Proceedings of the Annual Meeting of the American Statistical Association, August 5-9, 2001

SAMPLE DESIGN ISSUES FOR SURVEYS INVOLVING THE OBSERVATION OF DRIVERS

Gary Shapiro, Westat; Dennis Utter, NHTSA; Wende Mix, Joseph Croos, David Marker, Westat; and Nancy Bondy, NHTSA
Gary Shapiro, Westat, 1650 Research Boulevard, Rockville, Maryland 20850

Key Words: Shoulder Belt Use, National Occupant Protection Use Survey, Primary Sampling Units, Weights, Variance

1. Introduction

This paper pertains to surveys in which people stand on the street and observe characteristics of drivers or/and passengers of passing cars. It particularly concerns the National Occupant Protection Use Survey (NOPUS), but is also applicable to other observation surveys that relate to racial profiling or for which demographic characteristics of drivers are of interest. There are unique problems associated with such surveys. Kalton (1991) briefly discusses many of them. Also, Brick and Lago (1988) address how to handle many of these issues in earlier surveys.

NOPUS is concerned with the use of shoulder safety belt use. While there are many factors that contribute to the cause and outcome of a motor vehicle crash, one factor that can significantly reduce the injury and fatality consequences of a crash is the use of occupant restraints. Shoulder safety belt use reduces the risk of fatal injury to front-seat passenger car occupants by 45 percent and the risk of moderate-to-critical injury by 50 percent. For light truck occupants, safety belts reduce risk of fatal injury by 60 percent and moderate-to-critical injury by 65 percent. It is estimated that in 1999 the use of safety belts saved the lives of 11,197 persons over 4 years of age. Other research has shown that in passenger cars child safety seats reduce the chance of fatal injury by 71 percent for infants (age less than one year) and by 54 percent for toddlers (ages 1 - 4 years). An estimated 307 young lives were saved in 1999 by the use of child restraints.

A large part of the motoring public still does not buckle up while in their motor vehicles. In 1999, 62 percent of motor vehicle occupants that were killed in crashes and whose restraint use was known did not use their safety belt. And slightly more than one-half (53 percent) of infants and toddlers (ages 0 - 4 years) who were killed in crashes were not restrained. Reported restraint use among those injured is higher, with only 15 percent of all injured persons reporting not using a restraint and 16 percent of injured infants and toddlers were unrestrained.

Encouraging people to use their shoulder belts and to place their children in child safety seats is a U.S. Department of Transportation/National Highway Traffic Safety Administration (NHTSA) priority. And tracking changes in those use rates is a necessary part of that program. Since the early 1970's, NHTSA has observed and documented the public's use of safety belt systems. The 19-City Survey of Safety Belt Use (19-City Survey) was NHTSA's primary vehicle for monitoring trends in safety belt use from 1983 through 1991. The primary purpose of this survey was to track the relationship between regulated safety features and actual use of belt systems and to serve as an index of safety belt use to track usage trends in those 19 cities. NHTSA did not anticipate using the survey for other purposes, such as a national measure of safety belt use rates, and its design was not probability based. As the need for a National measure of safety belt use increase, NHTSA decided (in 1992) to cancel the survey and rely on state conducted surveys to provide a national measure of safety belt use.

Subsequent to its cancellation, numerous NHTSA offices expressed the need for more detailed information on safety belt use than that which could be provided by states. At that time most state surveys estimated only usage rates for drivers. Others also included front seat outboard passengers. Additionally, many methods were employed by the states in the design of these surveys and in their data collection methods. Information needed by NHTSA analysts and highway safety program managers, e.g., use rates by vehicle type, restraint type, age groups, sex, etc., is not collected. As a result, an Agency working group recommended that NHTSA develop its own probability based National survey of occupant restraint used to provide the needed information.

Consequently, NHTSA began conducting the National Occupant Protection Use Survey (NOPUS) in the fall of 1994 to obtain nationwide estimates of shoulder belt use and of characteristics of their users. The NOPUS was also conducted in the Fall of 1996, 1998 and 2000. It is composed of two separate studies: the *Moving Traffic Study*, which provides information on overall shoulder belt use; and the *Controlled Intersection Study*, which provides detailed information about shoulder belt use by vehicle type, characteristics of the belt users and child restraint use.

The Full NOPUS was designed as a multistage probability sample to ensure that the results would represent occupant protection use in the country as a whole:

First stage of selection. Counties were grouped by region (Northeast, Midwest, South, West), level of urbanization (metropolitan or not), and level of safety belt use in the state (high, medium, or low). Fifty counties or groups of counties (called primary sampling units or PSUs) were selected within the resulting strata. PSUs were selected with probability proportional to vehicle miles of travel. Vehicle miles of travel (VMT) are measures of the distance traveled by various types of motor vehicles on public roads. An estimate of motor vehicles on each designated road segment in the State is obtained by the State Highway Department or Department of Transportation. To obtain VMT, the vehicle estimate is multiplied by the segment's length. These measures are aggregated to the county level.

Second stage of selection. A sample of Census Tracts was selected within each of the sampled PSUs. Because Census Tracts are created to be approximately equal in population, tracts were selected with equal probability.

Final stage of selection. A probability sample of local roads was selected within each Census Tract. A sample of major roads was selected across the entire PSU, not restricted to the sample tracts. Major roads are those maintained by the State Highway Department. These were available from inventories maintained by these departments. The sample of local roads was obtained by enumerating on maps all roads within each of the Census Tracts that were on the State inventory.

Once the sample of roadway segments was selected, observational sites were identified. Observational sites depended on the type of sampled road segment: For segments identified on the interstate highway system or on controlled roadways it was an exit ramp; for major or minor highways which contained an intersection controlled by a stop sign or stoplight it was that intersection; and, for road segments that did not contain a "controlled intersection" it was a safe observation point or an uncontrolled intersection along the road segment. The roadway sample for the Full NOPUS Moving Traffic Study conducted in fall 2000 was 2,063 sites, with about 170,000 observed vehicles. The observation sites for the Controlled Intersection Study are those sites from the Moving Traffic Study controlled by a stop sign or stoplight. In the fall 2000 Controlled Intersection Study, observations were conducted at 640 sites. This paper is primarily concerned with the Moving Traffic Study.

Note that time-based estimates of shoulder belt use are what's of interest. That is, one wants to estimate (driver time spent wearing belts)/(total driver time). This has important implications for the design and for weights, especially in that it requires that long road segments should contribute more than short segments to the estimates.

The paper discusses a number of issues associated with observational surveys such as NOPUS. Section 2 discusses sample design issues regarding the first, second and final stages of selection. In particular, the section covers optimal number of sample PSUs, optimal number of sample sites, and an alternative to the present method of sampling segments. Section 3 discusses problems that low volume roads present. Section 4 discusses special problems with controlled intersection /limited access roads. The fifth section discusses various causes of large variations in sample weights. Finally, the last section is a brief conclusion.

2. Sample Design Issues

This section deals with issues regarding sample selection – the first stage selection of geographic areas, and the selection of segments within the first stage units.

2.1 First Stage of Selection

The NOPUS design uses stratification to group geographically similar PSUs (PSUs within same Census region and same urbanization) that are likely to have similar seat-belt usage rate (primary parameter of interest). Grouping PSUs with similar seat-belt usage improves the overall precision of usage estimates and grouping PSUs geographically ensures the sample to be diverse and representative of the nation.

The total survey cost depends on the number of PSUs, the number of sites, and the length of observation per site. The current NOPUS design has 50 PSUs and about 2,000 sites, and we have a good idea on the variable cost structure. The current NOPUS cost structure is such that per PSU variable cost is about 43 times the cost per observation period (or site). From a cost standpoint, it would seem preferable to observe a large number of sites in a relatively small number of PSUs.

We have noted a high between PSU variance component in the total variance of safety-belt usage estimates. This may have caused a larger design effect than originally intended. Therefore for estimates of level (of usage rate), it would seem to be desirable, despite high per PSU cost, to increase the number of PSUs beyond the current level of 50 PSUs. Based on the current NOPUS cost model we feel about 70 PSUs would be optimum. The increased cost of extra PSUs can be mitigated by a 40 percent reduction in the total observation periods.

But for estimates of change (between two NOPUS data collection periods), fewer sample PSUs would be optimum. This is because there is essentially no between-PSU variance component in the estimate of change since the data are collected on the same PSUs.

We are currently doing some data analysis on some simulated samples – samples with 30, 40, 60, and 70 PSUs – to compare the precision of the estimates (both level and change) to the current design under the same overall cost as the original NOPUS.

Other design changes we are studying include a better choice of measure of size to be used in the probability sample. The current measure of size, vehicle miles traveled (VMT), may not be a reliable measure of size in all parts of the country. For example, in vacation areas VMT may represent summer traffic rather than a yearly average. In this case, it may be better to use a function of population and/or a road-capacity rate instead of VMT. Road capacity rates are invariant over time, unlike VMT. Another problem with VMT is that it includes commercial trucks. Note that for an interstate highway through a sparsely populated part of a PSU, a high proportion of the traffic is commercial trucks.

2.2 Second Stage of Sampling

Under the current NOPUS, most of the cost is traveling between sites within PSUs and per PSU cost (such as training the team). On a typical work-day an observation team (one team per PSU) travel and observe multiple sites within the same PSU. A major part of the observation team's cost per site is in traveling and preparation.

The optimal allocation of number of PSUs, number of sites, and length of observation, under invariant cost, depend on the relative importance of several variance components. For estimate of level, the dominant variance component is the between-variance PSU followed by the between site variance. For the estimate of change, assuming we go to same sites within the same PSUs, the dominant component of variance is the within site variation.

For estimates of level, since the between site variation is an important component of the total variation, it may be desirable to increase the number of sample sites and therefore reduce the length of the observation period.

For estimates of change, since there is little between site variance, it is desirable to have relatively long observation periods at the expense of a smaller number of sites.

2.3 Sampling of Segments or Grids

The prevalence of Geographic Information Systems (GIS) and digital road databases facilitate development of the sampling frame for traffic observation studies. GIS allow us to define the sets of roads that are eligible for sampling, attach attributes such as volume or other measures of size to these roads, and output lists of road section information in a variety of formats for sample selection.

The most common approach to selecting locations to observe traffic has been a segment based approach. Roads within sampled areas are defined in segments, usually based on how a state department of transportation has defined their data collection segments. Classification by road type (limited access, major arterial, minor arterial, collector, local road, etc) is based on either the Census Feature Classification Codes (CFCCs) or assigned from the local area or statewide transportation plans. Generally, incomplete volume coverage for all road segments in a database necessitates determining average volume for segments within specific road types. Segment length is determined based on the database representation, or provided as a measured input from a state's inventory database.

Practical problems with implementing this approach are many. First, there is no spatial representation (hardcopy map or digital) of the transportation network that is complete (contains all existing roads) and accurate (for instance, has all roads correctly classified by type, contains only true intersections, etc.). Second, the term “road segment” has a variety of definitions that all must be reconciled. In terms of an observational survey, the preferred definition of segment is the section of road between two intersections. Drivers using this segment have little (driveways) or no opportunity to leave the segment, hence they drive from one end to the other. Thus, it does not matter where an observer is positioned for data collection. Road segments are defined by state DOTs in a variety of ways, based on projects, data collection needs, funding decisions, etc. In general, DOT criteria for defining segments result in road sections in the inventory database that usually are comprised of multiple adjacent segments of roads. Further, road segments in a statewide inventory system usually begin and end at intersections with roads of the same type or one class lower, that are part of the state system. Thus, a state highway that passes through an urban area will be defined where it crosses another state highway, US highway, or Interstate. Drivers who use the facility for local travel have many opportunities to enter and exit it, so that traffic observed near an end point or somewhere in the middle may be quite different.

GIS definition of segments depends on the underlying data model in use. Most spatial databases are enhancements of the original Census Bureau TIGER files which define nodes (intersections) when the line representing one feature crosses the line representing another feature. So, for instance, if the line representing a road crosses a utility line and a stream, then there is a unique road segment ID for that piece of pavement. Another issue with the GIS representation of segments concerns limited access facilities (like Interstates) with large medians separating travel in two directions. Each travel direction is represented as separate segments so an Interstate segment is twice as likely to be selected compared to a surface road where both travel directions are represented using a single line segment. In fact, changing the representation of Interstate facilities into single segments between interchanges is crucial prior to sample selection. Interstate facilities in spatial databases are extremely fragmented (by ramps, overlays with surface streets, water features, etc) so they are over-represented in the segment lists created for sampling. For example, using the Environment Systems Research Institute’s (ESRI) Streetmap digital database, the Innerbelt in Columbus, OH between I-670 and Broad St is represented by 10 digital line segments and between Broad St. and Main St. by 11 digital line segments. However, if any of these ten (or eleven) segments is selected, observers are given the same location to observe and will be looking at the same traffic regardless of which of the ten (or eleven) segments is selected.

One way to avoid the use of segments in the sample frame is to view a road network as a raster (or grid) representation rather than a vector, or line segment representation. GIS allows straightforward conversion from a line segment database to a grid database. Basically, the first step is to extract a transportation network that is eligible for sampling. The second step is to rasterize this database. Essentially rasterization means overlaying a very fine grid (pixels) on top of the road network and assigning a code of one for all cells in the grid containing a road and a zero for all cells not containing a road. (To represent roads by road type, for example, you can code cells with interstate roads as 1, cells with major arterials as two, cells with minor arterials as 3, etc. or you could create three separate binary raster representations). Sample selection then involves randomly generating row and column identifiers of road filled cells. Segment length becomes unimportant in sample selection if a very fine grid representation is used. Suppose a 36x36-foot grid is used. Then a road segment that is half a mile long and is precisely in an East-West (or a North-South) direction will be assigned about 73 grid cells and a road segment that is a mile long will be assigned about 146 grid cells. If any of the 73 (or 146) cells is selected, observers are assigned the same group of vehicles to count.

3. Low Volume Roads

In order to produce unbiased national estimates, all types of roads should be covered. Due to the extremely large number of local roads, many of them are included in any unbiased sample. Many of these roads have very little traffic, so sending observers to stand by a local road for 20, 30, or even 60 minutes may result in observing only a handful of vehicles. This is very costly on a per-vehicle basis.

Some local roads on sampling frames do indeed have heavy usage. The road may have periodic usage (for example near a recreational area or school) or may be miscoded and not really be a local road (e.g., the state has recently taken it over and widened the road). In such situations large volumes are associated with large weights (see Section 5), increasing the variability of estimates.

It has been suggested that a two-stage approach could be used, where local roads are quickly observed to see if they do indeed have low volume. If they do, these roads could be subsampled to reduce their frequency in the sample. This eliminates the risk of having a large weight associated with roads that turn out to be high volume. However, in many situations such an approach would only provide limited savings, since much of the cost associated with observational studies is in going to and from the sampled sites. Thus the “quick observation” will cost almost as much as a full measurement at a site.

Another related approach is to include local roads in the first implementation of a new selection for NOPUS. Those segments for which volume is low would be subsampled or eliminated entirely from future surveys. There are two difficulties with this approach. First, we wish to observe a given site at the same time for each implementation of the NOPUS so as to maximize the correlation over time. If this is strictly adhered to, the deletion of some sites would leave gaps in field staff schedules, and would thus result in little cost savings. Second, a site might not continue to have low volume in future years, and thus future high volume sites would be dropped from sample.

If the rate of seat belt usage and other characteristics on local roads are similar to those on other roads, it may be desirable to exclude such roads from a survey. For example, if drivers on local roads wear seat belts with similar frequency to those on other roads, then they could be excluded from seat belt usage surveys. The reduction in costs and increase in accuracy could be substantial from such a decision.

However, excluding local roads can have adverse effects on estimates. If the characteristics of local drivers are different (e.g., they are more likely to be minorities or to use safety belts) then excluding these roads will introduce bias into the estimates. Also, if there are many frame errors where non-local roads are coded as local, then such an exclusion will also eliminate many high volume roads.

There are alternatives that could be considered to either completely excluding or including local roads. Roads could be sampled proportional to volume or other measure that undersamples low volume roads. This results in large weights for local roads and low volume major roads, but reallocates resources to other roads where more vehicles can be observed. If done judiciously so as not to result in very large weights for low volume roads, such a design can improve the accuracy of estimates of driver characteristics.

An alternative procedure is to exclude certain types of low volume roads on a frame that are almost assuredly low volume. This would include roads in parkland, on Indian reservations, unpaved roads, and roads that do not have a name. These exclusions are likely to represent only a small percentage of drivers, but due to their remote locations may be the most costly data to collect.

4. Limited Access Roads

Driver observation is usually done by people standing by the side of the road as vehicles pass. On limited access highways, however, it is not safe to stand by the side of the road. Furthermore, it is difficult to make observations when vehicles are traveling at 60 mph or faster. The traditional method used for limited access roads is to do observations along exit ramps. This is feasible and safe but has several problems. One desires to estimate seat belt use for all vehicles approaching the exit ramp, not just those actually exiting. To obtain an appropriate weighting factor, a count is made of vehicles passing under the overpass by the exit ramp. For example, if one observed 15 vehicles during a 30 minute observation on a ramp and counted 92 vehicles during a 10 minute observation from an overpass, a weighting factor of (92/15)(30/10) = 18.4 would be used. This weighting factor results in appropriate expected values in terms of number of vehicles. However, the weighting factor can become quite large if relatively few vehicles exit at the observed ramp. In general, there is a risk of larger weights for limited access roads unless they are oversampled.

There are other problems with this methodology. For other types of roads, observers are instructed to observe all vehicles passing in the curbside lane. For limited access highway ramps, observers have been instructed to observe as many vehicles as possible. If there is only a single lane at an exit ramp, or if traffic is light, all or nearly all vehicles can be observed. Otherwise, however, there is some lack of control over which vehicles are observed and on the proportion of exit ramp vehicles observed. A more serious problem is that limited access vehicles represent different time lengths of travel and the weights do not reflect this. Suppose there are 2 cars that exit at the same ramp, 1 of which has been traveling on the limited access highway for 2 hours (getting on, say, 20 entrances before) and 1 of which has been traveling for 3 minutes (getting on at the previous entrance). Both of these vehicles have the same probability of being observed for their limited access driving, only if we are observing at the exit ramp when they exit. However, in 1 case the observation represents 2 hours of traveling and in the other only 3 minutes of driving. Since the estimates of seat belt use are intended to be the proportion of time that drivers and passengers wear seat belts, we would wish to use a weight for the 2 hour car that is 40 times that of the 3 minute car. Of course, this is impossible unless we stop a car and ask them when or where they got onto the limited access road.

Westat has used an alternative procedure for limited access roads that may be used for the redesigned NOPUS. Here, an observer car travels between 2 specific entrances on the road. The observer vehicle travels below the speed limit, say at 45 mph. Thus, nearly all vehicles can be expected to pass the observer vehicle. A passenger in the observer vehicle records whether the driver of each passing vehicle is wearing a seat belt. Drivers and passengers of any vehicles that the observer vehicle passes can also be observed. The observer car can traverse the specified section of road as many times as desired. This methodology overcomes the problems of ramp observation. It has the minor problem that any vehicle that is traveling exactly the same speed as the observer vehicle has no chance of selection. Also, vehicles driving different speeds have different probabilities of being observed, which should ideally be reflected in the weights. Vehicles driving only a little over 45 mph have lower probabilities of selection than do those traveling at higher speeds. For other road types, however, the same situation occurs – there is a higher probability of observing fast-traveling than slow-traveling vehicles. For limited access roads as well as other roads, this could be accounted for by using the estimated speed of individual vehicles to determine a weighting factor.

5. Large Weights

Weights are used in producing survey estimates to reflect variations in probabilities of selection. If all units do not have the same probability of selection and weights are not used, bias is likely. Weight adjustments are also commonly used for nonresponse adjustment and for controlling survey estimates to known population totals. Unfortunately, variation in weights leads to increased variance. In particular, unusually large weights can lead to large increases in variance. In travel surveys such as NOPUS, there are a number of factors that lead to variations in probabilities of selection and potentially large variations in weights. We list here some of the major factors present in the current design of NOPUS: Currently, the largest weights in NOPUS are about 50 times the mean weight.

A. Under-allocation to local roads. In general, local roads have lower volume than major roads. Thus, observation on a local road results in fewer vehicles being observed, and it is cost efficient to have relatively few observations along local roads. However, this results in larger weights for local road observations.

B. Sampling with probability proportional to vehicle miles of travel (VMT). In the current NOPUS design, major road sites were selected with probability proportion to their estimated VMT. (This was not done for local roads, where VMT estimates were not available.) This is cost efficient, since it results in fewer sample sites where few vehicles are likely to be observed. However, it results in sites with low VMT receiving large weights.

C. Long road segments. As part of the weighting, there is a factor for the length of the road segment. A car on a highway with lots of short road segments has numerous opportunities to be sampled (one for each segment), whereas a car on a highway with a few long road segments has few opportunities to be sampled. As mentioned above, major roads were selected with probability proportional to VMT in NOPUS. Length is a component of VMT, so weights for major roads are not affected by road length. However, for local roads, long road segments are not selected with higher probability than short road segments in the current NOPUS design.

D. Low vehicle speed. Also as part of the weighting, there is a factor for the estimated speed of the vehicle. Consider that for a fixed interval of time, a vehicle traveling at a fast speed traverses a larger number of road segments than a vehicle traveling at a slow speed. Therefore, the fast vehicle has a higher probability of being observed, and thus should be given a lower weight. In the current NOPUS, speeds of individual vehicles are neither determined nor used in the weighting. Rather, observers estimate the average speed during the observation as being in 1 of three categories: slow, 25 mph; medium, 40 mph; and high, 60 mph. Thus, vehicles on slow roads get weights that are 2.4 times as large as vehicles on fast roads.

E. Roads with many lanes. On a road with several lanes in each direction, observations are usually confined to the curbside lane. A lane adjustment factor is needed in the weighting, consisting of the ratio of the total number of lanes to the number of lanes observed. Thus, a road with 3 lanes in the observed direction where one lane is observed would get a lane adjustment factor of 3.

F. Low ramp volume for limited access roads. As discussed in section 4, vehicles on exit ramps are observed, and counts are made of overall interstate traffic. A factor is then applied as the ratio of the estimated volume of the interstate to the number of ramp vehicles observed. Thus, when there is a high volume interstate road, but the exit ramp is not being used much, a very high factor results. This can occur, for example, along a major interstate road where the exit ramp goes only to a small road or/and a small town.

We now discuss possible ways of reducing the weight variation for some of these factors. Regarding factor A, some local road sites that had lower probability of selection were in fact sites at which we are observing a large number of vehicles. The observations in these sites reduced the overall precision due to relatively large weights. One way to mitigate this effect is to make multiple visits to these sites. For example, if site A was observed twice, the corresponding site weights are one-half of the original site weight. This improvement is currently being explored. Regarding factor B, sampling with probability proportional to VMT, we are not planning to select sample for a redesigned NOPUS in the same manner as the current design. Our thinking at this point is to set up several broad strata, say for high volume, medium volume and low volume roads. We may use estimated average daily traffic or capacity rates rather than VMT for classifying road segments. This will result in somewhat variable sampling rates by stratum, but will not result in a real low sampling rate in the low volume stratum, so as to avoid large variations in weighting factors.

Regarding factor C, long road segments, we expect to sample grids and then points within the grid, as discussed in Section 2.3. If we use a small grid and define “road segments” to be within the grid, then there can be no very long road segments that require large weights.

Factor F, low ramp volume for limited access roads, will be no longer be an issue if we do moving traffic observations, as discussed in Section 3.

6. Conclusion

Surveys that are conducted periodically permit one to examine the weaknesses in the sample design and implementation and to then make improvements. This is precisely what NHTSA and Westat are presently doing for NOPUS. We hope to be able to make improvements along the lines indicated in this paper, which we will be able to report as they are incorporated into the survey.

7. References

Brick, M. and Lago, J. (1988). The design and implementation of an observational safety belt use survey, *Journal of Safety Research*, 19, 87-98.

Kalton, G. (1991). Sampling flows of mobile human populations, *Survey Methodology*, 17(2), 183-194.