A MODEL FOR SIMULATING INTERVIEWER TRAVEL COSTS FOR DIFFERENT CLUSTER SIZES

Julia L. Bienias, Elizabeth M. Sweet, and Charles H. Alexander
Bureau of the Census, Statistical Methods Division, Room 3725-3, Washington, DC 20233

KEY WORDS: Survey cost, modelling, interviewer travel, cluster sampling

BACKGROUND

In redesigning a complex sample survey, the field costs associated with different designs must be considered. For surveys conducted at least in part by personal visit, the interviewers' travel among sample units may account for a substantial portion of the total field cost. Typically, sample units are clustered geographically to reduce interviewer travel cost, especially when the increase in variance associated with clustering is not too great.

The cost savings associated with clustering are diminished when the interviewer is unable to obtain completed interviews from more than one sample unit on a particular visit to a cluster and must make more visits to the cluster. This may be becoming increasingly common for household surveys, as more people work outside of the home and are therefore less likely to be home when an interviewer visits.

In deciding among different possible cluster sizes for a design, one needs estimates of the differences in travel costs associated with different designs as well as an estimate of the proportion of total variable cost accounted for by travel. By total variable cost we mean total cost that is expected to vary as the sample size changes, excluding such "fixed" costs as administration. We are envisioning interviews as being conducted on one or more trips. A "trip" is the travel from a central point (e.g., the interviewer's home) to one or more sample units and back to the central point. Total travel is composed of home-to-cluster travel (the beginning and end travel for the trip), between-cluster travel, and within-cluster travel. In our discussion here we will assume within-cluster travel is negligible. This might not always be true (e.g., in rural areas).

In redesigning current surveys, data may be available on the travel and interviewing costs associated with the current sample design. However, there typically are no empirical data pertaining to an alternative sample size and/or cluster size. Costs for alternative designs or for new surveys may be estimated by subject-matter experts (e.g., regional field supervisors) and/or with a model.

A commonly-used travel cost model proposed by Hansen, Hurwitz, and Madow (HHM; 1953, p. 274) assumes the clusters are located at the intersections of a rectangular grid the size of the geographic area involved. Total travel is computed solely from the geographic size of the area and the number of clusters.

This model is unrealistic for many applications. Multiple visits to a sample unit ("callbacks") are often needed to obtain a completed interview. Modelling callbacks to clusters is complex because the travel associated with visits changes with cluster size. That is, the larger the cluster size, the more likely the interviewer can visit several units with one trip to the cluster. These extra visits can increase the number of interviews per trip, thus potentially decreasing the travel associated with making callbacks to obtain interviews.

HHM present a method for expanding their model to estimate the travel involved when callbacks are required (see HHM, p. 275). To use it, one must have knowledge of the proportion of clusters that require one, two, three, etc., visits to complete interviews for all the sample units in the cluster. These proportions are the probabilities of "completing the cluster," or completing all the (remaining) cases in a cluster, on a particular visit conditioned on the number of prior attempts. For example, the probability of completing a cluster might be .6 on the first visit, .5 on the second visit (given that it wasn't completed on the first visit), .3 on the third visit, etc. The travel associated with first visits is added to that associated with second visits and so on to obtain an estimate of total travel.

We cannot use the result in this form for our surveys, because we do not have the detailed data needed to estimate the various conditional probabilities of completing a cluster, even for the cluster size we now use in the field. In addition, it is assumed in HHM's expanded model as in their simple model that on each trip, all the clusters containing units requiring interviews are visited.

Assuming travel occurs in such complete "circuits" could result in modelling trips of unreasonable length, especially for surveys with very long interviews. In reality, for example, an interviewer might actually travel half a circuit (visit half the units on one trip), taking twice as many trips total. This is important because we wish to include home-to-cluster travel in our estimate of total travel, and the amount of home-to-cluster travel is directly affected by the number of trips required to complete the interviewing assignment. It is difficult to estimate an average value for home-to-cluster travel, especially for different cluster sizes, thus it is preferable to measure it directly from the model.

The HHM model has certain features that we wished to modify. First, the model assumes that clusters being visited on a given trip are uniformly spread around the interviewing area, rather than being grouped more or less closely as with a city and its suburbs. Second, in the expanded model, the exact locations of the units are not considered as fixed; an average distance is recomputed as though the design included only the remaining clusters uniformly distributed around the interviewing area. Third, the model does not allow for different patterns of trips or more home-to-cluster travel. Fourth, there is no adjustment made for different interview lengths. For a short survey such as the Current Population Survey, which takes approximately fifteen minutes to complete, it may be reasonable to expect an interviewer to attempt all units on one trip. However, for a longer survey, such as the Survey of Income and Program Participation (approximately 2 hours), it would be unreasonable to expect an interviewer to complete a cluster of size four, much less an entire caseload, without taking a break that would involve extra travel. Fifth, the model as presented by HHM assumes all units are interviewed eventually. For some surveys it may be reasonable to accept some never-completed units.

We propose using a simulation as an alternative to the HHM model and approach. For a particular simulation, we will use a fixed probability p that a sample case can be completed on a given visit. Such a probability p is, in effect, an average of the conditional probabilities required in the HHM model. We assume this p is independent of the number of previous attempts to obtain an interview and cluster size. For simplicity we assume that a case is entirely completed (for example, interviewed), on one particular visit. We include a fixed minimum amount of contact time for each visit that does not result in a completed case. The number of trips and the amount of time allowed per trip are additional parameters in the model.

Under this model, the travel associated with different cluster sizes and different probabilities of completing a case can be directly estimated through the use of a computer simulation. The travel associated with a particular set of input parameters is averaged over 1000 iterations of the program. Although such an approach carries with it its own set of assumptions, we believe it incorporates many important features of the problem. In particular, we randomly position the units for each iteration of the simulation, allowing the distance and the number of callbacks to vary more realistically by cluster size. Second, although our model also assumes the interviewer visits all the units still not completed before visiting the remaining units, we allow for multiple trips to complete a circuit. That is, we do not assume on each trip the interviewer visits each and every eligible unit.

The general approach is to estimate p from comparing available data on a particular design to simulated data and use this estimate to generate simulated data for an alternative design.

In this paper we present results from such a travel simulation program. Initially, we describe the assumptions of the simulation. Next, we compare our simulation to the HHM model and present the effects of varying many of the input parameters. Examining these gives insight into the complex interaction of various factors determining the relationship of cluster size and travel. Finally, we show the use of the travel simulation in actual cost modelling for the redesign of the National Crime Survey.

METHOD

The simulation, developed at the Bureau of the Census in Fortran 77 on a VAX minicomputer under VMS, uses a limited number of rules to model an individual interviewer's behavior while travelling around a particular interviewing area. The model uses a modified "nearest neighbor" algorithm; that is, the interviewer travels from the current position to the closest cluster in most situations.

Within a user-defined circular interviewing area of any size centered at the origin, the program positions the user-defined number of clusters of any size. These can be assigned at random points within the entire circle. Alternatively, clusters can be randomly assigned separately within two areas of the circle, where part of the circle is defined to have a greater density of clusters, such as one would find in an urban area.

Keeping track of time and distance, the imaginary interviewer travels from home, (the center of the circle) to a randomly-chosen cluster. We used such a random start to reflect the variability among (and within) individual interviewers' travel patterns. Total distance and total time required to reach the clusters are recorded. The distance between points (x,y) and (a,b) is calculated as $|x-a| + |y-b|$. We used thirty miles per hour for all simulations reported here.

Once in a cluster, each unit is "visited" (an interview is attempted) at all the units within that cluster if time permits. A visit could result in a completed case or not (see below). At each sample unit, time is recorded for either a completed interview (user input time) or an attempted interview (2 minutes). No time or distance is recorded for within-cluster travel.

The interviewer then proceeds to the next closest cluster, again recording time and distance. This process continues until the time it would take to complete another interview exceeds the user-defined trip duration (e.g., 3 hours). At that point, the interviewer returns home, a new trip begins, and the process begins anew. The simulation ends when the user-specified maximum number of trips is reached or when all units have been completed.

Unique to this model is the fact that visits to sample units are considered as independent Bernoulli trials. For each visit, the program compares a random number between 0 and 1 to a user-defined probability p of completing a sample case on a particular visit to determine whether or not a case is completed. By "completing" we mean both interviews and other forms of completing a contact with a respondent, such as a confirmed refusal. We implicitly assume total travel to refusals follows the same pattern as to completed interviews. Because our surveys have relatively low refusal rates, this assumption is not critical.

Failure to complete a case on a particular visit necessitates a callback and the unit is visited again later, time permitting. In most cases, the interviewer visits all units once initially and then visits all the remaining units once, and then re-visits all the units still remaining, and so on.[1] We realize an actual interviewer might not visit every sample unit once before attempting callbacks, but we felt this method of visiting all eligible units in "circuits" would be more realistic on average than any complex "optimal" behavior.

At the end of the user-allotted number of trips, all units are considered either "completed" or "never completed." All of the results reported here were based on the average of 1000 iterations of the simulation.

RESULTS

Plan of Analysis

Our computer program allowed us to vary the following parameters:

* the probability of completing a sample case on a particular visit (p)
* the total number of trips allowed (which could correspond to the number of days allowed for field interviewing for a particular survey; it would be greater if more than one trip is taken in a day)
* the maximum duration for any given trip (assumed constant for all trips; this does not include the time required to return home after visiting the last unit)
* the travel speed
* the length of time for each completed case (interview length)
* the density type (uniform or not)
* the area of the circle (interviewing area)
* the cluster size, and
* the number of clusters.

Ideally, to model a particular survey, we would prefer to have data on all the parameters for each interviewing area (as they undoubtedly vary by area), simulated the travel for each area, and summed over the areas using different cluster sizes. However, such data were not available for our surveys.

An alternative is to use average values for the parameters. It is not clear that a single simulation using such averages would accurately represent the total over different areas with different combinations of parameters. Moreover, we do not have reliable recent data for the average values of several key parameters. For a survey such as the National Crime Survey, we do have reasonable estimates for in-house interview length, the population density and square mileage of particular sample areas, the average cluster size, and the number of clusters. We can guess at a reasonable range of values for the total number of trips and duration of each trip, given we know the total travel time.

One critical piece of information we lack is the value of p (or more generally the HHM model's conditional probabilities, which in our model are implied by a value of p). The other critical missing piece is the average travel time. We do have data on total interviewer time but cannot split this up into in-house interviewing time, travel time, and other time (editing of forms, listing operations, etc.). For most of our surveys, time-and-motion studies have not been conducted to produce such data. The Census Bureau's Field Division is working to develop better information systems which we expect will make more information of this type available for future survey research.

Our present approach, however, was to simulate total travel for the cluster sizes of interest (principally 1, 2, and 4), varying the other parameters over their reasonable ranges. We hoped to find that the relative effect of cluster size on travel was fairly constant as other parameters varied. To make the problem manageable, we fixed several of the parameters for most of the simulations such as total personal-visit sample size (because this is roughly the same for all interviewers by design). We fixed the travel speed at 30 mph, a plausible value; because we planned to vary the other time and distance parameters, there was nothing to be gained by varying travel speed too. Total number of trips varied only over a small range, because this is fairly well determined by the number of days allowed to complete the survey. Also, much greater variation in the total number of trips gave unrealistically high or low proportions of "never completed" cases.

The principal variables were p, the area of the circle allowing the maximum number of trips which could be taken, the maximum duration of a trip, and the length of time to complete an interview.

For each combination of parameters we considered two (or three) cluster sizes. We considered primarily the following output from the model: total miles travelled and rate of never-completed cases. In addition, as we were particularly interested in the cost associated with changing from a design using clusters of size four to a design using clusters of size two, we considered the percent increase in travel over clusters of size four.

After discussing the case of the uniform population density area in detail, we turn to examining a case study involving an interviewing area actually in sample for the National Crime Survey (NCS).[2] Much of the following discussion deals with the problem of selecting simulation parameters so as to avoid unrealistic artifactual effects due to the interaction of the parameters.

Uniform Density Interviewing Area

In the simplest situation (described by HHM and above), clusters are uniformly distributed around the interviewing area, the probability p that an interview is completed on a given visit is one, and there is sufficient time to visit and interview all the sample units on one trip. In this situation, travel is inversely related to cluster size. For example, using the simulation with parameters that satisfy these assumptions, for an interviewing area of 905 square miles (a size used below to model the NCS), the distance travelled to complete twelve clusters of size 1 was 137 miles; for 6 clusters of size 2, 97 miles; for 3 clusters of size 4, 63 miles. The percent increase in travel associated with using clusters of size two instead of clusters of size four was

54%; for using clusters of size one instead of clusters of size two, 41%.[3]

When p is one and a uniform distribution of clusters is used, but there is insufficient time to visit and interview all sample units in one trip, the percent difference in travel associated with using one cluster size versus another is directly affected by the interaction of the trip duration, interview length, geographic area, and travel speed. For our purposes, we will define an "optimal" combination of these parameters as a set that yields trips during which the interviewer can visit all the units at each cluster visited (i.e., the interviewer never needs to go home while in the middle of visiting units in a cluster). At $p = 1$, only one visit is required for each unit to have a completed interview, so optimal travel for $p = 1$, implies that the number of visits per trip is an integral multiple of the cluster size. Such optimal travel has a direct effect on the travel associated with a particular cluster size and thus the recommendations that are drawn from the simulation.

At $p = 1$, the largest advantage for clustering occurs when comparing an optimal, more clustered design with a suboptimal one. Using the Akron, OH, parameters described later (with a sample size of 12) a trip duration of 2.5 hours results in exactly 1 visit to 1 cluster of 4 per trip and an increase in travel of 74%. For a 97 minute trip, which is optimal for clusters of size 2 and results in 2.05 visits per trip with clusters of size 4 (similar to 6 clusters of size 2) the increase is negligible (.4%).

When p is less than one, clustering tends to be less cost-efficient. On subsequent visits to a particular cluster the interviewer will only visit the sample units in the cluster still requiring interviews. This results in a less clustered design on the later trips.

As can be seen from Table 1, with our parameters the advantage of clustering, expressed as the percent increase in travel in changing from clusters of size four to clusters of size two, diminishes as p decreases.

Table 1
Percent Increase in Travel for Clusters of Size Two Compared to Clusters of Size Four For Several Values of P

P	Percent Increase
.25	14.16
.3	16.11
.35	18.60
.4	22.15
.45	22.27
.5	22.21
.55	25.34
.6	26.91
1.0	72.94

To further understand the performance of our model in a uniform density area, we varied the area size (50, 100, 250, 500, 2500 square miles) for a sample size of twelve. These area sizes represent actual Census household survey sample areas. For example, Washington, DC, is approximately 60 square miles; Green Bay, WI, is a little over 500 square miles; and the New Orleans Metropolitan Statistical Area is about 2400 square miles. We were interested in the relationship among total distance travelled, p, area size, and cluster size.

We realized that our standard 2.5-hour trip duration was not realistic for the larger areas, in which an interviewer might make fewer, longer trips. The shorter trip duration led to an unreasonable rate of never-completed cases. To model the situation more realistically, we found combinations of trip duration and number of trips that resulted in a rate of never-completed cases (NCR) very similar to that attained in the 50-square-mile area with 5 2.5-hour trips over 5 values of p (.2, .4, .6, .8, 1.0). This led to the following trip durations for the larger areas: 170 minutes for 100 square miles, 210 minutes for 250 square miles, 220 minutes for 500 square miles, and 330 minutes for 2500 square miles. This approach more closely approximates the actual conditions under which interviewers work, namely, with an expectation of completing a certain proportion of cases regardless of the area size.

Next, we gave the simulated interviewers an unlimited number of trips in which to complete their interviews, so that the NCR for all areas would be 0. This eliminated a possible confound of cluster size and NCR,[4] to let us look only at the difference in travel.

We wanted to find the best parsimonious fit to describe the variability in distance (dist) from our simulation results using the following variables and their interactions: p, square mileage of the geographic area (transformed as the square root to put it in the same units as distance), and an indicator variable I to represent the two possible cluster sizes ($I = 1$ for clusters of size 4 and - 1 for clusters of size 2). Because we would expect no distance to be travelled for an area of size 0, we did not include an intercept term in this model. Early analyses led us to the following model:

$$\text{dist} = 27.59 * \text{sqrt(area)} - 47.66 * p * \text{sqrt(area)} + 23.81 * p^2 * \text{sqrt(area)} + \text{error},$$

which described 97.46% of the variability in distance. To better isolate the effect of p and the cluster size, we reasoned that a more parsimonious model would be one in which we adjusted distance for the area in which it was measured, namely, by using distance/sqrt(area) as our dependent variable. Dividing sqrt(area) through the above equation, we obtain a basic model in p and p^2 plus a constant.

We wanted to include the effect of cluster size in the model. Examining the R-squares and plots of residual from several models, we determined that the following model fit best:

$$\text{dist/sqrt(area)} = 27.62 - 4.47 * I - 47.84 * p + 24.38 * p^2 + 10.09 * I * p - 6.38 * I * p^2 - 0.02 * p * \text{sqrt(area)} + \text{error}.$$

All coefficients were significant at $p < .006$ and R-square was .9954. Note that this model included a term reflecting the size of the area involved. That is, the distance travelled is a quadratic function of p, whose exact shape depends upon the cluster size used and the area.

To examine the effect of cluster size on travel, we next analyzed the variability in the percent increase in travel

$$(pct = |dist_2 - dist_4|/dist_4)$$

when using clusters of size 2 instead of clusters of size 4. When we look at this relative change, the standardizing divisor sqrt(area) cancels out. As we have an R-square of .9954 in the quadratic model above, and our data are basically error-free because they come from a simulation with a large number of iterations, we can consider the above model to adequately describe the variability in distance/sqrt(area) and consider the purely random error to be negligible. Then, a model to describe pct would be a ratio of the two quadratic functions that result from evaluating the quadratic equation above for the two cluster sizes.

We wanted to approximate this non-linear function with a linear model in p. A simple model resulted in:

$$pct = .41 - .57 * p + .51 * p^2 + error,$$

with all coefficients significant at $p < .0001$. This fit only accounted for 50.23% of the variability in pct. Examining the plot of the residuals by pct we found that there was a clear effect of area size, although our initial model for the variability in distance/sqrt(area) would not have predicted that.

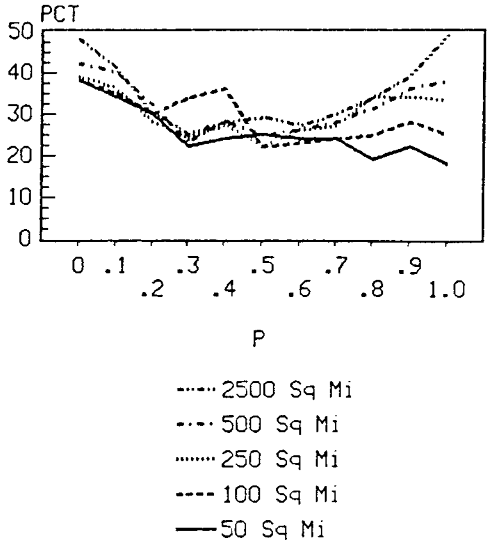
Analyzing the variability in pct further yielded the following model:

$$pct = .35 - .34 * p + .21 * p^2 + .002 * sqrt(area) - .01 * p * sqrt(area) + .01 * p^2 * sqrt(area) + error.$$

This model accounted for 81.48% of the variability in pct, with all coefficients significant at $p < .02$. This lower R-square is not unexpected because the dependent variable here is a measure of change rather than a standardized distance.

We also examined the simulation results graphically, expecting an increase in travel associated with using clusters of size two instead of clusters of size four, and that this increase would be greater for larger values of p. Inspection of the plot of pct by p in Figure 1, shows that there is always an increase in travel and the expected pattern occurs in most areas for $p > .2$. However, for high p and an area of 50 square miles, we see there is a decrease in the effect of using smaller clusters. The increase in travel associated with using clusters of size 2 increases greatly for an area of size 2500 square miles. Also, there is a greater increase in the effect of larger clusters for very low p as well.

Figure 1. Percent Increase in Travel (PCT) by Probability of Obtaining an Interview on a Given Visit (P) for 5 Geographic Areas



To better understand these phenomena we examined the number of visits per trip to see where "optimal" travel was occurring and how that might be affecting our conclusions about the impact of cluster size on travel. Our program calculates an average number of visits per trip, averaged across all iterations of the simulation for a given set of input parameters.

For $0 < p < 1$, we cannot formulate a simple rule for the average number of visits per trip by which to judge whether or not optimal travel took place. For example, an interviewer might visit all the units in 2 clusters of size four on the first trip (8 visits), obtaining 3 interviews and travelling optimally. If, on a later trip, the simulated interviewer visited only the 5 units in those 2 clusters still requiring interviews, that would still be optimal travel by our definition. However, the average number of visits per trip based on those 2 trips is 6.5. Because we cannot unequivocally claim optimal travel occurs for such nonintegral average numbers of visits, we can only meaningfully examine visits per trip for $p = 0$ and $p = 1$.

We found that the optimal travel effect explained part of the pattern shown in Figure 1. For an area of 50 square miles, the average number of visits per trip was found to be 4 for $p = 1$, regardless of cluster size. This indicates that our percent increase in travel measure was comparing the travel under two optimal scenarios.

By contrast, in the area of 2500 square miles at $p = 1$ we have 5.63 visits per trip for clusters of size 2 (suboptimal) and 6 visits per trip for clusters of size 4. Although the latter is not optimal by our definition, it does result in exactly 1.5 trips to each cluster, for a total of exactly 2 trips to complete all cases. (In fact, a modified definition of optimal travel might include any average number of visits per trip that divides evenly into the total sample size.) For clusters of size 2, 5.63 visits per trip implies over 2 trips were necessary. In such a large area, the extra home-to-cluster travel taken on the third trip with clusters of size 2 creates much greater travel than expected, resulting in a greater-than-expected advantage for clusters of size 4. This advantage is reflected in the value of *pct* in Figure 1.

For very low values of *p*, the assumptions about "tail-end" cases (those visited toward the end of the interviewing cycle) become very important but we don't know how to model such situations because they're unrealistic. We think that results for mid-range values of *p* are less affected by artificial interactions like the optimal travel effect or the tail-end effect. It is over these values that we observe our expected relationship between *pct* and *p*. However, it is clear that area size plays a role in the value of *pct*, so we need more precise information about *p* and area size to draw conclusions about travel for different cluster sizes.

A Case Study: Akron, Ohio, for the National Crime Survey

We chose Akron, Ohio, as our area to model because it is currently in sample for the NCS (U.S. Bureau of the Census, 1986) and because we were able to obtain reasonable-looking data on field costs. It was modelled as a circular area of 905 square miles with half the circle containing 79% of the sample units on average (see U. S. Bureau of the Census, 1982 & 1988). We ran the simulation for both 6 clusters of size 2 and 3 of size 4, as these were the sizes in which we were interested for the NCS. Trip duration was set at 2.5 hours, with a limit of nine or ten trips. Interview length is approximately 25-35 minutes.

Number of Trips and Trip Duration

If we want to complete all the cases in a given interviewer area using optimal travel, then for a fixed value of *p*, a lower bound exists for the duration of each trip and (conditional on the trip duration) for the number of trips. This lower bound can be most simply examined by discussing the case when *p* = 1. Recall when *p* is one an interview is obtained on the first visit. Because the trip duration restricts the number of visits in a trip, if we want optimality it is necessary to have enough time in a trip in order to visit and interview one entire cluster. And, given a particular trip duration, there must be enough trips in order to visit all units in order to keep the number of never-completed cases close to 0. The exact value of this minimal trip duration is dependent upon the size of the interviewing area, the interviewer's travel speed, and the length of the interview.

For trip duration, a reasonable upper limit might be eight hours, recognizing that an actual interviewer would take breaks and so might actually be away from home for more than eight hours.

The number of trips can be set as high as necessary to guarantee an interview at every sample unit. Realistically, interviewers have only a limited number of trips, determined by the number of days they are allotted to complete their interviewing (an interviewing cycle). For example, for a particular survey a reasonable maximum number of trips might be 15, taken in 14 days (if the trips were short enough that two could be taken on one day, perhaps one in the morning and one in the evening). With such a restricted number of trips, it might be the case that not all sample units receive completed interviews. However, having a rate of never-completed cases greater than 0 might be reasonable for a particular survey.

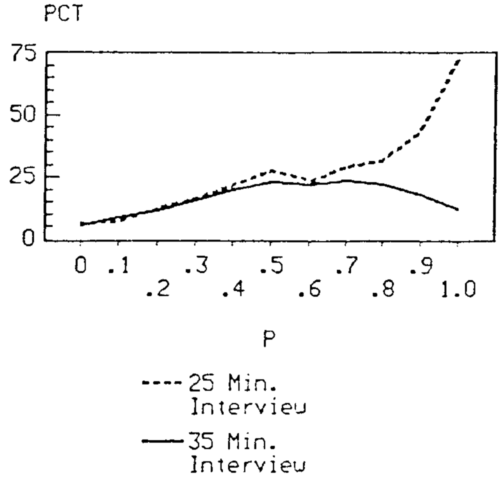
Using our Akron parameters, we produced simulations allowing a maximum of 9 trips and simulations allowing a maximum of 10 trips for various values of *p*. We expected that the number of trips taken would increase as *p* decreased (because more callbacks are required for lower values of *p*), which was confirmed by our data. For *p*=.2, the maximum number of trips was used in all cases.

As expected, the extra visits that can be made when a maximum of 10 trips was allowed resulted in a lower NCR than for simulations with a maximum of 9 trips allowed, regardless of cluster size. With the parameters given, once *p* is above .5 the NCR is similar for 9 and 10 trips because far less than the maximum number of trips were needed to complete all the cases.

Interview Length

Using the Akron parameters and fixing the maximum time per trip at 150 minutes and the number of trips available for use at 9, two lengths of time needed to complete an interview were examined: 25 minutes and 35 minutes. Nine trips were in fact sufficient in order to complete all interviews when *p* = 1 for both interview lengths. Examining the percent increase in travel going from a cluster size of 4 to a cluster size of 2, it was apparent that with both interview lengths there was always more travelling with a cluster size of two. With the shorter interview time (25 min.), our percent increase in travel rose as *p* increased, peaking at 73% for *p* = 1. With the longer interview length (35 min.) our percent increase never rose above 35% and in fact peaked at *p* = .7 instead of *p* = 1 (see Figure 2).

Figure 2. Percent Increase in Travel (PCT) by Probability of Obtaining an Interview on a Given Visit (P) for Two Lengths of Interview



The reason for the different relationship between the percent increase in travel and p for the two interview lengths becomes apparent when examining the number of visits per trip for both interview lengths and for both cluster designs.

With the longer interview length, not as many visits can be made per trip. The interviewer simply runs out of time. This can be seen especially with a high p, when each visit results in an interview (which uses up time). With the long interview length (35 minutes) in both cluster designs, only a little more than two sample units on each trip were visited for $p > .5$. For $p = 1$, 2.25 units were visited and interviewed per trip with clusters of size two and 2.45 with clusters of size four. For both designs, the interviewer must return to the cluster in order to finish interviewing the remaining units and is thus travelling suboptimally. The extra visit needed to finish interviewing all the units in a cluster results in an increase in the travel mileage. In our example, this lack of optimality resulted in more similar mileage for both cluster designs.

With the shorter interview length, however, optimal travel occurred for clusters of size four at $p = 1$, exactly four sample units (comprising one cluster) were visited/interviewed on every trip. When the cluster size was two, the interviewer visited 2.98 sample units per trip for $p=1$ which was not optimal. Thus, we compared a perfectly optimal travel pattern to a suboptimal one when assessing the percent increase in travel using clusters of size two instead of cluster of size four.

National Crime Survey Cost Model

Using the Akron parameters (including an interview length of 25 minutes and a maximum of nine trips) and actual sample sizes of clusters of size four, we compared the results of the travel simulation over a range of values of p to the data we obtained from the field for particular months. We attempted to match the never-completed rate from our simulation to the "no-one-at-home" rate from the actual data. Under the assumption that mileage reported in the actual data included miles used for listing of sample units, we expected our simulated total travel would be less than the reported travel for the correct value of p. We estimated the true value of p to be around .4., based on the above comparisons and an independent rough estimate by field staff that approximately 2.5 visits are required to complete an NCS case.

We ran the simulation for values of p ranging from .25 to .55 (containing our point estimate of .4). These runs were then compared with runs using clusters of size four. From these data the percentage increase in distance travelled was computed. There was at least a 15% increase in travel using clusters of size two instead of clusters of size four for p within this range. As p increased and thus more interviews occurred, the percent increase in travel between the two design also rose. We have reasonable confidence in this conclusion because these values of p are the same as those for which the graph of pct by p in Figure 1 was stable for different area sizes.

Combining the result that a design using clusters of size two requires at least 15% more travel and the fact that for the National Crime Survey, the intraclass correlations between survey items are low, we recommended clusters of size four rather than size two.

DISCUSSION

We have presented a complex model of travel that involves subtle interactions among features of field interviewing, including interview length, duration of time in the field, size of the interviewing area, and likelihood of obtaining interviews at sample units. There appears to be no single simple formula to describe the impact of cluster size on travel in all cases. In actual practice, there is great variation in these parameters; under certain conditions, one cluster size may have a special advantage over another because particular parameters "fit" together but there is no way to judge *a priori* how frequently this will occur. When the parameters do not fit in this way, there is "suboptimal travel," that is, the interviewer frequently runs out of time and breaks off in the middle of a cluster.

One feature in particular, the length of interview, is critical in assessing the impact of cluster size on travel. In practice, for a particular approximate trip duration (the value of which would be determined by the interviewer and depend upon the size of the geographic area being canvassed), optimal travel is more likely to occur with shorter interviews regardless of cluster size. That is, if an interview takes fifteen minutes, an interviewer is likely to attempt all units in a (reasonably-sized) cluster, because obtaining one or two more interviews than might have been expected does not extend the trip duration greatly. Our simulation as currently written does not allow for a "reasonable" extension of the trip duration in situations in which the interview length is short; if the time to complete an interview would extend the length of the trip past the fixed trip duration, the unit is not contacted. This inflexibility could be exaggerating the effect of optimal combinations of parameters.

However, with a long interview (such as 2 hours), an interviewer in the field would not be likely to attempt another unit in the same cluster close to the end of the planned trip duration. Our rigid trip duration is more realistic in this situation. We would expect to conclude that larger clusters are less advantageous because they do not lead to reduced numbers of trips to each cluster.

Even if a more flexible trip duration were incorporated into the simulation, it might not fairly represent reality; we have no basis by which to claim that interviewers always try to complete clusters which they have started. Clearly other tasks, such as listing of future sample units and editing of forms, take time in the field that might otherwise be used to ensure optimal travel.

In our discussion, we have presented a definition of "optimal" travel, in which an interviewer is able to attempt all eligible units at all clusters visited on each trip. This should lead to the least cost for any given cluster size. For $p=0$ and $p=1$, optimal travel involves visiting an integral number of full clusters on each trip. For $0 < p < 1$, the travel patterns are actually a mixture of models of travel to clusters of different sizes. In early trips to clusters of size 4, for example, an interviewer would visit 4 units at each cluster. As interviews were obtained, the interviewer would effectively visit clusters of sizes 3, 2, and 1. Optimal travel patterns would depend upon the value of p and the clustering of the remaining units.

In principle, our model could be used for determining the combinations of trip duration and number of trips that would be expected to produce optimal use of travel. Conceivably, this could be used to advise field staff in planning their time, especially after a change to a different cluster size than that to which they are accustomed.

In this paper we have attempted to model interviewer travel with very little actual data about frequency of callbacks, actual travel times, number and duration of trips, and other interviewer performance measures. We hoped to find differences between travel for different cluster sizes which were stable over the entire range of plausible parameters. To some extent this was achieved for the range of parameters which seemed to be applicable to the National Crime Survey. However, exact statements about the differences cannot be made without better data.

The greatest potential for this kind of simulation would be where it can be combined with good detailed data about interviewer performance for an existing cluster size. If the model could be shown to accurately describe existing travel patterns, it could give very believable predictions for the effect of changing cluster sizes.

NOTES

[1] We added an additional feature to allow the interviewer to make callbacks out of sequence if they were "close" (e.g., "on the way home").

[2] We also examined areas in sample for the Consumer Expenditure Diary Survey. These results are not presented because the conclusions drawn for the survey sample design involved many issues beyond the scope of this paper.

[3] Hansen, Hurwitz, and Madow (1953) assumed that the units are spread uniformly around a square area, whereas our simulation used a circle (which we felt more typically represented an interviewing area). Using the same parameters in our simulation with a square-shaped area yielded similar results. That was also true for other analyses reported here.

[4] We also limited the number of visits that could be made to a single sample unit to 10, to eliminate excessive travelling at the end of the interviewing cycle.

REFERENCES

Hansen, M. H., Hurwitz, W. N., & Madow, W. G. (1953). Sample Survey Methods and Theory, Vol. 1: Methods and Applications. New York: Wiley.

U. S. Bureau of the Census. (1982). 1980 Census of Population, Vol. 1, Chapter A: Number of Inhabitants, Part 37, Ohio. Washington, DC: U.S. Department of Commerce.

U. S. Bureau of the Census. (1986, September). PSUs Included in Current Survey Samples (Form 11-4). Washington, DC: U.S. Department of Commerce.

U.S. Bureau of the Census. (1988). County and City Data Book, 1988. Washington, DC: Government Printing Office.