MILEAGE AND OTHER COMPLEX ESTIMATES, AND SOME BIAS RESEARCH ON THE NATIONAL TRAVEL SURVEY

Joan Kahn, Dennis Schwanz, and Larry Carstensen, U.S. Bureau of the Census

INTRODUCTION

The 1977 National Travel Survey (NTS) was conducted April 1977–January 1978 as part of the Census Bureau's Quinquennial Census of Transportation. The 1977 NTS was a fourth in a series of National Travel Surveys, with earlier surveys being conducted in 1973, 1967, and 1972. This survey was designed to estimate characteristics of trips of 100+ miles (one-way distance) taken in 1977 by the civilian noninstitutional population of the United States. Tabulations were to be produced for each quarter of 1977 and for the full year, and for travel to and through each State as well as for the entire U.S. The major statistics for this survey, which will be described in more detail in later sections of this paper, included traveling/nontraveling households, household trips, traveling/nontraveling persons, person trips, person miles, person nights and auto/truck trips.

The estimation procedures employed for this survey were more complex than had been utilized for previous travel surveys. These methods of estimation incorporated several complicated features designed to reduce or eliminate bias resulting from special situations existing in the surveys. One of these bias–reducing features was introduced because there are so many major statistics of interest for each survey. If one weight were used to produce all of these statistics, as had been done in previous surveys, then it was very likely that biased estimates would have resulted for some of the statistics. Thus, the estimation procedure for each survey was in reality several estimation procedures, each one designed to produce the most appropriate weight for a given individudal or set of statistics.

Another feature of the estimation procedure resulted from the fact that trips could involve and be reported by either exactly one household member or more than one household member. Thus, trips involving only one member would be missed if that person were not interviewed; whereas, trips involving more than one member would not be missed if any of the participating household members were interviewed. Application of within-household noninterview adjustment to all trips, irrespective of the number of household members, involved, would result in an overestimate of trips since the trips involving a noninterviewed person and another interviewed household member would not have been missed as application of this adjustment would imply. Thus, the estimation procedure employed a within-household noninterview adjustment only for trips involving one household member.

Another bias–reducing feature was the use of ratio estimation to independent counts to account for missed trips resulting from the coverage problems of either missing households or missing persons within enumerated households. Use of one adjustment to account for both types of under-coverage for every statistic would result in biased estimates for those statistics which are not affected by both types of undercoverage. Thus, the estimation procedures employed several under-coverage adjustments, such that each statistic was properly adjusted for those aspects of under-coverage which affect that statistic. Another bias–reducing feature resulted from the longitudinal aspects of NTS. In NTS, households were interviewed four or five times during the year to obtain a complete calendar reading of the travel character-istics of household members. Only partial calendar information is available for persons who were non-interviews for part of the year or who moved in or out of a household during the year. Unadjusted inclusion of data collected from such persons would result in biased estimates for those statistics which were calendar year dependent. Thus, the rules for deciding whether or not to include partial year data were type–of–statistics dependent in the NTS estimation procedure to pro-duce estimates as unbiased as possible.

In addition to the bias–reducing features of the NTS estimation procedure, research was undertaken to reduce the bias associated with NTS data. One phase of this research involved NTS mileage cal-culation. Since reported trips would only be included in NTS tabulations if the one–way distance was 100 miles or more, accurate mileage estimation for trips near the 100–mile cutoff was very important. Since the planned method of mileage estimation had several sources of bias which could have caused trips to be incorrectly included or excluded, research was undertaken to develop a method of mileage estimation which could reduce these inaccuracies.

This paper after briefly describing the sample design for this survey will present a detailed explanation of the complex estimation procedure that was employed. In addition, the results of the mileage estimation will be presented.

SAMPLE DESIGN

National Travel Survey

The National Travel Survey sample consisted of approximately 24,000 housing units, of which 18,000 housing units had previously been interviewed for the Current Population Survey (CPS) and 6,000 housing units were currently being interviewed for the Quarterly Housing Survey (QHS). The CPS portion of the NTS sample consisted of a stratified multistage cluster national sample of 14,000 housing units in 376 primary sampling units (PSU's) and of a supplemental sample of 4,000 housing units from the CPS PSU's in 14 States, which was designed to increase the reliability of State estimates of travel.

The QHS portion of the NTS sample consisted of a stratified multistage cluster national sample of 6,000 housing units. The CPS households were divided into three panels of approximately equal size (6,000 households each) and different inter-viewing schedules were assigned to each panel. QHS households were denoted as Panel 4 households. In April 1977, households in all four panels, three CPS and one QHS were interviewed and asked to report trips of 75+ miles (one-way) since January 1, 1977.

Even though NTS was interested in trips of 100+ miles (one-way), it was decided to have a lower bound of 75 miles to minimize the number of 100+ mile trips that would be missed because the

respondent incorrectly though they were less than 100 miles.

Households in the first panel were visited again in May, August, November, and January 1978. Households in the second panel were visited again in June, September, and January 1978. The third panel was visited in July, October, and January 1978. The 6,000 QHS households were visited for NTS in the same months as households in panel 3. In all interviews following the initial interview, households were asked to report all trips of 75+ miles since their last interview. Households which were noninterviews in any visit were still asked to report trips taken since the last interview. For example, a panel 1 household interviewed in April but noninterviewed in May would be asked to report trips taken in April, May, June, and July when they were visited in August. A depiction of this interviewing schedule is presented in diagram 1 below.

Diagram 1 — NTS enumeration schedule

Month and Year of Enumeration	NTS Interviewed Panel(s)			
April 1977	1	2	3	4
May	1			
June		2		
July			3	4
August	1			
September		2		
October			3	4
November	1			
December				
January	1	2	3	4

NATIONAL TRAVEL SURVEY ESTIMATION PROCEDURE

The publication plans for NTS included a report presenting preliminary estimates of NTS data for January–March 1977 trips (i.e., the first quarter report) and a report presenting final estimates of NTS data for each quarter of 1977 and for the entire year (i.e., the annual report). As a consequence, it was necessary to develop two estimation procedures for NTS, a fairly simple and easy-to-implement procedure to expedite the preliminary estimates for the first quarter report, which is described in the first part of this section, and a more complicated procedure to produce the final estimates for the annual report, which is briefly described in the second part of this report.

First Quarter Report

The first quarter report for NTS covered travel for the months of January, February, and March and contained estimates for many different types of characteristics associated with travel. In April 1977, NTS sample households in all four panels were interviewed and reported travel for the period which covered the first quarter of 1977. Since households needed to be interviewed only once in order to obtain a full reading on the first quarter, problems of partial quarter data were not encountered in the first quarter weighting. However, since the first quarter report produced estimates of many major statistics, estimates produced for some characteristics would have been biased if all characteristics were assigned the same weight.

Each weight assigned to a characteristic constituted a different product of the components. Table 1 lists the specific product of components assigned to each characteristic. A description of each component and why it was needed to estimate

a specific characteristic will be described in the following paragraphs.

The 18,000 CPS households, which were part of the NTS sample, formed a national probability sample and the 6,000 QHS households, which were the other part of the NTS sample, also formed a national probability sample. Thus, each household in NTS had a known probability of selection from CPS or QHS separately. The basic weights represented the reciprocal of that probability. In certain instances, the basic weights assigned to households needed to be adjusted by the second component, the duplication control factors. These factors reflected the special subsampling done in the field when the number of observed housing units was much larger than expected and needed to be subsampled to reduce the interviewer workload to a reasonable size.

The household noninterview factors were applied to account for trips taken by members of occupied housing units which were eligible for the survey, but for which interviews were not obtained for any of the occupants of these units. These units, referred to as type A noninterviews, represented approximately 6 percent of the CPS units eligible to be interviewed in April for NTS and approximately 7 percent of the QHS units.

The within household noninterview factors were applied to account for persons who were not interviewed in April, but were living in households in which at least one person was interviewed in April. These persons were known as type Z noninterviews. Seven percent of the households eligible to be interviewed had at least one type Z noninterview. The within household noninterview adjustments were computed separately for CPS and QHS persons by region and cells which represented demographic characteristics of persons determined in the April NTS interview. This adjustment was equal to the following factor:

$$\frac{\substack{\text{Weighted Estimate of} \\ \text{Interviewed Persons}} + \substack{\text{Weighted Estimate of Non-} \\ \text{interviewed Persons in} \\ \text{Households in which at} \\ \text{Least One Person was Inter-} \\ \text{viewed (type Z noninterviews)}}}{\text{Weighted Estimate of Interviewed Persons}}$$

The number of type Z noninterviewed persons within an interviewed household was usually not more than one person. Thus, it was assumed that trips involving more than one person from a household would have been reported even if they involved a noninterviewed person within an interviewed household. However, trips that involved only a noninterviewed person in an interviewed household would be missed. For these reasons, the within household noninterview adjustments were only applied to trips and to characteristics of trips involving only one person from a household. Application of the within-household noninterview adjustment to all trips would have resulted in an overestimate of trips by adjusting for trips involving the noninterviewed persons, which would have been reported by interviewed household members. Thus, it was felt that applying this factor only to trips involving one person would result in more accurate trip estimates.

Although no exact estimate of the impact of this procedure was estimated, an approximate estimate can be obtained by using the overall within household noninterview factor of 1.04. If this overall factor has been applied to all household trips, the

Table I. Product of Components Assigned to Each NTS Characteristic for the First Quarter Weighting
(listed below are the definitions for each component)

Characteristic	Description of Characteristic	Product of Components Needed to Estimate Characteristic
1. Persons		(1) x (2) x (3) x (4) x (5) x (7b) x (8)
2. Households	Basic sampling and reporting unit consists of all persons occupying a housing unit such as a house, apartment or a room forming a separate living quarter.	(1) x (2) x (3) x (5) x (6) x (8)
3. Household Trips	A household trip is when one or more members of a household travel together to a place of 100 miles away from home and return together.	
a. One-person-household trips	Household trips involving only one member of a household.	(1) x (2) x (3) x (4) x (5) x (7b) x (8)
b. Two-or-more-person-household trips	Household trips involving two or more members of a household.	(1) x (2) x (3) x (5) x (6) x (8)
4. Person Trips	– A person trip is a trip taken by an individual.	
Person Miles	– Estimate of aggregate distance traveled by all persons who took a trip.	
Person Nights	– Number of nights spent by persons on a trip.	
a. One person-person trips	Person trips involving only one member of a household.	(1) x (2) x (3) x (4) x (5) x (7b) x (8)
b. Two-or-more-person person trips	Person trips involving two or more members of a household.	(1) x (2) x (3) x (5) x (7b) x (8)
5. Auto/Truck Trips	An auto/truck trip is when one or more households travel together in an auto/truck vehicle.	
a. One person-auto/truck trips	Auto/truck trips involving only one member of a household.	(1) x (2) x (3) x (4) x (5) x (7b) x (8)
b. Two-or-more person-auto/truck trips	Auto/truck trips, involving two or more members of a household.	(1) x (2) x (3) x (5) x (6) x (8)

(1) = Basic Weight
(2) = Duplication Control Factor
(3) = Household Noninterview Factor
(4) = Within Household Noninterview Factor
(5) = First Stage Ratio Estimate Factor
(6) = Second Stage Ratio Estimate Factor
(7) = Third Stage Ratio Estimate Factor
    a. Assuming type Z noninterviews* are good interviews
    b. Assuming type Z noninterviews* are noninterviews
(8) = Sample Survey Weight

* A type Z noninterviewed person is a person in an interviewed household for which trip information is incomplete but for which certain demographic information is known. Hence, for certain trip estimates the person can be considered as a good interview, while for other trip estimates the person can be considered as a noninterviewed person.

estimated number of household trips would have been 82,149,049 which is 2.7 percent higher than the published estimate of household trips. Thus, the 2.7 percent represents an estimate of the reduction in bias resulting from applying the within household noninterview factor only to trips involving one person.

Ordinarily, the distribution of the sample differs somewhat from the distribution of the total population from which the sample was drawn in terms of such characteristics as age, race, sex, residence, etc. Various stages of ratio estimation were employed for NTS characteristics in order to bring the distribution of the sample into closer agreement with the total population and thus increase the precision of the sample estimates. Ratio estimates were also employed to assist in correcting for known coverage problems in NTS, with regard to coverage of conventional new construction in permit-issuing areas and of new construction mobile homes, and within household coverage problems.

The first stage ratio estimate factor was applied only to households, persons or trip characteristics from NSR PSU's separately for CPS and QHS households. Its' purpose was to reduce the sampling variability resulting from the selection of a sample of non-self-representing (NSR) PSU's to represent NSR strata. For various categories of race and residence, ratios were calculated reflecting relationships between weighted 1970 census counts for all sample NSR PSU's in each region and the total NSR population of the region at the time of the census. All characteristics being estimated for the first quarter report received a first stage factor. The second stage of ratio estimation was designed to reduce sampling variability and to correct for known coverage problems in the NTS sample, with regard to the representation of conventional new construction and new construction mobile homes, by adjusting the NTS sample estimate of household and household trip characteristics to independently derived current household estimates for 12 residence-race of head-sex of head categories. This adjustment was applied to characteristics which would have been missed because of household undercoverage but not within-enumerated household undercoverage.

The third stage ratio estimates were designed to bring the distribution of sample persons from NTS into closer agreement with independent post-census estimates of the distribution of the population by various age-sex-race categories. Its purpose was also to reduce sampling variability and to correct for known household and within-household coverage problems in the NTS sample. The third stage ratio estimate for each specified category was equal to the following factor:

Independent Estimate of Civilian Noninstitutionalized Persons in Age-Sex-Race Cell for April 1977
_____
Sample Estimate of Civilian Persons in Corresponding Age-Sex-Race Cell for April 1977

These factors were applied to characteristics which would have been missed because of both household and within-household undercoverage.

As can be seen, each characteristic received either a second or third stage factor which accounted for only those aspects of undercoverage which affected that characteristic. If every characteristic had received a third stage factor, as had been done in previous NTS's, then biased estimates (overestimates) would have resulted for those NTS characteristics which were only affected by household undercoverage. Although no exact estimate of the impact of this procedure was estimated, an approximate estimate can be obtained by using the overall third stage factor of 1.0446. If this overall factor had been applied to all household trips, the estimated number of household trips would have been 80,269,269, which is .4 percent higher than the published estimate of household trips. Thus, this percent represents an estimate of the reduction in bias resulting from applying a second stage factor to characteristics affected only by household undercoverage and a third stage factor to characteristics affected by both household and within-household undercoverage.

Since the weighting procedure for the NTS first quarter report generated separate representative national estimates of NTS characteristics from the CPS sample units and the QHS sample units, the final weight for an NTS characteristic needed a component for combining CPS national estimates of characteristics with QHS national estimates of characteristics. This component was known as the sample survey weight and was denoted by W for CPS units and (1-W) for QHS units. The procedure used to estimate W and (1-W) was to estimate the proportion of total variance that might be contributed by CPS sample units and the proportion that might be contributed by QHS units. Estimates of variance produced by the NTS pretest were used to make these approximations. As a result, it was decided to apply a factor of $W = .775$ to each CPS sample unit and a factor of $1-W = .225$ to each QHS sample unit.

As can be seen from table I, the NTS first quarter estimation procedure utilized a different product of components for each characteristic in an attempt to produce the most accurate estimate for each characteristic. If one weight (i.e., the product of components corresponding to "Persons") had been used for all characteristics, as had been done for previous travel surveys, then biased estimates would have resulted for households, and household trips, person trips, person miles, person nights, and auto/truck trips for two-or-more-person trips. These estimates would have been biased since their weight includes adjustments for situations that do not affect these characteristics. However, this does not mean that the first quarter estimates were unbiased, since several sources of bias still exist. The noninterview adjustments procedure probably produces biased estimates since the assumption that the responses of interviewed and noninterviewed persons or households being similar is not necessarily true. Likewise, the different stages of ratio estimation also produce biased estimates. In addition, the different bias-reducing features of the first quarter estimation procedure may not be free of bias since the assumptions made in connection with these procedures may not be valid. (e.g., assuming that within-household noninterviews only affect one-person trip characteristics). Nonetheless, it can be stated that the first

quarter estimation procedure produced estimates with a lower bias then would have resulted if the bias-reducing features, described in the previous paragraphs of this section, had not been incorporated.

Annual Report

The annual report for NTS covered travel for the entire year and for each quarter. In order to cover the NTS travel year, households and persons needed to be interviewed four or five times depending on their panel. A household or person could have become a noninterview anytime after the household's first interview and remained a noninterview for the rest of the NTS travel year. Thus, NTS data would be available for only part of the year for these households. Persons interviewed in April could have moved out of the sample household anytime after April and because of the NTS mover rules would not have been followed and interviewed for the rest of the year. In addition, persons could have moved into a sample household from one interview to the next and thus such persons would have been interviewed for only part of the year. Due to these mover and noninterview problems, NTS data existed for only part of the year for such persons and households. Inclusion of partial year data for statistics which were calendar year dependent (e.g., a household or person distribution of the number of trips taken during the year) would result in biased estimates for such statistics. Thus, rules for deciding when to include partial year data for each of the statistics were developed for the annual estimation procedure. In addition, tabulations were to be produced for travel to and through each State as well as for the entire nation. Since the first quarter weighting was not designed to produce estimates of travel to and through a State, changes needed to be made in the first quarter weighting in order to produce more precise estimates of travel to and through each State.

The annual report estimation procedure employed bias-reducing features, similar to what had been used for the first quarter report estimation procedure, in an attempt to produce estimates as unbiased as possible. These included the application of a within-household noninterview adjustment to only those characteristics affected by this type of noninterview, the application of either a household undercoverage adjustment or a household and within household undercoverage adjustment to those characteristics affected by the respective type of undercoverage, and the creation and application of the "best" weight (i.e., the weight which would produce an estimate as unbiased as possible) for each major statistic. In addition, the estimation procedure also utilized some different bias-reducing and precision-increasing features as a result of situations not encountered in the first quarter. Data inclusion rules, which produced both estimates as unbiased as possible for all statistics and a fairly easy-to-use data users' tape, were incorporated for the annual estimation procedure to solve the problem of partial year data. Application of these rules resulted in some collected data being dropped, which necessitated the creation and application of sample ratio factors, precision-increasing factors reflecting the more reliable estimates that would have resulted if this data had not been dropped.

Sample survey weights and first stage ratio estimate factors were calculated and applied at the State level for the annual estimation procedure to produce more precise estimates of travel to and through each State. Second and third stage ratio estimate factors were also calculated and applied at the State level to reduce the bias created by undercoverage problems for each State separately. However, this is not to say that the annual report estimation procedure produced unbiased estimates since sources of bias still exist in connection with the use of the noninterview adjustments and ratio estimation for this procedure. In fact, given the way in which interviews and noninterviews were defined for this procedure and the numerous ratio estimation components, its much more likely that the estimates produced from this procedure are subject to some bias as result of these components. Nonetheless, the resultant estimates should be more reliable and less biased than if the bias-reducing and precision-increasing components, previously described, were not employed for the annual report estimation procedure.

MILEAGE ESTIMATION RESEARCH

In addition to the bias reducing features of the NTS estimation procedure, improved mileage estimation research was undertaken to reduce the bias associated with NTS data. The 1977 NTS had planned to utilized a computer system known as PICADAD - (Place Identification, Characteristics and Area, Distance and Direction) to estimate the mileage of trips reported by respondents. The PICADAD estimate of mileage was formed by calculating the straight line mileage for a trip based on the earth coordinates associated with the origin and destination of the trip, by multiplying this straight line distance by a circuity factor, and by adding this product to a side trip component. The circuity factor attempted to reflect that people do not travel in a straight line between two points since highways, train tracks and airplanes do not go in straight lines. A circuity factor of 1.25 was used for trips taken by train, a circuity factor of 1.15 was used for airplane travel, and separate circuity factors for each combination of between-State travel as well as separate factors for each State for within-State travel were used for auto, truck, and bus trips. The side trip component, which was equal to 25 percent of the straight-line distance, was only used for auto or truck trips and attempted to reflect the general tendency to take side trips or interesting routes on such trips.

Trips for which the PICADAD estimate of the round trip distance was 200 miles or more were to be included in NTS tabulations, and all trips for which this mileage estimate was less than 200 miles were excluded. Thus, it was very important that the PICADAD estimate of round trip mileage be accurate, especially for trips bordering the mileage cutoff of 200 miles, since the PICADAD estimate was the final determination of whether a trip was included or excluded in the NTS tabulations. Since it was felt that the PICADAD estimation procedure had several sources of error which could result in the incorrect inclusion or exclusion of a trip, improved mileage estimation research was undertaken for the 1977 NTS.

The improved mileage estimation research was conducted for trips taken by auto or truck which were reported in the April 1977 interview of NTS-QHS households. The focus of this research was to improve the estimates of mileage for those trips near the 200 mile cutoff, which were the most vulnerable with respect to PICADAD's errors influencing whether or not a reported trip is counted as an NTS trip. This is why the research included only trips taken by auto or truck, since a substantial portion of such trips are near the cutoff.

As part of the mileage estimation research, additional information was collected at the time of the NTS interview for those trips within the scope of this research. This additional information included the names or numbers of the main highways or routes taken on these trips, whether or not any side trips, that involved other main roads were taken, and the city and State to which each of these side trips was taken.

Utilizing this supplemental information for these trips, a clerical operation was implemented, to produce an alternative estimate of mileage for each of these trips. Specifically, this clerical operation involved the determination of the route taken between the origin and destination, based on this supplemental information, and the measurement of the mileage for this route based on an opisometer, a wheel-like device which was rolled along this route on a map. 1/

It was felt that this clerical estimate of mileage would be more accurate since this mileage estimate reflected the actual routes and side trips taken for the trip.

The results of a comparative analysis between the PICADAD and clerical estimates for each trip, indicated that the PICADAD estimates of mileage were significantly greater than the clerical estimates. Assuming the clerical estimates of mileage are the true estimates, the results indicated that approximately 17 percent of the auto/truck trips would have been erroneously included or excluded in the NTS tabulations if the PICADAD method were utilized. Thus based on this research, it was felt that mileage for all auto/truck trips should be clerically estimated. However, time and cost constraints eliminated this alternative for the 1977 NTS survey.

An attempt was then made to find a factor to be applied to the PICADAD mileage estimates which would reduce the percentage of trips misclassified as well as the difference between the percent erroneously included and the percent erroneously excluded. Application of a factor of .86 to the PICADAD mileage estimates for auto/truck trips reduced the percentage of trips misclassified as well as the difference between the percent erroneously included and excluded. With a factor of .86, 4.2 percent of the trips involved in the research were erroneously included and 4.7 percent of the trips were erroneously excluded. The estimated error rate with this factor was thus 8.9 percent with the difference between the two percentages being .5.

Thus the results of this research indicated that if a factor of .86 was applied to all PICADAD estimates of mileage, the bias associated with the estimate of 200 plus mile trips as a result of inaccurate mileage estimation would be substantially reduced. Since the clerical method of estimating mileage could not be incorporated in our survey, the PICADAD method with a factor of .86 was incorporated.

FOOTNOTE

1/ Credit should be given to Gary Shapiro of the Statistical Methods Division at the Census Bureau for originally suggesting this method of mileage estimation.