

Lawrence Berkeley National Laboratory

Recent Work

Title

INFORMATION VALIDATION: A WORKING PAPER

Permalink

<https://escholarship.org/uc/item/9hq6f31c>

Author

Horovitz, M.

Publication Date

1979-06-01



Lawrence Berkeley Laboratory

UNIVERSITY OF CALIFORNIA

ENERGY & ENVIRONMENT DIVISION

INFORMATION VALIDATION: A WORKING PAPER

Mark Horovitz and Carl York

June 1979

RECEIVED
LAWRENCE
BERKELEY LABORATORY

DEC 4 1979

LIBRARY AND
DOCUMENTS SECTION

For Reference

Not to be taken from this room

LBL-9237C.1

no batt

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

Information Validation: A Working Paper

LBL Project Staff:

Mark Horovitz
Carl York

Consultants:

Steve Alter
David Freedman
Jonathan Golovin
David Hopelain
Martin Stern

Lawrence Berkeley Laboratory
University of California, Berkeley

June 25, 1979

Information Validation: A Working Paper

Contents

I. Overview	1
II Introduction	1
III. The Need for Validation of Energy Information Systems	4
IV. A Validation Study	6
V. The Role of Field Work in a Validation Study	9
VI. General Problems Encountered in the LBL Pilot Study	12
a. Availability of Information about Systems	12
b. Assessment of User Needs	12
c. Constraints on the Rate of Expansion of a Project	13
d. The Need for Internal Evaluation of a Project by the Sponsoring Institution	15
e. The Need for OEIV to Provide Adequate Liaison for the Validation Study Team	16
f. The Role of OEIV In-house Experts	16
VII. Recommendations for the Future	18
VIII. Compilation of Report	21

Appendices

Appendix A.	A Brief History of the LBL Pilot Program
Appendix B.	An Outline of a Methodology for Validation of Energy Information Systems
Appendices C - G.	Executive Summaries of the Five LBL Validation Studies:
	C. Domestic Crude Oil First Purchaser System, LBL-7878

- D. Major Fuel Burning Installation System,
LBL-8434
- E. Monthly Fuel Consumption Report: Manu-
facturing Plants, LBL-8435
- F. Domestic Crude Oil Entitlements System,
LBL-8436
- G. Middle Distillate Price Monitoring System,
LBL-8437

Overview

- I. A one year pioneering project on validation of energy information systems is reviewed. Five such information systems were studied and methodology for validation was developed. The need for validation is presented; the content of a validation study, a strategy, and a framework for validation of information systems are described. Some methodology developed for validation is outlined and the role of field work in such studies is discussed. A brief history of the project is provided and some of the problems encountered during the course of the project are reviewed. Amongst the topics covered are problems of gaining access to data, assessment of user needs, constraints on the rate of expansion of projects, liaison between the study team and funding agency, and the role of technical experts in monitoring such projects. Recommendations are made concerning most of these items. Summaries of the findings of the five validation studies are appended.

Information Validation: A Working Paper

II. Introduction

This paper describes the lessons that were learned from a one year project performed under the auspices of the Energy Information Administration's Office of Energy Information Validation. It was a pilot project

whose goal was to "validate" five energy information collection systems and then attempt to extract from that experience the formulation of a general strategy, or methodology, for performing such validation studies. Validation studies of the five energy information systems were performed and reported in a series of interim reports¹. This paper describes the first attempt to generalize a strategy for such validation studies.

A validation study is the process of describing and evaluating the performance of an information system. The study is carried out "from the top down" and assesses the quality of the information collected and reported by the whole system. This approach sharply contrasts with an "audit" which is "from the bottom up" and is designed to catch people who fraudulently fill out reports. In working with Department of Energy information systems it is important to distinguish between a "compliance audit" and an "information system validation study." A compliance audit seeks to find specific violations of required reporting procedures. Such audits are performed by the two regulatory branches of DOE, the Economic Regulatory Administration and the Federal Energy Regulatory Commission and focus rather narrowly on the possibility of fraudulent or otherwise erroneous reports of data. In contrast, the validation study seeks to answer a much broader set of questions about the overall operation of the system. In addition to examining the accuracy of the reported data and the subsequent steps in compiling and adjusting data to prepare reports for policy makers, the question of methodology used for measuring data, the limitations to the accuracy of the data, and the impact of those limitations on conclusions that can be drawn from it, are all studied in the validation process. It also examines the government users'

requirements to assess whether the data are sufficiently (or excessively) precise to meet their needs and whether they are using the information appropriately.

One important similarity between audits and validation studies is that they both should be performed by an independent observer who is not directly involved in the operation or design of the system. The primary reason for involving an independent observer is to reinforce the credibility of the study.

An information system can be thought of as being composed of three parts, a "logical structure", a "subject matter" and a "reporting system". The "logical structure" of the system is the set of statements that describe the reasons for its creation and the objectives to be achieved. In the DOE systems these statements of purpose are to be found in the following kinds of documents:

- a legislative mandate (if any) or an executive order;
- the administrative rules and regulations which interpret the mandate;
- the design specifications for the information system;
- the intended outputs of the system; and
- the intended users of the system and the decisions they are to make using the system.

The "subject matter" of EIA information systems consists of the natural resources, the land, the technological methods for extraction and processing, and the business enterprises that are engaged in these activities as well as marketing the final products to their customers. The logical structure attempts to extract information about the subject matter for some stated set of purposes.

Finally, the "reporting system" is the actual system comprised of a collection of questionnaire forms with data and instructions, the sample frame, the sample, the data processing, editing and verification operations, and the final reported calculations and conclusions that are issued.

By an intercomparison of the logical structure, the subject matter and the information system, it is possible to determine:

- the modifications in the system, subsequent to the original design;
- the impact of external changes on the effectiveness of the system; and
- the ways in which the currently operated system does or does not meet the original design criteria.

Thus, the validation of an information system results in an assessment of the usefulness, accuracy, efficiency and maintainability of the currently operated system. Such a validation study may also produce a set of recommendations for improving the system.

III. The Need for Validation of Energy Information Systems

When the Department of Energy (DOE) was organized in 1977, the Administrator of the new Energy Information Administration (EIA) was assigned the responsibility "for carrying out a central, comprehensive, and unified energy data and information program." As the new organization evolved, 55 energy information systems were transferred to it. Because there had been a general concern throughout the government about the quality and performance of these systems, a Professional Audit Review Team (PART) was established in 1976. The first report of this group in 1977² indicated that the DOE information systems

were in very poor condition. They reported that "the credibility of OEIA's models has not been established because documentation, verification, and validation have been neglected." The five validation studies of LBL's pilot project confirmed that this observation could be extended to the entire information collection systems that were studied.

National Energy Policy depends on credible energy information. The performance of validation studies is one way to build the needed credibility, by documenting the systems, and by eliminating errors and inaccuracies in the data as well as correcting design flaws in the information system. One of EIA's important tasks is to bring the multitude of disparate information systems into an effective organization and to subject those systems to appropriate management and maintenance practices.

The systems themselves need to be validated to insure that they are serving the purposes for which they were originally designed and documented so that users can understand the limitations of the data. Or, given that changes in purpose may have led to modifications in the system, a validation study can determine whether or not the modified system satisfies the new demands placed on it. It is essential to determine whether system changes have been fully documented, because there is a relatively rapid turnover in operating staff. This can lead to inadvertant mishandling of the data and operating system, and has contributed to the loss of credibility which characterizes the public view of these systems.

IV. A Validation Strategy

Although it cannot be claimed that a single strategy can be applied to the validation of every information system under the purview of DOE, the three-component description of information systems introduced above is useful as a guide to mounting an orderly attack on the problem.

The logical structure of the system should be examined first. The relevant documents which mandate the establishment of the system must be assembled and thoroughly studied. Next the administrative rules and regulations which implement the mandate must be understood. The design of the system then must be mastered to insure that the outputs for users are consistent with those identified in the original mandates. These steps are relatively straightforward and provide a minimal factual base for the study.

An understanding of the subject matter is much more difficult to acquire, and much less well defined in scope. Sufficient knowledge of the operating characteristics and terminology of the specific part of the energy world to be studied must be accumulated to enable the system validator to communicate with specialists in the field and to comprehend the assumptions made about the subject by the logical design. There is ample evidence from the pilot studies that some assumptions used to design the information system were, in fact, invalid. The result was that in some cases the system could not logically give the desired outputs. The degree of understanding of the subject matter, as it relates to the system being examined, is a limiting factor on the success of the validation study.

Finally, the information system itself should be examined to determine whether or not it functions in a useful, accurate and efficient way. A detailed examination of the system will review its design and operation as a statistical survey. The details of this process are found in Appendix B. However, the examination of the information system is not limited to an investigation of its statistical aspects alone. The way in which the system results are used by decision makers should be assessed and compared with the initial intent of the legislative mandate. Estimates of the accuracy of the calculated results produced by the system should be attempted. If improvements of only a few per cent result from doubling the cost of data acquisition, then such an action would not be included in the recommendations of the validation study.

In Appendix B, a three phase approach to validating an energy information system is described. There it is recommended that the first phase be carried out in the spirit of an "exploration" of the system to describe it and focus attention on potential problem areas. This might be called a "diagnostic" phase of the work. It is assumed that the system documentation would be adequate and that this brief overview would take a relatively short time. If no significant problems were identified, then the system would be declared operationally satisfactory and the work would end. However, if significant problems were discovered, a second phase of the validation study would be launched. In this phase, data analysis and small pilot studies would be undertaken to verify or refute initial hypotheses about the identified problems. If those investigations led to the identification of further significant problems, then those

too would be explored. All of the work in this second phase would be directed toward setting priorities according to the seriousness of the identified set of problems. Finally, when the most serious problems are understood, a third phase of effort may be justified. At this stage extensive field work by specially trained analysts may be required to establish in detail the findings of the second phase. Once established and understood, a strong case for redesign of the system may be made to eliminate the problems.

In performing the pilot study of five energy information systems, the LBL staff completed the first two phases of the investigations described above, but did not undertake the last because the project was curtailed just as the significant field work was ready to begin. In at least two cases, the LBL study showed that after the diagnostic phase was completed, no further validation studies were justified. These systems were so poorly constructed that the only recommendation to be made was that the present systems be abandoned and a new design effort begun.

V. The Role of Field Work in A Validation Study

Field work is an essential feature of a validation study. A team of interviewers was assembled and trained early in the project's development. From the interaction of the field workers with the validation study teams emerged a perspective about the role of field work in such studies.

The term field work as used here means almost any information gathering activity performed by the validation study team members. As the term implies, these field work activities are pursued out in the field, as opposed to those done at the team's home base. There are two quite different types of field work that have played a role in the validation of information systems. "Exploratory" field work is to be distinguished from "confirmatory" field work and will be discussed here in relation to the overall strategy used in system validation.

Exploratory field work is a reconnaissance in the field and uses the techniques of personal interviews, telephone calls and letters to elicit information about the systems under investigation. It should be performed by the members of the validation team and has as its objectives a first hand learning experience about the system and the identification and definition of problem areas in the system. In the first phase of the validation study, it serves to acquaint the team members with the people and procedures involved in the system operation. The team member maintains a perspective which includes the entire system as he observes the actual questionnaire editing, data processing and other operational functions. In his discussions with the operating staff of the information system, he notes the problem areas in relation to the system as a whole.

If the questionnaire contains possible ambiguities, a limited telephone survey of respondents might be performed to identify the problem more specifically. A consultant who is a specialist in the subject matter of the information systems can be used to brief the validation team members on specific things to look for and ask about. This consultant can also help them phrase their technical questions so that the respondents can better understand these questions. These are, in fact, examples of field work techniques and are used by the members of the validation team in both the first and second phases of the strategy described in detail in Appendix B.

"Confirmatory" field work is a highly structured study of the specific questions identified by the validation team in the course of the first two phases of the study. Confirmatory field work requires careful planning to define the questions to be asked and the methodology of the analysis to be used. It will usually involve expensive, formal surveys that should permit statistical inferences to be drawn (i.e., to permit extrapolations of the results to larger populations). The surveys should also permit estimates of the errors on the derived results as well as some indications of possible biases in the results. The planning of such confirmatory field work should carefully weigh the relative merits and costs of using telephone interviews, questionnaires, face-to-face interviews, document analysis, forms analysis, and so on. This kind of field work is useful only when the data processing is well advanced so that respondents and non-respondents can be readily identified and consistency tests performed. Usually the design of the confirmatory field work will include a pretest or trial of any instruments to be used.

It was concluded on the basis of the LBL pilot program that a self-contained, semi-autonomous group of field workers, who were trained to do in-depth case studies, was not the best approach to this type of work. Field workers who are brought in as integral members of the validation team and are thoroughly briefed on the specific information system being studied, could be effectively used to augment a confirmatory field work effort. Before staffing for such an expanded effort in the field, the validation team members might be given some training in interviewing techniques, so that they could perform the pilot studies or pretests needed to justify a third phase, confirmatory study.

The most important conclusion to be drawn about the use of field workers in a confirmatory study is that to be effective, they must be thoroughly trained in the subject matter of the information system to be validated.

VI. General Problems Encountered in the LBL Pilot Study

a. Availability of Information about Systems

Each of the five studies encountered delays in obtaining system documentation as well as access to data stored in the system. It is impossible to evaluate data without knowing how they were collected, edited and stored. In several cases the data were of a privileged nature and access to them was complex. Even when the validation team had clear evidence from its analyses that more data were needed, their requests went unanswered or were significantly delayed. In effect this put the team in the position of not being able to complete its job in a timely way, because the DOE did not provide the necessary data.

To avoid this type of problem in the future it is recommended that OEIV clearly recognize the need to provide sufficient staff time to expedite the requests of the validation teams. Before undertaking a validation study, OEIV might very well establish working relationships with the relevant offices and individuals in DOE to speed up this fundamental liaison process.

b. Assessment of User Needs

When the users of the system are interviewed to determine whether they are satisfied with its output, limits to the inquiries need to be drawn carefully. If the inquiry includes questions about what information the user believes he needs, then the validation study could become a "needs assessment" or "market survey" for future designs. During the pilot program the importance of this limiting boundary on a validation study became a significant problem. The users' perceptions of their needs must be compared to the original mandates for the system. It is

also necessary to identify and differentiate between primary users and secondary users. Here we mean that those responsible under the mandate to take actions on the basis of the system output would be considered as primary users. However, many other policy makers and planners, both inside and outside of government may use the information data base and the system output. These are the "secondary" users.

We recommend that the identification of primary and secondary users, as well as the kinds of questions to be asked them, be carefully worked out with the OEIV program staff before this part of the validation study is performed. This part of the study should not be done before the first phase of the work is very nearly completed. It clearly will require the techniques of exploratory field work at that stage of the study.

c. Constraints on the Rate of Expansion of a Project

When a new research activity is launched, there are delays and limitations on the rate at which the activity can grow. The brief history of the LBL project, outlined in Appendix A and Tables 1 and 2, included rapid and abrupt changes in planned objectives and funding. Before validation studies on the first five information systems were completed, new goals and a greatly expanded scope of work were proposed by OEIV. LBL was asked to validate four new systems in addition to the five studies in progress and to prepare to initiate other new projects during Fiscal Year 1979. The newly formed research team which had functioned for three months at an expenditure rate of about one million dollars per year was to be expanded as quickly as possible to carry out a planned six million dollar per year program. Several LBL administrators and staff were temporarily

assigned to the project and a full scale recruitment effort for new personnel was launched.

From the beginning, the project was required to build an interdisciplinary team which had to function effectively in a new field of endeavour. This required that considerable attention and energy be devoted to learning, training, communication and structuring of the team. The demands of tight deadlines and rapid expansion divided the efforts of the project staff. At least three levels of organization were involved in meeting these demands:

- Individual learning. Whether recruited from within LBL or elsewhere, the staff was initially unfamiliar with the field of validation of energy information systems. Hence each team member needed to devote time and energy to learning about this new field.
- Team building. Time was required for individuals to adjust to working effectively with other team members with different professional training, attitudes and vocabulary.
- Interaction with other groups. The nature, scope and timing of the project placed unusual demands on LBL management and support services. A mutual learning process was necessary for effective interaction between these groups and the project staff and also between OEIV and the project.

Because all large institutions have specified procedures for hiring senior technical staff, a substantial period of time is required to bring such a group up to the staffing level proposed for the project.

The time required to recruit staff is strongly influenced by the standards set for the quality of the individuals who are sought. An organization like LBL, which works in a university setting, strives to recruit staff of the highest quality and to follow suitable personnel selection procedures. The search for exceptional individuals and the subsequent negotiations tend to be time consuming. The normal sequence of steps required to hire technical staff at most institutions requires that:

1. job descriptions be written up and posted;
2. if there are insufficient local candidates for the job, then the position is advertized in appropriate newspapers and technical magazines;
3. a file of qualified applicants be set up;
4. the file be rank-ordered by a selection committee;
5. the most promising applicants be brought to the Laboratory for a personal interview by the selection committee members and the relevant project and Laboratory administrators;
6. the documentation of a selected candidate is usually reviewed by Laboratory-wide committees for salary level and, where relevant, term of contract status; and
7. then the personnel office can make an offer to the candidate.

There are then additional delays before the offer is accepted, and the new staff member can relocate his family. He must then begin learning

the details of his new job in order to become a productive member of the validation team. A realistic estimate of the time required to perform each of these steps indicates that from three and one half months to seven months can be required to add an individual to the group. Delays of this magnitude set a constraint on the rate of expansion that can be anticipated for any project.

d. The Need for Internal Evaluation of a Project by the Sponsoring Institution

A new activity such as LBL's Energy Information Validation Project requires periodic assessment of the quality of it's work so that management can exercise proper control. The normal procedures for such quality controls at LBL include internal reviews of publications before their release and the convening of committees of experts in relevant disciplines to review the project's activities. Although both types of reviews were used, the rapid pace and evolution of the project made it difficult to conduct sufficiently thorough reviews in the time available.

e. The Need for OEIV to Provide Adequate Liaison for the Validation Study Team

It was noted above that exploratory field work by the validation team required the aid and support of OEIV staff. Facilitation of access to the DOE data also required substantial OEIV staff work. A substantial amount of staff time and effort is required to support the validation teams in their interactions with the federal establishment. A careful assessment of this work load and its timing, would be valuable.

f. The Role of OEIV In-house Experts

The role of the OEIV program officers was perceived by the project staff to be too intrusive into the day-to-day operational details of the project in Berkeley. This was contrary to the normal operation of projects at LBL and led to the impression that the technical independence of the project was being compromised. The OEIV project manager must clearly delineate his dual roles as an expert in the field from that of a manager of the project. Maintaining the delicate balance between a professional colleague sharing his ideas with his peers and a manager ordering a change in work scope for a project is essential to the health of any project.

VII. Recommendations for the Future

a. Access to information

The difficulty of obtaining information and documentation for the five information systems studied in the pilot program was discussed above. The same problems had been identified earlier by the PART team. The importance of a clear historical record as well as up to date documentation of an information system cannot be over emphasized. To improve this situation it should be recognized that the original systems designers did not include in their design criteria the need for later system validation. Similarly the systems operators were not specifically instructed to maintain current files and documentation of the system.

- We recommend that the design of each new system require ease of validation as one of the criteria for satisfactory performance and that one member of the operating staff be assigned the job of maintaining a complete file of the background material used in the logical design of the system and that a continually updated record of the changes in the system documentation be maintained.

In making this recommendation we feel that the responsibility for understanding and maintaining a system should fall directly on the shoulders of the operating staff. This staff might be required to perform periodically an internal validation of the operating system as a means of insuring that a current picture of the strengths and weaknesses of the system is available.

b. The liaison role of EIA/OEIV program staff

- The management of OEIV should carefully assess the amount of staff time that would be required to perform the necessary liaison between the external validators and the relevant government officials and agencies.

At least two functions require OEIV staff time. First there is the need to contact and explain the purpose of validation to relevant persons in other parts of DOE or outside agencies before the arrival of the validation team. Second, if access to data stored on a government computer is required, then sufficient OEIV staff time should be allocated to enable the validation team to do its job in a timely fashion.

c. Organization of Field Work

Because the exploratory field work is done by validation team members, it only required liaison and coordination with OEIV staff as recommended in (b) above. However, when supplementary field workers are brought in to help with the confirmatory phase of a validation study, they should be integrated with the validation team.

- Field workers, who are brought in to supplement the validation team, should be an integral part of that team and should be trained by that team in the work they are to do.

d. Constraints on the Rate of Project Expansion

The need for OEIV to expand its operation in new fields will almost certainly arise again. The problems that this caused in the LBL project provide an insight to one limit to growth, the appointment procedure for new technical staff. The time required for job posting, interviews, selection, relocation, and training is significant and should be assessed when new projects are planned.

- Before undertaking to expand a project, OEIV should request an analysis of the time required by the contractor to appoint new technical staff.

e. Institutional Project Evaluation

Because the validation study is a new and developing field, other institutions may be confronted with the need to assess the value and quality of their work in this area.

- OEIV should require, as a provision of its contract, that an internal evaluation mechanism be created by the contractor to assess the quality of the work.

f. The Role of In-house Experts at OEIV

- The role of in-house experts should be carefully reviewed by OEIV to avoid the tendency of program managers to shift from contract monitoring to technical involvement in the work of a contractor.

VIII. Compilation of Report.

This report has been generated by having each participant write down his thoughts on a given subject. These essays were distributed to the other members of the group and then discussed in a one day seminar held once a month in February, March and April. After the monthly meetings a synopsis of the ideas was prepared by a rapporteur. In this way a preliminary draft of this report was prepared and discussed. From this discussion, a draft was prepared and circulated for internal comment to members of the LBL and UC Berkeley staffs. Their comments have been integrated into this final report. Steve Alter, David Freedman, Jonathan Golovin, David Hopelain, Mark Horovitz and Martin Stern contributed numerous essays on the assigned topics and Carl York served as rapporteur for the group. Appendix B was composed by Jonathan Golovin. Mary Hart patiently typed and retyped the numerous manuscripts and supported the project by caring for the many logistic details that it involved. Each participant owes her a special debt of gratitude.

References

1. The five interim reports are:

Stern, M. O., et al, "Domestic Crude Oil First Purchaser System", LBL-7878, December 1978

Alter, S., et al, "Major Fuel Burning Installation System", LBL-8434, December 1978

Alter, S., et al, "Monthly Fuel Consumption Report: Manufacturing Plants", LBL-8435, December 1978

Nagarvala, P., et al, "Domestic Crude Oil Entitlements System", LBL-8436, December 1978

Hopelain, D. G., et al, "Middle Distillate Price Monitoring System", LBL-8437, December 1978

The Executive Summaries of these five reports are attached in Appendix C, below.

2. Professional Audit Review Team (PART), "Report to the President and Congress: Activities of the Office of Energy Information and Analysis, FEA", December 1977.

Note: This report lists 142 energy information forms under the purview of this office of FEA. Later other systems were added to this list to give approximately 185 forms which are now under the current DOE/EIA.

A BRIEF HISTORY OF THE LBL ENERGY INFORMATION VALIDATION PROJECT

This brief history is intended to provide some background, context and perspective for consideration of lessons learned. The LBL Energy Information Validation Project can be divided into the following major phases:

Preliminary exploration and negotiations	10/77 - 1/78
Main Study	1/78 - 1/79
Project close-down	1/79 - 2/79
Review	2/79 - 6/79
Completion of outstanding items	7/79 10/79

The major events which occurred during the project are identified in Table 1 and changes in the level of funding are listed in Table 2.

During the fall of 1977, EIA and LBL explored the desirability of establishing an energy information validation project, the scope of potential studies and the resources required to undertake such a project. Agreement to establish the program was reached in January 1978 and LBL then proceeded with planning, staffing and preliminary studies. In March 1978, OEIV proposed to expand the studies and reorient them to emphasize development of validation methodology. Reports on validation studies of five energy information systems were scheduled for completion in the fall of 1978. By the beginning of June 1978, five study teams were formed and exploratory work produced detailed plans for the five system studies. Preparations for extensive field work by a team of interviewers began in June. The five studies then proceeded towards production of interim validation reports; delivery of these reports to OEIV occurred in December 1978.

Discussions concerning the LBL program for Fiscal Year 1979 began in July 1978. OEIV proposed to expand the LBL project to a funding level of

six million dollars per year. Plans for completion of the five initial studies were modified by scheduling production of interim reports for the fall of 1978, with final reports on the first five studies due for completion during Spring and Summer of 1979. In addition, at least three studies started by Kindle Corporation were to be completed at LBL during 1979 and a comprehensive study covering all oil and gas production data systems was also to be carried out. Funding for this series of tasks was planned at the level of \$3.6 Million, one million dollars of which was received by LBL in December of 1978. A series of new studies more oriented towards research and development of methodology were also envisaged, with funding for these studies to be approximately \$2.4 Million for Fiscal Year 1979. These new studies were in the preliminary planning stage at the end of 1978.

On January 2, 1979, LBL was informed by OEIV that the LBL Energy Information Validation Project was to be terminated as quickly as possible. The project was closed down during January and February 1979 and documents from the LBL studies were shipped to OEIV at the end of February.

In February 1979, EIA and LBL agreed to a review project to be carried out during the period February through June 30, 1979. This review was intended to produce reports on lessons learned during the previous year of LBL validation studies; document data analyses carried out on the Crude Oil Entitlements System; present methodology for validation and, prepare a bibliography and list of sources of information parallel to Financial Reporting System Required Information for the Oil and Gas Industry. Four reports on these topics are due to be delivered to OEIV by June 30, 1979. The project will then be terminated and the LBL staff assigned to other work. The report on Validation Methodology is to be completed by the Fall of 1979.

LBL negotiated a subcontract during the Fall of 1978 with the Institute of Policy and Management Research. Under the direction of Dr. Aaron Wildavsky, perceptual biases in users of information on oil and gas reserve estimates were to be studied. In February 1979, the scope of work and level of funding for this project were reduced, due to the reduction of funds available to LBL from OEIV. A final report on this work, describing the history and perceptions of oil and gas reserve estimates for the U.S., is due at the end of September 1979.

TABLE 1

<u>DATE</u>	<u>MAJOR EVENTS</u>	<u>OUTCOME</u>
10/77	LBL management visits DOE	Mutual interest in validation studies at LBL is identified.
10/77 - 12/77	Exploratory discussion between EIA and LBL	Agreement to establish an energy information validation project at LBL.
1/9/78 - 1/10/78	Meeting at EIA	Agreement to start project. Funding at \$380,000 for FY'78; \$50,000 to be released immediately. Four systems identified for initial validation studies.
1/78 - 3/78	Project planning phase	Drafting of project plans and proposal. Selection of core study team. Addition of Crude Oil Entitlements study.
3/78	Meeting with C. S. Smith, EIA Assistant Administrator for Information Validation	Agreement on basic study approach. Plans to expand and reorient studies towards increased emphasis on development of methodology.
3/78 - 5/78		Schedule 189 proposal accepted. System managers and other personnel selected.
5/78	Study plans	Four study plans prepared.
6/78 - 8/78		Chapters of draft report prepared.
7/78	Planning for FY'79	Initial discussion of plans to expand project to \$6 million per year level during FY'79.
8/78		OEIV decides to schedule completion of interim report for Fall 78, final reports on 5 studies due spring/summer 1979. Four Kindle studies to be completed at LBL.
9/79	EIA forms survey	Telephone survey covering over 200 EIA forms carried out in Washington, D.C. by project team.

<u>DATE</u>	<u>MAJOR EVENTS</u>	<u>OUTCOME</u>
9/78 - 10/78	IPMR contract	Negotiations and drafting of work statement for study of perceptual biases in information use to be carried out by A. Wildavsky through the Institute for Policy and Management Research.
9/78 - 12/78	Writing interim reports	
10/78	Draft reports prepared	First draft of interim validation study reports on MFBI and MDS systems are prepared.
11/78	Draft reports prepared	Drafts of COE and COFP reports are sent to OEIV.
12/78	Interim reports delivered	5 Interim reports delivered to OEIV.
10/78 - 12/78	Discussion of study scope and criteria	Major disagreements surfaced between OEIV and the LBL study team concerning the formulation, boundaries and criteria for validation studies.
9/78 - 1/79	Recruitment for FY'79 project	Extensive search, selection and hiring program instituted to provide staff for planned expanded FY'79 project.
1/2/79	Decision to terminate project	OEIV informed LBL of the decision to close-down the LBL EIV project immediately.
1/79 - 2/79	Project close-down	Project staff transferred to other LBL activities or employment at LBL terminated.
2/28/79	Transfer of project documents to OEIV	Project documents officially transferred to OEIV and shipped to Washington.
2/1/79	Start of EIV project review	An EIV review project was started. This consisted of four tasks to be carried out during the period February through June '79.

<u>DATE</u>	<u>MAJOR EVENTS</u>	<u>OUTCOME</u>
3/30/79	Draft COE data analysis report prepared	A draft report reviewing data analysis on the COE system was sent to OEIV.
6/30/79	Review reports due for completion	Reports on lessons learned, methodology and FRS are to be delivered to OEIV.
6/30/79	Project close-down	Project is scheduled to close down on this date.
9/30/79	Reports due	Draft reports are due on Validation Methodology and results of IPMR study.

TABLE 2

Date	<u>Levels of Funding</u>			
	Total amount per fiscal year actually received by LBL		Potential funding planned by OEIV in discussions with LBL	
	Thousands of Dollars (Cumulative Totals)		Total for Fiscal Year Thousands of Dollars	
	<u>FY'78</u>	<u>FY'79</u>	<u>FY'78</u>	<u>FY'79</u>
January 1978	50		380	
March 1978			760	
April 1978			970	
May 1978	380			
June 1978	790			
August 1978	1,114			6,000
October 1978				3,600 (First increment)
December 1978		1,000		
January 1979				Project close-down
January 1979				1,425
March 1979		1,028		
April 1979		1,453		1,453

AN OUTLINE OF METHODOLOGY FOR VALIDATION OF
ENERGY INFORMATION SYSTEMSIntroduction

The following brief outline of a methodology for validation studies is presented. It is anticipated that a complete paper will be prepared on this subject with a few months.

A validation study using this proposed methodology is conceived as having three phases:

Phase I consists of exploratory work to describe the system and focus attention upon the potential problem areas (i.e., a diagnostic phase).

Phase II consists of data analyses and small pilot studies to test the initial hypotheses and confirm or refute them as well as discover any additional problems. All efforts in this phase are aimed at quantifying or resolving just how bad a suspected problem really is.

Phase III consists of all major field work/auditing needed to establish the Phase II findings in more detail. These are efforts requiring more manpower than the initial team staffing used in Phase I and II.

The detailed notes provided below indicate those topics that should be covered in Phase I alone. Time did not permit preparation of the details of Phase II and Phase III. This material will be included in the subsequent paper.

PHASE I

Section I - Review of the Regulations

A. Objectives:

To document the enabling legislation and associated regulations.

To document the mandated purpose or uses of the system.

To document the (mandated) data collection form or data to be collected.

To explore the following aspects of mandated and actual uses of the system:

data required for current decision making process (D.M.P.);

sensitivity of the D.M.P. to changes in the data;

alternative data sources used or available.

B. Questions to be Answered:

Description of the system and uses

What legislation/regulations established the system?

What are the mandated uses or purposes of the system?

Who are the primary and secondary users of the system?

What data does the system collect? Who must submit data?

What reports does the system publish?

For each use listed above explore the following questions concerning the D.M.P.

What specific uses of the data are made?

What is the actual D.M.P.?

What other information is used in the D.M.P.?

What are the key numbers, exception criteria, critical values used in the D.M.P.?

How sensitive is the D.M.P. to the input?

Is there a "well accepted" approach to making this (these) decision(s)?

How does the actual approach differ from the well accepted approach --and why?

What is the users' critique of the system (in general, in supporting their needs)?

What potential future uses exist for the data (anticipated uses/users)?

What uses can be supported by the data the system supplies?

What questions does it answer?

C. Methodology:

Interviews with users.

Review of the regulations/legislation.

Review of the literature pertinent to the D.M.P.

Interviews with experts on the D.M.P. The issues involved, data required.

D. Examples of Outputs:

Description of the system as mandated by the regulations and/or legislation.

Description of the current users/uses of the system outputs.

Table of the key uses/users and key values.

Section II - History of the System

A. Objectives:

To understand the rationale for the system as it was

designed; to understand any predecessor systems and why they were replaced/superseded.

B. Questions to be Answered:

Why was the system created?

Why was it necessary?

Who created it? What was its origin (originators, funders, intended users)?

What information systems did it complement and/or replace (how was this information available previously)?

What policies/political decisions/political environment existed at this inception?

Who were the key individuals during its history?

- designers
- maintainers
- "owners"
- users

What were the key design choices in going from the legislative mandate to the final design? Who made them?

What flexibility existed in the design choices (in going from the legislative mandate to the final design)?

What has been the system history or evolution?

- changes in the system design (data sources, data collected, universe definition, use of a sample, processing, users)
- changes in the use of the system

Why were these changes made - who affected the system's evolution?

How did these changes affect the system's usefulness, accuracy, purpose?

What changes occurred in the system's environment that affected its use?

e.g., oil embargo

OPEC price increases

Clean Air Act--environmentalists gain power

C. Methodology:

Interviews with

system maintainers

system designers

experts in the relevant areas

Review of the Hearings Literature Search

Section III - Description of the System

A. Objectives:

To document the system design and current operation
(to describe how the system works and what it does),
to detect weak points in the system (possible or actual),
to find parallel or cognate systems.

B. Questions to be Answered:

Universe/Frame/Sample

What is the universe definition?

How was the frame constructed?

How is the frame updated?

How stable is the universe?

How was the sample drawn?

Is the sample rotated?

The Form

How was the form (questionnaire) designed (e.g., use of experts, pre-test, etc.)?

How well is the respondent trained in completing the form?

What judgements are necessary by the respondent (i.e., are any terms not standard or well accepted or clearly defined)?

Are any definitions or questions ambiguous?

What differences exist in interpretation of the questions, between the intent of the collection agency and actual interpretation by a sample of respondents?

What difficulty would a respondent have in filling out the form (small, medium and large company; is that information usually collected; is it defined differently in standard industry practice)?

The respondent and the form

How do a few respondents collect the data required by the form? What measurement methods are used? How accurate are they?

How could these measurements be verified?

What accounting rules apply to the internal manipu-

lation (by the respondents) of the collected data,
e.g., guidelines given for FRS accounting?

Does the respondent know how the data will be used?

Is there any feedback to the respondent?

- immediate (in the editing function)
 - short range (dollar transfers, hearings, company regulation)
 - long range impact (e.g., industry regulation)
- that he perceives.

The data collected

What do the data collected really measure?

Can the data required be measured or is only a surrogate measure possible?

Editing of the data

What is the non-response rate?

What is done about non-respondents?

- telephone call
- periodic sample to check for non-response bias
- ignored

What is done about missing data elements?

What are the usual editing procedures?

- visual screening
- program error control routines and their logic,
e.g., outliers, consistency with previous month,
consistency with other firms.

What feedback is given to respondents and how might

it affect their reporting?

What is done with late data?

What is the usual keypunch error rate?

What type of keypunching verification is done?

Are the respondents ever audited or re-interviewed?

The programs and computer processing

Are the computer programs well documented?

How were they tested/validated?

Are they maintained and the documentation updated?

Is the data base secure?

Is there an audit trail for changes made to the data base?

What is the system and data base backup?

Estimates and models in the system

How is the estimate made?

What model is used in the calculations?

How does the model perform, e.g., sensitivity to input data?

How is the model used?

How is the model updated?

What are the procedures for updating its parameters or structure?

Is the model ever overruled or its output manually changed?

System support and maintenance

How does the user interact with the system?

- to request changes
- to use the output
- to request documentation

Is the user documentation adequate?

Who is responsible for each element of control of/in the system, e.g., data quality, system maintenance, user training, respondent training, updating of the frame, etc?

Allied systems

What allied systems exist?

- parallel systems, collecting the same data (or data element), or using another form.
- preceding or succeeding systems, collecting the data at an earlier or later point in its material flow.
- related systems, that use a similar universe (frame) or collect similar information.

What data does each collect? How does this differ from the system being studied? What relation should the data bear?

Auditing and investigations

How is the decision made to audit a respondent?

Findings of audits to date: what effect have the findings had on the system? What changes do they suggest should be made?

Is there any routine, rotating verification auditing done?

Incentives for Respondents

What incentives do respondents have to bias the data?

How do respondents think this system affects them?

How could the data be biased and what would the advantage be?

What would the biased data look like (patterns of bias vs. random mistakes)?

How could bias be detected?

- analyses
- tips
- intuition

What are the current procedures in place (analyses, checks, audits) for detecting biases?

Who is affected by misreporting and to what extent?

C. Methodology:

Review of the system documentation

Interviews with key system personnel

Interviews with experts

Expert review of the form

Respondent interviews

Review of the edited forms, observation of the editing process.

D. Output from this Section:

Description of the system - design and operation.

Flow chart of the physical flows that the system measures.

Flow chart of the information system.

Table of the related or allied information systems with their differences and the expected relation between the data in each system.

List of the known problems with the system.

List of the potential, suspected weak points in the system.

Table of the data elements

Collected	Source or Measurement Point	Possible Error or Bias Technique	Verification Possibilities
-----------	-----------------------------------	---	-------------------------------

E. Additional Output from this Phase

Areas for further investigation based upon;

- problems in usage in Section I
- suspected or actual weaknesses in Section III
- taxonomic or specific risk factors relating to this system

Study design for each area

Associated budget and schedule for Phase II

0 0 0 5 4 0 2 2 7 2

Appendix C

DOMESTIC CRUDE OIL FIRST PURCHASER SYSTEM VALIDATION STUDY

EXECUTIVE SUMMARY

PREFACE

Both Congress and the Executive Branch have been greatly dissatisfied with the degree of meaningfulness, accuracy, and consistency found in energy information. Particular concerns include the quality of oil and gas reserve and resource estimates, financial data from integrated energy producers, and projections made using econometric models. The lack of credible energy information is due in large part to the lack of "independently developed or independently verified data."

Energy information includes all statistical data collected and collated by the U.S. Department of Energy (DOE), as well as statistics acquired by DOE from secondary sources (for example, the American Petroleum Institute) and projections and other results of models used to analyze these statistics.

The LBL Data Validation team studied the accuracy and meaningfulness of the data generated by the Domestic Crude Oil First Purchaser System (COFP). The COFP System was developed to meet the following three general objectives:¹

- 1) To collect and present the information necessary to monitor and assess the weighted average price of the first sale of domestically produced crude petroleum.
- 2) To collect and present information sufficient to assist in identification of potential violations of domestic crude oil pricing regulations and to assist follow-up activities.
- 3) To provide information sufficient to assess crude oil pricing policy measures on both prospective and historical bases.

¹FEA Draft Requirements Analysis for the Domestic Crude Oil First Purchaser Reporting System, February 9, 1976.

This data validation review seeks to determine the actual meaning and accuracy of broad classes of data. It focuses on the methodology used in measuring and reporting data and its effect on the meaning and accuracy of the resulting statistics. It examines the government user's requirements and processing procedures to determine whether the data are necessary, sufficiently (or excessively) precise, and used appropriately. Finally, it attempts to determine if any other data system is similar enough to permit the requirements of both systems to be met by a combined effort.

Preliminary results of the LBL Validation of the COFP system are contained in this report.

EXECUTIVE SUMMARY

I. SYSTEM IDENTIFICATION

A. CURRENT IDENTIFICATION

System being validated: Domestic Crude Oil First Purchaser
(COFP) System

Statutory Authority: Energy Policy and Conservation Act
(1975)

Energy Conservation and Production
Act (1976)

Authority expires May 1979 unless
extended.

FORM: FEA-P-124-M-1

B. PURPOSE OF SYSTEM

To implement the intent of the Energy Policy and Conservation Act of 1975, which imposed ceilings on the price of oil. COFP was authorized in order to collect information necessary to permit reliable calculation of a monthly weighted average crude oil composite price based on actual first purchase transactions. This price is then compared with statutory price guidelines.

C. HISTORICAL INFORMATION

Antecedent systems include:

- Cost of Living Council form CLC-90 dating from August, 1973.
- Federal Energy Administration form FEA-90.

- Federal Energy Administration form FEA-P-124-M-0.
- Form FEA-P-124-M-1 was approved by GAO on July 8, 1977. Approval expires July 31, 1980.
- Background information on the COFP System can be found in Section I.C. of the Interim Report.

D. SYSTEM DESIGN

1. The Universe: All domestic first purchasers of crude oil (estimates vary from about 300 to about 600).
2. The Sample Frame: Assuming compliance with reporting regulations, the sample frame would consist of all companies purchasing crude oil. There is evidence that the sample frame is incomplete.
3. The Sample: The sample consists of all companies in the sample frame that first purchased more than 150,000 bbls of crude oil per year.

For a discussion of the Universe and Sample frame, see Section I.D. of the Interim Report.

4. Information requested: Table 1 shows the information reported on each of the five schedules in FEA-P124-M-1, those required to file each schedule and the filing frequency. Any firm purchasing domestic crude oil must report.

Table 1

NATURE OF INFORMATION AND REPORTING REQUIREMENTS

	Annual	Monthly
Schedule A Company Identification and Certification	Class I, ^a II ^b	Class I
Schedule B Summary of monthly first purchase transactions: tier, ^c volume (bbls), total price paid	-	Class I
Schedule C First purchase by producer, state, tier, volume (bbl), price paid	-	Class I
Schedule D Other purchases, additions, sales, exchanges, and reduc- tions, identifies parties to other transactions	Class I, II	-
Schedule E Producer operator identification	Class I, II	-

- a. Class I firms are those first purchasing 150,000 bbls or more annually.
- b. Class II firms are those first purchasing less than 150,000 bbls annually.
- c. Tiers dealt with separately are upper, lower, stripper, Naval Petroleum Reserve (NPR) and Alaskan North Slope. Speaking roughly, old oil is crude produced by a property up to its 1972 production level; additional crude produced is classified as new oil. Stripper oil is oil from wells producing less than 10 bbls/day, and NPR oil is crude from Naval Petroleum Reserves.

The forms are supplied by the Energy Information Administration (EIA) to respondents upon request or to new companies required to respond for the first time. Monthly respondents are estimated to account for about 99.7% of the reported transaction volume. Companies reporting

monthly must mail Schedules A, B, and C on the last day of the month following the month for which transactions are due (e.g., on March 31 for February transactions). Yearly schedules are due in EIA within 90 days after the close of the year. See Section I.D.3 for further detail on FEA-P124-M-1.

E. SYSTEM IMPLEMENTATION

EIA collects the data from 336 known first purchasers and performs an "eyeball" check. The forms are then forwarded to the Division of Crude Oil Supply Allocation (DCOSA) in ERA which supervises more extensive verification, followed by preparation of punched cards and tapes. Within a month of receipt of the data by DOE, computer reports based on the updated master file, and generated at Optimum Systems, Inc. (OSI) in Rockville, are available from EIA.

Verification, key punching and data processing are done with the help of outside contractors retained by EIA. A contractor was also retained to create and maintain the processing programs. (For a discussion of the reporting process, see Section I.E.1 of the Interim Report.)

1. Status of the Schedules

Schedules A and B were being routinely processed by DCOSA at the beginning of our study (May 1978). Examination and verification of Schedule C had just begun. Only January 1978 data were available in semi-final form. By now (November 1978), Schedule C is also being processed routinely as it is received. The master files for both B and C data are kept on-line at Optimum Systems, Inc. (OSI) for two months for

possible revisions. After that, the information is frozen and archived. Schedule D is now in the early stages of verification. It will take about six months to process the data reported on that very complex form. Schedule E, which asks purchasers to list names and addresses of producers/operators they have dealt with during a calendar year, has not been examined so far by either ERA/DCOSA or by us. The annual Schedule D data for 1977 are currently available, but have not been keypunched.

F. USES OF OUTPUT

1. Composite Price Model

The Composite Price Model uses COFP data to calculate the current value of the Domestic Composite Price of crude oil and to forecast prices based on various policy assumptions. For an Analysis of the Model, see Section I.F.1. of the Interim Report.

2. The Energy Information Administration (EIA) publishes a series of reports based on information contained in Schedules B and C. A list of reports that can be generated, their description and a distribution list can be found in Section I.F.2. of the Interim Report.

3. COFP System Users

- Primary users: Office of Fuels Regulation/Energy Regulatory Administration (OFR/ERA) personnel
OFR is charged with setting domestic crude oil price ceilings on the basis of the observed domestic composite price and the statutory guideline.
- Secondary users: Office of Enforcement/Energy Regulatory Administration (OE/ERA) personnel
Enforcement may use the data to seek at least statistical evidence of ceiling price violations.

- Tertiary users: Those DOE personnel or others who use selected COFP data in periodicals, technical, economic, or policy-oriented analyses. COFP data are often used as a surrogate for production data.
- Publications using COFP data:
 - Monthly Energy Review (EIA/OEDI)
 - Quarterly Report to Congress (EIA/OEDI)
 - Monthly Petroleum Statistics (EIA/OEDI)
 - Weekly Report of Oil and Gas Exploration and Development Statistics (DOE/RA/OONGS)
 - National Energy Plan (DOE/PEA)

G. RELATED SYSTEMS

COFP is the only information system gathering crude oil first purchase data. For the purpose of external validation, the following systems collect comparable data:

- Crude Oil Entitlements System--COE (refinery receipts)
- Crude Oil Production System--COPS(production data) based on production data corrected by individual states.
- American Petroleum Institute (production data)

See Section II.D. of the Interim Report for comparisons.

II. SYSTEM VALIDATION

Based on an initial evaluation of perceived problems in the COFP System, the Data Validation Team focused on completeness of the sample frame and accuracy in reporting, and on verification of input information and processing. Procedures used in the Validation Study included:

1. Analysis of Survey Instrument (Section II.A.1. of the Interim Report).

2. Analysis of data collection, verification and processing (Section II.A.2 of the Interim Report).
3. Analysis of the forecasting model (Section II.A.3 of the Interim Report).
4. Internal Validation methods focused on:
 - characterization of sample frame
 - doublecounting
 - reporting consistence
 - tier price outliers^a
5. External Validation methods focused on comparison between COFP and
 - State production data
 - Crude Oil Entitlement Data (Section II.A.5 of the Interim Report).
6. Analysis of the sensitivity of system outputs to Data quality (Section II.A.6 of the Interim Report).
7. Field Validation (partly completed) of users and respondents (Section I.A.7 of the Interim Report).

III. CONCLUSIONS

A. REPORTING

1. Survey Instrument (Form P124-M-1)

The P124 Schedules and accompanying instructions were examined for possible ambiguities that could lead to reporting errors. A number of such instances were found, most of them in Schedule D whose main function is to provide checks on internal consistency and a way of keeping the sample frame up-to-date. None of these instances, in our opinion, could have resulted in a clear, intentional or unintentional bias in response. (See Section II.A.1 of the Interim Report.)

^aTransaction prices above prescribed tier ceiling prices.

B. INTERNAL VALIDATION

1. Reporting Consistency

Schedule D was reviewed for agreement in quantities bought, sold, or exchanged as reported by partners in total annual transactions during 1977. Eight of the 10 largest oil companies were reviewed. Substantial disagreements (more than 5% were found about half the time between reports of partners). Discrepancies of as much as 8 million (out of 12 million) barrels are evident. Reporting errors and inconsistencies in response attributable to ambiguities and/or lack of accounting realism in the schedule are expected to be mainly responsible for the disagreements. See Section II.A.4 of the Interim Report.

2. Tier Composition

Ambiguities and changes in tier definitions have led to variations in the type of oil reported as belonging to particular tiers. Where the definitions are clear and stable, the validation team has not uncovered any errors in reporting. However, with the exception of NPR production records, there are no independent sources of data against which to check the data in P124. Thus, tier certification becomes primarily an audit problem. See Section II.A.4 of the Interim Report for further discussion.

3. Price Compliance

A study was made of price outliers by tier, to assemble statistical information permitting a judgment on COFP's value as a price compliance tool. It was found that average new oil prices in West Virginia were consistently and substantially (more than 10¢/bbl) above national ceilings. In addition, transaction prices for individual companies purchasing oil were compared to national and state price ceilings. A number of outliers appeared for national ceilings and many more transaction prices exceeded state ceilings. These results are still being analyzed and will be followed up by field interviews. (See Section I.A.4 of the Interim Report.)

4. Incompleteness of Sample Frame

The sample frame is not equal to the universe because companies first purchasing less than 150,000 bbl/yr. need not report monthly, and indeed do not report first purchases annually either, although they do report other transactions. This is causing an estimated 24,000 bbl shortfall in the daily totals. In addition, however, ERA has found that the 1977 schedules D include some 350 names of companies that buy or exchange with reporting companies but that do not, themselves, file a report. The effect of this on sample frame incompleteness cannot be estimated at present.

C.External Validation

The degree of incompleteness of the sample frame was deduced by comparing COFP first purchase data to state production data ("COPS") and Crude Oil Entitlement (COE) refinery receipts. Where substantial disagreement was evident, an attempt was made to investigate possible causes (e.g. double counting, misreporting, etc.) in order to account for the differences. It was found initially that:

- COFP state totals in California and Louisiana were ~7% smaller than corresponding state production totals.
- COFP totals for Texas were 2% smaller than state totals.
- COFP totals for Florida, Wyoming and Alaska were greater than corresponding state totals.
- national totals showed COFP at least 2.4% below production totals.
- COFP totals were 1% above COE totals.

It was subsequently found that

- Discrepancies in Alaska and Florida were due to chronic double counting in COFP of large transactions involving the same purchases month after month (discovered by LBL and ERA/OFR independently). These errors have been corrected.
- The state data base in Wyoming was incomplete (discovered by EIA/OED). Estimated corrections bring totals (COFP and COPS) into close agreement.
- In California ~21,000 bbls/day of NPR oil currently go unreported in COFP in addition to an estimated 50,000 bbls/day used for tertiary recovery at the site - which is largely unreported. These two effects are believed to explain the discrepancy.
- A COFP deficit of ~132,000 bbls/day in Louisiana can be explained, in part, by unreported USGS oil sold to small refiners (~45,000 bbls). The remaining deficit has not been explained.
- In Texas the discrepancy of ~70,000 bbls/day may be due only in part to differences in COFP and State data regarding inclusion of plant condensate.

- COE totals were in excess of COFP by about 50,000 bbls/day after corrections for double counting in Florida and elsewhere, and some other corrections, had been made. What is this disagreement likely to be due to?
 - some of the discrepancy may be due to plant condensate correctly excluded from COE reports and mistakenly included in COFP.
 - 40,000 bbls/day are almost certainly exported to Canadian refineries after a domestic first purchase largely unreported to COE.
 - the Joint Petroleum Reporting System (JPRS) of the EIA/OED estimates that 14,000 bbls/day are diverted to fuel blending. Our impression is that the correct figure is much higher. Field work may improve this estimate.
 - since 160-odd small crude oil purchases were omitted from the COFP sample, 24,000 bbls/day go unreported in the system.
 - the JPRS reports that 8,000 bbls/day which were exported from the West Coast are probably unreported to COE for the most part.

Based on the results of the investigation thus far, the COFP sample frame is estimated to be about 1.7 percent short.

D.Sensitivity of the System to Data Quality

A sensitivity analysis of the COFP system model and the physical facts underlying it (See Section II.A.6 of the Interim Report) revealed the following:

- Random errors of measurement due to faulty gauges or readouts are extremely unlikely to have any impact on the requirements of COFP users whether primary, secondary, or tertiary.
- Systematic errors in total amounts transferred are also unlikely to be of importance, since such errors can be policed through the market place.
- Systematic underrecording of old oil combined with compensating overrecording of new oil from a single mixed-tier lease cannot be ruled out completely.
- Reported first purchase price errors if unaccompanied by tier reporting errors (intentional or not) are unlikely to have a significant effect on the primary use of COFP. Furthermore, the Entitlements program, covering more than 95 percent of crude going to refineries, has a corrective effect on aberrant market prices.
- Incompleteness of the frame, unbiased with regard to tier composition, has virtually no effect on the national composite price.
- Biased incompleteness of the frame (e.g. more unreported lower tier than upper tier oil) or tier misreporting could impinge on the primary use of COFP if sufficiently large. Our preliminary conclusion is that it is not.
- Any incompleteness in the frame would be of concern to secondary and tertiary users. Our preliminary conclusion is that some concern is warranted on this score. (See Section II.A.5 of the Interim Report.)

IV. RECOMMENDATIONS

A. LIMITATIONS AND CAVEATS

Our preliminary conclusion is that COFP currently reports crude oil at 2.4% below the amount produced. Included in this shortage are

- 0.7% unreported lease oil by definition, not included in COFP.
- 0.3% known sample incompleteness
- 1.4% due to a combination of incomplete reporting and countervailing doublecounting.

The system is meaningful for primary users since current flaws will not affect the accuracy of the national composite price.

Some improvements could be made in reporting format and content for secondary users. (See Recommendations for System Improvements.)

For tertiary users, the COFP system is not a completely reliable surrogate for production data and would not be--even if all unknown errors and omissions were corrected. COFP, by definition, excludes lease oil used at the production site, and will always be short of production by that amount. The shortfall due to lease oil use appears to be increasing over time.

B. RECOMMENDATIONS FOR SYSTEM IMPROVEMENTS

- Corrections to previous estimates should be kept separate from current data. The dubious distinction between mistaken estimates and errors should be dropped.
- One should investigate the possibility of issuing a schedule of prices to be used in reporting transfers within integrated companies. The schedule should be updated frequently to reflect changes in prices shared in arms-length transactions.
- The state code should be made to depend clearly on location of the well or the site of first purchase, not both.
- The request for information about the tier of oil exchanged should be omitted from Schedule D because valid information on this point is usually not available.
- Schedule D should be changed to include first purchases of oil companies reporting only annually. This will tend to make the sample frame more complete.
- Reporting requirements should be brought into line with standard accounting procedures. (See Section II.A of Interim Report.)

C. RECOMMENDATIONS FOR FURTHER WORK

- The sample frame should be thoroughly checked for completeness. This can be accomplished by: 1) questioning a sample of producers to discover the firms to which they sell; 2) checking Schedule D reports to determine if any listed purchasers are not in COFP; and 3) finding a list of operators not named in Schedule C and questioning a sample of these about firms to which they sell.
- Further analysis of price outliers by tier should be conducted, followed by field interviews.
- Field interviews should be conducted, especially in California, to determine the extent of, and the degree to which, lease oil is included in the system.
- Field interviews should be conducted to determine quantities of domestic crude oil put to non-refining use.
- Field work should be done to evaluate the cost burden of the system to the private sector and to government in its present form, and to assist in estimating the consequences if suggestions for improvements in the system were implemented.

PROJECT CHRONOLOGY

<u>DATE</u>	<u>TASK</u>
May 1 - June 30	Project Definition
June 1 - June 15	Personnel Acquisition
June 30	VALIDATION TEAM ASSEMBLED
June 15 - July 10	COFP Sample Data Acquisition
July 15 - September 30	External Validation
July 15 - August 1	Legislative History Research
July 15 - August 30	Internal Validation
August 15 - August 30	Sensitivity Analysis
August 15 - October 1	Field Questionnaire Preparation
November 30	DRAFT INTERIM REPORT
December 15	INTERIM REPORT
December 15 - January 1	Doublecounting Investigation
January 1 - March 1	Sample Completeness Checks
January 1 - May 1	Field Interviews
February 1 - March 1	Price Compliance Analysis
February 1 - April 1	Further State Data Analysis
May 1 - June 30	Final Report Preparation
June 30	FINAL REPORT DUE

2. The Forecasting Model

This model is not central to the function of the P124 System. However, it was reviewed and found to have some idiosyncrasies in the way that it is presently imputed/used to generate the Implied Domestic Composite Price (IDCP):

- a) The discontinuity between the tier oil fractions predicted by NEP for the future and those obtained from P124 up to the present causes inaccurate IDCP forecasts.
- b) The valuation of Alaskan North Slope (ANS) oil at the wellhead price and its inclusion in the IDCP calculation depresses the value of the IDCD.
- c) Current updating methods change the positions of crude oil price ceiling for different regions and different qualities in a way that may cause inequitable distortions in the prices at which oil is traded.

0 0 0 0 5 4 0 2 2 3 1

Appendix D

INTERIM VALIDATION REPORT

MAJOR FUEL BURNING INSTALLATION SYSTEM

EXECUTIVE SUMMARY

PREFACE

The purpose of this study is to examine the meaningfulness and accuracy of all information in the data base initiated by the 1975 Major Fuel Burning Installation (MFBI) survey. The data base remains the principal source of information about MFBIs in the U.S.

The information collected by the original survey [performed to implement the Energy Supply and Environmental Coordination Act (ESECA) of 1974] was intended to identify all MFBIs in the U.S. and to indicate which were potential candidates for orders prohibiting the burning of natural gas and oil. This interim validation report describes findings and recommendations to date regarding the meaningfulness and accuracy of the information collected by the original survey.

EXECUTIVE SUMMARY

I. SYSTEM IDENTIFICATION

A. CURRENT IDENTIFICATION INFORMATION

System being validated: Major Fuel Burning Installations (MFBI)

Form: FEA-C-602-S-0

- EIA Standard Series Number: currently unavailable
- Clearance: General Accounting Office, mid-April 1975
- Expiration: June 30, 1977

Statutory Authority: The reporting requirements were established to implement Section 2(a) of the Energy Supply and Environmental Coordination Act of 1974 (ESECA) Pub. L. No. 93-319, 88 Stat. 24.

B. HISTORICAL IDENTIFICATION INFORMATION

No antecedent information systems have been identified. A new MFBI information system is being developed to implement the Power Plant and Industrial Fuel Use Act of 1978, Pub. L. No. 95-620, __Stat.__. Related, supplemental or successor systems are the EIA-97, Boiler Manufacturers' Report; the form FEA-C-607-S-0, MFBI Early Planning Process Report; and the FEA-C-602-S-1, supplement to form FEA-C-602-S-0.

C. IDENTIFICATION OF PURPOSE

The purpose of the MFBI system was both to identify all major fuel burning installations^a and to provide information for specifying which

^aRegulations developed under ESECA defined a major fuel burning installation as "an installation or unit other than a power plant that has or is a fossil fuel fired boiler, burner or other combustor of fuel, or any combination thereof, at a single site, that has individually or in combination a design firing rate of 100 million Btus per hour....Gas turbines and combined cycle units were excluded from this classification."

MFBIs were potential candidates for orders prohibiting the burning of natural gas or oil. The information required was derived from the regulations in ESECA. ESECA stated that the following criteria should be applied in issuing prohibition orders to particular MFBIs:

- The combustor should be capable of burning coal.
- The burning of coal should be practical.
- The burning of coal should be consistent with the purposes of ESECA.
- Coal and coal transportation equipment should be available.
- Other factors could be considered, such as the location of the installation, the amount of oil or natural gas presently burned, etc.

D. IDENTIFICATION OF INFORMATION COLLECTED

The form FEA-C-602-S-0 contained information in three categories -- general plant information, combustor information, pollution control device information. The numerical information on the form involved 1973 and 1974 fuel use. All information on the form was related to combustors in particular installations. Other than the fuel use information, most of the information on the form consisted of nominal or yes-no responses. Several questions called for narrative descriptions, e.g., identification of impediments to using more coal at the installation.

A second form, FEA-C-602-S-1 was sent to those installations that contained coal-capable combustors and reported using more than 150,000 barrels of oil or 921,000,000 cubic feet of natural gas (or the equiva-

lent combination) in 1974. Information collected on this form has not yet been analyzed.

E. IDENTIFICATION OF COLLECTION AND COLLATION PROCESS

The target universe of the MFBI survey was all Major Fuel Burning Installations that existed in June, 1975. The frame is an ad hoc construction consisting of names drawn from several trade association lists, and individuals who responded to Federal Register announcements. Response to the survey was mandatory. Responses were submitted by mail.

The MFBI survey was performed by the Office of Coal Utilization (OCU), which also checked both the incoming forms and edit listings of the information. Control Data Corporation keypunched and verified the responses. CACI developed a computer system for storing the information on tape, producing edit listings, and generating a series of standard reports. The MFBI tape is currently maintained by the Office of Data Services (ODS).

F. USES OF OUTPUT

The decision to perform the MFBI survey was made by the Office of Coal Utilization. According to the formal request for clearance of the form, the purpose of the form was to "identify the coal conversion candidate universe and provide necessary information for the selection of specific installations for further in-depth analysis before the issuance of mandatory conversion orders." During 1976-1977, the information from the form was used by the Office of Coal Utilization in the process of issuing orders prohibiting the burning of oil or natural

gas. It was used primarily as a screening tool for identifying which combustors should be examined in detail. The MFBI data base is currently being used only infrequently in regulatory activities.

The computerized MFBI data base has been and is being used for analytic purposes by many secondary users (see Appendix F). The MFBI data base (edited to protect confidentiality) has been made available for public distribution, and has been cited in reports written by secondary users.

G. RELATED SYSTEMS

The Environmental Protection Agency (EPA)'s on-going National Emission Data System (NEDS) contains information that overlaps with the information collected in 1975 by the MFBI system. Common data elements include:

- parent company name
- SIC codes (first four digits)
- combustor capacity (for boilers)
- primary energy source
- stack height

The NEDS target universe is identified in terms of point sources of pollution rather than MFBI's exceeding a designed firing rate cutoff.

II. CONCLUSIONS AND RECOMMENDATIONS

A. CONCLUSIONS

The information in the MFBI data base is a basically unsound representation of the universe of MFBI's and the characteristics of that universe. The meaningfulness and accuracy of this information are severely impaired by numerous design and implementation flaws. Nonetheless, the MFBI data base is probably the best currently available source of information about MFBI's.

The meaningfulness is seriously restricted by the unquantified non-coverage bias. No systematic procedure or rigorous definition exists to identify the elements of the frame. The quality of coverage of the universe by the frame remains unquantified. (Further LBL research will estimate the coverage and assess the impact of the non-coverage on the information.) Available evidence strongly suggests the possibility of non-response bias as well. No determination of the extent or estimate of the effect of the non-response exists. (Further LBL research will generate such an estimate.)

In addition, the recorded information about the engineering characteristics of the MFBI's is not specific enough to demonstrate coal-conversion capability. This lack of focus of the recorded information is compounded by the large quantity of extraneous information collected. (Section II.A.2, main text.) The lack of bounds for the non-coverage and non-response biases and the inadequacy of the information collected render the MFBI data base unreliable as a comprehensive

source of information about MFBI's in the U.S.

The condition of the data base itself is also highly suspect. At least 699 of 4,199 records are missing, without explanation. About 29 percent of the remaining records contain identified logical inconsistencies or errors. The information in the data base is out of date; most records are 3 to 6 years old. Nothing short of a total reconstruction is likely to cure the inaccuracies and inadequacies of the information in the MFBI data base.

B. RECOMMENDATIONS

1. Caveats

- The information is 3 to 6 years old.
- Coverage of the frame is uncertain (LBL studies will explore the frame coverage).
- The data base contains many internal inconsistencies and errors.

2. System Design

Recommendations about what specific information should or should not be collected to support the regulatory process are not being made at this time.

C. WORK TO BE DONE

Additional work by Lawrence Berkeley will focus on the development of methodologies for creating the frame for a new MFBI survey. This work will make it possible to develop a statistical estimate

of the coverage of the current MFBI data base. A secondary research area will involve the identification of mechanical features of different types of combustors that can be used to determine whether conversion to the use of coal is practical. Research on the information collected by the FEA-C-602-S-1 can also be undertaken if this work promises to provide useful results for OEIV.

FUTURE WORK TO BE DONE TO COMPLETE THE SYSTEM STUDY

Major Fuel Burning Installations

1. Investigation of the information required to appraise both the short-term and the long-term fuel switching capability of an installation, in terms of specific economic and technical factors.
2. ~~Field work by experts on the accuracy of the submitted forms.~~
3. Discussion of consistency with other systems (including those under development).
4. Exploration of coverage of the frame.
5. Examination of relationships between the MFBI system and the Power Plant and Industrial Fuel Use Act of 1978.

0 0 0 0 5 4 0 2 2 8 6

Appendix E

MONTHLY FUEL CONSUMPTION REPORT

MANUFACTURING PLANTS

EXECUTIVE SUMMARY

PREFACE

The Energy Information Administration maintains a group of coal information systems which provide information on production, consumption, and stocks of coal. One of these systems is the Monthly Fuel Consumption Report: Manufacturing Plants (MFCR), which gathers information on coal consumption and stocks of both general industrial users, accounting for about 9 percent of total coal usage, and steel and rolling mills, accounting for less than 1 percent.

The purpose of this interim validation report is to describe progress to date in assessing the accuracy and usefulness of the information that is collected and distributed by the MFCR system. Work to date has focused only on the general industrial coal consumers. Due to changes in the MFCR system in late 1977, 1976 is the most recent year for which complete information is available. Much of the data analysis to date uses 1976 information. On-going work will extend this analysis to include the years 1973-1975. Work to date has not attempted to evaluate the potential impact of planned changes in the MFCR system.

EXECUTIVE SUMMARY

I. SYSTEM IDENTIFICATION

A. CURRENT IDENTIFICATION INFORMATION

System being validated: Major Fuel Burning Installations (MFBI)

Form: FEA-C-602-S-0

- EIA Standard Series Number: currently unavailable
- Clearance: General Accounting Office, mid-April 1975
- Expiration: June 30, 1977

Statutory Authority: The reporting requirements were established to implement Section 2(a) of the Energy Supply and Environmental Coordination Act of 1974 (ESECA) Pub. L. No. 93-319, 88 Stat. 24.

B. HISTORICAL IDENTIFICATION INFORMATION

No antecedent information systems have been identified. A new MFBI information system is being developed to implement the Power Plant and Industrial Fuel Use Act of 1978, Pub. L. No. 95-620, __Stat.__. Related, supplemental or successor systems are the EIA-97, Boiler Manufacturers' Report; the form FEA-C-607-S-0, MFBI Early Planning Process Report; and the FEA-C-602-S-1, supplement to form FEA-C-602-S-0.

C. IDENTIFICATION OF PURPOSE

The purpose of the MFBI system was both to identify all major fuel burning installations^a and to provide information for specifying which

^aRegulations developed under ESECA defined a major fuel burning installation as "an installation or unit other than a power plant that has or is a fossil fuel fired boiler, burner or other combustor of fuel, or any combination thereof, at a single site, that has individually or in combination a design firing rate of 100 million Btus per hour....Gas turbines and combined cycle units were excluded from this classification."

D. IDENTIFICATION OF INFORMATION COLLECTED

- Reporting requirements: As of December 1977, respondents are required to complete all parts of form EIA-3.
- Numerical information requested (on a monthly basis):
 - (1) Quantity of fuel consumed in the following fuel categories: bituminous coal, sub-bituminous coal, lignite, anthracite, oil, gas, furnace coke, foundry coke, and other.
 - (2) Quantity received during the month and average price per unit in each of the fuel categories.
 - (3) Beginning and ending stocks in each of the fuel categories except gas.
 - (4) Average Btu content per ton in each of the fuel categories except oil and gas.
 - (5) Average sulfur content in each of the fuel categories except gas.
 - (6) Estimated shipment time and distance for each major coal source.
 - (7) Quantity of waste ash removed after consuming coal.
- Level of resolution: Monthly consumption, receipts, and beginning and ending stocks at individual manufacturing plants or other commercial establishments.
- Descriptive information requested:
 - (1) Plant location.
 - (2) Plant standard industrial classification
 - (3) Bureau of Mines district of coal origin.
 - (4) Justification of any claim that the information on the form is confidential and will cause substantial competitive injury if released.

E. IDENTIFICATION OF COLLECTION AND COLLATION PROCESS

1. Universe Identification and Sample Design

- Universe: The universe is not clearly defined. Various aspects of system documentation imply each of four definitions that can be referred to as the "form title" definition, "form instructions" definition, "published" definition, and "in-house" definition.

The title of the reporting form is "Monthly Fuel Consumption Report: Manufacturing Plants." The instructions for form EIA-3 imply that the universe is "all firms and establishments that consume bituminous coal and lignite," and state that establishments such as airline terminals and administrative offices should be included. Footnotes published between 1958 and 1977 refer to a baseline estimate based on consumption and stocks of users directly involved in manufacturing or mining. The in-house definition of the universe appears to be all establishments other than electric power utilities and coking plants that buy coal directly from retailers.

- Frame: The frame primarily consists of a list of establishments that voluntarily submitted the BOM 6-1400-M-1 before reporting became mandatory in December 1977 (retroactive to January 1977). The 1976 data base contained responses from 747 firms, 106 of which did not burn coal. Additional establishments added during 1977 and 1978 have increased the size of the frame to over 1,000.
- Census of frame elements: Although no census of the elements of the frame has been attempted in conjunction with the MFCR system itself, the Census of Manufactures reports information that is directly related to the MFCR frame.
- Sampling: The MFCR system collects responses from the entire frame.
- Means of information collection: Form EIA-3 is completed on a monthly basis and sent to the Coal Statistics Branch of the Division of Coal and Electric Power Statistics in the Office of Energy Data and Interpretation in EIA. Response was voluntary before December 1977, when it became mandatory retroactive to January 1977.
- Length of reporting period: One month
- Frequency of data collection: Monthly

2. System Implementation

- Agency collecting the information:

Coal Statistics Branch
Division of Coal and Electric Power Statistics
Office of Energy Data and Interpretation
Energy Information Administration

- Processing the information: The completed forms are manually edited by the Coal Statistics Branch and then sent to Optimum Systems Inc. (OSI) for keypunching. Several iterations of manual and computerized checks are performed by Computer Sciences Corporation (CSC) in the process of updating the system's master file, which is stored on tape.
- Information volume: During 1976 (the most recent year for which complete information was available), a total of 722 respondents reported. The average monthly response rate was 86 percent. The frame has been increased to over 1,000 during 1977-1978.
- Processing time: Forms are due fifteen days after the close of the reporting period. The editing process requires approximately six weeks. Preliminary estimates of national consumption and stocks by fuel category are published ten weeks after the end of the reporting period. These preliminary estimates are revised one month later. A final estimate for the reporting month is published one year later, i.e., approximately 15 months after the reporting period.
- Custodians of computer files: Optimum Systems Incorporated (OSI).

F. USES OF OUTPUT

1. Direct Uses of the Information

- Regulatory uses: No regulatory uses of the information collected by form EIA-3 have been found.
- Use in specific analyses and models: Further research is required in order to identify specific analyses and models that use information collected by form EIA-3.

2. Publication of Information

- Agency that publishes the information: EIA
- Reports published:
 - (1) EIA Energy Data Reports, Weekly Coal Report
(previously called Bureau of Mines Mineral Industry Surveys, Weekly Coal Report)

- (2) EIA Monthly Energy Review
- (3) DOE Annual Report to Congress
- (4) Minerals Yearbook
- (5) Mineral Facts and Problems

- Primary Format of Reports: Tables
- Recipients of reports: Examination not yet complete. The Energy Data Reports, Weekly Coal Report is sent to anyone who requests it. Presently, about 2,000 individuals or organizations receive the report. These include government agencies, investment firms, libraries, research groups, and firms associated with the coal industry.

G. RELATED SYSTEMS

Examination of the relationship between form EIA-3 and related information systems is in progress. The related systems can be classified as follows:

- (1) EIA coal information systems - EIA forms 1 through 7, 20, and 210.
- (2) Other coal information systems - the Annual Survey of Manufactures and Census of Manufactures of the Bureau of Census.
- (3) Other fuel information systems - the Major Fuel Burning Installations Survey (FEA-C-602-S-0), the National Emissions Data System of the Environmental Protection Agency, and the Annual Survey of Manufactures and Census of Manufactures of the Bureau of Census.

II. CONCLUSIONS AND RECOMMENDATIONS

A. CONCLUSIONS

1. Usefulness

- Preliminary user interviews indicate that the published outputs of the MFCR system are not generally considered useful by individuals concerned with monitoring or

analyzing long-term trends in coal demand and availability. Many users and potential users of coal information state that the MFCR outputs are not timely, are too highly aggregated and are perceived to be inaccurate. (Additional work will be done to confirm initial findings.)

- If the MFCR system were accurate, it might provide potentially useful information that would, however, constitute only part of the information required for monitoring or analyzing long-term trends in coal demand and availability.

2. Accuracy

- The accuracy of time series produced to date by the MFCR system is highly suspect for the following reasons:
 - (1) Ill-defined universe: The universe is not clearly defined. The definition of the universe implied by the instructions for the EIA-3 conforms neither to the baseline cited in published statistical estimates nor to the "in-house" definition used in collecting the information.
 - (2) Inappropriate and outdated baseline: The baseline cited in published statistical estimates from 1958 through 1977 does not correspond precisely to the quantity apparently being estimated. During these years, the Weekly Coal Report references a 1954 baseline including only manufacturing and mining firms. However, the information gathered by the EIA-3 is currently used to estimate coal consumption of a group of users including establishments that are not manufacturers. Although several adjustments to the baseline have apparently been made, at least one (for inclusion of institutional users of fuel) has not been documented. A new baseline is now being developed by the Bureau of Census to improve the accuracy of future estimates.
 - (3) Technical bias: The linked relative procedure used to estimate national consumption and stocks is most appropriate when the population of respondents does not change. Although estimates based on this procedure would reflect changes in coal consumption within an established population of respondents, these estimates would not reflect the impact of entry to and exit from the respondent population. Consequently, significant trends first away from coal use (in the 1960s and early 1970s) and later toward coal use (in the mid and late 1970s) should be underrepresented in estimates generated by the MFCR system.

- (4) Disagreement with Bureau of Census estimates: The statistical estimates generated by the MFCR system in the years 1974, 1975, and 1976 are 24%, 37%, and 31% higher than comparable estimates published in the Annual Survey of Manufactures (1974, 1975, 1976).
- (5) Disagreement between the data base and the information published: Monthly percentage changes in consumption for 1976 were computed by applying the linked relative procedure to the 1976 information in the data base. These percentage changes differed from percentage changes published in the Weekly Coal Report. The cumulative effect of these differences was quite significant. The ratio of December 1976 consumption to January 1976 consumption in the published report is 1.237; the comparable ratio computed from the data base is .998. Thus, the published December 1976 consumption is 24% higher than the 1976 information in the data base would suggest. (A similar analysis for previous years is in process.)
- (6) Possible non-coverage bias: The following combination of LBL findings and past DOE estimates raises suspicions about, but does not conclusively prove the existence of, possible non-coverage bias, related to the effects of seasonality. The information in the data base for 1976 covers approximately one half of total estimated bituminous coal consumption in 1976 (26 million of a total 55 million tons). Histograms of 1976 information indicate that consumption is highly skewed. Approximately 25% of the coal users who responded consumed 75% of the total amount consumed. April 1978, DOE estimates of the size of the 1978 MFCR universe (3700 to 4000) indicate that the 1976 data base (containing 616 coal users among 722 respondents) may have missed on the order of 80% of coal users. Assuming that these users used the 29 million tons not accounted for in the data base, it is likely that the majority of these users are small. Since preliminary comparisons of seasonal consumption patterns suggest that coal usage by small users has stronger seasonality than that of large users, it is possible that the published information systematically underrepresents the seasonality of coal usage.
- (7) Possible selection bias: Until December 1977, reporting was voluntary. It is not known whether this affected the estimates that were generated by the system.

- (8) High variance of the estimate: A mathematical formula has been derived for a lower bound on the variance of an estimate generated using the link relative method for M periods. Substitution of 1976 population statistics into this formula indicates that the application of the link relative method for twenty years would result in monthly estimates for the last year whose standard deviation is at least 32 percent of the estimated value.

B. RECOMMENDATIONS

1. Caveats

- The information generated by the MFCR system is highly suspect for reasons cited above. Whenever possible, annual consumption information generated by the Bureau of Census should be used in its stead.

2. System Design

- The purpose of the MFCR system is to provide information rather than to support a specific regulatory process. Consequently, recommendations concerning system design cannot be made until a thorough study of the need for coal information has been performed.

3. Continuing Research

- Completion of data quality investigations
- Design and execution of a study of needs for coal information. Such a study is required in order to make recommendations about the redesign of the MFCR system.

Appendix F

EXECUTIVE SUMMARY

DRAFT INTERIM VALIDATION REPORT:
DOMESTIC CRUDE OIL ENTITLEMENTS SYSTEM

PREFACE

In 1974, acting through the Federal Energy Administration, the President promulgated regulations creating the Domestic Crude Oil Entitlements (DCOE) Program to allocate equitably the benefit of access to low-cost old oil among domestic refiners. The mechanism for achieving equitable allocation of low-cost old oil is the entitlement. Each entitlement stands for the right to process one barrel of deemed old oil a month. Basically, the system requires refineries with above-average old oil inputs to buy entitlements from refineries that are old oil-deficient.

This is a draft interim report on the validation study of data for the DCOE Program. It consists of research conducted at the Lawrence Berkeley Laboratory through November 1, 1978. The purpose of the study was to assess the validity of the data collected by and used in the DCOE System and the validity of the systems and procedures used in processing this information in order to meet the stated goals of the program.

In order to conduct a validation study such as this one, access is required to the large amount of data collected by the system over time. Initial use of this data is by nature "exploratory"--a time when initial or preconceived notions and hypotheses are tested, often discarded, and new lines of inquiry initiated. Unfortunately, the study team was unable to obtain the free access to data necessary to conduct this exploratory research in an efficient and timely manner. The results reported in this draft interim report should therefore be considered preliminary.

EXECUTIVE SUMMARY

DOMESTIC CRUDE OIL ENTITLEMENTS PROGRAM

VALIDATION STUDY

I. SYSTEM IDENTIFICATION

A. HISTORICAL IDENTIFICATION

The Domestic Crude Oil Entitlements (DCOE) Program (10 C.F.R. Section 211.67 (1978)) was established in December 1974, pursuant to the Emergency Petroleum Allocation Act of 1973 (EPAA). Pub. L. No. 93-159, 87 Stat. 627 (November 27, 1973), as amended, codified at 15 U.S.C.A. Section 751 et seq. (1978).

B. IDENTIFICATION OF PURPOSE

Section 4(a) of the EPAA required the President "to promulgate a regulation for the mandatory allocation of crude oil, residual fuel oil, and each refined petroleum product in amounts specified in (or determined in a manner prescribed by) and at prices specified in (or determined in a manner prescribed by) such regulation...(for all such products) produced in or imported into the United States." 15 U.S.C.A. Section 753(a) (1978). (See Appendix A.) Acting through the Federal Energy Administration, the President promulgated regulations creating the DCOE Program to allocate equitably the benefit of access to low cost "old oil among all sectors of the petroleum industry, including independent and small refiners, and thereby to assure that domestically refined petroleum products are sold at equitable prices by all distributors of petroleum products." 39 F.R. 31650 (August 30, 1974).

C. THE FORMS AND THE INFORMATION COLLECTED

The FEA promulgated requirements for the collection of information necessary to run the DCOE Program codified at 10 C.F.R. Section 211.66.

Information for the program is collected in a series of up to five mandatory monthly reports by refiners and eligible firms^a (for a more detailed discussion of these reports, see Section I.B.2).

1. Entitlement Transactions Report

All program participants are required to submit the Entitlement Transactions Report on Form FEA-P103-M-0 (GAO B-181254 (R0110), approval expires January, 1979) pursuant to 10 C.F.R. Section 211.66(i). This form collects information concerning a respondent's sales and purchases of entitlements and its net entitlements position at the end of the previous month. This form has been in use since the beginning of the DCOE Program. It will be replaced by Form ERA-116. Non-refiner importers are not required to submit reports for months in which they did not receive entitlements.

2. Refiners Monthly Report

All refiners are required to submit the Refiners Monthly Report on Form ERA-49 (OMB 38R0215, approved July, 1978), pursuant to 10 C.F.R. Section 211.66(h). Form ERA-49 collects information concerning respondent's crude oil receipts, costs and runs to stills, and the volume of residual fuel oil produced and sold by the respondent in, or into, the East Coast market. Form FEA-P102-M-0 was the original form used to

^aSee glossary for a definition of eligible firms.

collect monthly data from refiners. Forms FEA-P102-M-1, FEA-P102-M-2, and ERA-49 were successive replacements for this form.

3. Report on Imports of Residual Fuel Oil

All importers of residual fuel oil into the East Coast market are required to submit data on the volume of such imports on Form FEA-P113-M-0, Report of Oil Imports into the United States and Puerto Rico (GAO B-181-254CRO386), approval expires August 30, 1979, pursuant to 10 C.F.R. Section 211.66(j). Data on residual fuel oil imports are reported at Schedule T, column m of the form. Previously, this data was collected on a separate form (FEA-P126-M-0). This form will be replaced by ERA-60.

4. Naphtha Imports Monthly Report

All importers of foreign naphtha into Puerto Rico for use as a petrochemical feedstock are required to submit the Naphtha Imports Monthly Report on Form FEA-P129-M-0 pursuant to 10 C.F.R. Sections 211.66(h)(4) and 211.67(d)(5)(iv). This form collects information on the volume of foreign naphtha imported for such use, the total volume of imported naphtha and the volume of export sales of naphtha. OMB approval is not required because less than ten companies are required to submit this form. This form has been used since the DCOE Program was modified to provide additional entitlements for naphtha, effective July 1976.

5. Report of Imports to the Strategic Petroleum Reserve

All refiners and eligible firms delivering crude oil to the Strategic Petroleum Reserve (SPR) are required to submit a letter to the

TABLE ES - 1

DOMESTIC CRUDE OIL ENTITLEMENTS PROGRAM

DATA COLLECTION

DOE Form No.	Respondents	Major Data Items	Frequency of Collection and Level of Resolution
P103-M-0 (ERA 116)	All Participants	Entitlements buy/sell transactions	monthly/company
ERA 49	Refiners	Crude oil receipts, crude oil costs, crude runs to stills, residual fuel oil, California oil	monthly/company
P113-M-0 Schedule T, Column m (ERA 60)	Importers of Residual fuel oil into the East Coast Market	Imports of Residual fuel oil	monthly/company
P-129-M-0	Importers of foreign Naphtha into Puerto Rico	Naphtha imports and costs	monthly/company
Letter	Importers of Strategic Petroleum Reserve Crude	SPR imports and costs	monthly/company

DCOE Program office, setting forth the total cost, weighted average cost, and volume of the imported crude oil accepted for delivery by the United States Government.

D. IDENTIFICATION OF THE COLLECTION AND COLLATION PROCESS

1. System Design

The DCOE target universe is all domestic refiners and importers of eligible products.^a Domestic refineries include all refineries in any of the fifty states plus Puerto Rico, Virgin Islands, and Guam.^a Eligible firms are importers of residual fuel oil into the East Coast market.^a Data are also collected from naphtha importers to Puerto Rico, and from importers of crude oil to the Strategic Petroleum Reserve.

All refiners are included in the DCOE frame. Since it is normally advantageous for importers and new, small refiners to participate, the program is as concerned with the exclusion of non-eligible participants as it is with the inclusion of eligible firms.

Data are collected each month from all elements in the frame, i.e., the frame is not sampled. Response is mandatory. A current census of all participants is maintained in the program's computerized Administrative Log.

Data normally submitted by U.S. Mail. Occasionally when the data are late in arriving, or when clerical errors are discovered in the data submitted, data will be accepted by telephone. In the latter case,

^aEntitlements Program Handbook, Economic Regulatory Administration, U.S. Department of Energy.

a certified hard copy of the data or changes is still required by mail.

2. System Implementation

The data are collected and processed by the Economic Regulatory Agency (ERA). ERA is aided in the processing of data by an EIA computer operator and three private subcontractors. These contractors are responsible for key punch and machine verification operations, rented computer capability, and computer program development and maintenance.

Approximately 4,000 pages of information are submitted to ERA each month. Every set of information is file-stored as hard copy in two forms, as an original unmarked copy, and as a "working copy" with ERA changes and annotations. The month's "Masterfile" (the complete computerized version of this information as used in the month's calculations) is archived with the computer subcontractor. A second, shortened version is archived with other DOE data on a data base management system. This second system, known as the Data Collation Project, is almost operational.

Data for a particular month are reported to ERA by the fifth day of the second month following the reporting month. The Entitlement Notice is published ten days later, i.e., by the fifteenth. The transactions required by the notice must be completed by the end of the month. ERA must be notified of these transactions by the tenth day of the next month, i.e., the third month following the period for which the data were reported.

E. USES OF OUTPUT

The DCOE Program regulations (10 C.F.R. Sections 211.66, 211.67 (1978)) require the collection of necessary data, and specify the procedures to be used in developing the entitlement positions of participants. Pursuant to 10 C.F.R. Section 211.67(i), the ERA publishes the output, known as the Entitlement Notice, each month in the Federal Register. The Office of Energy Data, EIA, DOE, publishes the Entitlement Price, the National Old Oil Domestic Supply Ratio, and the Crude Oil Entitlement Benefit in The Monthly Energy Review.

F. RELATED SYSTEMS

Four other DOE data systems gather data that appear related to the information collected for the DCOE Program. Several states also collect similar information. These related systems and the common data elements are shown in Table 11 of the accompanying report. This table also lays out the status of the acquisition of this information, and includes comments on the differences between the various systems.

No model or forecast mechanism has been identified that can forecast the entitlement positions of the participants, or can substitute for the present system.

II. CONCLUSIONS AND RECOMMENDATIONS

This is a draft interim report on work done through November 1, 1978, on the validation of data for the DCOE System. During the course of the research, access to required data was difficult to obtain, therefore the conclusions reached are limited to those which could be drawn from available data, and are preliminary.

The Domestic Crude Oil Entitlements information system is basically sound.

Several potential problem areas are discussed below, and more may surface with additional research. At this stage, there is no reason to believe that the basic integrity of the system will be invalidated.

A. CONCLUSIONS

The major problem areas identified to date include:

- Lack of internal structural consistency:

For the period of June 1977 through May 1978, crude runs to stills exceeded refinery receipts by about 105 million barrels. This internal inconsistency is being examined further. (See Section II.E.)

- Lack of consistency with other related data series:

The examination of the consistency of DCOE data with other information has not proceeded far enough to be conclusive. (See Section II.E.)

- Errors in manual calculations:

A number of errors are introduced during the manual calculations performed by ERA analysts. The most significant of these are corrected almost immediately. Smaller errors are not always discovered or corrected. (See Section II.D.)

A detailed sensitivity analysis (see Section II.E) indicates that the system is reasonably robust, i.e., inaccurate reporting by a single firm does not significantly affect other participants, though it does affect the entitlement position of the respondent.

The DCOE System was developed in response to specific legislation passed by Congress. The system is meaningful relevant to these goals, which could not otherwise be attained. Despite some nominal overlap, the Joint Petroleum Report System (JPRS) cannot substitute for the DCOE System, since it does not collect its information by oil tiers, the basic modus operandi of the DCOE System. Additionally, the DCOE System collects information necessary to verify certain pieces of data (e.g., processing agreements) that are not collected elsewhere. The information is collected and processed within sixty days of the end of the reporting period, as required by the regulations.

B. RECOMMENDATIONS

Based upon the investigations to date, current flaws in the existing system should not seriously impact the purpose for which the system was developed, i.e., an equalization of the crude input costs for all refiners and other selected participants.

There is an apparent tendency for the system also to encourage other petroleum production activities, such as synthetic fuels, SPR storage, and most recently, the increased lifting of California crude oils. Should this tendency become a dominant factor, the design of the information system might need to be reexamined.

The monthly summaries of the DCOE System contain a great deal of information not available elsewhere. This is available only on an informal ad hoc basis. A systematic method of making available the non-proprietary parts of this data to interested users, both inside and outside DOE, should be investigated. Since some of the time series data are 'contaminated' by the rolling forward of corrections (rather than the a posteriori corrections normally made), the DCOE data should be so flagged for this extended audience.

The DCOE Program is subject to constant review, with a number of new provisions introduced or 'revisions' made each year. In spite of the programming loads associated with this changing system, the DCOE Program has developed fast and efficient procedures for completing their monthly assignment. One major area, the manual calculation of amendments in order to correct original data, needs to be addressed as soon as possible. These manual operations consume a significant portion of the analysts' time, and invite the possibility of large errors. ERA has had to correct a number of large errors resulting from this process, and a number of smaller errors have gone unnoticed. A computerized system has been developed to carry out these computations and should be implemented as soon as possible.

Due to the interim nature of this report, certain lines of investigation were begun but not completed. Research in the following areas needs to be continued:

- Cross-checks of the data with other data series. A number of 'parallel' data series have been identified in the report.

These comparisons will be done initially on aggregated national data and, if necessary, for a small sample of firms. (See Section II.E.)

- Complete the internal structural consistency checks identified in the report. There are reasons to believe that there may have been either a structural change or a change in reporting behavior over time. It may be necessary to investigate a small sample of companies to understand the reasons for this. (See Section II.E.)
- Complete the checks on the manual calculations done for the amendments process. Those calculations are carried out in three basic steps. Two of these steps have been checked to date. (See Section II.D.)
- Develop a scheme for computerized monitoring of the input data to aid the present 'eyeballing' used by the analysts. This capability would greatly enhance the present manual, case-by-case approach. (See Section II.B.)

PROJECT CHRONOLOGY

May, 1978

Planning the validation effort,
finding personnel for the study group

From June through November 1,
1978, the following tasks were
undertaken:

1. Documentation of the system.
2. Identifying the legislation and regulations.
3. Economic impact analysis.
4. A study of the exceptions and appeals based on Federal Energy guidelines.
5. Forms analysis.
6. Analysis of algorithm and the code.
7. Sensitivity of calculations.
8. Internal consistency checks.
9. External consistency checks.
10. A study of the amendments.
11. Oil industry accounting.
12. Measurement practices.

Three field trips to Washington, one trip to Los Angeles, and several in the Bay Area.

Begun October 25, 1978

Work on draft interim.

Completed November 30, 1978.

DRAFT INTERIM REPORT

Further research is still to be done (between January and April 1) in the following areas:

1. Internal consistency checks.
2. External consistency checks.
3. A study of the amendments.

In addition to the above, the following new areas are going to be included in the second phase of the validation study of the DCOES.

1. A study of the adjustments process.
2. Pretest field interview questionnaire, update interview form and complete field interviews.

3. Analyses of information obtained through interviews with the oil companies.
4. Four more trips to Washington.
5. Write final report.

0 0 5 4 0 2 2 9 9

Appendix G

INTERIM VALIDATION REPORT MIDDLE DISTILLATE PRICE MONITORING SYSTEM

EXECUTIVE SUMMARY

PREFACE

This is an interim report which covers the work performed, to date, by the Energy Information Validation Project at the Lawrence Berkeley Laboratory on the Middle Distillate Price Monitoring System (MDS). The report states the project's findings and conclusions and, where appropriate, sets forth recommendations to improve the accuracy and usefulness of information processed by the system.

Briefly, the MDS collects data on prices and gross margins for No. 2 heating oil from a sample of refiners, resellers and retailers in the petroleum industry. The data, collected and aggregated by the Energy Information Agency (EIA), is used by the Energy Regulatory Agency's (ERA) Office of Fuels Regulation (OFR) to evaluate the level of competition and the reasonableness of prices in the heating oil market. If, on the basis of this information, OFR determines that the objectives of the Emergency Petroleum Allocation Act (EPAA) are not being fulfilled, it may recommend that DOE take remedial actions which may include the reimposition of mandatory price and/or allocation controls for No. 2 heating oil.

One conclusion is that information on average prices and gross margins does not provide a basis for determining whether a market is competitive. Accurate descriptive information on prices and gross margins could be useful for a preliminary analysis--trends or anomalies in these series could prompt further investigation. However, this information could not provide any specific focus for such an inquiry. A second con-

0 0 5 5 4 0 2 3 0 0

clusion is that there is serious doubt as to the accuracy of the information collected by the system.

EXECUTIVE SUMMARY

I. SYSTEM IDENTIFICATION^a

A. CURRENT IDENTIFICATION INFORMATION

- Name of System Being Validated: Middle Distillate Price Monitoring System
- Number of Form as Cleared: EIA-9
- EIA Standard Series Number: Not yet known
- Form Clearance Information: Submitted to the Office of Management and Budget for clearance on September 13, 1977. It is scheduled to expire on August 31, 1980.
- Statutory Authority Under Which Reporting Requirements Were Established:
 1. Federal Energy Administration Act of 1974 (P.L. 93-275, Sect. 7 and 13, May 7, 1974, as amended by P.L. 94-332, June 30, 1976; P.L. 94-385, August 14, 1976; P.L. 95-70, July 21, 1977; and P.L. 95-91, August 4, 1977).
 2. Energy Supply and Environmental Coordination Act of 1974 (P.L. 93-319, Sect. 11, June 22, 1974, as amended by P.L. 94-163, December 22, 1975; and P.L. 90-50, July 21, 1977).
 3. Emergency Petroleum Allocation Act of 1973 (EPAA, P.L. 93-159, November 27, 1973, as amended by P.L. 93-511, December 5, 1974; P.L. 94-99, September 29, 1975; P.L. 94-133, November 14, 1975; P.L. 94-163, December 22, 1975; and P.L. 94-385, August 14, 1976.
- Other Current Standard Reference Numbers or Names Identifying the Form: None

B. HISTORICAL IDENTIFICATION INFORMATION

- Names and Numbers of Antecedents to the Present System: Middle Distillate Monitoring System; FEA-P-112-M-1. This system was used during the 1976-77 winter.
- Planned Successors to the System: None

^aEach item in Section I of the Executive Summary is explained in the corresponding section of the text. For example, "Uses of Output" are summarized in Section 1.F. of the Executive Summary, while a full discussion is contained in Section 1.F. in the main body of the report.

C. IDENTIFICATION OF PURPOSE

- Purpose of the System: To provide the information necessary for ERA's Office of Fuels Regulation to determine whether the objectives of the EPAA, regarding market competition and equitable prices, are presently being fulfilled in the market for No. 2 heating oil.

D. IDENTIFICATION OF INFORMATION COLLECTED

- Reporting Requirements : Firms are required to complete all parts of EIA-9.
- Numerical Information Requested:
 1. Monthly sales volume for No. 2 heating oil
 2. Average unit price for No. 2 heating oil
 3. Monthly domestic purchase volume for No. 2 heating oil
 4. Average unit costs for domestic purchases of No. 2 heating oil
 5. Monthly imported purchase volume for No. 2 heating oil
 6. Average unit costs for imported purchases of No. 2 heating oil
 7. Beginning monthly inventory of No. 2 heating oil
 8. Estimated storage capacity for No. 2 heating oil
- Degree of Resolution for Numerical Information Collected: Items (1) and (2) are requested for each State in which heating oil is sold. Items (3) through (8) are firm aggregates.
- Descriptive Information Requested:
 1. Primary line of business (refiner, reseller, retailer, or reseller/retailer)
 2. Classification of reporting unit (consolidated or unconsolidated)
 3. Various address and firm identification information.

E. IDENTIFICATION OF COLLECTION AND COLLATION PROCESS

1. Universe Identification and Sample Design

- Universe: All refiners, resellers, and retailers of No. 2 heating oil within the political boundaries of the United States.
- Frame: Those firms in the "1974 Market Share Historical File" who were engaged in the No. 2 heating oil business as determined by a census taken through the use of forms P-305 and P-308. The frame was revised slightly in 1977 to include outlets divested by AMOCO. There are 9560 firms in the frame.
- Sample Selection: The frame was stratified into 216 cells by the 1974 size of the firm, location, and location of customers. In each cell the firms were arrayed by 1974 size, and a random start systematic sample was drawn. There are 1463 firms in the sample. This panel reports monthly for an entire year.
- Means of Collecting Information: Mail Survey.
- Reporting Interval: One month.
- Frequency of Data Collection: Monthly.

2. System Implementation

- Collection Agency: Office of Energy Data/EIA is responsible for collecting the information.
- Processing Agency: Data Technology Industries (DTI) receives completed EIA-9 forms from firms. DTI checks, processes, and enters the data on computer files, which are given to EIA. EIA checks the information again and publishes the results. The contract is with EIA/Energy Data.
- Information Volume: DTI collects and processes EIA-9 forms from approximately 1200 firms each month during the heating season, a total of about 3000 records.
- Processing Time: Forms are due 20 days after the reporting period has ended. Data are published 45 days after the reporting period has ended.
- Custodians of Computer Files: Optimum Systems Incorporated (OSI) under contract to EIA/ADP Services.

F. USES OF OUTPUT

1. Prescribed Uses of Data

- Regulatory Requirements: Volume 43, Federal Register, pp. 2917-23, January 20, 1978, establishes requirements regarding data use for the following groups:
 1. ERA's Office of Fuels Regulation: OFR is required to examine MDS data in order to determine if the EPAA objectives of market competition and equitable prices are being fulfilled in the market for No. 2 heating oil.
 2. The ad-hoc Subcommittee of DOE's Fuel Oil Marketing Committee: The Subcommittee is required to review MDS data and to make recommendations regarding benchmark prices and gross margins used by OFR in its analysis in Item (1).
 3. ERA's Office of Administrative Review is required to hold an evidentiary hearing on the marketing and pricing of No. 2 heating oil during the 1977-78 winter. After considering OFR's conclusions and any comments by consumers and industry, OAR is required to forward its recommendations on any needs for further regulatory action, regarding No. 2 heating oil, to ERA's Administrator.

The MDS is not explicitly required by the Code of Federal Regulations. These regulations merely exempt No. 2 heating oil from price controls. These prescribed uses were established by ERA, in order to fulfill their obligations to see that the objectives of the EPAA continue to be met. (See Appendix A for details)

- Regulatory Decisions Supported by Information: If, based on the conclusions reached by OFR, the ERA Administrator deems it necessary, he may take the following steps:
 1. Audits of individual firms
 2. Public hearings regarding the price of No. 2 heating oil
 3. Voluntary price restraints
 4. Re-imposition of mandatory price and/or allocation controls

2. Publication of Data

- Agency Which Publishes Data: Prices, Costs and Marketing Section of EIA
- Reports Published:
 1. Energy Data Report: Heating Oil Prices and Margins (monthly) (EIA 0031)
 2. Monthly Energy Review (EIA 0035)
 3. Monthly Petroleum Product Price Report (EIA 0032)
 4. Quarterly Report to Congress (EIA 0008)
 5. Monthly Report to the President
- Primary Format of Reports: Tables
- Recipients of Reports: Not yet examined

II. CONCLUSIONS AND RECOMMENDATIONS

A. INFORMATION USEFULNESS

- Information on average prices and gross margins does not provide a basis for determining whether a market is competitive. Accurate descriptive information on prices and gross margins could be useful for a preliminary analysis -- trends or anomalies in these series could prompt further investigation. However, this information could not provide any specific focus for such an inquiry.

B. INFORMATION ACCURACY

With a well-constructed sample survey, a user should be able to determine how much confidence to place in estimates: their uncertainty should be quantified. The Middle Distillate Price Monitoring System does not appear to satisfy this criterion. It is difficult or impossible to judge the accuracy of the estimates derived from this system.

- The frame appears to be seriously incomplete. Preliminary estimates of undercoverage indicate that the frame misses roughly one-third of the firms in the target universe. The impact of the non-sampling error this creates may be large, but it is presently unknown.

- The definitions of No. 2 heating oil used by firms in different segments of the market are not consistent. Again, this creates a non-sampling error which may be large, but is presently unknown.
- The sample design is not suited to the computation of standard errors. Therefore, the impact of sampling error on the estimates can be assessed only by making untestable assumptions. With other equally practical designs, standard errors can be computed directly from the sample.
- The non-response rate, whether measured by number of firms or volume of oil, is about 10 percent -- despite the fact that responses are mandatory.
- In some respects, Form EIA-9 and its instructions are confusing to respondents.
- Answers are sometimes internally inconsistent, and edit routines are not designed to detect these inconsistencies.

Recommendations follow for improving the quality of the information collected by the system. They are discussed in more detail in the body of the report (Section IIC).

- The frame should be revised periodically. Firms that have gone out of business should be dropped, new firms that have come into the market should be added. After revision of the frame, a new sample should be drawn.
- Either the definition of No. 2 heating oil should be clarified, or information should be collected on a better-defined product class, such as: all middle distillate fuel, or all No. 2 oil (heating and diesel).
- The sample should be redesigned to facilitate computation of standard errors, and reduce the burden of reporting.
- More effort should be put into contacting the non-respondents.
- The questionnaire should be revised.
- More sophisticated computer edit routines should be implemented to identify inconsistent responses.
- A small random sample of forms should be taken each month for close review.

FUTURE WORK TO BE DONE TO COMPLETE THE SYSTEM STUDY

Middle Distillate System

1. Investigate the information required to judge market competitiveness and price equitability. Review relevant paradigms and their associated problems.
2. Investigate the incentives for bias and the process of respondents filling out the forms. Investigate effects of errors introduced in this process.
3. Describe the heating oil market.
4. Investigate coverage and bias due to non-coverage.
5. Investigate bias due to non-response.
6. Estimate sampling error.

This report was done with support from the Department of Energy. Any conclusions or opinions expressed in this report represent solely those of the author(s) and not necessarily those of The Regents of the University of California, the Lawrence Berkeley Laboratory or the Department of Energy.

Reference to a company or product name does not imply approval or recommendation of the product by the University of California or the U.S. Department of Energy to the exclusion of others that may be suitable.

TECHNICAL INFORMATION DEPARTMENT
LAWRENCE BERKELEY LABORATORY
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720