

Appendix M: References for Petroleum Marketing Program

Bradsher-Frederick, Howard, [A Methodology for Evaluating Sufficiency of Survey Frames](#), American Statistical Association Conference, Seattle, Washington, August, 2006. The Energy Information Administration (EIA) has continued its strategic planning efforts to maintain and improve data quality and its movement toward employing a greater number of performance measurements. As part of that effort, EIA evaluated its own survey frames for “sufficiency.” This effort involved the development of a set of evaluation criteria that called for collecting a large quantity of information and data for each of 34 EIA master frames. An inter-office team developed the evaluation criteria and was responsible for the data and information collection and also applied the criteria to evaluate the 34 master frames. The team eventually decided that all but four of the frames were sufficient; the insufficient frames were deemed “insufficient” for a variety of reasons.

Cox, Brenda, [Petroleum Marketing Program Evaluations](#), American Statistical Association Committee on Energy Statistics, Washington, D.C., 2006. This report summarizes survey method and provides recommendations for improvement of the Petroleum Marketing Program. These recommendations are intended to improve: (1) program and survey documentation, (2) survey procedures, (3) Consistency across surveys in program, (4) efficiency, and to prevent or mitigate risk. The agency selected to divide these recommendations into the following categories: must (those changes which are essential and affordable and can be accomplished immediately); should (those changes which are essential and either not currently affordable or not able to be accomplished in the foreseeable future); and would (those changes which are not essential but desirable sometime in the future).

Cox, Brenda and Kirkendall, Nancy, [Developing A Survey Program Evaluation Process](#), 2004 Fall meeting of the American Statistical Association Committee on Energy Statistics, Washington, D.C., October 28, 2004. The Performance Assessment Rating Tool (PART) was instituted to encourage rigorous performance assessment to boost the quality of Federal government programs. For survey programs, an important component of performance assessment is the evaluation of each of the program’s survey components and then the evaluation of the effectiveness of the overall program. This paper summarizes the results to date of a study to develop templates for evaluating individual surveys and to evaluate a program composed of a family of surveys.

For this investigation, we will develop two templates, one for the evaluation of an individual survey and one for the evaluation of a family of surveys that form a survey program. The goal will be to develop templates that an external group could use as guides to conduct survey program evaluations. To validate and refine the templates, we will use the templates to evaluate EIA's Petroleum Marketing Program and its 11 component surveys. Then, the templates will be updated to create the penultimate templates for survey and program evaluations.

Harper III, Robert G., "[Comparisons of Independent Petroleum Supply Statistics](#)" from *Petroleum Supply Monthly*, Washington, D.C., October, 2005. This compares final petroleum data published in the PSA with similar petroleum data obtained from other sources. Data comparisons are presented for 1994 through 2003. Archived versions available.

Hassett, Nancy, [Combining Data to Produce Timely Estimates](#), *Proceedings of the American Statistical Association, Survey Research Methods Section*, August, 1995. The Energy Information Administration (EIA) publishes prices for crude oil and petroleum products in the *Petroleum Marketing Monthly*. Crude oil prices are obtained from three EIA surveys. Petroleum product prices are based on the EIA-782 survey, a monthly survey of refiners, resellers, and retailers of various petroleum products. Due to the nature of the surveys and the required data processing time, results are not available until over two months after the reference period. Alternate price series and methodologies are examined as a means of producing more timely price estimates at the national and regional levels. One such method is the use of time series models using available prices from a variety of sources.

Estimates are desired at both the wholesale and retail level for several petroleum products. This paper will focus primarily on wholesale regular unleaded gasoline prices. Forecasts are required at both the national level and the Petroleum Administration for Defense District (PADD) level 1.

Hallquist, Theresa, et.al, [Forecasting Petroleum Product Prices](#) *Proceedings of the American Statistical Association, Survey Research Methods Section*, August, 1996. The Energy Information Administration (EIA) wishes to produce price estimates of selected petroleum products in a more timely manner. Currently, estimates for petroleum product prices are published 2-3 months after the end of the reference period. Our customers would like to have the information earlier than we usually publish. Evidence of this was seen in the results of the EIA customer surveys done in 1995 and 1996. The information from the two customer surveys shows high levels of satisfaction with customer service

provided by EIA staff. However, the surveys show lower levels of satisfaction with timeliness of our data release relative to the other areas the surveys measured. As a result, EIA has targeted timeliness as an area needing improvement. The goal of this project is to publish prices within 2 weeks of the end of the reference month. The accuracy of the estimates should be within 1cent. These estimates would then be finalized through the current process and final numbers would be published according to the current schedule.

O'Colmain, Benita, et.al. , [Variance Estimation for EIA-877 Propane Prices](#), *Proceedings of the American Statistical Association, Survey Research Methods Section*, August, 1994. The Energy Information Administration (EIA) receives No. 2 fuel oil and propane price data from various State Energy Offices on a semi-monthly basis during the winter months as part of the EIA-877 State Heating Oil and Propane Program (SHOPP). The individual State Energy Offices telephone a sample of propane and heating oil companies and request current price data for each reporting period. State and regional aggregate price estimates are published in the EIA Winter Fuels Report.

The price estimates derived from the survey are subject to sampling error. The heating oil sample is drawn using a stratified random sampling technique and variances of heating oil price estimates are computed using a classic stratified random sample variance formula. The propane sample is drawn using a stratified systematic sampling technique and requires a modification to the classic variance formula. This paper describes the proposed variance estimator for propane prices and presents the results of tests using this formula in comparison with the classic variance formula.

Saavedra, Pedro, et. al., [An Improved Imputation Methodology Derived Through Regression Trees](#), *Proceedings of the American Statistical Association, Survey Research Methods Section*, Washington, D.C., August, 2009. The EIA collects monthly information on the balance between supply and disposition of crude oil and petroleum products through a family of surveys. The process requires all imputed values to be available before responses are received, but uses values for only select cells. Previous analysis led to the recommendation for some surveys to implement an imputation method using historical values obtained through exponential smoothing and trend adjustments from a weekly survey. One survey was particularly difficult to resolve because of more extensive dimensions of the survey, the many cases of zero values, and fewer comparable cells in the weekly survey for trend adjustment. To group cells, a regression tree method (CART) was used to obtain groups for which the

same smoothing coefficient could be used. Simulation analyses were then conducted to identify an optimal coefficient for each group.

Saavedra, Pedro, et. al., [Comparison of National and Regional Gasoline Prices from Two Surveys](#), *Proceedings of the American Statistical Association, Survey Research Methods Section*, August, 1994. The U.S. Department of Energy's EIA-782 is a national monthly price and volume survey of refiners, resellers, and retailers of various petroleum products including gasoline. The EIA-878 is a national weekly survey of regular unleaded gasoline prices at the pump. Gasoline prices based on the EIA-782 are published monthly in the *Petroleum Marketing Monthly*; however, due to the nature of the survey and the required data processing time, the results are not available until over two months later. The Energy Information Administration (EIA) of the Department of Energy would like to produce more timely estimates of national and regional gasoline prices using alternative approaches. One such approach is the use of the EIA-878 survey data to produce the estimates. This paper compares EIA-782 price estimates versus those derived from the EIA-878 for the period October, 1992 through September, 1993.

Saavedra, Pedro, et. al., [Implicit Stratification and Sample Rotation Using Permanent Random Numbers](#), *Proceedings of the American Statistical Association, Survey Research Methods Section*, August, 1998. There is a class of sampling techniques which can be collectively described as order sampling. In this class of techniques a random number is assigned to each member of the frame and the sample is in some way drawn so that among units with similar characteristics (stratum, size, etc.) those with the smallest numbers will be selected. When the random number is preserved in order to control the overlap of the sample with a second sample from an overlapping frame, we speak of a Permanent Random Number (PRN). While the techniques presented in this paper were motivated by the use of Permanent Random Numbers, the first of these techniques is relevant to any form of order sampling, whether or not the numbers are used to preserve the overlap with other surveys. However, since there are better ways to achieve the objective than using order sampling when PRNs are not needed, the technique is particularly relevant to order sampling. The second technique is relevant when Permanent Random Numbers are used to control overlap and one wishes a particular kind of overlap (say close to 50%) at all levels of the sample (i.e., regardless of size or stratum). Both of these techniques were used in the new design of the EIA-782 and are especially useful in an entire class of sampling designs.

Saavedra, Pedro, et. al., [Imputing Price as Opposed to Revenue in the EIA-782 Petroleum Survey](#), *Proceedings of the American Statistical Association, Survey Research Methods Section*, August, 1993. The EIA-782B is a monthly price and volume survey of petroleum resellers and retailers. Every month, preliminary results are published for the current month and final results for the previous month. Missing data and data that fail the edits are imputed using a combination method which involves predictive ratios of forecasted volumes and revenues. The predicted value is obtained using exponential smoothing for volume and revenue. This approach has the disadvantage that the same alpha coefficient affects both revenue and volume, while there is some evidence that more recent values are better predictors for price than for volume, thus suggesting that different alpha coefficients for price and volume would be more efficient. An alternate approach predicting price and volume separately, and using an imputed price for exponential smoothing when the volume is zero was simulated under different conditions and compared with the current approach.

Saavedra, Pedro J, Linking Multiple Stratifications: Two Petroleum Surveys. *Proceedings of the American Statistical Association, Survey Research Methods Section*, New Orleans, Louisiana, 1988, page 777-781.

Saavedra, Pedro, et. al., [The New Design for the EIA-878 Gasoline Price Survey](#), *Proceedings of the American Statistical Association, Survey Research Methods Section*, New York City, New York, August, 2002, pp. 3019-3023. The EIA-878 is a survey of motor gasoline outlet prices that produces estimates of national and regional level prices, as well as separate estimates for several states and cities, two formulations and three grades of gasoline. Up to recently, this survey has used a monthly survey of resellers and refiners as phase I of a multi-phase sample, subsampling the sample units of the monthly survey that report the specific outlet sales category. A new design extends coverage to independent stations, and targets additional states and cities. The design is an area sample that uses data from several sources, allocating stations to counties and sampling stations from the selected counties. Weights make use of the number of places at which gas may be pumped as a proxy for volume and the proportion of gas by grade in the state. Several data sources are used for analytic purposes and to obtain allocations and size measures.

Saavedra, Pedro, et. al., [The Use of Variant Poisson Sampling to Reduce Sample Size in a Multiple Product Price Survey](#), *Proceedings of the American Statistical Association, Survey Research Methods Section*, Anaheim, California, August, 1997. The EIA-782 Monthly Petroleum Product Sales Report collects state level prices and volumes of petroleum products by sales type from all refiners and a sample of resellers and retailers. The data collected are aggregated to produce approximately 30,000 estimates and are published in the Petroleum Marketing Monthly. The basic sample design of the survey has been in effect since 1984, but has been modified in minor ways to reflect new information or changes in the market. Samples have been rotated during that time to reduce the individual company burden. The current sample is the eleventh such sample.

The basic design made use of two groups, a certainty group and a noncertainty group. For each of eight targeted product/end-use categories, the noncertainty group was stratified by sales volume and urbanicity and then sampled within each stratum. A select set of state level average prices was targeted at a 1% Coefficient of Variation (CV) for determining sample sizes. These price CVs roughly correspond to volume CVs of 10% or 15%, depending on the petroleum product. Neyman allocation was used to determine the sample size required for each targeted product/end-use category. A triennial survey of all sellers of petroleum products provided state level sales volumes at the targeted levels and was used as the sampling frame and basis for stratification.

Sample selection was carried out using a linked sample procedure. In this process a respondent was selected randomly from the frame and used simultaneously to satisfy the required allocation in each of the targeted products. If the respondent's stratum had already reached the required allocation for one or more target variables, but not all, the respondent was considered to be a volunteer or visitor for those variables. In the target variables for which the respondent helps to satisfy the allocations, the respondent was considered to be in the basic sample. The linked selection reduced the overall sample size by using each selected respondent to satisfy multiple requirements. Because the selection was not independent, the probability of selection for a sampled unit could not be calculated directly. Instead, the probabilities were derived by simulating 1000 sample selections and counting the number of times each respondent was selected. The inverse of the frequency of selection divided by the number of simulations was used as the sample weight for estimation. In this design, the desired C.V's, therefore, drove the sample allocation process, calculated separately for each basic sample.

With a reduced budget in 1996, the focus shifted to reducing survey operations' costs significantly. These costs were directly associated with the sample sizes because operational efficiencies were said to have already been fully realized. It was determined that the expected budget in 1997 would be sufficient to operate a sample of approximately 2000 companies, compared to the current sample of approximately 3000 companies. This sample was, therefore, named Sample 2000. In that an operational sample 66% as large as the previous sample represented a tremendous decrease, it was expected that some variables would no longer be targeted and design CVs would be increased for other variables.

Also, the requirement for state level estimates for all states, all variables, would have to be loosened to reach the reduced sample size. The various combinations of CVs for the target variables at various geographic levels that were possible to achieve a total sample of approximately 2000 were numerous. The linked selection procedure, however, did not lend itself to easy computation of individual target variable contributions to total sample size to compare the variety of scenarios. In addition, because frame data were available for the first time for one of the petroleum products by end-use type, the number of target product/end-use variables was expanded to ten.

Saavedra, Pedro, et. al., [Using Order Sampling to Achieve a Fixed Sample Size After Nonresponse](#), *Proceedings of the American Statistical Association, Survey Research Methods Section*, Miami Beach, Florida, August, 2011. There are situations when a study requires a fixed sample size, either for contractual reasons or because the cost of collecting data for too many cases is prohibitive. This makes the preferred practice of oversampling and then adjusting for nonresponse impractical. Under certain conditions a simple random sample can be obtained by randomly sorting the frame and selecting the first n in the random order. This yields a fixed initial sample size, but a variable respondent sample. In a case where potential respondents beyond the targeted number of completes can be approached in sequential order (exhausting contact attempts before going to the next unit), the sampling process can continue until the desired number of completes is obtained. Nonresponse adjustments can then be made as if the combined set of respondents and nonrespondents constituted an initial sample. A similar approach to the one described above could be used to achieve a fixed number of completes using Sequential Poisson Sampling or Pareto Sampling. Here the probability of selection is changed, but the difference may be minimal. Simulations using SRS, SPS and Pareto were conducted to examine this practice.

U.S. Energy Information Administration, [Comparison of Selected EIA-782 Data With Other Data Sources, 1998 - 2008](#), Office of Petroleum and Biofuels Statistics, Washington, D.C., April 1, 2010. This article compares annual average prices reported from the EIA-782 survey series for residential No. 2 distillate, on-highway diesel fuel, retail regular motor gasoline, refiner No. 2 fuel oil for resale, refiner No. 2 diesel fuel for resale, refiner regular motor gasoline for resale, and refiner kerosene-type jet fuel for resale with annual average prices reported by other sources. In terms of volume, it compares EIA-782C Prime Supplier annual volumes for motor gasoline (all grades), distillate fuel oil, kerosene-type jet fuel and residual fuel oil with annual volumes from other sources. Archived reports available from 1993.

U.S. Energy Information Administration, [Petroleum Marketing Annual Explanatory Notes](#) in 2009 [Petroleum Marketing Annual](#), Office of Petroleum and Biofuels Statistics, Washington, D.C., August 6, 2010.

U.S. Energy Information Administration, Quality Assessments (internal reports), Statistics and Methods Group, Washington, D.C., March, 2005. These reports summarize interviews with survey manager seeking to identify strengths and areas of improvement for survey operations.

U.S. Energy Information Administration, [EIA Survey Forms](#), Washington, D.C., May 22, 2012. Contains EIA survey forms and instructions.

Waugh, Shawna, [Planning and Integration of EIA Survey Program](#), *Proceedings of the American Statistical Association, Survey Research Methods Section*, Seattle, Washington, August, 2006. One of the challenges facing the Energy Information Administration (EIA) is planning and integrating survey programs. One initiative in 2006 was to utilize recommendations from survey and program evaluations to redesigned and integrate surveys in the Petroleum Marketing Program. This paper provides information on conditions and motives for using these recommendations to redesign survey instruments and instructions and information concerning the results of redesigning the Petroleum Marketing Program. This paper also summarizes lessons learned, including the primary reasons recommendations were not adopted by survey managers. As a result of this project insights concerning the importance of achieving balance through trade-offs among cost, timeliness, and quality.

Waugh, Shawna, [Achieving Information Quality via Continuous Quality Improvement](#), *Federal Committee for Statistics and Methodology (FCSM) Conference*, Washington, D.C., January, 2012. This paper focuses on a recent initiative to redesign the Petroleum Marketing Program (PMP), using a hybrid framework for information quality and data quality and the Continuous Quality Improvement (CQI) cycle. A CQI cycle was implemented to achieve data and information quality in the Office of Petroleum and Biofuels Statistics (PBS) of the U.S. Energy Information Administration (EIA). The CQI cycle consists of four stages - plan, implement, monitor and evaluate.

When designing a new survey or program, one starts with the planning phase, however for existing programs the first step is the evaluation phase. An evaluation was conducted in 2011 of the PMP to identify and select recommendations to implement in the 2013 PMP during the Office of Management and Budget (OMB) forms clearance process. Many of the recommendations identified from several evaluations have been adopted during the planning phase, in preparation for implementation of the program in 2013. Additional recommendations may be implemented in the future as the CQI process continues from one survey cycle to the next one.

PBS manages two information data collections – the PMP and the Petroleum Supply Program (PSP). Combined these two programs provide weekly, monthly and annual statistics pertaining to petroleum supply, demand, and price, including prices for crude oil and petroleum products in the United States.

A comprehensive review of the entire program is conducted every three years during the Office of Management and Budget (OMB) forms clearance process. This process started with compiling, reviewing, evaluating and prioritizing recommendations intended to enhance information quality and data quality. A review was conducted of recent and previous reports of the program and associated surveys, both internal and external. The review included reports from outreach activities with internal and external customers, internal self-assessments by survey managers and external evaluations of the program, including peer review of the survey questionnaires.

A hybrid framework combining existing information quality and data quality frameworks was used to sort recommendations into several dimensions of quality related to the program, to the products and to the surveys. OMB's Information Quality Guidelines was used to sort recommendations pertaining to the program. The seven dimensions of information quality introduced by Gordon Brackstone from Statistics

Canada were used to sort recommendations pertaining to publications. The framework from OMB's Statistical Policy Working Paper Number 31, Measuring and Reporting Sources of Error in Surveys, was used to sort recommendations pertaining to sampling and nonsampling errors.

A team of survey managers, methodologists and contractors evaluated which of the 200 potential recommendations to adopt, of which many would have been adopted had resources not been a constraint. Forty of the recommendations were adopted, many of which involved modifications to the survey questionnaire and instructions to enhance consistency across surveys in the program.

There are several challenges when planning and implementing the 2013 PMP. One challenge involves integrating ten surveys into a comprehensive program and another integrating this program with the PSP which provides an overview of the petroleum flow in the United States. Another challenge is providing publications for a variety of customers - policy makers and analysts with federal/state/local government agencies, petroleum and other industries, along with the media and public. These customers use select data for a variety of purposes.

Consequently, a new product proposed for 2013 is a Product Profile for each PBS publication. A separate Product Profile will supplement the existing technical notes in the publications. The Product Profile will provide additional documentation regarding sources of data and their limitations, survey methods and appropriate uses.

Weir, Paula, [A Comparison of the Use of Telephone Interview to Telephone Audio CASI in a Customer Satisfaction Survey](#), *Proceedings of the American Statistical Association, Survey Research Methods Section*, Indianapolis, Indiana, August, 2000. For the last six years, the Energy Information Administration (EIA) has conducted an annual satisfaction survey of customers calling the National Energy Information Center (NEIC) during a three-day period. In this survey, volunteer staff conducted interviews and asked callers to rate EIA's products and services on five attributes, such as courtesy and timeliness. This year the satisfaction survey was conducted by volunteer staff interviews, as well as, by Telephone Audio Computer Assisted Self-Administered Interviewing (CASI). The Telephone Audio CASI, TA CASI, used a pre-recorded interview in which respondents pressed the appropriate buttons on their telephones to respond to the same set of questions. It was hoped that this relatively inexpensive use of technology would not only free up staff time in the future, but more importantly, would allow the

survey to be conducted over a longer time period, thereby producing larger sample sizes and a greater ability to distinguish statistically significant changes in customer preferences and satisfaction. This paper compares the results of the two modes for conducting this customer satisfaction survey.

Weir, Paula, et.al, [A Comparison and Evaluation of Two Survey Data Collection Methodologies: CATI vs. Mail](#), *Proceedings of the American Statistical Association, Survey Research Methods Section*, August, 1999. In order to assess and evaluate the relative effectiveness of administering a traditionally mail based attribute frame survey using computer assisted telephone interview (CATI) as the primary data collection mode, EIA conducted a pilot study. The pilot study used a matched pair stratified random sample. Each sampling unit was matched on characteristics thought to be related to the ease of responding by mail versus telephone. Two separate surveys were then conducted and tracked by the primary collection mode, mail or CATI. The results of the pilot are presented and compared for two sets of measures, cost and data quality, where data quality is measured through response rates, response time, and response edit failures. The implications and recommendations for applying the lessons learned to a full survey data collection effort are also described.

Weir, Paula, [Conversion of Multiple Survey Systems' Edits and Imputation to StEPS](#) *Proceedings of the American Statistical Association, Survey Research Methods Section*, New York City, New York, August, 2002. The Energy Information Administration (EIA) operates over 100 different surveys that collect data on various types of energy at various points in the distribution flow from producers to users. These data address supply and demand issues by measuring production, imports, storage, sales, and consumption. Associated with these data are individual processing systems developed to accommodate the specific survey needs. These systems operate in multiple environments--mainframe, LAN and PC--and multiple languages and databases. While these surveys and systems have evolved over time, most are old and have been patched a number of times. Some of these systems have become very difficult to operate, causing problems with greater and greater frequency. It is clear that most need to be rebuilt. A more integrated approach to rebuilding the systems was considered for one fuel group of surveys. Resource limitations are the main factors driving the need for a generalized system. Even more preferable, if possible, is a generalized system that has already been developed, tested and is fully operational. As a result, EIA began considering the use of the Standard Economic Processing System (StEPS) developed by the US Census Bureau. This processing system includes modules for specifying parameters for the specific users and survey, modules for data collection

activities including mailing, receipt and check-in, as well as modules for post collection such as editing and imputation. At this time, EIA is in the process of loading survey specifications and data for one of its surveys into StEPS installed at EIA. The first of these surveys was chosen because of its relative simplicity in methodology and procedures. It is expected though, that much of the learning derived through the conversion process will be useful in loading the second, and following surveys. The edit and imputation requirements of the first survey are straightforward and similar to other surveys, even though reduced and simplified. This paper will focus on those two processes for this survey.

Weir, Paula, et.al., [The Graphical Editing Analysis Query System](#), *Proceedings of the American Statistical Association, Survey Research Methods Section*, August, 1996. In 1990 the Data Editing Subcommittee of the Federal Committee on Statistical Methodology released the Statistical Policy Working Paper No. 18, "Data Editing in Federal Statistical Agencies." The paper presented the subcommittee's finding that median editing cost as a percentage of total survey costs was 40% for economic surveys. The committee felt that the large proportional cost was the direct result of over identification of potential errors. Hit rates, the number of identified potential errors that later result in a data correction divided by the total number identified, were universally low. As a result, a lot of time and resources were spent that had no real impact on the survey results. The report cites research by the Australian Bureau of Statistics concerning the use of graphical techniques to find outliers at both the micro ,and macro level. A similar graphical approach to editing used by the U.S. Bureau of Labor Statistics (BLS) for the Current Employment Survey is also described. The BLS Automated Review of Industrial Employment Statistics (ARIES) system identifies true errors quicker and results in fewer man-hours to edit the data than the previous hard copy error listing report method.

Subsequent to the efforts of the Data Editing Subcommittee, a working group of analysts, research statisticians and programmers was formed within the Bureau of Census to examine the potential use of graphics for identifying potential problem data points in surveys. It was felt that the existing procedure of flagging cases failing programmed edits and reviewing each edit on a case-by case basis, had three main disadvantages: 1) analysts see neither the bigger industry, picture nor the impact of the individual data point on the aggregate estimate, 2) analysts, therefore, examined more cases than necessary, 3) edit parameters or tolerances were derived from previous surveys, implying constant relationships over time. The group felt that the tools of exploratory data analysis combined with subject matter specialists' expertise were well suited for identifying unusual cases. The working group developed a prototype that

made use of box plots, scatter plots and some fitting methods, as well as transformations. Two other systems, the Graphical Macro-Editing Application at Statistics Sweden, and the Distributed Edited Deposits Data System Editing Project (DEEP) of the Federal Reserve Board, have further demonstrated the efficiency of graphical editing. The recommendations of the subcommittee included focusing on the need for survey managers to evaluate the cost efficiency and timeliness of their editing practices and the implications of technological developments such as microcomputers, local area networks, and various communications links, in conjunction with traditional subject matter specialists' expertise.

Weir, Paula, et.al., [Separating the Wheat from the Chaff: The Search for the Best Imputation Methodology](#), *Proceedings of the American Statistical Association, Survey Research Methods Section*, Washington, D.C., August, 2009. Developing an imputation methodology is often wrought with basic data issues. Yet, the interpretation and treatment of the data have bearing on the methods to be considered and the performance of competing estimators. In this study, the data suffer from a recent change in both the data elements collected and the processing system, confusion over truly missing versus zero values, and reliability of edit failed data elements. In addition, the implemented imputation must be performed without access to concurrent reports, as the system requires immediate imputation before data from other respondents are available. Exogenous data from a related survey are considered, and a number of different estimators are compared through an exploratory approach to determining an imputation model that is compatible with both processing requirements and data characteristics.

Weir, Paula, [Superimposition of a Geographical Stratification on a Complex Design](#), *Federal Committee on Statistical Methodology Research Conference*, 1990. The purpose of this paper is to present some of the work done determining the effect of the geographic location of the survey respondent on sample rotation. Such effects could lead to discontinuity in a published data series such as, the EIA-782B, reseller/Retailers' Monthly Petroleum Product Sales Report". Furthermore, this paper will address incorporation of geography in the design of the seventh sample cycle of that data series. The work presented here is preliminary, and subsequent work has pursued this same issue to continue with the geographic modification.

Weir, Paula, et.al, [The Evolution of the Weekly Gasoline Price Survey through Changes in Design and Frame](#) The Evolution of the Weekly Gasoline Price Survey through Changes in Design and Frame, *Federal Committee on Statistical Methodology Research Conference*, Location, November 16, 2005.

Weir, Paula, et.al., [The Impact on Data Quality of the Transition to Clean Burning On-Highway Diesel](#), *Proceedings of the American Statistical Association, Survey Research Methods Section*, Washington, D.C., August, 2009. The U.S. Environmental Protection Agency has required that beginning in June 2006 refiners and importers of petroleum must ensure that at least 80 percent of the volume of on-highway diesel fuel they supply be ultralow sulfur diesel (ULSD). By December 2010, all on-highway diesel is required to be ULSD. Between 2006 and 2010, both ULSD and low sulfur diesel (LSD) may be offered for sale at retail locations outside of California, with some diesel fuel outlets carrying both fuels and others choosing to sell only one or the other. Until January 2007, EIA has collected the price of on-highway diesel fuel without distinguishing the sulfur level. This paper describes how the weekly diesel price survey was modified to account for the transition to ULSD. Evaluations of the variance using a bootstrap method and sensitivity analysis to explore the impact of alternate assumptions are presented.

Weir, Paula, et. al, [Two Multiple-Phase Surveys that Combine Overlapping Sample Cycles at Phase 1](#) *Proceedings of the American Statistical Association, Survey Research Methods Section*, August, 1998. The EIA-888 is a survey of diesel fuel outlet prices that produces estimates of national and regional level prices. The EIA-878 is a survey of motor gasoline outlet prices that produces estimates of national and regional level prices, as well as separate estimates for four formulations and three grades of gasoline. Both of these weekly surveys have used a monthly survey as phase I of a multi-phase sample, subsampling the sample units of the monthly survey who report the specific outlet sales category. Recently phase I of both of the weekly surveys has used a combination of two overlapping sample cycles of the monthly survey as phase 1, adjusting the Probability Proportional to Size (PPS) size measures to account for sample units present in both sample cycles.

Zhang, Bin, et. al., [Improvement of Data Quality Assurance in the EIA Weekly Gasoline Prices Survey](#), *Proceedings of the American Statistical Association, Survey Research Methods Section*, Washington, D.C., August, 2009. The EIA weekly survey of retail gasoline prices collects prices using a Computer Assisted Telephone Interview system with interactive data editing embedded to assure data quality. Edit performance statistics, however, showed that the data editing criterion sometimes missed true outliers at one end of the price change distribution and falsely over flagged outliers at the other end of the distribution, especially during times of large price change seen in the last three years. In order to improve the efficiency of the data editing criterion, a new data editing criterion based on price change

relative to market change was developed. In addition, a new post-collection data validation procedure for screening price and price change outliers by region and grade that makes use of all respondents collected.