the list was then assigned a number between 1 and n . A second number z between 1 and 10 was then selected randomly. Enumerators referred to a kish grid that ensured a 6/10 probability of selection to determine how many and which individuals were to be interviewed.

3.5 Sampling Probabilities for the LHS and SHS

Within a given stratum, the probability of selection for an individual respondent in the LHS and SHS depends on three elements: 1) The probability that village k is selected, based on proportional probability sampling in each strata; 2) the probability that household j (given village k is selected), with households sampled according to a fixed number; and 3) the probability that individual i (given household j in village k selected), where one individual per household was selected. The probability of selection for an individual i in household j in village k can therefore be represented as:

$$\Pr(ijk) = \Pr(k) \times \Pr(j|k) \times \Pr(i|jk)$$

In the survey, $\Pr(k)$ was set to $1/8$. $\Pr(j|k) = 5/n$ for LHS, $2/n$ for SHS rural, and $8/n$ for SHS urban (with adjustment to account for deviations in the actual number of surveys conducted in LHS and SHS areas) where n is the estimated number of households in the village (*desa*) or urban neighborhood (*kelurahan*). $\Pr(i|jk)$ is $1/(\text{eligible respondents in household for each survey-type})$.

For household level probabilities, the probability of selection is simply:

$$\Pr(jk) = \Pr(k) \times \Pr(j|k)$$

For ex-TNA probabilities, the probability of selection is:

$$\Pr(ik) = \Pr(k) \times \Pr(i|k)$$

Where $\Pr(k)$ still equals $1/8$ and $\Pr(i|k)$ is set equal to $6/10$.

The sampling weight for an individual is the reciprocal of the probability that the individual is selected to be in the sample.

4 REPLACEMENT PSUs

In some cases, surveys were not able to be conducted in originally sampled villages. In general, two types of problems were encountered:

1. Villages did not exist or existed in different subdistricts.
2. Ex-TNA surveys were forbidden in some districts.

If enumerators entered a village and household surveys could not be completed, the procedure involved randomly selecting a replacement village from within the same strata. Table 4 presents the list of randomly sampled replacement villages. In cases where the ex-TNA survey has forbidden, enumerators generally were able to complete village head and LH or SH surveys.

Table 5 presents the list of villages where ex-TNA surveys could not be completed.

5 TABLES

Table 2: Survey modules in each survey

Module #	Module	Long Household Survey	Short Household Survey	Village Head Survey	GAM Survey
I	Survey Identifier Information	✓	✓	✓	✓
II	Household Roster	✓	✓		✓
III	Household Wealth	✓	✓	✓	✓
IV	Individual Level Behavior And Attitudes	✓	✓		✓
V	Collective Action	✓			
VI	Project Perceptions	✓			
VII	Recruitment [March 1998 – August 2005]	✓	✓		✓
VIII	TNA Module				✓
	Village Head Modules			✓	
	Provides data for:	BRA-KDP evaluation	GAM study	BRA-KDP/GAM study	GAM study

Table 3: Comparison of World Bank/BPS and RMU Population Data

Kabupaten	World Bank Data		2007 RMU data		Compared	
	# of villages	Total pop	# of villages	Total Pop	RMU-WB Villages	<u>RMU-WB Population</u>
ACEH BARAT	274	98740	321	185477	47	86737
ACEH BARAT DAYA	132	116799	132	116799	0	0
ACEH BESAR	604	297614	610	298693	6	1079
ACEH JAYA	126	64614	176	93531	50	28917
ACEH SELATAN	256	211469	258	213648	2	2179
ACEH SINGKIL	192	158803	192	154828	0	-3975
ACEH TAMIANG	213	242542	212	244614	-1	2072
ACEH TENGAH	269	176123	276	176196	7	73
ACEH TENGGARA	236	144334	236	144334	0	0
ACEH TIMUR	497	341915	504	367998	7	26083
ACEH UTARA	832	498017	832	497800	0	-217
BENER MERIAH	232	118778	232	119078	0	300
BIREUEN	576	334757	561	364511	-15	29754
GAYO LUES	129	98024	129	98024	0	0
KOTA BANDA ACEH	86	176881	90	49605	4	-127276
KOTA LANGSA	51	137586	51	28739	0	-108847
KOTA LHOKEUMAWE	67	162423	28	74273	-39	-88150
NAGAN RAYA	169	97005	223	54025	54	-42980
PIDIE	952	507558	952	492421	0	-15137
SIMEULUE	135	80139	138	73861	3	-6278
Grand Total	6028	4064121	6153	3848455	125	-215666

Table 4: Replacement Villages

No	Original Locations			Replacement Locations			Reason
	District	Subdistrict	Village	District	Subdistrict	Village	
1	Aceh Jaya	Teunom	Blang Ramee	Pidie	Mila	Daya Andreue	Blang Ramee was a restraintment village. Sampled Alue Jang as a replacement, but survey teams had already left the area. We then sampled from an under-represented subdistrict (due to integer issues) in an uncompleted area. Within Pidie, population stratum 19161 has 15 elements of which only 1 had been sampled for a sampling probability of 1/15, well below average. Dayah Andeue was randomly drawn from this stratum, bringing the sampling probability to 2/15, only moderately above average.
2	Aceh Utara	Lhoksukon	Kreung Mbang	Aceh Utara	Lhoksukon	Cot Asan	Krueng Mbang does not exist. Replaced with Cot Asan in same strata.
3	Aceh Selatan	Kluet Selatan	Paya Dapur	Aceh Selatan	Kluet Selatan	Gelumbuk	Paya Dapur in Kluet Timur, not Kluet Selatan. Replaced Paya Dapur with Gelumbuk; ACN completed surveys in both.
4	Aceh Tenggara	Bambel	Trt. Megaral Pasir	Aceh Tenggara	Bambel	Trt Megara Pasaran	Trt Megaral Pasir misspelled in original dataset.
5	Aceh Tenggara	Babul Makmur	Gunung Pak Pak	Aceh Tenggara	Leuser	Gunung Pak Pak	Leuser is a new subdistrict.
6	Gayo Lues	Rikit Gaib	Kenyaran	Gayo Lues	Pantan Cuaca	Kenyaran	Kenyaran moved to Pantan Cuaca after BRA-KDP.
7	Gayo Lues	Blang Peugayon	Kampung Jawa	Gayo Lues	Blang Kejeren	Kampung Jawa	Note: Kept kecamatan code for Blang Peugayon
8	Gayo Lues	Teripe Jaya	Persada Tongra	Gayo Lues	Terangon	Persada Tongra	Kecamatan split
9	Pidie	Bandar Dua	Blang Kuta	Pidie	Bandar Dua	Paya Pisang Klat	VH forbade all interviews in Blang Kuta. Replaced with Paya Pisang Klat.
9	Pidie	Titeu/Kemala	Daya Meunara	Pidie	Titeu	Daya Meunara	Kecamatan split
10	Pidie	Titeu/Kemala	Paloh Teungoh	Pidie	Keumala	Paloh Teungoh	Kecamatan split
11	Pidie	Titeu/Kemala	Pulo Cahi	Pidie	Keumala	Pulo Cahi	Kecamatan split
12	Simeulue	Simeulue Timur	Kahad	Simeulue	Simeulue Timur	Lugu	VH refused surveys in Kahad. No surveys conducted in replacement either.

Table 5: Places where Ex-Combatant Survey could not be Completed

No	District	Subdistrict	Village	Replacement Village?	Surveys Completed	Surveys Not Completed	Reason
1	Aceh Timur	Banda Alam	Seuneubok Bayu	N	HH and VH	TNA	VH not cooperative.
2	Aceh Timur	Banda Alam	Seuneubok Kandang	N	HH and VH	TNA	VH not cooperative.
3	Aceh Timur	Bireum Bayeun	Paya Peulawi	N	HH and VH	TNA	Require KPA letter.
4	Aceh Timur	Rantau Selamat	Bayeun	N	HH and VH	TNA	VH not cooperative.
5	Aceh Timur	Rantau Selamat	Damar Siput	N	HH and VH	TNA	VH not cooperative.
6	Aceh Timur	Sungai Raya	Geulumpang Payong	N	HH and VH	TNA	VH not cooperative.
7	Bener Meriah	Bandar	Bukit Wih Ilang	N	HH and VH	TNA	VH not cooperative.
8	Bener Meriah	Bandar	Janarata	N	HH and VH	TNA	VH not cooperative.
9	Bireuen	Gandapura	Geureugok	N	HH and VH	TNA	VH not cooperative.
10	Bireuen	Gandapura	Lingka Kuta	N	HH and VH	TNA	VH not cooperative.
11	Bireuen	Gandapura	Pulo Gisa	N	HH and VH	TNA	Require KPA letter.
12	Bireuen	Gandapura	Samuti Aman	N	HH and VH	TNA	Require KPA letter.
13	Bireuen	Jeumpa	Blang Seunong	N	HH and VH	TNA	Require KPA letter.
14	Bireuen	Jeumpa	Paloh Seulimang	N	HH and VH	TNA	Require KPA letter.
15	Pidie	Bandar Baru	Mns. Blang Sukon	N	HH and VH	TNA	Panglima Sagoe forbade interview
16	Pidie	Bandar Dua	Paya Pisang Klat	Y	HH and VH	TNA	VH not cooperative, despite Camat letter
17	Pidie	Geulumpang Tiga	Meunje	N	HH and VH	TNA	Panglima Sagoe forbade interview
18	Pidie	Muara Tiga	Sagoe	N	HH and VH	TNA	Panglima Sagoe forbade interview
19	Pidie	Mutiara Timur	Paloh Baro	Y	HH and VH	TNA	Conflict between VH and ex-TNA in Mon Ara led to sampling Paloh Baro as a replacement.
20	Pidie	Titeu	Pulo Loih	Y	HH and VH	TNA	Kecamatan split.
21	Simeulue	Simeulue Timur	Lugu	Y		HH, VH, TNA	No surveys conducted