Summary of Research Proposal/Résumé de la proposition de recherche

Introduction. Surveys in war zones, other difficult areas, or Low and Middle Income Countries (LMIC) have various aims, including estimation of mortality and other harms, population needs for food, water and shelter, and vaccination coverage. Yet such surveys may pose considerable challenges for researchers: there may be little information on which to base the sampling method, the information available may be out of date, etc. Also, interviewers may have little education or training, funds may be limited, and results are often needed quickly. To overcome these problems several methods, such as the WHO's Expanded Program on Immunization (EPI) approach, have been used, even though they are not ideal. Mostly, they use Probability Proportional to Size sampling to select clusters – e.g., towns or villages - and various methods are used to select units (households or individuals) within the clusters. Since complete enumerations of towns rarely exist and it may be unsafe for interviewers to conduct such enumerations, selection of units is not properly random and adherence to a specific method of sampling may be difficult or impossible.
Recently, we have used a new method, based on global positioning systems (GPS) and satellite photographs, which is truly random sampling, but has its own limitations. The important question is whether the methods, especially the new one, are 'good enough', that is, whether they produce estimates with little bias and good precision. We plan to use simulations of the different methods of sampling to better understand their properties, and to identify which are optimal in specific circumstances.

Objectives. **Our overall objectives are (i) to conduct simulation studies to determine the bias and precision of different sampling methods in difficult settings and (ii) to determine the circumstances in which one method may be preferred over others**.

More specifically, we will explore **three questions**: (1) Can analytical methods be developed to improve estimates of sampling probabilities in the GPS/satellite photo approach? (2) How do variations in the GPS/satellite method affect estimates' bias and precision? (3) Are there circumstances in which some method(s) perform better than others?

Methods. A virtual population in several towns will be 'created' with known characteristics and distributions of both categorical and continuous variables such as gender, age, disease status, or health related quality of life. The towns will reflect real-life situations – e.g., different sizes of towns, variations in population density within towns, different distributions of characteristics in different areas of towns, and 'pocketing' (the tendency of some outcomes such as presence of infectious disease to be clustered in very small areas). We will incorporate spatial autocorrelation among the variables and will build in some associations between variables. We will conduct simulations picking numerous samples by each method from the populations. We will compute the mean of sample means or proportions and variances for each method of sampling, and measures of association between variables. These results will be compared to the true population values. This process will be repeated varying the characteristics of the populations, e.g., changing the prevalence of the variable of interest, or the geographical distributions of the population.

Lay abstract/Résumé non scientifique

There is often a need for health surveys in Low Income Countries, war zones or areas suffering from disasters. The surveys may measure the health status of the population, such as the proportion of children who have been immunized for infectious diseases. Yet it can be difficult to conduct the surveys because often there is very limited information to allow a proper sample to be selected. Researchers have developed several ways around this problem, but none of these methods is ideal; they all have some flaws, so the final results may be at least somewhat incorrect or have a large margin of error. This project aims to find out which sampling method produces the best results in different situations.

It will do so by using a computer to create virtual populations. The criteria programmed into the computer will try to make the populations as realistic as possible, while varying characteristics such as the proportion of children immunized. The computer will take samples from the virtual populations using the different sampling methods. Many thousands of samples will be taken for each method, so that we can understand how well each method performs depending on the population characteristics.

The results will help governments and NGOs to get the best information possible when they direct resources to improving the health of particular populations.

H.S. Shannon. **Sampling methods: a simulation study.** $121,467
Response to previous reviews

We thank the reviewers and committee for their thoughtful and constructive comments. We have framed our response under headings based on the limitations noted in the Scientific Officer notes. We deal with all other concerns raised by reviewers under these headings. As Reviewer 1 noted explicitly, many concerns stem from a lack of detail about our proposed simulation.

Lack of detail in some areas (could have described simulation study more)

We agree that we did not provide enough detail, in part due to our lack of experience with simulations. As Reviewer 1 recommended, we have added to the team someone who has done many simulations, Dr Ben Bolker. We believe the proposal has significantly improved with his input.

We will geocode buildings from satellite photos of relevant areas and use well-established methods to estimate the parameters of spatial point process models (p.7). These fitted models will then form the basis for simulating virtual populations, which will thus reflect real distributions of residences.

Similarly, we will use the information we have on families in four surveys to generate realistic household memberships (p.8). We have noted that we will incorporate 'pocketing' (local clusters of disease) by modeling a contagion process or by simulating presence of pockets at specific geographic locations (p.8).

To create the different populations, we have decided to use the Latin hypercube approach (p.9). This manages the number of possible populations, since a fully factorial design would be unfeasible given the number of parameters we wish to vary. It also allows us to understand interactions between the factors, which would have been only partly possible with the method of our previous submission.

We will also ensure that the datasets containing the virtual populations will be constructed to allow, e.g., efficient searching for houses close to a reference point (p.9).

Satellite pictures will not be feasible in particular situations (e.g., war situations).

We acknowledge that the satellite photos will not always be available, although in some circumstances commercial organizations can take the required photos for a cost. However, even if they cannot (e.g., because of cloud cover) or the cost is too high, the GPS component of our new method can still be used. Interviewers could go to the randomly chosen location and conduct an enumeration of the households in a given area and randomly choose one household to be part of the sample. We accept that this does remove one of the advantages of the method, namely that preparatory work using satellite photos reduces interviewer effort in the field. If the area to be enumerated is a square (and we now prefer this to the circle that we have used before), the GPS unit would ensure the interviewers can readily find the boundaries of the area and thus map the households therein. When conditions are dangerous, it may still be possible for interviewers to enumerate a very small area (since this might be done unobtrusively), and we will examine the consequences of using smaller squares. We would not make it too small, as this could lead to many 'empty' squares, and force the interviewer to visit more GPS locations. In practice there will be a balance between the effort of conducting enumerations against the effort of visiting more locations. We will explore this in our simulations (p.9). In any event, we will still be able to estimate any bias in methods that do not require enumeration, and this could be part of any decision on which type of survey to conduct.

Unclear how approaches will be compared

We have noted some criteria which we will need to integrate to compare the different sampling methods (p.11). They include the time/cost to reach a town (cluster), time/cost to get to locations within towns, time/cost to complete interviews, number of sampling units per cluster, number of return visits needed to find person sampled, and likelihood of errors in enumeration. Some can be quantified directly, others may be more difficult to do so, and a more qualitative judgment may be required. These costs will have to be traded off against the Mean Square Error (MSE – a combination of bias and variance) of each sampling method. As well, the experience and technical expertise of the survey team may influence a decision on the optimal method in any setting. We anticipate developing a decision tree to provide guidance to potential surveyors.

Some criteria may not be easily quantifiable, e.g., number of return visits needed, likelihood of enumeration errors. So while we can list the criteria, we probably cannot specify in advance exactly how they will be used to compare the methods

Cost analyses needed more clarity

We believe this issue is related to the previous one – and we noted relevant costs that would be taken into account (p10/11). Several of the costs listed above can be reasonably estimated when planning a survey. We have found only one paper that compares methods based on cost – Macintyre (1999) described two surveys of views of family planning six months apart in the same area. We will repeat her analyses to the extent possible (she had real costs from the two surveys) and will allow the unit costs to vary somewhat to allow for local circumstances.

Real life application would be helpful

We agree that applying the approaches to real-life datasets would be ideal. However, to do that, we would need a population for which census data with geographical coordinates on each household is available. We know of no such dataset – indeed, the methods we are comparing are needed in exactly those settings where this background information is lacking. As we note, if such a dataset becomes available to us, we will conduct simulations on it to compare the different methods. Even then, since it will be specific to a particular population, it will not allow us to explore how variations in populations might lead to different decisions on the optimal sampling approach.

KT plan is generic

Our previous plan noted we would present results at scientific meetings and in journal papers, as well as produce documents in lay language with guidelines for those doing these surveys.

We have added to this plan (p.13): we anticipate developing a decision tree to help guide surveyors. As well, we expect to produce diagrams (small multiply conditioned plots -- similar to scatterplot matrices) showing the relationship between various factors and the MSE vs (quantifiable) cost trade-off. We will further consult with those who do surveys to better understand their needs and if there are preferred ways for us to present our results.

Budget

We have reduced the duration of time for which we will hire the programmer. We agree with the Scientific Officer notes that the time needed for this person to complete the work is roughly 18 months. However we want to keep this person for an additional 2 months during report preparation to help with the writing, and especially to do extra analyses that will likely be required. The total budget over 2 years is thus reduced by about $33,000.

Introduction

Surveys in war zones, other difficult areas, or Low and Middle Income Countries (LMIC) have various aims, including estimating mortality and other harms and population needs for food, water and shelter, and vaccination coverage, as well as monitoring and evaluating the effect of public health interventions. Yet such surveys may pose considerable challenges for researchers; for example, there may be little information on which to base the sampling method, or the information available may be out of date. As well, funds may be limited, interviewers may have little education or training, and results are often needed quickly. Several methods have been used to overcome these problems. Even though they have known flaws, the important question is whether they are good *enough*; that is whether they allow estimates with sufficiently little bias and sufficiently good precision to be obtained within the budget and timeframe available.

Whatever sampling method is used, quantifying the bias and precision of the sample requires a) determining the probability that a sampling unit was included in the sample; and b) computing the design effect (i.e., allowing for the sampling process). These computations allow calculation of the sampling weights to be applied and of proper point and variance estimates. In difficult settings, various methods have been proposed. However, compromises are often be so that, for example, sampling may not be properly random.

The broad aim of this project is to conduct simulation studies to determine which methods are indeed 'good enough' and to identify the circumstances in which one method may be preferred over others. Our focus is limited to the sampling, and does not include other features of surveys such as questionnaire design.

Sampling methods in difficult settings

Simple or stratified random sampling is not feasible when there is no enumeration of the target population and indeed may be impractical if the population is spread over a wide geographical area. ('Enumeration' entails counting, mapping and listing all sampling units, such as eligible people or households.) Multi-stage sampling - sampling successively smaller groups, e.g., area, then town, then household, then individual – may overcome limitations due to lack of information on the target population. The method may incorporate clusters, which reduces the cost of interviewing, since the time and expense of travel is reduced compared with simple random sampling. Statistically, a penalty is paid since the clustering almost always increases the standard errors of the parameter estimates – that is, there is an effective reduction in the sample size. Still, provided the similarity among people within a cluster is not very high (as measured by the intra-cluster correlation, ICC), this approach can make better use of the resources available –the necessary increase in sample size may be more than balanced by reduction in cost per interview. Notably, this method does not require complete enumeration of all sampling units in the population. It does require dividing the population into mutually exclusive and exhaustive clusters, typically based on geography.

A multi-stage cluster design might first select regions of a country, then towns within each region, followed by districts within each town, subdivisions within districts, households within subdivisions, and finally individuals within households. At each step, the units from which the sampling is being conducted must be enumerated, so that probabilities of selection can be computed. Even when information is limited, there is often data for at least the first steps in this process, yet at the later steps, research teams must conduct the enumeration on site.

The EPI method. If clusters are small, all units (individuals, households, etc) may be included in the sample. Usually, however, some method of sampling from within the cluster is needed. In a study that formed the basis of the "30 x 7" methodology (see below), Henderson et al. (1973) assessed vaccination coverage, vaccination scar rates and smallpox scarring in five areas of West Africa. After selecting villages based on probability proportional to size (PPS) and making adjustments to the list of villages based on local knowledge, the assessors stood in the centre of the village. A direction was randomly chosen. (This was done using random numbers; in other surveys interviewers have spun an object such as a pen to select the direction, hence the term 'spin-the-pen' method.) All dwellings from the centre to the edge of the site along a line in the chosen direction were counted, and one was chosen randomly. All members of the dwelling were interviewed. If the number of people was less than the target sample size for the cluster (village), in this case 16, the assessor went to the next household along the line away from the centre, and interviewed its members. If assessors reached the edge of the village before completing the sample, they moved clockwise to the next household and moved back towards the centre of the village, conducting interviews along the way.

The authors acknowledged some limitations with the method. In practice, the estimates of population size are likely subject to inaccuracy, hence the first stage of sampling is better labelled probability proportional to *estimated* size (PPES) rather than PPS. The authors also pointed out that the method was biased to selecting houses close to the centre, but that the bias could be minimized by counting dwellings within a sector radiating from the centre rather than on a line. Also, they noted that if some households were further from the village's vaccination site, it could affect likelihood of vaccination. Since villages were small, though, they considered that to be of little concern. They also noted that they could not be sure that distance from the village centre was unrelated to likelihood of vaccination.

Henderson and Sundaresan (1982) later discussed the "30 x 7" method used by the Expanded Programme on Immunization (EPI), based on the survey methodology in the 1973 paper. It entails sampling 30 clusters and evaluating immunization status in seven eligible subjects per cluster. Taking the design effect (allowance for clustering) to be 2, the authors showed how the EPI method could ensure that any estimate of vaccination coverage obtained from surveys was within 10% of the population value with a reasonable degree of certainty, i.e., the 95% Confidence Interval (CI) should be no wider than ± 10% from the sample estimate. Using results from 60 surveys and computer simulations, the authors concluded that the 30 x 7 method serves its purpose well enough.

In the EPI approach, a different method is used to choose the second and subsequent households in a cluster after the first one was identified. The interviewer visits the 'next nearest' household, defined as the closest one (front door to front door) to the previously visited one. If no one in a household is eligible for the survey, the interviewer simply proceeds to the next nearest house. The process continues until the target sample size is reached

Further computer simulations of the performance of the EPI approach were reported by Lemeshow and colleagues (Lemeshow, Tserkovnyi, Tulloch, et al., 1985). They created 'virtual' clusters, with particular levels of population density, patterns of household distribution, immunization rates and immunization patterns. They too concluded that the method met the aim of obtaining 95% CIs within 10% of population values for vaccination coverage estimates.

Problems with the EPI approach. In that paper and another (Lemeshow and Robinson, 1985), though, some possible problems with the EPI method of selecting households were described. If a roster or 'on the spot' enumeration of households is available, a sample can be picked using random number tables or equivalent. When this is not possible, the method described above is recommended (WHO, 1991). Lemeshow and Robinson stated that in urban areas, this may be very difficult. They also noted that "there is no particular advantage in using 30 clusters", pointing out that the balance between the number of clusters and the sample size per cluster depends in part on the intra-cluster correlation (ICC). Since this varies from setting to setting, no rule can be universally optimal. One factor that affects the determination is the expected prevalence of the variable being measured. If it is very low, larger sample sizes (more clusters and/or more units per cluster) are desirable – a sample of 30 clusters of 30 units is sometimes used.

Lemeshow and Robinson noted that the selection within clusters is not a simple random sample, and any spatial clustering - in small areas within the cluster - of the characteristic being studied may lead to under- or over-estimation of the prevalence in the cluster. Indeed, in simulations this phenomenon, known as 'pocketing', among vaccinated children within a cluster created bias within clusters (Lemeshow, Tserkovnyi, Tulloch, et al., 1985). Further, while not noted by the authors, any household spatially separate from other households in the cluster would only be included in a sample if it were selected as the starting household, since it would never be the nearest household to any other.

Lemeshow and Robinson further argued that there is opportunity for interviewer discretion to create bias. This could occur in several ways. If the cluster is large, it may be easier for the interviewer to identify the starting house for convenience, not at random - for example, without counting all the households to the edge of the cluster. As well, the decision on which is the nearest household to the one just visited might be determined by interviewer choice (e.g., quality of housing) not objective distance.

The authors also mentioned two other issues: the EPI practice of not returning to households where no one was home on the initial visit and using respondent *report* of vaccination status rather than documentation. These, though, apply to any survey and while they can lead to bias they are not specific to the EPI methodology.

Improvements to the EPI method and alternative approaches. Some attempts have been made to modify or improve on the EPI method. Bennett and colleagues considered preferable ways to select the sample within clusters, noting that in the absence of a random sample it is 'better to choose, say, the fifth nearest household'. (Bennett et al., 1991: 100) They also argue in favour of having more than one starting point in different parts of large communities so the sample is spread out. Both of these amendments aim to reduce pocketing.

Brogan et al. (1994) also reminded readers of the concerns about whether the sample within clusters is properly random and the extent of non-response. They also pointed out that since it is households that are sampled at the second stage of selection, one should use the number of households per cluster rather than the total population per cluster in the first stage of sampling. (They also noted that in practice cluster sizes will likely not be accurate, leading to some degree of bias.) Like Henderson et al. (1973), they observed that households closer to the centre of the cluster have an increased probability of being sampled and they reinforced the concern of Lemeshow and Robinson (1985) that interviewer judgement can create bias.

Concerned to achieve a probability sample (preferably an equal probability sample) Brogan et al. argue for enumeration of the clusters. However, recognizing that in practice this may not be feasible, they suggest segmenting the initial (larger) clusters, choosing one segment based on probability proportional to the estimated sizes of the segments, and enumerating that segment. (If necessary, sub-segmenting could be done to allow enumeration.) They argue that estimates of the segments could be approximate if it is not feasible to provide accurate numbers, and that it is possible to estimate selection probabilities, and hence sample weights, for each individual in the sample. One can then obtain an equal probability sample of households in the segment using simple random sampling, systematic random sampling or cluster sampling. Brogan et al. note that the cluster sampling could be achieved by randomly sampling one household and then choosing the next households on the enumeration list. This would circumvent the problem with the standard EPI method, since all households would have an equal probability of being selected first, and interviewer judgement would be removed when selecting the other households. The authors also discuss the importance of calling back at least several times to contact households where no one is present at the initial visit.

Turner and colleagues (1996) acknowledged that rigorous cluster sampling methods are often not feasible in LMICs, but wished to improve on the EPI design, maintaining a degree of simplicity while overcoming some of the limitations of the EPI method. They proposed maintaining the PPES of the EPI method and using the following approach for the next stage: sketch-mapping the sample clusters, creating segments of roughly equal size (that is, equal across all selected clusters), randomly choosing one segment per cluster, and interviewing *all* eligible persons within the segments chosen. To execute this, the required total sample size N is estimated, taking into account the prevalence of variables of interest. The size of each segment is then calculated by dividing N by the number of first-stage clusters. When the estimates of cluster sizes are up-to-date or if all clusters have grown at the same rate since the last census, all households included have the same probabilities of being chosen, so weighting is unnecessary and this method is preferable to the standard EPI design. Still, the method requires knowledge of clusters and their sizes to a fairly fine level – they cite a national survey of Bangladesh with PPES sampling of mauzas and mahallahs (administrative subdivisions) containing roughly 250-300 households.

Milligan et al. (2004) conducted two surveys of vaccination coverage in western Gambia just a few months apart; in one they adopted the EPI plan, in the other they used segmenting. They found the methods gave similar results, but recommended segmenting on the grounds it is less susceptible to poor quality fieldwork and can give estimates of population totals (rather than just proportions and means) to guide planning.

A commentary on papers from a 2006 conference on survey methods in difficult situations noted that research to develop these methods 'has largely stagnated in the past two decades' (Bostoen et al., 2007: 1). One conference paper from Grais et al. (2007) compared the 'spin-the-pen' selection of the first household with two other approaches. The first superimposed a grid on a map of the cluster, randomly chose x- and y-coordinates on the grid, and identified the closest compound (houses in the setting tended to be in walled compounds). The second used GPS coordinates to identify a randomly chosen point, and the nearest compound to the right when facing north at the point. The new methods had advantages. Survey teams found them easier to implement than the spin-the-pen method. The teams were most enthusiastic about the GPS method, while the grid approach was fastest. Grais and her colleagues noted, though, that with both alternative methods there were higher probabilities of selection of households in low density areas of the clusters.

Roberts (2001) noted the problem of differing household density. He first identified sampling points based on randomly choosing a point in the south of the main north-south highway in Katana, Democratic Republic of the Congo (DRC), and then a grid of locations at fixed intervals to the north and to the east and west of that point. For each point the five closest households were selected. Density was allowed for by estimating the 'radius of the sampling point' as distance between the point and the furthest of the five households. The report was short and Roberts did not describe how the radius was incorporated into the analysis (Roberts, 2001).

New method based on GPS and satellite photographs. Shannon and colleagues (2011 – see version under review in additional materials) described an approach that uses recent technology, in a situation where it was considered unsafe for interviewers to conduct an enumeration of the population in the field. They obtained geo-coded digital satellite maps (photographs) of chosen towns. Locations within the town were randomly sampled using Global Positioning System (GPS) coordinates. The points were located and a circle around the point was drawn on the photograph. The radius represented 20 m on the ground. Buildings within the circle were numbered, and one was randomly chosen. If there was no building within the circle, sampling continued. Interviewers subsequently used the maps to find the building. If the building was multi-residential, one household was randomly sampled. One adult was randomly chosen from the household to be the respondent. The conditional probabilities of selection can be computed at each step, and the product provides the overall probability of selection. (The fraction of the area of each town covered by the circles chosen is built into the calculation of probabilities of selection. Circles that do not contain any buildings are also included in this estimate.)

Strengths of the method are several: it reduces the work in the field for interviewers, minimizes their discretion in choosing buildings and is safer for them. It incorporates population (household) density, allows random selection with known probabilities, and minimizes the effect of 'pocketing' within clusters by spreading out the sample within the cluster. Enumeration of buildings is needed for only very small areas, a task that can be done before going into the field; and interviewers only need to enumerate households for multi-residential buildings. However, concerns may arise because a building could be sampled from different circles that contain different numbers of buildings so the probability of sampling the building may be very difficult to estimate. Also, circles surrounding the randomly chosen points may overlap and affect the probabilities of selecting any building. Boundary problems can occur if the GPS points and/or houses are close to the edge of town so that probabilities of selection are affected. Statistically, since the method is not self-weighting, sampling weights can vary so that variances of estimates are likely to be inflated.

We should note that others have explored the use of satellite photos to draw samples. For example, Lowther et al. (2009) geo-coded a 25 km^2 area in Lusaka, Zambia, attempting to identify all residential buildings. They identified over 16,000 such structures, and took a simple random sample of buildings. This process took roughly two weeks of work. This time and the associated cost can be balanced against the advantage of being able to draw a simple random sample.

Bostoen et al. (2007) noted that there had been little research on survey methods in difficult settings for two decades and a call for more research 'to further validate existing methods and to explore possible alternatives' was made by the Working Group for Mortality Estimation in Emergencies (2007: 8). More specifically, Checchi and Roberts (2008: 1027) stated that '[f]urther research, *particularly based on mathematical simulation*, is needed to optimise sampling choices' (emphasis added).

Simulation studies to date. As noted above, some simulations have been reported previously. Simulations (e.g., Harris, Lemeshow, Lwanga et al.,1991) show the EPI method is flawed when there is pocketing - and realistically it is often present. Thus the earlier simulations which ignored it may have provided false reassurance that the method 'worked'. Two studies have conducted simulations of sampling when census data have been available. Bennett and colleagues (1994) found that the performance of EPI (and several variants) depended on which of several outcomes was considered - for some there was little effect on precision or bias, but other variables showed more inefficiency. Yoon et al. (1997) computed efficiency ratios as the mean square error divided by total distance travelled. On this basis the EPI method performed well. However, the authors cautioned users about the bias associated with this method. Brogan (1994) also noted that simulations could not properly mimic interviewer judgment in choosing households. Some preliminary simulations of the GPS/satellite sampling method found little bias or variance inflation (Ahsan, 2011). However, it was based on only one virtual town, was limited in the range of characteristics considered, and only looked at household characteristics, not the individuals in those households.

One simulation study has explored relationships between variables using a modification of the EPI method comparing it with simple random sampling. Harris and Lemeshow (1991) estimated relative risks using a sampling method proposed by the WHO Global Programme on AIDS. When pocketing of infection was included, some differences were noted, but the authors considered them insufficient to recommend one method over the other. These papers are now fairly old. Despite the call noted above from Checchi and Roberts (2007) for further simulation studies, our updated literature search has found no papers that have done such studies.

In summary, various methods for sampling in difficult situations have been proposed. Cluster sampling is the basis for most (often using PPS or PPES) although methods of selecting units within clusters vary – in part because of local circumstances or available data. All the proposed approaches are known to have theoretical problems. For example, some bias is present for many methods, while others do not use properly random sampling. While limited simulation studies have shown that EPI performs well under ideal circumstances, in many realistic situations it is biased or inefficient. The new GPS/satellite method shows promise, but needs further testing. As well, several authors of simulation studies have noted that carrying them out took much computer time, but these were studies done some time ago. The increased computing power available today will allow much more complex - and realistic - situations to be examined.

Hence, **our overall objectives are (i) to conduct simulation studies to determine the bias and precision of different sampling methods in difficult settings and (ii) to determine the circumstances in which one method may be preferred over others.**

More specifically, we will address **three questions**:

(1) ***Can analytical methods be developed to improve estimates in the GPS/satellite approach?*** The first component to the study will not involve simulations. Initial indications are that analytical solutions to estimating the probabilities of selecting households may become intractable when circles from which they are chosen overlap. We will explore this further.

(2) ***How do modifications to the GPS/satellite method affect the bias and precision of estimates of population characteristics?*** It was noted above that sampling weights for selected households or individuals in the sample vary, leading to increases in estimates of variance. We will conduct simulations to determine whether modifications to the sampling method can lead to greater homogeneity of weights and lower variance estimates.

(3) ***Are there circumstances in which different methods perform better than others?*** The GPS/satellite method is new, and has not been directly compared with existing approaches. We will carry out simulations to identify the preferred method. We anticipate that this may depend on features of the population, and no one method may be universally optimal.

Methodology

Question 1: Can analytical methods be developed to improve estimates in the GPS/satellite approach?

The statistical properties of estimators are key to judging how effectively they provide reliable and stable estimates of the population characteristics of interest. They also supply quantitative indicators to establish their comparative merits when two or more estimators are available. The two main criteria are bias and variance. The aim is always to construct unbiased estimators with small variance. Naturally each of these quantities depends on the specific sampling design used, the size of the sample and the particular form of the estimation formula.

From the theory of sampling, we know that a general unbiased estimator for a population total is the Horvitz-Thompson estimator. The estimator can be readily adapted to estimate population means and proportions. An estimator of its variance is also available (e.g., Thompson, 1992: 49; Cochran 1977: 259-260). In applying this method, the challenge is to calculate the inclusion probability π_i for every population unit i. For the variance estimator, we also need to obtain π_{ij}, the probability of including both units i and j in the sample.

We will further develop analytical methods for calculating π_i and π_{ij} for the GPS/satellite photo sampling method based on the random selection of points for the geographic area of sampling and the radius r considered around the points selected. The selection of points will be completely at random using a uniform planar distribution. If the exact inclusion probabilities prove too complicated, analytical approximations to an appropriate order will be considered and the effect on bias and variance will be established. Likewise, we will obtain the same measures for the competing sampling methods such as EPI.

Question 2: How do modifications to the GPS/satellite method affect the bias and precision of estimates of population characteristics?

This includes writing a more detailed protocol than can be described here.

Creating virtual populations. The first step to answering both Questions 2 and 3 is to create virtual populations to reflect different situations. The broad outline of our approach is as follows: each population will consist of 1,000,000 people, subdivided into 250 units (cities/towns/villages) of varying sizes. We will (i) develop different populations with varying sizes of the main cities; (ii) locate buildings geographically in towns; (iii) place people in the households; (iv) allocate demographic characteristics to those individuals and (v) allocate (health-related) outcomes to individuals on the basis of their location and demographic characteristics.

We begin by determining the distribution of people in cities, towns and villages. Urban and rural populations will be incorporated with proportions in urban areas taking values of 15, 30, 40, 50, and 60%. These values are based on Uganda, Vietnam, Mozambique, Haiti and Botswana, which had urban proportions in 2010 of 13%, 30%, 38%, 52% and 61%, respectively (UNFPA, 2011).

We will use satellite imagery to incorporate realistic spatial distributions of buildings. We plan to geo-code residential buildings from satellite photographs of Low Income Countries or refugee camps. Each such building will have its location defined by East-West and North-South coordinates. Lowther et al. (2009) identified over 16,000 stuctures that appeared to be residential. When team members went to the locations, they found that 96% of them or the nearest neighbour to the right were indeed residences. They did not report if they checked whether their process excluded any residential buildings, but noted that their broad inclusion criteria ensured they would not miss people living in ‘unusual housing structures’. For our purposes, we believe this level of accuracy is sufficient.

We will use Google Earth for the relevant satellite imagery. It is free and readily accessible via the internet. In addition, importing randomly-chosen points into Google Earth tools can be conducted easily with open access free software (e.g., GPS Visualizer: http://www.gpsvisualizer.com/map_input?form=googleearth). Based on experience with a survey in southern Lebanon, there may be some difficulty with poorer resolution in rural areas. We found that privately-purchased images were significantly better than Google Earth’s photos, so for areas for which it is necessary, we will purchase high-resolution photographs from, e.g., DigitalGlobe Services, and cross-reference the maps for accuracy. Once we have the coded images, we will quantify patterns of the spatial distribution of houses (based on the centroids of buildings presumed to be residential). Using well-established methods (Baddeley et al. 2005; Diggle, 2003), we can estimate the parameters of spatial point process models from the set of building centroids – these models define building density, tendency of buildings to be clustered or regularly spaced, and larger-scale spatial variability in density. These fitted models will then form the basis for simulating virtual populations – sets of new households with the same spatial parameters, possibly over a much larger area than the original imagery sample. We will identify different values for parameters from maps of different areas, e.g., urban vs rural, or in different countries. We will use these different values in generating the set of virtual populations.

In urban areas, we will include multi-residential buildings, basing the proportions of such buildings on available datasets. These datasets are from south Lebanon, Haiti, Lebanese refugee camps and Gaza. (Even in the refugee camps, there are multi-residential buildings.)

We will also vary the mean household size, which will affect the number of buildings. We can use the datasets noted above to build in realism when we do this – our program will create households that have similar compositions to the real samples. The mean size of households in those studies is roughly 5, and we will use mean sizes of 4, 5, and 6.

Having created the population of buildings, we will 'fill' them with people. Different household compositions will be created, and each household member will be 'given' an age and sex. To do this we will use our available datasets to calculate the mean age (and standard deviation, SD) of adult men and women in the households, and the mean and SD of the difference in ages of men and women in the household. We will compute the mean and SD of the difference in mean age between the adults and oldest child within households. We will also examine the age-spacing of children within households. This information will give us appropriate values to use for generating the household members when there are one or two adults in the household.

We appreciate that there will be complications when, e.g., there are more than two adults in a household. For our purposes, we will assume that 'couples' are heterosexual, and will consider the male and female adults closest in age to be the couple of the household and use their data for our computations. We will note the proportion of times that households include more than two adults, and will incorporate this into our virtual populations. Similarly, we will account for, e.g., occasions when children are raised by their grandparents, so the age gap between the adults and the oldest child will be large. We will then allocate characteristics to each individual. We will include both categorical and continuous variables. For example, a categorical variable may be whether or not an individual has a chronic disease or has experienced a particular illness in the recent past. We will take the population proportions to be 0.01, 0.05, 0.1, 0.2, 0.3 and 0.5. These variables will also have differing degrees of 'pocketing', reflecting, e.g., the infectiousness of the illness. The pocketing will vary within the town, depending on housing density, and can be modeled either by a contagious process -- increasing the probability of infection of people living close to an infected person – or by simulating the presence of pockets at specific geographic locations. Within households this probability will be increased even more. Since mortality is often measured in surveys, we will include people who have recently 'died', varying the proportion of the population that has died – values will be 0.005, 0.01, 0.025, 0.05, 0.075, and 0.1.

Continuous variables will have different distributions. Thus, income or wealth will have a skewed distribution (e.g., log normal), while a measure such as height (height-for-age in the case of children) will have a normal distribution. We will vary, e.g., the variance of income to allow for societies with greater or lesser degrees of inequality. We will incorporate spatial autocorrelation among the variables to allow for a gradient in the prevalence of characteristics, and will build in some associations between variables. This will allow us to model relative risks between 'exposure' and 'outcome' variables.

Some pairs of binary variables will have either no relationship (relative risk = 1) or a known relationship (relative risks = 1.5, 2.0 and 3.0). Pairs of continuous variables will be created with bivariate normal distributions showing no, small, medium and strong correlations (ρ = 0.0, 0.2, 0.5, 0.8, respectively). Checks will be made to ensure that the populations indeed show the characteristics we wish to apply to them.

Burton et al. (2006) state that factors (proportion with disease, strengths of relationships between variables, spatial point process parameters, etc.) are usually examined in a fully factorial arrangement. In our case, this would lead to too many virtual populations (in the six figure range) for our study to be feasible. One possibility would be to compare the different sampling methods for different populations resulting from creating a 'standard' population, using the mid-levels of each factor, and then varying the levels (values) of factors, one factor at a time. This would have identified factors whose level did not affect the relative merits of each sampling method, so that only one level (e.g., mid-point) of that factor would need to be considered. However, this would not identify interactions among the factors. Instead, we propose to use a Latin hypercube approach, in which the k factors and their levels constitute a k-dimensional space, and sampling is done from that space in a way that ensures adequate marginal coverage of each factor, but still allowing assessment of potential interactions (e.g., Blower and Dowlatabadi, 1994).

We will construct the virtual population files to reduce the computer time required. For example, we do not want to have to search the whole file (or even all households in a town or city) to identify buildings close to a given point. Rather we will situate buildings within a grid system, so that grid squares within a certain distance of the reference point can be quickly identified, and any search would only have to consider buildings in the (limited number of) squares close by.

Simulations of sampling. We will take 10,000 samples from each virtual population using the GPS/satellite photo method. This will select clusters via PPS, and sample households within the chosen clusters. Overall sample sizes will range from 210 (the value used in the standard EPI approach) to 900 (the value typically used in EPI-like sampling when the expected prevalence is low, consisting of 30 clusters of 30 units). We will also take samples of 1,500 and 3,000 units to understand the effect of a larger sample size than is typically used.

We will also do these simulations using several modifications of the method. For example, when housing density is higher or lower than the norm in certain geographic areas of the virtual population, we will use circles of smaller or larger diameters, respectively, so that the sampling weights are more homogeneous. Another variant, which builds on other methods, will superimpose a grid of squares on the virtual photograph and randomly sample a certain number of squares. We will then sample all buildings within the squares (as a cluster) or randomly choose one building.

In practice, this last method could also be used in situations where it is safe for interviewers to conduct enumerations of the population. Given the affordability of phones or other instruments with GPS capability, interviewers could be directed to enumerate all households within a square defined by GPS points. In practice, the determination of whether a building is 'inside' a specified area on a photograph is made by finding the 'mid-point' of the building. There is likely some small error in doing this. Information on all households could be entered into the unit, and the unit could choose one at random to be surveyed. This mimics the approach in which the enumeration is done by the research team before interviewers go into the field, and could also be used when satellite photographs are not sufficiently up-to-date, as in refugee camps or many post-disaster settings. Vanden Eng and colleagues (2007) have successfully used such technology – teams used handheld devices equipped with GPS units to map all households in selected areas; the devices were programmed to choose a random sample and direct the teams to the households picked. In practice small squares might be enumerated unobtrusively even in more dangerous conditions, and we will explore the effect of sampling smaller squares.

Analysis: For each sampling method, the following will be done. We will calculate sampling weights. Sample proportions (means for continuous variables) and their variances will be computed, allowing for weights and any clustering. We will also compute the empirical sampling variance from the means of the simulated samples as well as the proportion of times when the confidence interval around the sample mean includes the actual mean for the virtual population. We will check the efficiency of the method by comparing the mean of the variances of the samples with the theoretical value for simple random sampling (the variance of the virtual population divided by the standard error of the sample size). Examining these results across the different sampling methods will show which, if any, of the sampling methods is optimal in showing the least bias and/or smallest variance. Bias and variance will be combined to estimate Mean Square Error (MSE). This will assist comparisons when one method has smaller bias, but the other has smaller variance.

We will not examine costs of doing the surveys for Question 2, since the variants of the GPS/satellite photo method are likely to involve very similar costs. This, though, will not be so for the methods examined in Question 3, and later we discuss ways to examine the trade-offs.

Question 3: Are there circumstances in which different methods perform better than others?

Creating virtual populations. We will use the same virtual populations as for Question 2.

Simulations of sampling. The sampling methods will include all those we described in the literature review, which will be compared with the optimal GPS/satellite photo approach identified in answering Question 2. Specific sampling methods are as follows (several are from Bennett et al.,1994):

- Simple random sampling of buildings
- Simple random sampling of people
- The standard EPI approach
- The EPI approach taking the 3rd nearest neighbor as the ‘next’ household at any step
- The EPI approach taking the 5th nearest neighbor as the ‘next’ household at any step
- Each community selected is divided into 4 quadrants, and the EPI method is used to select a quarter of the sample from each quadrant
- Half of the sample is selected at the centre of the community and half at the periphery - specific instructions are given in Bennett et al. (1994)
- The two grid approaches used by Grais and colleagues (2007)
- The approach of Roberts (2000)
- The optimal GPS/satellite photo approach (or more than one depending on results from Question 2)

We have distinguished between simple random samples of buildings and people, as enumerations in the field will typically list buildings, not people. As well, while the results from simple random sampling of people should be completely predictable from basic theory, we will simulate such sampling as a check that the programming has been properly done.

We recognise that some of these sampling methods (e.g., taking the 5th nearest neighbor) rely on interviewer judgment in the field. It will not be possible to program that directly into the simulations. Rather, we will use an algorithm to choose households. To the extent that interviewer judgement creates bias, any systematic approach that we adopt is likely to show less bias than the real-life setting. We will interpret our results accordingly.

Analysis: The analysis will follow the same overall approach used to answer Question 2. For each of the sampling methods, the parameters of interest (means/proportions and their variances, and total error) will be computed. We will estimate the proportion of times when the confidence interval around the sample mean includes the ‘true’ value. The sampling efficiency will be calculated. All methods will be compared to establish which is optimal. We anticipate that no one method will be universally optimal. For example, if there is no pocketing, or the degree of pocketing is small, the standard EPI method may be acceptable, but another method may be superior if there is pocketing.

In making recommendations, we will account for the simplicity and potential cost of each approach. Thus even if one method might have slightly less bias and smaller variance than another, the latter may be preferred if it is easy and cheap to implement and the survey must be done quickly and on a limited budget.

There is a trade-off between the costs of a survey and the mean square error (MSE - a combination of bias and variance). We anticipate that our simulations will produce good estimates of the MSE. The costs are a mix of factors such as time required to travel to and within clusters; time to complete an interview (partly depending on the number of items in the questionnaire); number of clusters and number of sampling units per cluster; etc, which may be estimable in advance of the survey. Macintyre (1999) used actual costs of two surveys done a short time apart in Ecuador asking about family planning acceptance; one was a 'regular' national survey, the other a 'rapid assessment'. She found as might be expected that the latter was much more cost-efficient, but biased. (We have found no other formal analyses of costs vs quality of information.)

As well there are some costs that may not be quantifiable *a priori* including likelihood of errors in any enumeration or number of return visits needed to complete an interview. Also difficult to quantify are such factors as the simplicity of the overall procedure, especially in relation to the skills and experience of the research team and its ability to analyze properly data from more complex survey designs. We think it unlikely that we can produce definitive advice on the balance between these factors. Nevertheless, we expect that we can use diagrams (small multiply conditioned plots -- similar to scatterplot matrices) showing the relationship between various factors and the MSE vs (quantifiable) cost trade-off.

Software and Hardware: For Questions 2 and 3, we plan to use MATLAB and R as appropriate, since they can be readily programmed and have excellent capabilities for conducting simulations. We plan to access SHARCNET (Shared Hierarchical Academic Research Computing Network), which will have the power to conduct the very intensive computing required. The Appendix to this proposal includes an e-mail from the system administrator for SHARCNET at McMaster.

Further remarks on the simulations

We chose 1,000,000 as the population size since it is large enough to represent many populations in which surveys are done. As a sensitivity analysis, we plan to create a much smaller population to understand how the sampling methods perform in them. Since our programming will be generic, it will be straightforward to do this, and the computer time required to create this smaller population will be much less than for our main analyses.

Our choice of the number of samples to select - 10,000 - is somewhat arbitrary and at the higher end of most simulation studies. Burton et al. (2006) suggest basing the number on the desired level of accuracy for the estimate of interest. However, we will be comparing methods, and using various types of distributions. Further, most of the methods apply cluster sampling so, while we could assume certain values for the design effect *deff*, we cannot with any certainty incorporate a value for the variance. Thus, we prefer at this stage to err on the side of caution, and propose taking more rather than fewer samples than might be necessary.

Burton et al. (2006) also discuss the starting seeds for random number generation (RNG). The packages we plan to use, R and MATLAB have very rigorous and well-tested methods for RNG, so we are confident that this will be properly done. We plan to store all the random number seeds so we can reconstruct the simulations if necessary.

Strengths and limitations of our study

We will conduct the most comprehensive examination to date of the sampling methods. Indeed, many have not been compared with others, and computing power has increased greatly since the last reports of simulations. This will enable us to explore a much wider variety of situations than was possible before. At the same time, we recognize that no simulations can be entirely realistic. We have limits on the number of parameters we can incorporate into the virtual populations. Still, by judicious choices we can create realistic settings which should mirror actual populations. We will experiment with our creation of the virtual populations, and expect that there will be some additional levels of realism/complexity that we can introduce.

Ideally, we would conduct our simulations on real populations. Two of the simulation studies noted in the literature review used census information, albeit on relatively small sub-populations - 4,320 eligible children in Bennett et al. (1994) and 4,297 eligible children in Yoon et al. (1997). However, these datasets do not appear to be available any longer. If we become aware of suitable datasets we will contact their owners to ask if we can test the various sampling methods with their real data.

Importance/Relevance of this work

As noted in the literature review, the EPI method was developed to estimate vaccination coverage, an important piece of knowledge for global health, especially to understand how well the United Nation's Millennium Development Goal 4 to reduce child mortality (United Nations, 2000) is being met. Similarly, to understand Goal 1 - eradicate extreme poverty and hunger (United Nations, 2000) - requires surveys of food security. In both cases, the surveys are often done in difficult settings, i.e., areas of the world where information on the population is limited.

An example of the importance of high quality evidence follows from the January 2010 earthquake in Haiti and its aftermath. Shortly after the quake, the Haitian government reported that 230,000 people had died as a result of the quake. They later stated that the number was closer to 300,000. One of us (HSS) was a co-investigator of a survey in Haiti shortly after the quake to establish mortality as well as the food and other needs of the population. We estimated that in the capital Port-au-Prince alone there were about 156,000 excess deaths (95% CI 134,000 – 178,000) during the quake and in the six weeks after (Kolbe et al. 2010). A reporter Melissen (2010) claimed that the death toll was under 100,000 based on 'observation and research on the ground'. A much more recent study for the US Agency for International Development (USAID) claimed the death toll was between 46,000 and 85,000 (Daniel, 2011). The methods for the USAID study were criticized by us in a Los Angeles Times OpEd, in particular because the sampling after the event was flawed (Muggah et al., 2011). Importantly, the USAID report also claimed that the number of people still living in tent camps was much lower than had been previously thought, which had major implications for aid and the reconstruction efforts. While these other estimates were not based on the methods we will compare in the simulations, they illustrate the serious consequences of survey estimates. It is crucial to use the best possible methodology to ensure aid money and the limited resources of aid groups are optimally spent.

This project will contribute to Canada's efforts and leadership in global health, including especially the work to improve child and maternal health.

Timeline

Months 1-6: Analytical work for Question 1.

Months 1-2: Prepare detailed protocol for simulations; hire programmer.

Months 3-8: Orient programmer to project; develop population parameters based on prior datasets; create virtual populations using Latin hypercube sampling; test populations to ensure criteria for creation are fulfilled.

Months 9-14: Run simulations for Question 1; identify main effects and low-level interactions. Complete simulations. Begin writing report.

Months 15-20: Run simulations for Question 2; identify main effects and low-level interactions. Complete simulations. Continue report from Q1, begin report from Q2.

Months 21-22: Continue writing papers and reports, including additional analyses by programmer. Draft guidelines and documentation for practice.

Months 23-24: Complete writing of papers. Develop guidelines and documentation for practice.

Research team

Dr Harry Shannon is an experienced biostatistician. He has worked with colleagues in Michigan using the GPS/Satellite photo sampling method – his responsibility covered the statistical aspects of the survey, including design of the sampling approach, estimation of the sample weights, and analysis of the data. In this project, as Principal Investigator, he will be involved in all aspects of the work. He will take primary responsibility for the simulations for Questions 2 and 3, including supervising the statistical programmer.

Dr Román Viveros-Aguilera is a statistician with over 20 years of experience in developing methodology for biostatistical questions in the modeling and analysis of survival data as well as in industrial statistics. Relevant practical experience includes the development of a sampling design for the Ontario potato industry and analysis of survey data from a large survey on drug users in the US. He will develop analytical methods to calculate weights, biases, variances and confidence intervals from GPS/satellite photo sampling and other sampling methods studied.

Dr Ben Bolker is a mathematical biologist and biostatistician with extensive and diverse experience in spatial simulation and spatial statistics. He has worked on computational and analytical models of childhood disease in developed countries, including extensive deterministic and stochastic simulation modeling; analysis of computational models of community dynamics in forests and carbon flow in terrestrial ecosystems; and spatial dynamics and spatial statistics of ecological communities. He will advise on the development of practical and efficient simulation methods, and co-supervise the statistical programmer.

Training

We plan to bring in a graduate student (or maybe two students, one per year) in statistics or biostatistics who will carve out pieces or extensions of the work for his/her thesis. Possible topics include: (a) linking the GPS coordinate system with information available in Google and other electronic sources to facilitate the application of the GPS sampling methodology, and (b) study quantitatively the effect on the GPS methodology when sampling is done WITH replacement and WITHOUT replacement, particularly in small populations. The student will be introduced to global health, sampling, and simulations, as well as presenting work in non-technical language.

Knowledge Transfer

We anticipate that each Question in this project will lead to one or two scientific papers as well as presentations at relevant meetings. We will consult with NGOs, etc. who do these types of surveys for their views on what information and presentations are most useful for them.

We plan to develop guidelines for those who conduct surveys in difficult settings, e.g., non-governmental or quasi-governmental organizations. The guidelines will describe which method is best in which circumstances, for example whether it is safe for interviewers to enumerate the population. The guidelines could be summarized in a decision tree, which will lead potential surveyors through a series of questions to the preferred method. In some instances the final choice may be conditional on other factors, so our aim will be to help the surveyor limit the possible choices. We also expect to produce diagrams (small multiply conditioned plots -- similar to scatterplot matrices) showing the relationship between various factors and the MSE vs (quantifiable) cost trade-off. The documentation will outline in lay language the exact steps required to carry out a survey using the chosen method(s). As well, we intend to develop non-technical presentations explaining the results and their implications for surveys in the field. The resulting products will be made available on a web site (and some hard copies will be printed).

REFERENCES

Ahsan Z (2011). Simulation study on new sampling method when data is limited. Master's thesis. Hamilton: McMaster University.

Baddeley A, Turner R (2005). Spatstat: An R Package for Analyzing Spatial Point Patterns. Journal of Statistical Software 12: 1-42.

Bennett S, Radalowicz A, Vella V, Tomkins A (1994). A computer simulation of household sampling schemes for health surveys in developing countries. International Journal of Epidemiology 23: 1282-1291.

Bennett S, Woods T, Liyanage WM, Smith DL (1991). A simplified general method for cluster-sample surveys of health in developing countries. World Health Statistics Quarterly 44: 98-106.

Blower SM, Dowlatabadi H (1994). Sensitivity and Uncertainty Analysis of Complex Models of Disease Transmission: An HIV Model, as an Example. International Statistical Review / Revue Internationale de Statistique , 62: 229-243.

Bostoen K, Bilukha OO, Fenn B, Morgan OW, Tam CC, ter Veen A, Checchi F (2007). Methods for health surveys in difficult settings: charting progress, moving forward. Emerging Themes in Epidemiology 4: 13.

Brogan D, Flagg EW, Deming M, Waldman R (1994). Increasing the accuracy of the Expanded Programme on Immunisation's cluster survey design. Annals of Epidemiology 4: 302-311.

Burton A, Altman DG, Royston P, Holder RL (2006). The design of simulation studies in medical statistics. Statistics in Medicine 25: 4279-4292.

Checchi F, Roberts L (2008). Documenting mortality in crises: what keeps us from doing better? PLoS Medicine 5:7, e146.

Cochran WG (1977). Sampling techniques, 3rd edition. New York: Wiley.

Daniel T (2011). Death toll in Haiti far lower than originally estimated: U.S. report. http://www.thestar.com/news/world/article/999651--death-toll-in-haiti-far-lower-than-originally-estimated-u-s-report. Accessed 9 February 2012.

Diggle PJ. 2003. Statistical Analysis of Spatial Point Patterns. 2nd ed. Hodder Arnold.

Earn DJD, He D, Loeb MB, Fonseca K, Lee BE, Dushoff J (2012). Effects of school closure on incidence of pandemic influenza in Alberta, Canada. Annals of Internal Medicine 156: 173-181.

Grais RF, Rose AMC, Guthmann J-P (2007). Don't spin the pen: two alternative methods for second-stage sampling in urban cluster surveys. Emerging Themes in Epidemiology 4: 8.

Harris DR, Lemeshow S (1991). Evaluation of the EPI survey methodology for estimating relative risk. World Health Statistics Quarterly 44: 107-114.

Harris DR, Lemeshow S, Lwanga SK. Chin J. Duchesneau R. (1991). Evaluation of a standardised survey design proposed for use in epidemiological research on AIDS. International Journal of Epidemiology 20:1048-1056.

Henderson RH, Davis H, Eddins DL, Foege WH (1973). Assessment of vaccination coverage, vaccination scar rates, and smallpox scarring in five areas of West Africa. Bulletin of the World Health Organization 48: 183-194.

Henderson RH, Sundaresan T (1982). Cluster sampling to assess immunization coverage: a review of experience with a simplified sampling method. Bulletin of the World Health Organization 60: 253-260.

Kolbe A, Kolbe AR, Hutson RA, Shannon H, Trzcinksi E, Levitz, N, Puccio M, James L, Miles B (2010). Mortality, crime and access to basic needs before and after the Haiti earthquake: A random survey of Port-au-Prince households. Medicine, Conflict and Survival 26: 281-297.

Lemeshow S, Robinson D (1985). Surveys to measure programme coverage and impact: a review of the methodology used by the expanded programme on immunization. World Health Statistics Quarterly 38: 65-75.

Lemeshow S, Tserkovnyi AG, Tulloch JL, Dowd JE, Lwanga SK, Keja J (1985). A computer simulation of the EPI survey strategy. International Journal of Epidemiology 14: 473-481.

Macintyre K (1999). Rapid assessment and sample surveys: trade-offs in peecision and cost. Health Policy and Planning 14: 363-373.

Melissen HJ (2010). Haiti quake death toll well under 100,000. http://www.rnw.nl/english/article/haiti-quake-death-toll-well-under-100000. Accessed 9 February 2012.

Milligan P, Njie A, Bennett S (2004). Comparison of two cluster sampling methods for health surveys in developing countries. International Journal of Epidemiology 33: 460-476.

Muggah R, Kolbe A, with Hutson R, Shannon H (2011). Haiti: Why an accurate count of civilian deaths matters. http://www.latimes.com/news/opinion/commentary/la-oe-muggah-haiti-count-20110712,0,7935632.story. Accessed 9 February 2012.

Roberts L (2000). Mortality in eastern DRC. Results from five mortality surveys by the International Rescue Committee. http://www.rescue.org/sites/default/files/resource-file/IRC_SUR_MOR_DRC_2000.pdf. Accessed 9 February 2012.

Shannon HS, Hutson R, Kolbe A, Stringer B, Haines T (2011). Choosing a survey sample when data on the population are limited: a method using Global Positioning Systems and satellite photographs. Submitted for publication.

Thompson SK (1992). Sampling. New York: Wiley.

Turner AG, Magnani RJ, Shuaib M (1996). A not quite as quick but much cleaner alternative to the Expanded Programme on Immunisation (EPI) cluster survey design. International Journal of Epidemiology 25: 198-203.

UNFPA (2011). State of World Population 2011. United Nations Population Fund. http://foweb.unfpa.org/SWP2011/reports/EN-SWOP2011-FINAL.pdf Accessed 9 February 2012.

United Nations (2000). Resolution adopted by the General Assembly – United Nations Millennium Declaration. http://www.undemocracy.com/A-RES-55-2.pdf. Accessed 9 February 2012.

Vanden Eng JL, Wolkon A, Frolov AS, Terlouw DJ, Eliades J, Morgah K, Takpa V, Dare A, Sodahlo YK, Doumanou Y, Hawley WA, Hightower AW (2007). Use of handheld computers with global positioning systems for probability sampling and data entry in household surveys. American Journal of Tropical Medicine and Hygiene 77: 393-399.

World Health Organization (1991). Training for mid-level managers: the EPI coverage survey. Geneva: WHO Expanded Programme on Immunization. WHO/EPI/MLM/91.10.

Working Group for Mortality Estimation in Emergencies (2007). Wanted: studies on mortality estimation methods for humanitarian emergencies, suggestions for future research. Emerging Themes in Epidemiology 4: 9.

H.S. Shannon. Sampling methods: a simulation study. Research Proposal Appendix. $121,467

Yoon SS. Katz J. Brendel K. West KP Jr (1997). Efficiency of EPI cluster sampling for assessing diarrhea and dysentery prevalence. Bulletin of the World Health Organization 75: 417-26.