

# The Use of Random Geographic Cluster Sampling to Survey Pastoralists

*Kristen Himelein  
Stephanie Eckman  
Siobhan Murray*

The World Bank  
Development Research Group  
Poverty and Inequality & Computational Tools Teams  
September 2013



## Abstract

Livestock are an important component of rural livelihoods in developing countries, but data about this source of income and wealth are difficult to collect because of the nomadic and semi-nomadic nature of many pastoralist populations. Most household surveys exclude those without permanent dwellings, leading to undercoverage. This study explores the use of a random geographic cluster sample as an alternative to the household-based sample. In this design, points are randomly selected and all eligible respondents

found inside circles drawn around the selected points are interviewed. This approach should eliminate undercoverage of mobile populations. The results of a random geographic cluster sample survey are presented with a total sample size of 784 households to measure livestock ownership in the Afar region of Ethiopia in 2012. The paper explores the data quality of the random geographic cluster sample relative to a recent household survey and discusses the implementation challenges.

---

This paper is a product of the Poverty and Inequality and Computational Tools Teams, Development Research Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The authors may be contacted at [khimelein@worldbank.org](mailto:khimelein@worldbank.org).

*The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.*

## The Use of Random Geographic Cluster Sampling to Survey Pastoralists

Kristen Himelein, Stephanie Eckman, and Siobhan Murray<sup>1</sup>

*Keywords:* GIS, cluster sampling, pastoralists, livestock surveys

*JEL Classifications:* C83, C90

*Sector Boards:* Agriculture and Rural Development (ARD), Poverty Reduction (POV)

---

<sup>1</sup> Author affiliations: Kristen Himelein (AFTPM) and Siobhan Murray (DECRG) of the World Bank, Stephanie Eckman of the Institute for Employment Research (IAB).

Acknowledgements: The authors would like to thank their partners in the Ethiopia Central Statistics Agency, in particular Samia Zekaria, Biratu Yigezu, Habekiristos Beyene, Abate Sidelel, Jemal Ali, Abdulaziz Shifa, and the other CSA staff that supported this project. We would also like to thank Alemayehu Ambel and Jon Kastelic of the World Bank for their facilitation and technical assistance, as well as Sarah Walker, Svenja Wippich, Ruben Bach, Angus Cameron, Mike Brick, Keith Rust, the participants at the 2012 International Conference on Methods for Surveying and Enumerating Hard-to-Reach Populations conference, and four anonymous reviewers for their comments on earlier concept notes and drafts. Finally, we would like to thank Asmelash Haile Tsegay for his critical work on all levels of the project. Funding for this project was provided by the Bill and Melinda Gates Foundation Trust Fund for Improving the Quality and Policy Relevance of Household-Level Data on Agriculture in Sub-Saharan Africa and the Knowledge for Change programs at the World Bank. All views are those of the authors and do not reflect the views of the World Bank or its member countries.

## 1. Introduction

Livestock ownership comprises a large part of rural wealth and well-being in the developing world, serving diverse functions from food source to savings and investment vehicle. The sector, however, has recently come under increasing pressure from a number of sources, including increased demand for meat and dairy products from the expanding middle class, climate change, and loss of traditional pasture land to development. Efforts to understand these evolving dynamics, and their impact on the welfare of livestock-owning households, are hampered by a lack of high quality data on which to base analyses. Beyond the general data collection issues of definition and quantification, livestock statistics present particular challenges due to the nomadic and semi-nomadic nature of many pastoralists.

The most common sample selection methodology for household surveys in the developing world is a multi-stage stratified sample (Grosh and Munoz, 1996). In the first stage, primary sampling units are selected from census enumeration areas. In the second stage, dwellings are selected from a housing unit frame, usually compiled through costly in-field listing. However, semi-nomadic households who are temporarily absent, as well as fully nomadic households without fixed dwellings, are undercovered by this approach. In areas where a large portion of the poor and vulnerable population engages in pastoralist activities, this undercoverage could lead to substantial bias in livestock and welfare estimates.

This paper considers the use of an alternative approach to collecting data from livestock-owning households, Random Geographic Cluster Sampling (RGCS). Similar methods are commonly used by developed world agricultural statistics agencies, such as the United States Department of Agriculture, to measure agricultural production and livestock (USDA, 2010), and have also been used by researchers to study farms in Scotland and livestock in Somalia, South Africa, Thailand, and Laos (Emerson et. al., 1995; Cameron, 1997; Soumare et al 2007; von Hagen, 2002). They are also common in forestry surveys (Husch 1982; Roesch et al 1993). This paper describes a pilot project to test the RGCS methodology in the Afar region of Ethiopia, carried out in collaboration by the World Bank Development Economics Research Group and the Ethiopian Central Statistical Agency (CSA).

In an RGCS design, the study area is stratified using data from Geographic Information Systems (GIS) sources. Within each stratum, points (latitude and longitude) are randomly selected, and then a circular cluster of a given radius is created around the point. All eligible respondents found within this cluster are selected for the survey. The main advantage of this design is that it captures everyone who resides in the selected circles at the time of the interview, including those who do not have a permanent dwelling or who are temporarily away from their dwelling. Properly implemented, this design eliminates undercoverage resulting from mobile populations.

There are other alternative methodologies for measuring livestock ownership which we do not use directly in this study. The CSA used a flyover survey in 2004 to estimate the total number of livestock for areas in the Ethiopian Somali region not covered in the agricultural census due to security concerns. In addition to high costs and difficulties in implementation, flyover surveys do not allow researchers to link livestock to households, which severely constrains the use of the data for socioeconomic analysis. Water point surveys are also common, but are biased as they exclude all livestock not found at a known watering point. Adaptive sampling is an approach often used for wildlife studies that are rare and unevenly distributed. However, such a design requires ongoing and close supervision by a sampling statistician, and often multiple trips to the same area, neither of which was possible in this project (Thompson 1990; Thompson 1991; Thompson and Seber 1996). Other geographic sampling methods use

a grid or hexagon design to eliminate overlap, but are more difficult to implement in the field or would require more expensive GPS hardware (Reams et al 2005).

We developed the RGCS approach to address the shortcomings of the other available data collection methodologies while taking into account the limited technical capacity of the implementing partner. We note some advantages of the RGCS over a traditional household-based survey, but also report the many challenges encountered. Unfortunately, some of the difficulties in implementing the design seem to be due to interviewers' failure to implement the procedures. Though unforeseen challenges, such as natural disasters and ethnic violence, also played a role, it is also possible that the design, which at times required interviewers to cross long distances on foot in very harsh conditions, is not feasible in terms of what it is realistic to require of an interviewer. We conclude with thoughts on the limitations of RGCS specifically in the drylands context but also discuss more generally its potential use in surveys of persons.

## 2. Background on Afar Region

To test the RGCS approach in the field, we carried out a survey in July and August of 2012 in the Afar region of Ethiopia. This region was selected for the pilot project for a number of reasons. First, the CSA had conducted an agricultural and livestock household survey, the Ethiopia Rural Socioeconomic Survey (ERSS), six months prior to the implementation of the RGCS field work. In Afar the ERSS included a module on pastoralist issues. We had therefore expected to be able to use the ERSS data as a point of comparison for our RGCS results. Unfortunately, we have concerns about the ERSS data as a benchmark, as discussed below.

The second factor in our choice of Afar for this project was the high-quality existing GIS infrastructure at the CSA compared to other potential study areas. The CSA has compiled GIS data layers for the entire country and has several trained staff members. The agency also maintains a stock of GPS devices suitable for the specialized fieldwork. We expected that the CSA's previous experience with the technology used during planning, sample selection, and data collection would be beneficial to the project outcomes.

Third, the Afar region also offered geographic advantages over other pastoralist areas in the region. Afar covers a land area of approximately 72,000 square kilometers located in the north of the country, and is relatively isolated. It is bordered to the north by Eritrea, to the east by Djibouti and the Ethiopian state of Somali, to the south by the state of Oromia, and to the west by the states of Tigray and Amhara. In addition to being a well-guarded national frontier, the northern border retains many landmines placed during the Eritrean war for independence. To the west, the country of Djibouti and the Ethiopian state of Somali are ethnically distinct, and have a history of livestock raiding and armed conflict with the Afar people. The states of Oromia, Tigray, and Amhara, on the western border, are largely agricultural and traditionally hostile to the Afar. The Tigray and Amhara borders are also mountainous, further limiting contact. For these reasons, the migration patterns of the Afar people are not likely to take them outside of the Afar region, which simplifies comparability between the RGCS and ERSS data sources.

The Afar region is divided into five administrative zones. The companion ERSS survey covered only Zones 1 and 3. As the RGCS survey was designed to make comparisons to this survey, these two zones were taken as a basis for the new approach. However, since seasonal migration patterns take regular residents of Zones 1 and 3 into Zones 4 and 5, these two zones were also included. (See Figure 1 for a map of the usual migration patterns.) Zone 2 in the far north of the region, is excluded from both

surveys due to extreme weather conditions, recent violence against Western nationals, and its self-contained migration patterns. According to the ERSS, 55 percent of respondent households in the Afar region that own livestock indicated that they had taken their livestock outside of the village to graze for at least one night during the previous season, and 41 percent indicated that they had similar plans for the upcoming dry season. Most respondents (56 percent) made only one trip in the previous year, with an additional 24 percent making two trips. Of those making trips with their livestock, less than one percent travelled outside of Afar and no one reported migrating to Zone 2. These results also support our choice of Afar for this pilot project. See figure 1 in the appendix for further details on traditional migration routes.

### 3. Study Design

#### 3.1 Stratification

We divided Afar into five strata before selecting points to improve the statistical and cost efficiency of the project. The five strata were defined by the expected likelihood of finding herders and livestock, based on an assumption that herds congregate around limited water sources and available pasture in the driest part of the year. Spatial datasets describing land cover, land use, and other geographic features were used as input to delineate five discrete, mutually exclusive strata.

The first stratum consisted of land in or near towns, defined by population density measures in the AfriPop dataset (Tatem, 2010). The second stratum consisted of permanent agriculture, under the assumption that livestock would be largely excluded from these areas. Boundaries were defined based on the interpretation of five meter resolution SPOT Imagery from 2006 from the CSA's Land Cover Mapping project, and included commercial agriculture as well as some small individual farms. Area placed in the first two strata was then excluded from remaining strata definitions.

The third stratum consisted of land within two kilometers of a major water source, including the Awash River and its permanent tributaries, and which also met criteria for pasture based on the average annual mean and range of the long-term normalized difference vegetation index (USGS Earth Resources Observation and Science Center, 2012). This stratum was considered to be the most likely to contain livestock. The fourth stratum consisted of land between two and ten kilometers from a major water source which met criteria of pasture land. The remainder of the land was placed into the lowest probability stratum. See Figure 2 in the appendix for a map of the five strata.

A total of 125 points were selected from these five strata for the survey. The total number of points selected and the allocation between strata was based on sample size calculations from the previously collected data from the 2008/2009 Agricultural Sample Survey, the expected number of households to be found and interviewed in each stratum based on the results of the pretests, and the available budget for the pilot project. The number of selected points was higher in the strata where we expected the highest concentrations of potentially nomadic households and livestock (stratum 3), and lower in areas of lower expected density (stratum 5). Points were selected in areas with low-likelihood of finding pastoralists, towns and settled agricultural areas, because excluding these areas would bias the total livestock populations. The radii for the circles also varied across the strata. In areas where we expected higher densities, we drew smaller circles to keep the workload reasonable. In areas where we expected few or no livestock, we expanded the circle radius to the largest feasible dimensions to maximize the probability of finding animals. See Table 1 for the definition, sample size, and radius used in each of the five strata.

Table 1: Stratification of Afar Region

Stratum	Description	Radius (km)	Points Selected	Total area (km <sup>2</sup> )	Percent of total landscape
1	High likelihood: towns	0.1	10	33	<1%
2	Almost no possibility: settled agricultural areas / commercial farms	0.5	15	930	2%
3	High likelihood: within 2 km of major river or swamps	1	60	3,538	6%
4	Medium likelihood: within 10 km of major river or swamps	2	30	6,921	12%
5	Low likelihood: all land not in another stratum	5	10	45,152	80%
Total			125	56,574 <sup>a</sup>	100%

<sup>a</sup> The total area in the table does not match the total area of Afar due to exclusion of Zone 2 from our study.

### 3.2 Survey Implementation

To develop the framework protocols for the RGCS approach in Afar, two pretests were conducted, the first in December 2011 and the second in June 2012. The first focused on equipment and field practices, and qualitative research into seasonal migration patterns. The second finalized the protocols and tested the survey instrument.

The resulting methodology was designed to be relatively straightforward to implement in a low-capacity field environment. Each interviewer was given a GPS device to which the selected points, and the circles around them, had been preloaded. In addition to the usual zoom and pan features, the device always displayed where the interviewer was in relation to selected area and was set to sound an alarm when the interviewer entered the circle. The interviewer teams were to drive as close as possible to the circle and then travel the rest of the way on foot, if necessary. Figure 3 shows an example of a point and circle. The selected circle has a radius of one kilometer and includes both land and water (on the eastern edge). To assist in locating the area, each interviewer was also provided with printed maps like that shown in Figure 3 in the appendix.

Once inside the circle, the team was assigned to canvas the area and interview all livestock holding households. The device recorded the interviewer's path of travel within the circle so that he could navigate back to the starting point.

When a team member encountered a household (or a group of people travelling together) inside the circle, they attempted to complete three questionnaires. The first was a household roster, completed with a household informant, which captured basic demographic information about each member of the household, such as name, age, schooling, and health information. The second questionnaire, also for the household informant, gathered data about the goats, cattle, and camels currently travelling with the household (those away for a day to graze were included in this roster). This questionnaire also asked who owned the livestock and whether the owner was currently travelling or staying with the group. The third questionnaire was administered to each individual livestock holder in the household and contained more information about the animals in his or her possession. Following the completion of each selected circle, the supervisor filled out a cover sheet indicating how many persons or households were found in the circle.

A one-week training for supervisors and enumerators was conducted in the city of Awash in southern Afar in early July 2012. A total of 22 field workers, five supervisors, one field coordinator, and one CSA branch head participated. All participants were recruited by the CSA and some had prior survey experience. The training stressed questionnaire administration, sampling protocols, safety, instruction with the handheld GPS devices, and the use of field guides. Data collection took place from July 10 to August 9, 2012. Interviewers worked in teams consisting of four interviewers and one supervisor.

The use of local field guides was strongly encouraged in this study. During pretesting, we found the most helpful available guides were young men from the local area with extensive knowledge of the terrain and the people living there. The data collection budget included funds to hire such guides whenever necessary. The guides played essential roles in determining the best route from road and river access points to the circle boundary, and acting as intermediaries between the government data collection teams and a suspicious and occasionally hostile local population.

All questionnaires were administered on paper and were provided in Amharic. The interviewers used local translators and the local guides to translate the questionnaire into Afar when necessary. The interview lasted on average 20 minutes per household, though there was substantial variation based on the household size and livestock holdings. At the end of fieldwork, all of the household and holder questionnaires, as well as the supervisor questionnaires, were returned to the CSA headquarters in Addis Ababa where data entry took place.

The fieldwork was facilitated by a survey coordinator who participated in the second pilot, conducted the training, and performed selected field visits. The survey coordinator was contracted independently of CSA, and had extensive experience with primary data collection projects in Ethiopia. The survey coordinator visited the teams throughout the course of fieldwork, accompanying each team to between three and five circles. His visits were not randomized, though he attempted to cover the distribution of teams, zones, and strata.

As the study area encompasses some of the harshest terrain in the region, and the methodology was novel both for the research and implementation teams, a number of unexpected difficulties were encountered. First, the timing of the fieldwork, which was originally designed to coincide with the dry period, unfortunately fell during the annual Ramadan fasting period. As most field guides and respondents were observant Muslims, they were reluctant to participate in activities during daylight hours. Second, the seasonal rains started earlier than had been expected, which created access problems such as flooding of roads and land bordering the rivers. The access issues necessitated longer walks for enumerators, including one incident where a team had to walk 15 km to reach the selected site. Other obstacles, such as national park boundaries, active volcanoes, and militarized areas further restricted access. Third, ongoing strained relations between local communities and the national government led to a few isolated security incidents, including minor assaults on drivers and fieldworkers, and the (brief) kidnapping of the survey coordinator. Team supervisors repeatedly cited these challenges to explain their lack of progress in completing assigned field tasks

### 3.3 Weighting

The probabilities of selection for such a design are in principle rather straightforward. Setting aside the issue of stratification for a moment, say we select  $c$  points and draw an  $r$  kilometer radius around each one, selecting all households that fall within the circles. To get the probabilities of selection of a given household  $i$ , we invert our reasoning and consider the set of all points such that, if any of those points were selected, household  $i$  would be interviewed (see Roesch et al, 1993 and Thompson

and Seber, 1996, p.108 for a similar approach). Call this set  $A_i$ . For most households,  $A_i$  is simply a circle with radius  $r$  centered at household  $i$ . (For households near the boundary of the study region, the circle may be cut off a bit, but we ignore this issue for the moment.) Then the probability of selection of household  $i$  is one minus the probability that no point in the area surrounding that household is ever selected, across all  $c$  selections (based on Särndal et al 1992, p.50).

$$\pi_i = 1 - \left(1 - \frac{\pi r^2}{\text{total area}}\right)^c$$

However, due to the stratification used in this study, the probabilities of selection of the interviewed households are more complex. Because the strata are quite co-mingled (see Figure 2) a circle drawn around a point selected in one stratum could extend outside of the boundaries of that stratum and include land in another stratum. For example, consider a household that lies in stratum 2 near the boundary of strata 1 and 2, as shown in the left panel of Figure 4. Household  $x$  can be selected if points inside stratum 2 are selected but also if points inside stratum 1 are selected. In terms of the notation developed above, the selection region for household  $x$ ,  $A_x$ , contains land in both stratum 1 and stratum 2: this region is indicated with shading in the right panel. See figure 4 in the appendix for an example.

This issue with stratum boundaries is not trivial. In our study, all land in the town stratum (stratum 1) is within five kilometers of the low probability stratum (stratum 5) and thus was also selectable from that stratum. In fact, more than 90 percent of all land area in strata 1, 2, 3, and 4 falls within the selectable range of points in stratum 5, due to the very large radius of stratum 5. Thus, even though each household itself lies in only one stratum, many households were selectable from more than one stratum. The probability of selection of a household  $i$ ,  $\pi_i$ , is equal to the probability that the points in  $A_i$  that lie in stratum 1 were selected, plus the probability that points in  $A_i$  that lie in stratum 2 were selected, and so on for the  $H$  strata. Define  $\pi_{i,h}$  as the probability that household  $i$  is selected from stratum  $h$ . The overall probability of selection of household  $i$  is then:

$$\pi_i = \sum_{h=1}^H \pi_{i,h} + \sum_{j=2}^H (-1)^{j+1} \left[ \sum_{h_1 < h_2 < \dots < h_j} \prod_{h_1}^{h_j} \pi_{i,h} \right] + (-1)^{H+1} \prod_{h=1}^H \pi_{i,h} \quad (1)$$

where the terms after the first adjust for overlapping probabilities. However, because most households are selectable from only one or two strata, many of these terms are zero and the probabilities simplify a good deal.

Let  $S_h$  be the land within stratum  $h$ , and let  $A_i \cup S_h$  be the land in the selection region of household  $i$  that lies within stratum  $h$ . Let  $|S_h|$  and  $|A_i \cup S_h|$  be the areas of these two sets of land. Then the constituent terms in equation 1, the probability that household  $i$  was selected from within stratum  $h$ , are each:

$$\pi_{i,h} = 1 - \left(1 - \frac{|A_i \cap S_h|}{|S_h|}\right)^{c_h}$$

that is, one minus the probability that none of the land in stratum  $h$  that is within the selectable range of household  $i$  is selected, across all  $c_h$  selections in stratum  $h$ . The GIS tools allow us to calculate the areas

of  $A_i$  and  $S_h$  precisely. Using the actual areas, rather than the areas of the circles with radius  $r_h$ , addresses the issue of lower probabilities of selection for households near the boundary of the study area (see Barrett 1964 for a discussion of “edge effect bias”). (It is also possible to conceptualize the sampling technique used in this study as a form of indirect sampling. Such an approach would also lead to appropriate probabilities of selection and weights (Lavallée 2007). We have chosen not to take such an approach here because of the difficulty of dividing the study area into slices of land which lead to selection of unique sets of households. See Roesch et. al. (1993) for such an approach in the context of RGCS.)

Although 125 circles were selected, only 102 were visited by interviewing teams, as discussed below. For the  $c_h$  values, we use the number of visited circles rather than the number of selected circles, under the assumption that the circles within a stratum that the teams did not visit are missing at random. The initial weight for each household is then the reciprocal of its overall probability of selection:  $w1_i = \pi_i^{-1}$ .

We make one adjustment to this weight, for the unobserved portions of the selected circles. The fieldwork protocol stipulated that interviewer teams should systematically observe the entire circle, however, this was not always possible due to the challenges discussed above (and possibly also due to low effort by the interviewing teams, which is discussed in more detail below). We calculate an alternative set of weights that adjusts for the portion of each circle that was not observed. The GIS technique of Viewshed analysis uses the tracks recorded by the GPS devices as the interviewers traveled within the circle, along with an altitude map, to determine what the interviewers were able to observe, that is, the area that was in their line of sight as they travelled around the circle. Figure 5 in the appendix shows an example of a map produced by the Viewshed analysis. The white tracks are the paths taken by the interviewing team members in circle 134 and the land within the circle that they could observe from those paths. We see that although the interviewers walked only a small portion of the circle, they were able to observe the majority of the area, 72.4 percent in this case. Across all of the visited circles, the observed coverage percentages range from 14.1 to 99.0, with a mean of 84.0.

The multiplicative weight adjustment is the reciprocal of the percent observed in each circle. If we believe that there are households within the unobserved portions of the selected circles, and that these households are similar to those interviewed in the observed portion, then the adjusted weight is appropriate and improves estimates. If, alternatively, we believe that the areas that were not observed were missed because they could not possibly contain any livestock, due, for example, to flood water or vegetation too thick to traverse, the adjustment to the weights is not necessary. We use both the unadjusted and the adjusted weights in the results section. No further adjustment to the weights for household nonresponse was made as the field teams did not report any issues with participation. Household surveys in rural areas of the developing world, and in particular Ethiopia, have historically had high response rates.

The weights require one further caveat. The probabilities of selection on which the weights are based are accurate only if the people and livestock that the survey aims to capture do not move during the study period. If a man and his camels are selectable in more than one circle over the data collection period, then they have more than one probability of selection, which greatly complicates the weighting. To minimize this complication, we constrained the data collection period to one month. We also asked three questions in the survey regarding the past and future movements of the respondents. The first item asks whether the respondent had traveled during the dry season with his livestock outside of the area where the interview took place, the second asks if the livestock had traveled separately outside the area where they were currently, and finally if the respondent planned to travel outside of the current area during the dry season. In approximately six percent of the cases, the respondent had travelled with

their livestock to a different area during the current dry season prior to the survey. In about ten percent of cases, the respondent's livestock had travelled separately to another area previously. Additionally, 11 percent of respondents indicated that they were planning to move with their livestock during the current dry season. Therefore while mobility remains an issue with this method, in this particular context it is unlikely that it led to substantial bias. We note this issue of case mobility also affects similar area designs such as adaptive sampling.

Using these weights, we apply the Horvitz-Thompson estimator of the mean (Särndal et al 1992, p.111). To estimate variances, we use the bootstrap method with 1000 replications. In each replication, we select a sample of  $c_h$  circles with replacement from the  $c_h$  selected circles within each stratum and recalculate the mean. The estimated variance of the mean estimate is the variance of the replicated means around the full-sample mean (Kolenikov 2010).

## 4. Results

### 4.1 Field Work Results

As mentioned above, of the 125 points selected, 102 were visited. Of those visited, 59 circles (58 percent) contained at least one livestock. In total, the interviewers collected information from 793 households which owned livestock, though nine of these households were shown by their GPS points to be outside of the circle boundaries and are therefore excluded from the analysis, leaving a total sample size of 784. The number of interviewed households per circle with livestock owning households ranged from one to 65, with a mean of approximately 15.

Table 2: Field Work Results

Stratum	Description	Selected Points	Visited Circles	Households in Circles	Circles without Livestock
1	High likelihood: towns	10	10	69	4
2	Almost no possibility: settled agricultural areas / commercial farms	15	14	113	8
3	High likelihood: within 2 km of major river or swamps	60	49	229	24
4	Medium likelihood: within 10 km of major river or swamps	30	22	182	6
5	Low likelihood: all land not in another stratum	10	7	191	1
Total		125	102	784	43

It was also necessary to replace four circles during the course of the fieldwork. These replacements were made at the discretion of the survey coordinator, with input from the CSA and World Bank teams, for locations that fell within restricted areas.

In total, 3,698 individuals living in households owning livestock were identified as part of the survey. Of these, 127 reported that having no permanent dwelling, which weights up to an estimate of 4,701 (95% CI: 94, 9,307), or two percent of the livestock-holding population in Zones 1, 3, 4, and 5. All but five of the individuals without a permanent dwelling lived in households in which all members are completely nomadic. The inclusion of households without permanent addresses in the survey was a main objective of the original research agenda, as this group is traditionally undercovered in dwelling-based surveys. There are, however, very few of them in the study, not enough to perform independent analyses.

## 4.2 Means and Totals

To assess the RGCS approach, we compare weighted estimates of means and totals from the RGCS survey to estimates from the ERSS, a household survey carried out by CSA during the 2011/2012 agricultural season. The ERSS used the traditional stratified two-stage cluster design to select households and completed interviews with more than 4,500 households throughout Ethiopia. The Afar portion of the survey included 144 households in 12 clusters, ten of which were rural and two of which represented small towns. As discussed above, only two zones were covered in the ERSS survey, Zones 1 and 3, and therefore we limit our comparisons to these two areas. Among the interviewed households, 83 percent (weighted) reported owning livestock, and were administered an additional livestock questionnaire in November and December 2011 (Central Statistics Agency and World Bank, 2012). Thus there is a short time gap between the ERSS livestock survey and our RGCS study for which we compensate in the analysis. Though each survey has its shortcomings, our expectation was that the surveys should agree in the aggregate.

To construct comparable measures, we use retrospective questions about livestock in the RGCS to derive the number of livestock the household owned at the time of the ERSS survey. These questions account for slaughter, loss, death, purchase, and birth over the six months prior to the survey. We calculate two weighted estimates, one using the base weights and the other using the weight which adjusts for the proportion of the selected circle actually observed, as discussed above.

Table 3: Weighted Estimates of Total Livestock in Study Area and Average Livestock Held by Household (Conditional on Ownership), by Animal Type

	Mean (SE)			Total (SE)		
	RGCS (unadjusted weights)	RGCS (adjusted weights)	ERSS	RGCS (unadjusted weights)	RGCS (adjusted weights)	ERSS
Cattle	10.4 (1.5)	10.8 (1.8)	15.3 (3.3)	153,505 (34,384)	186,164 (51,283)	1,092,752 (367,307)
Camels	8.1 (1.4)	7.7 (1.4)	6.2 (1.9)	92,009 (25,893)	139,608 (37,186)	237,568 (116,430)
Goats	20.2 (3.1)	19.7 (3.0)	20.7 (3.1)	566,139 (146,182)	815,310 (222,853)	2,095,876 (488,027)

Compared to the ERSS survey, there are no statistically significant differences in the mean number of animals found per household across the three groups, the RGCS unadjusted and adjusted,

and the ERSS. These numbers are also in the range of secondary source estimates of herd size and composition, though the available estimates are dated (see Sabates-Wheeler et al. 2013, Getachew 2004, and Said 1994 for further discussion). There are, however, large differences with regard to the totals. While the ERSS and RGCS estimates are not statistically different for camels due to the wide confidence intervals on the estimates, the ERSS estimate is more than 70 percent higher than even the higher of the two RGCS estimates. The gap is even wider for goats, where the ERSS estimate is more than 1.5 times higher than the adjusted RGCS, and for cattle, where the ERSS estimate is nearly 5 times higher (Table 3).

We hypothesize two issues that could have led to these discrepancies. The first possible explanation is the *interviewer effort hypothesis*: RGCS interviewers did not make efforts to reach all portions of the circles that they could have and/or did not interview all households in the circles and all holders at those households, and thus systematically excluded many livestock from the survey. The lower levels of effort could be attributed to the weather, which was extremely hot during this period, flooding, which would have made access more difficult by requiring interviewers to take long detours on foot or forge swollen rivers, and also the Ramadan period, which would have limited access to local guides to assist the teams. Low effort by the interviewers could lead to undercoverage of livestock and thus to underestimates of totals in Table 3. The second possible explanation is the *ERSS over-estimation hypothesis*: implementation issues with the ERSS upwardly biased the livestock totals. The following two sections explore these two hypotheses in more detail.

### 4.3 Test of Interviewer Effort Hypothesis

To further explore the interviewer effort hypothesis, we estimate three regression models in which measures of effort are the dependent variables. The covariates in each of the models are similar and are of two types: measures about the area and the land, and measures about the workload and the interviewers. The models are all run at the level of the circle, rather than on the household or holder level.

The first two models use a logistic regression in which the dependent variable is whether a selected circle was visited (1) or not (0) by a field team, regardless of whether any livestock households were found. Recall that there were 23 circles that were selected but never visited, and this failure to complete assigned work load is one measure of interviewer effort. Since the unit of observation is the circle neither weighting nor stratification are necessary to include. In the model,

$$\Pr(Y = 1) = \frac{1}{1+e^{-\eta}} \text{ where } \eta = \alpha + \beta X + \varepsilon$$

$\alpha$  is a constant term,  $X$  is a vector of relevant household and team characteristics, and  $\varepsilon$  is the error term. In the first model, the variables included in vector  $X$  are the distance of the center point of the circle to the nearest paved road, the distance from the center to the nearest locality, the distance from the center to the nearest large body of water, the relief roughness of the terrain (the maximum elevation minus the minimum elevation divided by site radius, Meybeck et al, 2001), the radius of the circle, a historical mean vegetation index (NDVI, a measure of ‘greenness’) value, and supervisor (or team) level fixed effects. The distance measures are included to capture how difficult it was for interviewers to access the selected circle; interviewers may have been less likely to visit circles which were further from the road or from a town. Similarly if the circle was situated in rough terrain, it may

have been more difficult to access. Because it was not possible to calculate the NDVI value at the time of the attempt for those sites that were not visited, 10-year historical average NDVI value for that area is used. Also, in lieu of strata-level fixed effects, we include the circle radius, which, along with the distance to a major water source and the long-term NDVI mean values, constitutes the strata definitions. These results are presented in column 1 of table 4.

In addition to the information included in the first regression, we also know that in the 20 cases where the survey coordinator was present, the sites were always successfully visited. The model is re-run to exclude those 20 sites in which the survey coordinator was present. Those results are presented in column 2 of table 4.

The third model is a standard OLS model in which the dependent variable is the proportion of the circle observed, measured between 0 and 1 according to the Viewshed calculations discussed above. Again, walking more of the circle and observing the area is a sign of greater effort by the interviewers. This model is conditioned on the interviewers having visited the circle, and thus includes only 102 data points. Here the covariates included are the distance to a main road, distance to nearest locality, distance to a major water source, relief roughness, the radius of the circle, historical mean NDVI values, total rainfall in the week prior to the survey, current mean NDVI values, the supervisor fixed effects, and the indicator of the coordinator's oversight. There are two additional variables included in this model, total rainfall in the past week, which is added to further explore the teams' assertion that flooding was the main obstacle to coverage, and current NDVI values, to test if perhaps dense vegetation hampered observation rates.

Table 4 in the appendix presents the results of the three models of interviewer effort. The supervisors had reported that flooding and rough terrain were the main reasons they could not access or fully observe the selected areas, but the models reveal limited support for these claims. The first column in Table 4 shows that the closer a circle is to the main road, the more likely interviewers are to visit it ( $\hat{\beta} = -0.140$ ,  $p = 0.014$ ). The interpretation of this result is a bit ambiguous, as it could be due to the need to travel long distances off-road to reach the circle, leaving the teams vulnerable to flooding or other terrain hazards, or it could be interpreted as a lack of willingness by the interviewers to attempt to access these sites. The negative and significant estimated coefficient on the radius size ( $\hat{\beta} = -0.634$ ,  $p = 0.090$ ) suggests that teams preferentially worked the circles that were smaller and thus easier, which supports the lack of effort hypothesis. If flooding or rough terrain at the sites themselves were the problem, we would have expected to see negative coefficients on the distance to river variable or relief roughness, but none was found.

Repeating the model excluding those overseen by the survey coordinator, we see similar effects, though their magnitude is larger. Teams are less likely to visit larger and more remote circles in both the full and restricted models ( $\hat{\beta} = -0.174$ ,  $p = 0.007$  in the first model, and  $\hat{\beta} = -0.933$ ,  $p = 0.030$  in the second model, respectively). They are also less likely to visit circles with historically higher rainfall totals ( $\hat{\beta} = -6.154$ ,  $p = 0.067$ ) when not accompanied by the survey coordinator. This is in contrast to the assertion from the field teams that those circles closest to the river were the most difficult to access due to flooding. Unsupervised, teams were less likely to visit historically drier circles, which would most likely be located in the harshest terrain.

In the final model, of the percent of the circle directly observed, the only measure related to geography that is significantly related to the percent of the circle observed is the circle radius ( $\hat{\beta} = -0.057$ ,  $p = 0.001$ ), indicating that larger circles have lower coverage percentages. In addition, the historical NDVI value is also weakly significant, perhaps again indicating interviewer unwillingness (or inability) to spend long periods of time in harsh climates. The relief roughness, distance to river

variables, total rainfall in the past seven days, and current NDVI value, which correspond to the reasons cited by the interviewers as explanations for not observing the whole circle, are not significant.

Perhaps the most striking finding across the three regressions is the consistent significance of the supervisor effects. In terms of the number of sites visited, the teams led by supervisors 2, 4, and 5 are consistently lower compared to supervisors 1 and 3. This is particularly true for the cases in which they were not accompanied by the survey coordinator. In terms of the percent of the circle observed, the results are harder to interpret. Supervisors 2 and 3 observed smaller proportions of their assigned circles, which would seem contradictory to the findings in the previous two regressions which identify Supervisor 2 as low effort and Supervisor 3 as high effort. It may be difficult in this case to separate what is a lack of effort and what is the inability to completely observe a relatively inaccessible site that other teams would not have extended the extra effort to visit. Finally, the last row of Table 4 indicates that the presence of the survey coordinator was not significant in terms of the area of the circle observed.

Taken together, the three regressions present a picture of what occurred during field implementation and why some areas were not thoroughly worked. Though some evidence on a lack of interviewer effort is confounded by actual obstacles to task completion, such as the distance the team must travel from a paved road, a general lack of significant findings related to flooding and terrain, the two main difficulties cited by supervisors, point toward a low effort interpretation. The substantial findings of supervisor-level effects and the survey coordinator effect further support the low effort hypothesis but also demonstrate that effort level varied across teams. Low field effort can in turn explain why our collected data seems to capture too few livestock, relative to the ERSS household survey, as shown in Table 3.

#### 4.4 Tests of ERSS Quality Hypothesis

In addition to undercoverage by the RGCS, another possible explanation for the discrepant totals in Table 3 is some degree of over-estimation in the numbers produced by the ERSS for Afar. There may be some distortion due to the small sample size, which, as noted in the CSA and World Bank Basic Information Document (2013), is insufficient to support region-specific estimates for the less populous regions, including Afar. The recommendation to group several regions into one “other regions” category for analysis was not possible for this exercise, so the regional results presented here must be interpreted with caution.

There have also been quality issues in data generated by CSA in the past (see Dercon and Hill, 2009 for more detail). The Afar field office is particularly vulnerable to data quality issues as it is a remote region of the country, has fewer staff members, and generally lower levels of skills and training compared to headquarters or other regional field offices. Communication between headquarters and the Afar office is difficult, and communication between the field office and teams is even more complicated due to frequent power outages and unreliable cell phone networks. These issues are exacerbated by long distances and a limited road network which make field supervision challenging. In the review of the ERSS data quality, Afar was one of the regions with the highest incidence of problems, including incorrect listing forms, missing questionnaires, and incomplete information in administered surveys. In particular, the release of the livestock data was delayed for almost a year following the end of fieldwork while data cleaning was completed.

We find support for the hypothesis that the ERSS over reports livestock in the data collected in a survey of the neighboring Somali region. In 2003, as part of the Agricultural Census, USAID contracted a consulting firm based in London and Nairobi to conduct an aerial surveillance estimation of seven of the nine zones in the Somali region that could not be covered due to remoteness and security

considerations. This methodology is limited in that it cannot provide any information at the household or holder level, but it can produce high quality data on livestock totals for a given area (see CSA 2004 for details on estimation techniques). As both Afar and Somali have a largely pastoralist population base and similar climates, we would expect the density of animals to be broadly similar in the two areas.

Table 5: Livestock Density (n/km<sup>2</sup>)

	Afar Regions (Zones 1 & 3)			Somali Region (Aerial Survey 2004)						
	RGCS (unadjusted)	RGCS (adjusted)	ERSS	Afder	Degehabur	Fik	Gode	Korahe	Shinile	Warder
Camels	2.5	3.8	6.4	2.6	3.6	0.8	2.8	4.2	2.6	8.4
Cattle	4.2	5.1	29.7	3.0	1.4	0.5	4.1	0.7	5.3	0.8
Goats	15.5	22.3	56.9	13.2	19.6	4.2	24.3	19.5	21.7	31.6

Table 5 compares estimates of livestock per square kilometer from three different sources. The first three columns give estimates from the RGCS (using both the unadjusted and adjusted weights) and the ERSS survey. The last seven columns give estimates for the seven zones in Somali in 2004. We see that the RGCS estimates are within the range of those from the aerial surveillance, while the ERSS estimates are substantially higher. While it should be stressed that these estimates are not directly comparable, as they are for different areas in different time periods, we would expect the ranges to be similar for the reasons stated above. This increases our confidence in the accuracy of the RGCS estimates over those produced by the ERSS, at least with regard to livestock totals.

## 5. Discussion & Conclusion

This pilot project of the RGCS technique to collect livestock data in the Afar region of Ethiopia demonstrated that the implementation of such a design is feasible; however, questions remain as to whether it is the best available method. The project showed that sufficient GIS information is available, often through the public domain, to create strata for the probability of finding livestock, and to select points within those strata. With maps and relatively inexpensive GPS devices, teams can navigate to points and identify eligible respondents within these clusters. These respondents can then be interviewed regarding their households socioeconomic conditions and livestock holdings, creating the linkages necessary to perform poverty analysis on these populations. In addition, using standard statistical methods, it is possible to calculate weights that take into account the varying probabilities of selection and sufficiently address overlap probabilities. Also, information generated as part of the GPS field implementation, such as the Viewshed results, can be used to estimate the area observed by individual interview teams, and account for undercoverage, if necessary. And finally, the methodology was able to do what it was designed to do -- capture households without permanent dwellings who would have been missed by a traditional dwelling-based sample design. The location and interviewing of these persons is a major benefit to the RGCS technique over the traditional household based approach to survey sampling.

A number of questions remain as to whether this method should be considered the best practice for collecting this type of information. The RGCS has demonstrated some advantages over the traditional household-based survey methodology, such as eliminating the need to conduct a cluster-listing exercise and allowing data collection to be completed in a single step. The methodology was also successfully implemented in a low capacity environment and avoided overly technical issues, such as those one would have with an adaptive sampling method. However, problems were still noted in the

implementation. Interviewers did not visit all of their assigned areas, and did not observe the entire area when they did visit. Some supervisors required supervision themselves by the survey coordinator. Overall, monitoring in this study was difficult compared to the standard household survey as, without a household listing operation, the paper trail on total cluster size was limited, and the populations are mobile, which limits the usefulness of repeat visits to verify the data collected. The project was also highly dependent on the cooperation of local guides, which are outside of the management structure and may be unreliable in some areas. Perhaps most importantly, the terrain in which the survey was implemented is difficult. The weather was extremely hot and numerous natural obstacles to successfully completing the survey tasks occurred. Although these factors affect any survey in Afar, they were particularly troublesome for the RGCS, which required a good deal of driving and walking to reach the selected areas. 'High effort' was required from supervisors and interviewers throughout the project to implement the design as developed. Unfortunately we are not able to thoroughly evaluate how much these factors impacted the quality of the data collected. We have concerns that weighted estimates from the data do not accurately capture the number of livestock held in Afar, but we have no reliable comparable standard against which to compare our numbers.

Based on our experience in Afar with the RGCS, we have a number of suggestions that would improve the implementation. In the particular, we recommend more careful planning to avoid conducting the survey during the Ramadan period and the completion of data collection before the onset of the seasonal rains. In addition, training should better explain to supervisors and interviewers the goals of the survey in order to elicit more 'high effort' fieldwork. Providing incentives to teams that complete more circles and observe greater percentages of assigned circles may further increase effort. Though it is possible that the RGCS approach may work better in a different country context, the limited capacity of statistics bureaus and the potentially dangerous terrain of the study area are common to nearly all pastoralist areas in the developing world.

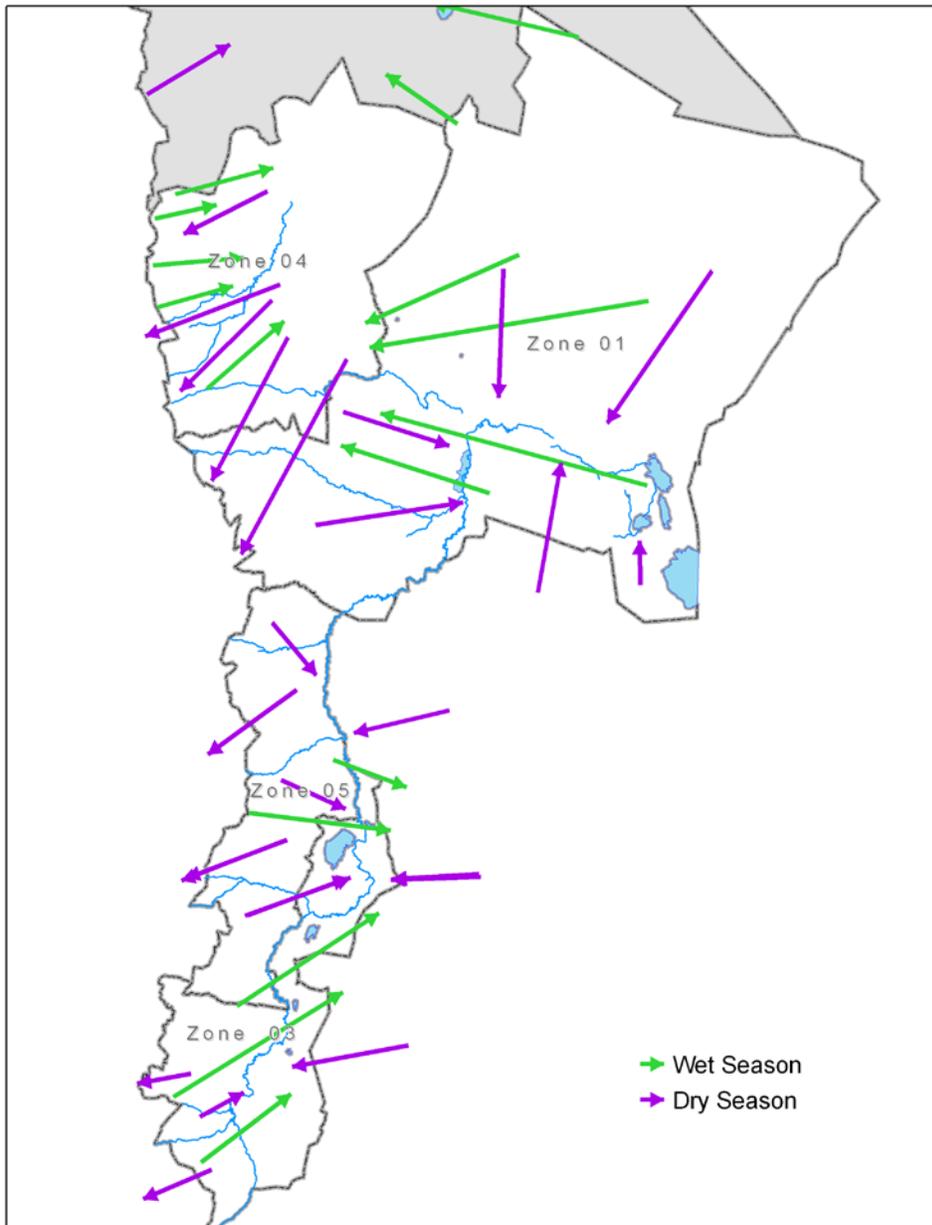
Despite the limitations noted above, drylands areas remain difficult to survey and the RGCS offers a viable alternative to traditional approaches. The Ethiopia CSA has decided to extend the method and include it as part of the data collection method in pastoralist areas nationwide as part of the upcoming Agricultural Census. Beyond the specific livestock in drylands context, we wonder if this technique might have applications to other contexts, such as the measuring of homeless persons. We believe that the RGCS approach deserves more study in both the developed and the developing world.

## References

- Barrett, J.P. (1964). Correction for Edge Effect Bias in Point-Sampling. *Forest Science*, 10 (1), 52-55(4).
- Cameron, A.R. (1997). *Active Surveillance and GIS as Components of an Animal Health Information System for Developing Countries- Thailand and Laos as Examples*. Queensland: University of Queensland.
- Central Statistical Agency (2004). *Livestock Aerial Survey in the Somali Region*. November 2003. <<http://www.dppc.gov.et/Livelihoods/Somali/Downloadable/Livestock%20Aerial%20Survey%20in%20the%20Somali%20Region%20November%202003.pdf>> (Accessed 3 July 2013).
- Central Statistical Agency (2012). *Population Projections*. January 2012. <[http://www.csa.gov.et/index.php?option=com\\_rubberdoc&view=doc&id=226&format=raw&Itemid=561](http://www.csa.gov.et/index.php?option=com_rubberdoc&view=doc&id=226&format=raw&Itemid=561)> (Accessed 24 June 2013).
- Central Statistical Agency & World Bank (2012) (2013). *Living Standards Measurement Study- Integrated Surveys on Agriculture: Ethiopia Rural Socioeconomic Survey - Basic Information Document*. December 2012.
- Dercon, S., and Hill, R. V. (2009). Growth from agriculture in Ethiopia: Identifying key constraints. In IFPRI's ESSP-II policy conference 'Accelerating agricultural development, economic growth and poverty reduction in Ethiopia', Hilton Hotel, Addis Ababa (pp. 22-24).
- Emerson, H. and MacFarlane, R. (1995). Comparative Bias between Sampling Frames for Farm Surveys. *Journal of Agricultural Economics*, 46(2), 241-251.
- Ahmed, E. F. (2007). *Epidemiological survey and assessment of optional therapy for ovine fascioliasis around Middle Awash River Basin, Afar, Ethiopia*. Bangkok: Kasetsart University.
- Getachew, K. N. (2001). *Among the Pastoral Afar in Ethiopia: Tradition, Continuity and Socio-economic Change*. Utrecht: International Books.
- Grosh, M. E. and Munoz, J. (1996). *A Manual for Planning and Implementing the Living Standards Measurement Study Survey*. Living Standards Measurement Study (LSMS) Working Paper NO. LSM 126. Washington, D.C.: The World Bank. <<http://documents.worldbank.org/curated/en/1996/05/438573/manual-planning-implementing-living-standards-measurement-study-survey>> (Accessed 4 January 2013).
- Husch, B., Miller, C.I., and Beers, T.W. (1982). *Forest Mensuration*. New York: Wiley.
- Kolenikov, S. (2010). Resampling Variance Estimation for Complex Survey Data. *Stata Journal*, 10(2), 165-199.
- Lavallée, P. (2007). *Indirect Sampling*. New York : Springer-Verlag.
- Meybeck, M., Green, P., and Vörösmarty, C. (2001). A New Typology for Mountains and Other Relief Classes. *Mountain Research and Development*, 21(1), 34-45.

- NASA Land Processes Distributed Active Archive Center (2011). ASTER Global DEM V2 data. USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota.<[https://lpdaac.usgs.gov/get\\_data](https://lpdaac.usgs.gov/get_data)> (Accessed 14 January 2013).
- Reams, G.A., Smith W.D., Hansen, M.H., Bechtold, W.A., Roesch, F.A. and Moisen, G.G. (2005). The Forest Inventory and Analysis Sampling Frame. In: Bechtold, W.A. and Patterson, P.L. (eds). The Enhanced Forest Inventory and Analysis Program – National Sampling Design and Estimation Procedures, 11–26. Asheville, North Carolina: USDA Forest Service, Southern Research Station.
- Roesch, F.A., Green, E.J.Jr., and Scott, C.T. (1993). An Alternative View of Forest Sampling. *Survey Methodology*, 19, 199-204.
- Sabates-Wheeler, R., Lind, J. and Hoddinott, J. (2013). Implementing Social Protection in Agro-Pastoralist and Pastoralist Areas: How Local Distribution Structures Moderate PSNP Outcomes in Ethiopia. *World Development* 50, 1-12.
- Said, A. (1994). Pastoralism and the state policies in Mid-Awash valley: the case of the Afar, Ethiopia.
- Särndal, C.-E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*. New York : Springer-Verlag.
- Soumarea, B., Tempiab, S., Cagnolatic, V., Mohamoudb, A., Van Huylenbroeckd, G. and Berkvensa, D. (2007). Screening for Rift Valley Fever Infection in Northern Somalia: A GIS Based Survey Method to Overcome the Lack of Sampling Frame. *Veterinary Microbiology*, 121(3-4), 249-256.
- Tatem, A.J. (2010). Ethiopia AfriPop Data 2010 (alpha version). Gainesville, Florida: Emerging Pathogens Institute, University of Florida.  
<[http://www.clas.ufl.edu/users/atatem/index\\_files/Ethiopia.htm](http://www.clas.ufl.edu/users/atatem/index_files/Ethiopia.htm)> (Accessed 15 June 2012).
- Thompson, S.K. (1990). Adaptive Cluster Sampling. *Journal of the American Statistical Association*, 85, 1050-1059.
- Thompson, S.K. (1991). Stratified Adaptive Cluster Sampling. *Biometrika*, 78, 389-397.
- Thompson, S.K., and Seber, G. A. F. (1996). *Adaptive Sampling*. New York: Wiley.
- United States Department of Agriculture Area Frame Section.  
<<http://www.nass.usda.gov/research/AFS.htm>> (Accessed 8 November 2010).
- USGS Earth Resources Observation and Science Center (2012). eMODIS NDVI Africa (monthly means)..  
<<http://earlywarning.usgs.gov/fews/africa/index.php>> (Accessed 20 June 2012).
- USGS Earth Resources Observation and Science Center (2012). eMODIS NDVI Africa (pentadal)..  
<<http://earlywarning.usgs.gov/fews/africa/index.php>> (Accessed 30 October 2012).
- von Hagen, C. (2002). Using an Area Sampling Frame to Calculate Livestock Statistics in the Gauteng Province, South Africa, within a GIS. *Directions Magazine*.

Figure 1 : Traditional migration routes in Afar, showing the four of the five administrative zones. Zone 2 is represented by the grey area at the top of the map. The green lines represent rainy season migration and the purple lines represent the return migration in the dry season. Lengths of lines are indicative of migration distances but are not drawn to scale.



Source: Ahmed, Endris Feki (2007)

Figure 2: Stratification Map

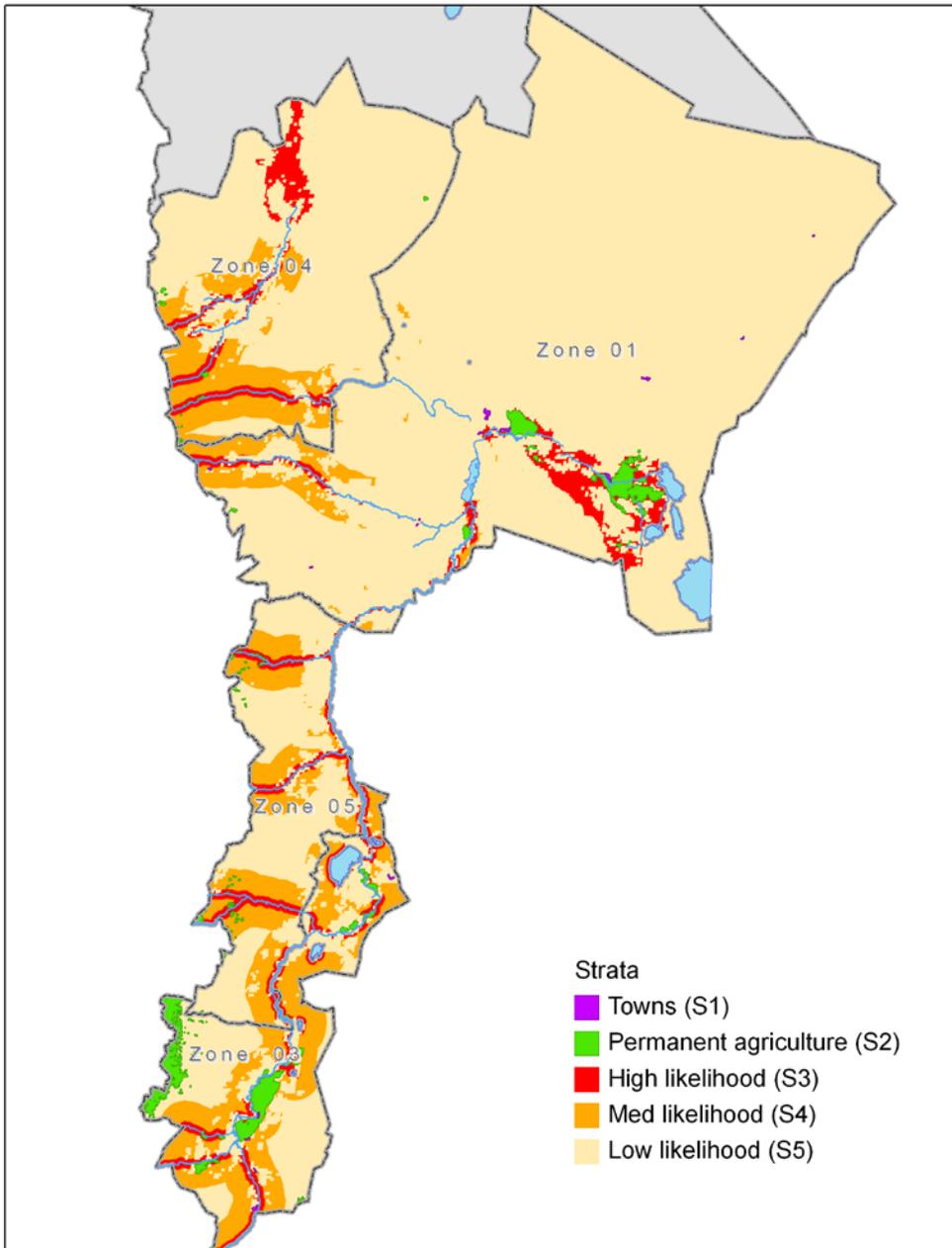


Figure 3: Example of Selected Point and Circle



Figure 4: Overlap between circles in different strata

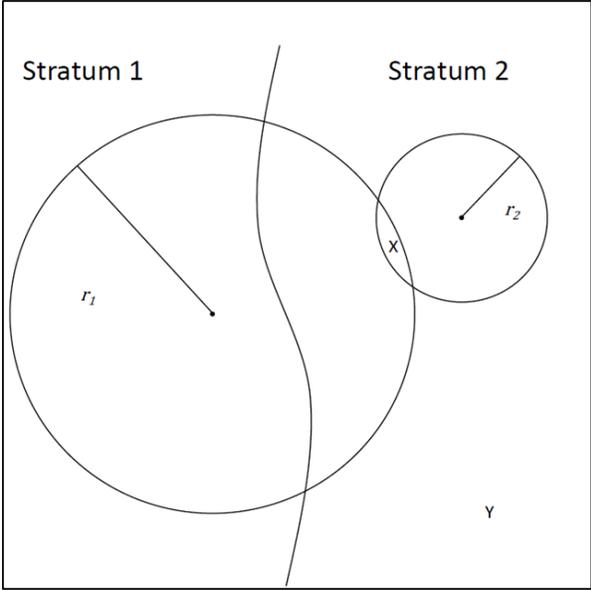


Figure 4.1: Household X, in stratum 2, can be selected by points selected from stratum 1 or 2.  $r_1$  is the selection radius used in stratum 1;  $r_2$  is the selection radius in stratum 2.

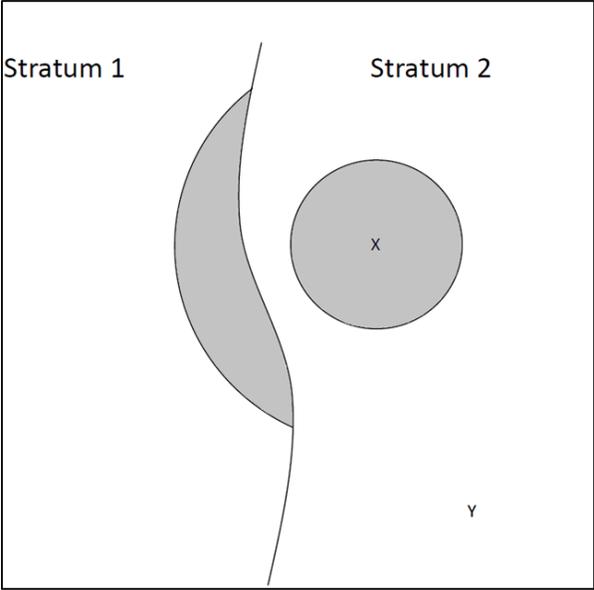
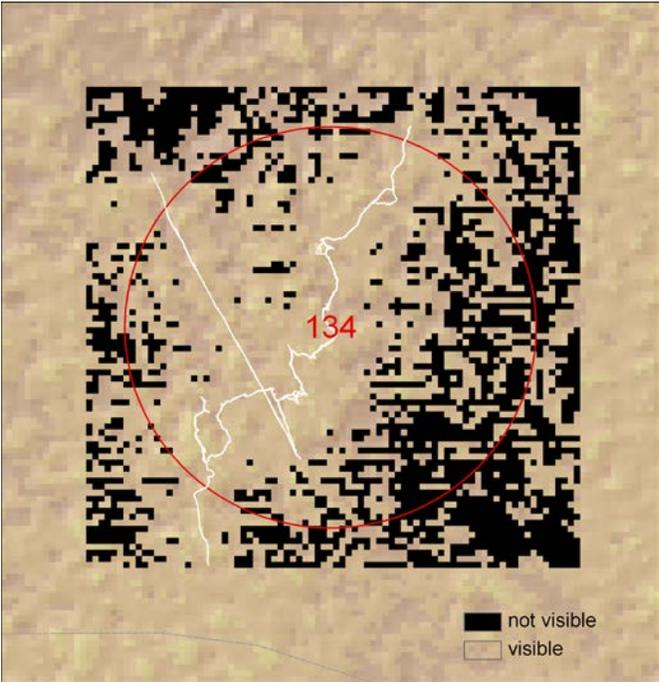


Figure 4.2: Grey shading indicates the set of all points that, if they were selected, would lead to the selection of household X.

Figure 5. Viewshed Analysis



**Table 4 : Regression Results**

	(1)		(2)		(3)	
	Selected site visited (logit model)		Selected site visited (logit model) excluding overseen		Percent of circle observed (OLS model)	
	$\hat{\beta}$	Std. Error	$\hat{\beta}$	Std. Error	$\hat{\beta}$	Std. Error
Kilometers to main road	-0.140**	0.057	-0.174***	0.064	-0.005	0.004
Kilometers to nearest locality	-0.070	0.136	-0.116	0.159	-0.002	0.009
Kilometers to river	0.010	0.033	0.000	0.032	-0.000	0.001
Relief roughness	0.006	0.005	0.007	0.005	-0.000	0.000
Circle radius	-0.634*	0.374	-0.933**	0.430	-0.057***	0.016
Historical mean NDVI value	-4.575	3.076	-6.154*	3.354	-0.441*	0.231
Total rainfall week prior to survey					-0.000	0.000
Current mean NDVI value					0.146	0.160
<i>Reference : Supervisor 1</i>						
Supervisor 2	-2.416**	1.207	-3.197**	1.413	-0.085**	0.036
Supervisor 3	-0.249	1.646	-0.086	1.898	-0.215**	0.094
Supervisor 4	-3.211***	1.148	-4.021***	1.383	-0.011	0.045
Supervisor 5	-1.771	1.215	-2.740**	1.395	-0.051	0.037
Overseen by survey coordinator					-0.056	0.046
Constant	6.549***	1.912	8.357***	2.371	1.131***	0.076
Number of observations	125		105		102	
Pseudo R <sup>2</sup> / R <sup>2</sup>	0.264		0.314		0.515	

note: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1