

IMPROVING THE QUALITY OF DATA FROM MONTHLY GASOLINE PURCHASE DIARIES

Wendel L. Thompson, U.S. Department of Energy
Lynda T. Carlson, U.S. Department of Energy
Thomas H. Woteki, U.S. Department of Energy
Kenneth A. Vagts, U.S. Department of Energy

The Office of the Consumption Data System, Energy Information Administration, has the responsibility for collecting data on energy consumption in all sectors of the economy. The data is to be used for monitoring, modeling, and policy evaluation. One area of major concern is the energy consumed by the residential sector in the use of personal vehicles.

In our original plan for the Residential Energy Consumption Survey, we included a series of questions on vehicle use and household transportation. These questions related to the type of vehicles used, number of miles driven for long distance and local driving and miles per gallon (MPG) obtained by the vehicle. Although previous surveys in the U.S., England, and Canada indicated the data would be of questionable validity, they were included anyway due to the importance of energy consumption in transportation. In fact, respondents from the National Interim Energy Consumption Survey (NIECS) provided so little useful data that the missing data could not reasonably be imputed from other data sets or from the sample itself.

When we asked households for their annual mileage estimates, the range was from 300 to 129,000 miles per year with most estimates clustered between 9,000 and 13,000. Information on the number of gallons consumed, which we inferred from respondents' estimates of their mileage, ranged from 24 to 9,055 gallons per year. Analysis of the data indicated that, for the most part, households did not have an accurate perception of their driving or gasoline consumption.

Because we realized the problems we would have with household self-reporting of miles driven and gallons consumed, we had always planned to have a small panel of households reporting accurate consumption data which would allow us to adjust the yearly estimates for the entire sample of households. The first avenue we explored was that of metering vehicles in the panel to obtain consumption and mileage information. However, we found this approach to be expensive, cumbersome, frequently unreliable, and likely to have impacts on biasing the sample vehicles. Therefore, we planned a small, limited transportation panel in which sample households from the residential survey would maintain vehicle diaries to be used to adjust the self-reported estimates of households.

After examining the NIECS data we realized it would be impractical to adjust the NIECS self-reported data and that we would be better off enlarging the transportation panel sample to provide the annual estimates of consumption, miles driven, and miles per gallon directly. Since we were going to use the transportation panel as the actual data set, we determined to make it as useful as possible by developing a sampling and data collection pattern which would reflect the seasonality in miles per gallon and

miles driven (miles driven are traditionally expected to be higher in the summer months, whereas cold weather conditions are supposed to lower the miles per gallon achieved by a vehicle). Therefore, we developed a pattern to capture the seasonality component as well as to test whether in fact there were differences within a household in driving patterns and vehicle efficiency in gasoline consumption during different points in the year.

The plan devised to obtain these accurate annual estimates was to have households from the NIECS sample join the panel for two months, keeping separate diaries for each vehicle for each month. The household then would leave the panel for four months and return for two months. This pattern would allow us to capture the seasonality component as well as provide a large enough base to develop annual estimates of consumption by vehicle and household. The plan was to pretest the panel in April and May of 1979. In addition, a second pretest was set for May and June to determine if we could get households to obtain a full tank of gas at the beginning and end of the month.

During the pretests, a major gasoline crunch developed with long lines at gas stations, limitations in the number of gallons sold per vehicle, and limited availability of certain fuel types in certain geographic areas. It became apparent that the panel could be utilized to serve several additional purposes. First, we could monitor month to month changes in consumption, miles driven, and miles per gallon. Second, we could collect limited data on the presence of gas lines and types of fuel availability. Third, we could obtain data to meet EPA's need for an indicator of fuel switching from unleaded gasoline to leaded gasoline. Therefore, starting in June 1979, the pretest became an actual data collection with a monthly panel of 500 households (which increased to 1,000 households in January 1980).

Description of the Panel. The Household Transportation Panel is a national sample. In each Panel household, the principal driver of each vehicle is asked to report on every purchase of fuel during the month. The driver keeps the diary in the vehicle and, at the time of each purchase, records the odometer reading, the number of gallons purchased, price per gallon, the type of fuel, and the total expenditure. After turning on the engine and allowing the fuel gauge to reach its maximum position, the driver records the fuel gauge reading. This data on the fuel gauge is used to estimate how many gallons were in the tank initially.

Each driver is given \$2 at the beginning of each month as an incentive to maintain the diary through the coming month. Within a week after the end of the month, a telephone call is made to collect the data recorded in the fuel diary.

The caller also asks a few topical questions (the subject during the first year was "waiting in line for gas").

Diary Completion Rate. The rate of diary completion was of utmost concern to us from the beginning of the project. The completion rate was the subject of an extensive national pre-test in April and May 1979. The test was designed to indicate what effect a one- versus two-month reporting period and a telephone versus personal contact would have on the completion rate. Based on the pretest, we expected the procedures put into effect in June 1979 to produce a cooperation rate of about 60 percent. The first three months of full-scale data collection (June, July, and August 1979) produced completion rates of 66 percent, 58 percent, and 53 percent, respectively, for eligible vehicles in each of the three months. This set of figures does not represent an increasing trend of noncooperation. Rather, it appears to represent the range from high to low in the level of cooperation we can expect to find. The August figure of 53 percent may represent the low point when persons take vacations and don't want to be bothered with a diary. More recent data for February through May 1980, showed a success rate of 61 percent.

Panelists are asked to report for two consecutive months. Cooperation in the second month is expectedly lower than it is in the first month. The experience with the panels that first reported in June and July showed the following decreased participation in the second month.

Percentage of Eligible Vehicles With Completed Diaries for the Month

	<u>First Month</u>	<u>Second Month</u>
June Panel (Base)	66% (1052)	54% (1041)
July Panel (Base)	62% (1084)	55% (784)

These levels of participation are too low for us to be satisfied with the Panel as it is currently operated, especially since the Panel is a subsample of a national survey (NIECS) that had its own completion rate of 85 percent of all eligible households.¹ Clearly, the panel procedures must be modified to improve the rate of cooperation.

Our thoughts about what to do to improve cooperation fall into four groups: (1) use of larger incentives; (2) creating a greater feeling of involvement on the part of the panelists; (3) changing the policy about following movers; and, (4) redesign of the diary form.

Larger Incentive. Tests on the effect of using a \$5 incentive rather than a \$2 incentive were carried out in February and March 1980. The tests were conducted on randomly assigned subsamples of the national panel and showed that we could expect a significant increase in the response by switching to a \$5 incentive. The \$2 incentive we have been using produced a response of 58 percent among the 412 panelists in February and March. Compared with this, the \$5 incentive produced a higher rate of

70 percent response among 406 panelists. The \$5 incentive represents a considerable increase in the costs of conducting a panel. If it were to be adopted for all households, the total cost for the panel could increase by 10 to 15 percent. The incentive need not be raised from \$2 to \$5. A figure in between would be expected to produce improved participation over the present \$2 incentive.

Another somewhat less costly approach would be to continue with the present \$2 incentive at the beginning of each of the two months reporting and send \$5 to each driver at the end of the second month when the second diary information was received. This procedure would be less costly than the \$5 incentive since the \$5 reward is given only to drivers who complete the diary and not to everyone. This less costly approach produced a response rate of 66 percent among 408 panelists receiving that experimental treatment. This response is statistically improved over the 58 percent response in the group receiving the present \$2 incentive. However, with the present sample size, the 66 percent for the \$5 reward is not statistically different from the response of 70 percent for the \$5 incentive. The experiment was continued through April and May. If the difference between the \$5 reward and \$5 incentive remains at 4 percentage points throughout those two additional months, the larger sample size over the four-month period may mean the \$5 incentive produces a statistically improved response rate. We are evaluating the costs involved on a yearly basis to determine how soon we can increase the incentive to households.

Greater Involvement. A second method for improving cooperation is to create greater involvement in the project among the panelists in the hope that they will become more personally committed to the objectives of the study. Two actions directed to this end are (1) to get a verbal commitment to keep the diary in the recruitment phase and (2) to provide some results of the study in the form of a pictorial brochure. At present, the respondent is not asked to commit herself/himself to complete the diary. An expectation of cooperation is assiduously maintained and materials are sent to each eligible household except those who refuse to participate. We propose to ask each respondent "Will you help us?" Those who are reluctant will be questioned and prepared lines of argument followed to deal with the objections. For example, if someone thinks it will take too much time to keep the fuel purchase diaries, we can propose a reduced amount of record-keeping such as reporting only the gallons purchased and the odometer readings. Failing that, we might even retreat to accepting odometer readings at the beginning and ending of the month. Even this bit of information will be useful in reducing the uncertainty surrounding the behavior of those households not presently participating in the panel.

Movers. A third approach to increasing cooperation is to change our policy on following moving families. We have attempted to follow a household when a majority part has moved. Following a household that moves is problematic

and not always successful. One proposal is to define our sample as the housing unit rather than the household which resides in it. When a move occurs, we would not attempt to follow the mover, but would instead take the new resident household as the sample household. This procedure would probably mean an increased success rate since the location of the household is already known. If contact with the new household could not be made through the mail or by telephone, an interviewer would be sent to the address to make the initial contact. This policy has one disadvantage; it reduces the continuity of households that report over a two-month period and a later two-month period. As presently planned, a household reports for two months, and then reports for another two-month period after a four interval. Thus, a family that stays with the panel provides mileage data over an eight-month period and consumption data in two different seasons of the year. A policy to stay with the housing unit when a move occurs would reduce the number of same households available for a time-series comparison over different seasons.

Diary Form. A fourth plan is to change the present TV Guide-shaped diary into a business envelope-shaped diary and to provide a plastic holder with pencil that could all be attached to the sun visor in the panelist's car. The easy access and ready visibility may serve to remind the driver to record the fuel purchase information.

Salvaging Incomplete Records. About 10 percent of each month's diaries are incomplete. However, they do contain a minimum amount of data which makes their salvage possible. In view of the low response rate, this salvage operation is very worthwhile. A diary is considered complete for the month if it contains total mileage, total gallonage, and a reasonable miles per gallon figure (MPG). The MPG figure is calculated from the information provided in the record; it is not provided directly by the respondent.

Some records are obviously incomplete, leading one to suspect that the odometer might be inoperative or that someone recorded the odometer readings each time fuel was purchased but did not record any data from the fuel pump. However, there occurs another kind of incompleteness which is harder to detect--that of the missing purchase. Missing purchases are detected when we compare the MPG figure calculated at each purchase with the overall MPG for the month.

For example, assume a vehicle had an average of 20 MPG over the month based on the miles and gallons reported in the diary. We also calculate a MPG between purchases. If this purchase MPG is within 25 percent of the overall MPG for the month, the record is accepted as complete.² If the purchase MPG differs from the monthly MPG by more than 25 percent, the record is visually inspected. The inspection results in a judgment that the record contains "good gallons," "good miles," or that the record is not usable. A record that contains "good miles" or "good gallons" is completed by using a MPG to calculate the missing data element. This MPG is not the same as the one used earlier to detect missing purchases. This time the MPG is calculated

from the diary record itself only if there are least 300 miles between consecutive purchases and the variation of the purchase MPG's is small. If the diary record cannot be used, the MPG is imputed from a distribution of MPG figures for other vehicles of the same class (or size) and model year for which complete diary records are available.

Using an MPG to salvage these incomplete records, we have increased the response rate, and have been able to include a group of vehicles which are used more on the average than vehicles with complete records. The figures below show greater use of vehicles when the vehicles with imputed records are included.

Averages per Vehicle ³	Total Sample With Imputed Records	Total Sample Without Imputed Records
Miles Driven In:		
June 1979	834	805
July 1979	883	855
August 1979	827	805
Gallons Consumed In:		
June 1979	55.8	54.0
July 1979	58.6	56.5
August 1979	56.2	54.1

This fact leaves us with the nagging thought that the vehicles for which we have no records may also be used more on the average than those in the panel with complete records. If so, this would constitute a serious bias in our results. Further investigation of this issue has high priority for us in our efforts to understand more about the data we are collecting.

In addition to these data salvage procedures, we are actively considering development of a CATI system--Computer Assisted Telephone Interviewing. With this system, the telephoner would enter the data directly on a computer terminal while the respondent is on the phone. Instantaneous computer editing of data entries would provide feedback for telephone interviews who could immediately question respondents about incomplete or inconsistent entries. This system promises to decrease the processing time and may help to salvage more records.

NOTES

¹An additional 5.3 percent of eligible households were contacted by mail, but these households were not brought into the Transportation Panel. Although the figures for completion rates are variously quoted as based on vehicles or households, the fuel purchase diary is completed for a vehicle, not a household. The completion rate for households is a few percentage points higher than the completion rates for vehicles. This occurs because multi-vehicle households participate more often. Also, multi-vehicle households have "complete" records only if reports are available for half or more of the household vehicles.

²Variation in MPG can also be caused by the unreliability of our estimates of gallons consumed between purchases. This is especially true when the tank is not filled up at each purchase and we must rely on fuel gauge readings to estimate the gallons consumed. For this reason, if fewer than 150 miles were driven between purchases, the purchases are cumulated until a minimum of 150 miles is reached and the MPG calculated over the cumulated distance.

³Vehicles that were not driven during the reporting month were not included in the base in calculating these averages.