

GAO

Report to the Ranking Minority Member,
Committee on Governmental Affairs,
U.S. Senate

June 1995

HEALTH AND SAFETY

DOE's Epidemiological Data Base Has Limited Value for Research



**Resources, Community, and
Economic Development Division**

B-260637

June 6, 1995

The Honorable John Glenn
Ranking Minority Member
Committee on Governmental Affairs
United States Senate

Dear Senator Glenn:

During the 1980s, the dual role that the Department of Energy (DOE) played by both producing nuclear weapons and assessing the potential health hazards associated with this production raised serious concerns about the credibility of the results of DOE's research on the health of people working at or living near DOE's facilities. In early 1990, the Secretary of Energy announced several initiatives to address these concerns based on recommendations from a special panel of experts—the Secretarial Panel for the Evaluation of Epidemiologic Research Activities.¹ One of these initiatives was the development of a data base to store and retrieve data from DOE on the demographics, health, and exposure of its workers and the communities near its facilities. The data base, to be developed under the guidance of the National Academy of Sciences, was expected to be a valuable, comprehensive resource for those conducting long-term epidemiological and other health studies. For the first time in DOE's history, these data would be accessible to independent researchers.

In 1992, DOE began releasing the data used in its past research on health effects to outside researchers through a system it called the Comprehensive Epidemiologic Data Resource (CEDR). However, you were concerned that this system was not as comprehensive as originally envisioned and might be of limited use. Consequently, you asked us to determine (1) whether the current system functions as the comprehensive repository of epidemiological data² about DOE's workers and the communities surrounding the Department's facilities envisioned by the Secretarial Panel and the National Academy of Sciences and (2) whether it

¹The panel, chaired by Kristine Gebbie, M.N., the then-Secretary of Health for the state of Washington, includes professors from schools of public health, epidemiology, and law; several directors of state health agencies; and representatives of the United Auto Workers union and the American Cancer Society.

²Epidemiological data include the medical, demographic, exposure, environmental, and other data necessary to support many kinds of research activities, such as health surveillance and monitoring, screening programs, studies of the incidence of diseases (morbidity studies), and long-term studies of death rates (mortality studies). For individuals, such data can be drawn from employment history and from information on demographics, health and medical history, and occupational and other exposures, such as smoking and diet. Follow-up studies also provide data on individuals.

meets their intended objectives of accessibility and utility for outside researchers. You also asked us to determine DOE's future plans for this system.

Results in Brief

The Comprehensive Epidemiologic Data Resource that DOE developed is not the comprehensive data base for epidemiological research envisioned by the Secretarial Panel and the National Academy of Sciences. The system lacks uniform data on the exposure of DOE's current laboratory workers to radiation and other hazardous substances that might affect their health, as well as data on the health of these workers and residents near DOE's facilities. These data have not been routinely collected or maintained throughout the DOE complex. DOE is trying to standardize the way its facilities collect and maintain these data and to develop a more comprehensive health surveillance program on its employees, as recommended by the Secretarial Panel, but is at least 3 years from accomplishing these goals. Without these data, researchers cannot make the kinds of comparisons that lead to findings on health effects. The data that currently appear in the system are primarily the results of past DOE studies of workers' deaths and are of limited value for original research.

While the Comprehensive Epidemiologic Data Resource is easily accessible, few independent researchers have used it because problems with the data currently in the system limit its usefulness for new research. Problems include the absence of updated or original data, the extent to which some personal identifiers have been removed to protect the privacy of DOE's workers, missing and inconsistent data elements, and inadequate documentation by the researchers who provided the data. Consequently, new researchers have had to examine original records at DOE's facilities, where they have encountered some problems in obtaining these records.

DOE is uncertain whether the system will ever be the comprehensive data base envisioned by the Secretarial Panel and the National Academy of Sciences. DOE has not developed specific long-range plans that identify the tasks, milestones, and resources necessary to develop a system that would maintain and disseminate uniform data on the demographics, exposure, and health of the Department's workers and residents near its facilities. Furthermore, DOE has not assessed alternatives to the current system and does not know whether there is a more cost-effective and practical means of providing independent researchers with access to data from its epidemiological studies.

Background

Over the past 50 years, as a result of producing tens of thousands of nuclear weapons, DOE's facilities have also produced radioactive and other toxic substances that pose potential health threats to DOE's workers and the communities located nearby. These substances include the radionuclides uranium, plutonium, and cesium; toxic metals; organic solvents; and chlorinated hydrocarbons. Epidemiological research—research on the incidence, distribution, and control of disease in a population—provides a scientific evaluation of the health effects of exposing workers and the public to such potentially harmful materials. Such research uses health, exposure, environmental monitoring, and personnel records to analyze health effects and evaluate methods to protect people and prevent harm. As such, epidemiological research is essential to a comprehensive occupational and environmental health program.

DOE and its predecessor agencies have a long history in epidemiological research, starting with studies of the survivors of the atom bomb. In the past, much of this research was conducted by DOE or its contractors in secret and concentrated on the correlation between the rates of cancer-related deaths of workers at DOE's nuclear weapons complex and their exposure to ionizing radiation. A number of separate mortality studies—studies of death rates—have been conducted on approximately 420,000 workers over the past 30 years. However, because the records that researchers needed to study the health effects of working in DOE's facilities were maintained differently at each facility and were difficult to locate, the types and quality of epidemiological research that could be conducted were limited. To alleviate these problems and facilitate epidemiological research on the health effects of exposure to radiation and other hazards, the Secretarial Panel recommended that DOE continue developing CEDR as a comprehensive repository of data on its workers.

In addition, to break down what was perceived as “a wall of secrecy” and to help establish the credibility of and maintain independence in the conduct of DOE's epidemiological research, the Secretarial Panel recommended opening this research and its supporting data to external investigation and scrutiny. Among other things, the Secretarial Panel recommended that DOE execute a memorandum of understanding with the Department of Health and Human Services (HHS), making HHS responsible for long-range, analytic epidemiological studies, while DOE remained

responsible for descriptive epidemiology.³ As a result, much of the epidemiological research on DOE's facilities is now managed by HHS. Within HHS' Centers for Disease Control and Prevention, which implemented this memorandum of understanding, the National Institute for Occupational Safety and Health was made responsible for occupational health research (i.e., research on workers employed by DOE and its contractors), while the National Center for Environmental Health was made responsible for research involving the environment, including communities near DOE's facilities.

The Secretarial Panel also called for greater outside scrutiny by recommending that the National Academy of Sciences (NAS) play a key role in overseeing and monitoring the development of CEDR. In response to the Secretarial Panel, as well as a concurrent request from DOE to provide general scientific advice on the status and direction of DOE's epidemiological programs, NAS established a Committee on DOE Radiation Epidemiological Research Programs.⁴ In 1990, this committee issued a report making a number of recommendations about access to data for researchers outside DOE, the types of data to be included in CEDR, and its future development.⁵ The report also noted that use of CEDR will depend on ease of access to the information it contains and researchers' perception of its value.

Beginning in 1990, a DOE contractor facility, the Lawrence Berkeley Laboratory, in Berkeley, California, constructed a prototype, known as preCEDR, to serve as the basis of CEDR. In 1992, DOE made data available through this system. In August 1993, DOE published a catalog of data available in CEDR to assist current and potential users in identifying data sets⁶ for potential use and to provide instructions on how to obtain access to these data. Through fiscal year 1994, DOE had received \$14.35 million in appropriations for CEDR, of which it had spent \$9.45 million for CEDR and

³Analytic epidemiological studies are designed to test causal hypotheses; for example, the correlation between exposure to specific substances and illness among groups of people. Descriptive epidemiology uses basic data on exposure, demographics, work history, and other factors to identify patterns of illness and exposures among groups of people without determining a specific causal relationship.

⁴The committee's full title is Committee on DOE Radiation Epidemiological Research Programs, Board on Radiation Effects Research, Commission on Life Sciences, National Research Council.

⁵Providing Access to Epidemiological Data: First Annual Report, Committee on DOE Radiation Epidemiological Research Programs, National Research Council (National Academy Press: 1990).

⁶A data set is a collection of logically related data files. CEDR contains two types of data sets: (1) working data sets that contain data extracted by researchers from original records, such as payroll, personnel, or dosimetry records, and (2) analytic data sets that contain composites of working data that have been merged and analyzed by researchers to answer specific questions. (See app. I.)

related expenses and redirected the remaining \$4.9 million to other activities.⁷ CEDR is budgeted at \$1 million for fiscal year 1995, of which \$500,000 was funded as of February 1995.

Lack of Important Epidemiological Data Limits CEDR's Value

DOE does not have available the uniform demographic, exposure, medical, and environmental data that would make CEDR a comprehensive and valuable epidemiological resource for independent researchers. The Secretarial Panel recommended in 1990 that DOE define a minimum set of data necessary for epidemiological research and routinely maintain and collect these data at all DOE facilities. As part of this effort, in May 1992 DOE requested that each of its facilities, within 3 years, complete an inventory of 123 specific types of records that the Department believed were important for conducting epidemiological studies. We reported on this and other DOE efforts to manage records in a May 1992 report.⁸ DOE officials told us that when completed, this records inventory would be included in CEDR and would more easily identify for researchers where these specific types of records are located. Meanwhile, DOE is waiting for its facilities to complete their records inventories, which may take until 1996, before it takes steps to routinely collect and maintain the types of records it has already identified as important.

In addition, the NAS committee stated that CEDR should be capable of supporting many kinds of epidemiological studies, including long- and short-term health surveillance, monitoring studies, screening programs, and long-term mortality studies. However, as we reported in December 1993,⁹ DOE probably will not establish a comprehensive health surveillance program until at least 1998. Such a program would standardize the documentation of workers' occupational exposures to radiation and other industrial hazards—such as chemicals, gases, metals, and noise—and could identify trends in workers' illnesses and injuries that might be related to these exposures. Until such a program is in place, the comprehensive data on health effects and exposure needed for important epidemiological research will not be available for placement in CEDR. Moreover, DOE's Assistant Secretary for Environment, Safety, and Health

⁷Of the \$9.45 million spent, \$7.12 million was used for direct expenses for CEDR, and \$2.33 million was used for "related expenses."

⁸DOE Management: Better Planning Needed to Correct Records Management Problems (GAO/RCED-92-88, May 8, 1992).

⁹Health and Safety: DOE's Implementation of a Comprehensive Health Surveillance Program Is Slow (GAO/RCED-94-47, Dec. 16, 1993).

told us in October 1994 that standardization of data at DOE's facilities was a problem that would take several years to resolve.

Without the important data necessary to support many types of epidemiological research, CEDR today mainly contains the limited data from DOE-sponsored mortality studies of workers at DOE's facilities at Oak Ridge, Tennessee; Rocky Flats, Colorado; Hanford, Washington; and elsewhere. Of the 37 data sets in CEDR, 36 contain the retrospective information—data on past incidents—used to conduct these studies. (See app. I.) Some new data will be included when certain ongoing studies are completed. These studies include mortality studies of DOE's workers at the Idaho National Engineering Laboratory and the Portsmouth Gaseous Diffusion Plant in Ohio; a study of cancer incidence among workers at Rocky Flats by the National Institute for Occupational Safety and Health; and studies from the National Center for Environmental Health, including estimates of the effect of the radiation from Hanford on the air and water in the surrounding area. While adding the results of these studies will make some of the data in CEDR more current, the system will still lack the comprehensive data discussed above that would make it the valuable resource that the Secretarial Panel and NAS recommended.

According to many NAS committee members and CEDR users we spoke with, the current lack of comprehensive epidemiological data limits CEDR's value for research. The Secretarial Panel cautioned DOE that retrospective data would have limited value for future research. Also, members of the NAS committee told us that the data on mortality that CEDR currently contains limit the types of studies that can be done and have minimal value for future research on health effects. NAS noted in its 1994 report that the scope of the data currently in CEDR limits the type of research that can be conducted.¹⁰ The data restrict researchers by defining the groups that can be studied, the variables that can be examined, and the analytic methods that can be applied. Officials at the National Institute for Occupational Safety and Health and the National Center for Environmental Health also stated that CEDR would be of greater value if it contained data on chemical exposures and health effects. These data will not be available until DOE's health surveillance program is completed. Since CEDR contains only limited retrospective data, researchers who need more information must still locate records at DOE's facilities, where the records are not consistently maintained. However, despite CEDR's limited value for health effects research, several NAS experts, current users, and

¹⁰Epidemiologic Research Programs at the Department of Energy: Looking to the Future, Committee on DOE Radiation Epidemiological Research Programs, National Research Council (National Academy Press: 1994).

DOE officials believe that it has significant value as a teaching tool for students of epidemiology.

CEDR Is Easy to Access, but Limitations Impair Its Utility to Researchers

DOE has made data from its mortality studies easy for outside researchers to access through CEDR, and thousands of people have accessed the system to see what basic data are available. However, few researchers have used the data for original studies on health effects. In addition, some members of the NAS Committee on Epidemiological Research and some researchers we interviewed noted problems that impair the usability of the data. Difficulties include a lack of data that have not been previously modified by other researchers to meet their specific research needs, data that are hard to work with because they have been edited to protect the privacy of the workers, and data that are not current. In addition, some researchers have encountered problems with the quality of the data, including missing and inconsistent data and inadequate documentation of the studies included. For these reasons, some CEDR users need to review original records at DOE's facilities but find the records difficult to obtain.

CEDR Is Easy to Access

For the first time in its history, DOE has made the data used to support its epidemiological research accessible. DOE has created a system that allows researchers easy access to the epidemiological data that were used to conduct its mortality studies, as recommended by both the Secretarial Panel and NAS. In addition to data from past studies, CEDR contains summary information, such as the 1992 annual summary of epidemiological surveillance data from Brookhaven National Laboratory. Potential users of CEDR can obtain basic information about the system's contents and file structure (but cannot access the actual data) through DOE's published catalog of available data or via a computer link with CEDR directly or through the Internet.¹¹ The summaries, which do not provide detailed research data, are available to all Internet users. We were able to access CEDR directly from personal computers using communication software and found the instructions relatively easy to follow. According to the CEDR staff at the Lawrence Berkeley Laboratory, computer logs show that thousands of people have accessed CEDR to find out what basic data are available.

¹¹The Internet is an interconnected web of thousands of computer networks, cooperating to transport a variety of information to millions of users worldwide. Authorized users can also access CEDR by using their computers to dial directly into the telephone connections at Lawrence Berkeley Laboratory.

To view or obtain the actual data on DOE's workers, a user must receive authorization from DOE. Getting such authorization is a relatively simple process. The required forms, including confidentiality agreements, are provided in the CEDR catalog. Authorization generally takes about a month. Approved users can obtain data from the Lawrence Berkeley Laboratory via electronic tape or diskette, or through direct transmission if they have specialized equipment. Users we talked with reported no major problems in obtaining data from CEDR.

Few Researchers Are Using CEDR

Despite the system's accessibility, few independent researchers have sought approval from DOE to become authorized CEDR users. In addition, some authorized users have never obtained data from CEDR. DOE provided us with a list of 22 primary users as of September 1994.¹² Some of the users listed, however, were not independent researchers but worked for DOE or its contractors. Some of these users were involved only in loading, testing, and maintaining the system. We identified 13 independent researchers who were primary users and may have obtained data from CEDR. (See table 1.) We confirmed that nine independent researchers had obtained data from CEDR. Three of these users worked on studies funded by the National Institute for Occupational Safety and Health, three worked on university research projects, two conducted research for public health institutes, and one was a private consultant.

Table 1: Primary CEDR Users as of September 1994

Type of user	Number
DOE employees and contractors	4
GAO evaluator	1
Independent researchers	
Researchers using CEDR data ^a	9
Researchers not using data	4
Researchers not contacted ^b	4
Total	22

^aThese nine researchers represented seven projects, two of which had two primary users each.

^bWe were unable to contact some users despite repeated attempts and did not attempt to contact those located in Europe.

¹²A primary user establishes a CEDR account and receives data from the Lawrence Berkeley Laboratory. A primary user is allowed to share data with assistants on the same research project, who are authorized as secondary users on the same CEDR account. We interviewed primary users listed as of September 1994. In October 1994, DOE told us that the number of primary users had increased by 10, to a total of 32.

Usefulness of Data in CEDR Is Limited

Researchers using CEDR have encountered a number of problems with the data in the system, limiting the value of these data for their research. Although four of the nine researchers we spoke with found the quality of the data satisfactory for their research purposes, the other five researchers reported the following problems:

- Original data, not previously edited by other researchers, are not available through CEDR.
- To protect workers' privacy, key data elements important for certain research have been removed.
- The data in the mortality studies are frequently old and have not been updated.
- Research is hindered by problems with the quality of the data, including missing and inconsistent data and inadequate documentation of studies by prior researchers.

Data as Originally Recorded at DOE's Facilities Are Often Unavailable

It is difficult to conduct research beyond DOE's initial studies or to fully validate the results, according to many of the researchers we spoke with, because CEDR may not contain data as they were originally recorded at DOE's facilities. Instead, it generally contains data that have been assembled and edited by prior researchers to answer specific research questions. Some independent researchers using data in CEDR stated that they need the original records to conduct their studies. Two CEDR users conducting studies under contracts with the National Institute for Occupational Safety and Health stated that their research was hampered because the working data sets available in the data base were not original data but had already been edited by prior researchers. Answering new research questions would require obtaining the original records directly from DOE's facilities. Another CEDR user conducting research for a public health institute told us that the best data for research are the original records found at DOE's facilities. An official of the National Institute for Occupational Safety and Health, as well as a member of the NAS committee, stated similar views.

Researchers Have Difficulties With Data That Have Some Personal Identifiers Removed

The extent to which some personal identifiers have been removed from the data in CEDR to protect the privacy of workers has made it difficult for some CEDR users to do more precise calculations or compare records. For example, DOE replaced identifying data elements, such as names and social security numbers, with pseudo-identifiers. DOE also rounded some key dates in workers' files, such as birth date, hiring date, and death date, if applicable. In contrast, an official from the National Institute for Occupational Safety and Health stated that while the Institute replaces

identifying data elements, such as the name and social security number, in data that it releases to the public, it does not truncate dates.

Researchers funded by the National Institute for Occupational Safety and Health noted that truncating key dates makes it difficult to do precise calculations of exposure, for which it is necessary to know the exact numbers of days a worker is exposed to a hazard. In addition, replacing identifying data elements makes it difficult to compare various records on workers by, for example, consulting a state or national cancer registry. Consulting such registries is often necessary to obtain a worker's complete health history.

Mortality Data Are Not Updated

Several NAS committee members and current CEDR users told us that CEDR would be more useful for follow-up studies if mortality data were updated, especially data on those exposed to radiation. The mortality studies included in CEDR were conducted on various workers who were employed between 1942 and 1988 at different DOE facilities. In many of these studies, the most recent mortality data are more than 10 years old. Researchers are unable to follow up on the results of the mortality studies without significant additional work. Researchers we spoke with explained that because the chronic effects of exposure to low doses of radiation may not occur until decades afterwards, workers who have been exposed to radiation should be studied over lengthy periods. One epidemiologist, a member of the NAS committee, stated that unless the workers in a study are monitored until the cause of death has been determined, the results of the study are not conclusive. Other epidemiologists and health physicists from the Centers for Disease Control and some DOE contractors also agreed that the data in CEDR would be more useful if the information on mortality were updated. DOE's Assistant Secretary for Environment, Safety, and Health said that while she considers it the responsibility of the Department to update these radiation studies, she is not sure that the funding necessary to do this will be available, given the current emphasis on funding research on the occupational health effects of hazardous chemicals rather than radiation.

Quality of Some Data Is Questionable

Some researchers working with CEDR have encountered additional problems with the quality of the data. Five primary users we interviewed had encountered missing, inconsistent, or inaccurate data. Measuring exposure was a major problem for these users. Examples provided by the data base manager of a research project sponsored by the National Institute for Occupational Safety and Health included the following:

-
- In one file, the researchers identified data on 115 workers that conflicted with other information in the file about the amount of radiation to which these workers had been exposed. The researchers could not determine which data were correct.
 - In another file, researchers found 1,000 people listed as never having been monitored for plutonium exposure. Nevertheless, a date was entered in the field for “first date monitored for plutonium exposure.” The researchers could not tell which information was correct.

One CEDR user, who had served on the NAS committee, expressed concern that inexperienced researchers could draw erroneous conclusions on the basis of the data currently in CEDR. In her opinion, DOE should not widely publicize access to CEDR for research until some of the problems with its data have been addressed.

In an attempt to identify problems with the quality of the data, DOE is setting up a computer bulletin board for CEDR users to communicate with each other and point out problems they have uncovered. DOE cannot be sure, however, that users will take the time to point out these problems.

Studies Are Inadequately Documented

The Secretarial Panel noted that an important element of epidemiological studies is documentation from the original researcher explaining the study’s methodology, assumptions made, and limitations of the data. While both the Secretarial Panel and the NAS committee recommended that all studies provided to CEDR should be supported with documentation, some researchers using CEDR have found insufficient documentation, making the studies difficult to reconstruct. In one case, a university researcher had to go to the facility that was the subject of the study to resolve problems with the documentation. Researchers using CEDR for the two studies sponsored by the National Institute for Occupational Safety and Health also noted problems caused by inadequate documentation.

The staff at the Lawrence Berkeley Laboratory responsible for developing CEDR told us that the researchers who provided the studies often did not comply with documentation guidelines. DOE has recently issued revised guidelines in an attempt to improve compliance. However, this measure will not correct inadequate documentation of those studies already in CEDR, and it is unknown whether future data providers will be more responsive to this revised guidance.

Records Are Hard to Obtain From DOE's Facilities

Because of the limitations of the data in CEDR, some researchers seek to obtain original records from DOE's facilities, but they report encountering difficulties. Researchers using CEDR for the two studies sponsored by the National Institute for Occupational Safety and Health reported that difficulties in obtaining original records are inhibiting their research. The two researchers told us that when requesting such records from DOE sites, they encountered either uncooperative contractor staff or a lack of adequate staff resources to service their requests.

According to DOE's Assistant Secretary for Environment, Safety, and Health, CEDR is not really intended to be the sole source of data for epidemiological researchers from the National Institute for Occupational Safety and Health, who are likely to require the original records from DOE's facilities. She was aware that these researchers and others have had difficulties obtaining records from some DOE sites, and she was attempting to work with the contractors to resolve specific problems on a case-by-case basis.

Future of CEDR Is Unclear

Although DOE is adding to the contents of CEDR, doubt remains whether the data base will become the system that NAS and the Secretarial Panel envisioned, containing uniform and useful demographic, exposure, medical, and environmental data. The DOE Assistant Secretary responsible for the CEDR program acknowledged the system's current limitations and told us CEDR may not become this comprehensive data base. Moreover, DOE has not attempted the long-range planning needed to achieve this vision.

The Secretarial Panel had recommended that DOE, under the guidance of NAS, establish a clear statement of CEDR's intended goals and uses and an orderly plan for implementing the system. Such a plan would define the steps to be accomplished, milestones for completing the work, and resources needed. NAS committee members told us they were not aware of any long-range planning for CEDR. DOE officials with the Office of Epidemiology and Health Surveillance told us they did not have any long-range plans that identified the specific tasks, priorities, time frames, or resources necessary to develop CEDR into a comprehensive data base containing the types of data that NAS had recommended. DOE currently does not know when comprehensive epidemiological data will be available to put into CEDR, how much it will cost to place these data in CEDR, or how many researchers will potentially use these data.

DOE is making progress toward standardizing and maintaining data on the exposure of its current laboratory workers to radiation and other hazards that might affect their health. Rather than develop CEDR into a comprehensive data base, the DOE Assistant Secretary said DOE may consider that the data base's current function of providing the public with access to its existing epidemiological research data is sufficient. In addition, the Assistant Secretary told us in October 1994 that the budget for CEDR—\$1 million in fiscal year 1995—will be reevaluated if usage does not increase substantially. Even with increased usage, however, it is not clear whether CEDR is the most cost-effective and practical means of accomplishing the more limited objective of providing access to DOE's epidemiological data and data gathered under the memorandum of understanding with HHS. Some researchers and others we spoke with suggested that a far less expensive clearinghouse arrangement might meet this need just as effectively. For example, a clearinghouse might simply list the name of the study, the type of data it contained, and the location of the data. These data would remain at the facility where they were collected.

Conclusions

CEDR was originally intended both to help dispel public fears about secretive research at DOE and to be a valuable resource for independent researchers studying the long-term epidemiological and other health effects of working at or living near DOE's facilities. The current system has removed the "wall of secrecy" surrounding DOE's epidemiological research by making some of the data available to outside researchers. However, as it now stands, CEDR has limited utility as a research data base. DOE is years away from routinely collecting and maintaining the epidemiological data on its workers that are needed to help make CEDR a comprehensive resource.

Consequently, CEDR appears to be at a crossroad, and an overall assessment of the system would help DOE better ensure that it is spending its limited funds wisely. If DOE decides to pursue the original vision for CEDR, it cannot be assured of an orderly implementation without a long-range plan that sets forth the required time frames, resources, and costs and takes into account the ongoing efforts to uniformly collect and maintain epidemiological data throughout DOE's facilities. If DOE decides not to develop a comprehensive epidemiological data base, it could either maintain or abandon the current system. However, maintaining the current system may not be the most practical and cost-effective means of providing the epidemiological data used in DOE's past studies and those currently being conducted by HHS. Resolving the problems impairing the

usefulness of the data in the current system could cost DOE still more. Finally, if DOE decides to abandon the system, continued openness and public access to its health effects research cannot be ensured without identifying alternative means of collecting and disseminating epidemiological data.

Recommendations

We recommend that the Secretary of Energy, in consultation with the Secretary of Health and Human Services, the National Academy of Sciences committee, and representatives of the research community, determine whether the Comprehensive Epidemiologic Data Resource is the most practical and cost-effective means of providing epidemiological data for research on health effects. The assessment should cover the costs, benefits, and time frames for including more comprehensive data on health effects in the data base, as well as alternative means of making these data available to outside researchers.

If the Secretary determines that the Comprehensive Epidemiologic Data Resource is not the most practical and cost-effective means of compiling epidemiological data, DOE should determine whether continued funding is appropriate.

Agency Comments

As requested, we provided a draft of this report to DOE for comment. Although DOE did not provide a written response, the Acting Director of the Office of Epidemiology and Health Surveillance did express her views on the report.

Overall, she agreed with the problems we identified with the data. However, she maintained that such limitations are inherent in data collected from historical studies and that these data on former workers are nevertheless important and useful. She noted that DOE is making efforts to update and review these data to resolve inconsistencies. She further noted that DOE is required to remove personal identifiers to protect the identities of individual workers. We fully agree that workers' privacy must be protected. Nevertheless, as we stated in our report, unlike the National Institute for Occupational Safety and Health, DOE truncates (abbreviates or shortens) key dates, an action that can limit the usefulness of the data.

Regarding the need to include data on current workers and residents in CEDR, the Acting Director agreed that the information is vital and will be

included as new studies are completed. However, while adding the results of these studies will make some of the data more current, the system will still lack the comprehensive data—such as uniform health, exposure, environmental monitoring, and personnel data—that would make it the valuable resource for new research on health effects that the Secretarial Panel and NAS recommended.

The Acting Director also expressed concern about our recommendation that the cost-effectiveness of CEDR be evaluated, noting that most of the costs for CEDR have already been incurred. However, these costs are the costs of the present data base, which contains historical information. DOE does not know what it will cost to include the types of health surveillance data in CEDR that the Secretarial Panel and NAS recommended. If CEDR will not include these data, even the costs of maintaining the current system may not be justified.

Finally, the Acting Director told us that DOE has added five primary users of the data base since we completed our audit work and has added over 100 files in the last year. We did not verify or evaluate this information.

We also discussed the facts presented in this report with CEDR program officials at the Lawrence Berkeley Laboratory, who generally agreed that these facts were accurate. They provided updated information on users of CEDR and data sets in the system, which we incorporated into the report.

We performed our review between February 1994 and May 1995 in accordance with generally accepted government auditing standards. In performing this review, we interviewed officials at DOE headquarters, including the Assistant Secretary for Environment, Safety, and Health. We also interviewed the personnel at the Lawrence Berkeley Laboratory, Berkeley, California, responsible for designing and operating CEDR. We spoke with eight of the nine members of the NAS committee responsible for monitoring progress on CEDR, officials at the National Institute for Occupational Safety and Health and the National Center for Environmental Health, and all authorized CEDR users we were able to contact. (See app. II for details of our scope and methodology.)

As arranged with your office, unless you publicly announce its contents earlier, we plan no further distribution of this report until 30 days after the date of this letter. At that time, we will send copies to the Secretary of

Energy and other interested parties. We will also make the report available to others on request.

Please call me at (202) 512-3841 if you or your staff have any questions. Major contributors to this report are listed in appendix III.

Sincerely yours,

A handwritten signature in black ink, appearing to read "Victor S. Rezendes". The signature is fluid and cursive, with the first name "Victor" and last name "Rezendes" clearly distinguishable.

Victor S. Rezendes
Director, Energy and
Science Issues

Contents

Letter	1
Appendix I Data Sets Included in CEDR	20
Appendix II Scope and Methodology	23
Appendix III Major Contributors to This Report	24
Tables	
Table 1: Primary CEDR Users as of September 1994	8
Table I.1: Data From DOE-Sponsored Studies on Workers Available Through CEDR as of November 1994	21

Abbreviations

CEDR	Comprehensive Epidemiologic Data Resource
DOE	Department of Energy
GAO	General Accounting Office
HHS	Department of Health and Human Services
NAS	National Academy of Sciences

Data Sets Included in CEDR

The Comprehensive Epidemiologic Data Resource (CEDR) provides a repository of data that have been used to support epidemiological studies conducted on workers at Department of Energy (DOE) facilities. DOE has funded studies on various groups of workers of DOE or its contractors from the 1940s through the 1990s at facilities involved in the production of nuclear weapons. (See table I.1.) More than one study has been included in CEDR for several of these facilities.

As of November 1994, CEDR contained a total of 37 data sets, or logically related data files. Table I.1 lists the 36 data sets covering DOE-sponsored studies on workers; an additional data set covers a 1990 study of atom bomb survivors. Of the 36 data sets in CEDR as of that date, 29 are analytic data sets from past studies at DOE's facilities and 7 are working data sets. Of the 29 analytic data sets from DOE sites or facilities, 28 are from mortality studies. The remaining set came from a morbidity study that examined the incidence and cause of respiratory disease among workers.

Appendix I
Data Sets Included in CEDR

Table I.1: Data From DOE-Sponsored Studies on Workers Available Through CEDR as of November 1994

Facility or site	Analytic data ^a				Latest mortality data	Working data Number of data sets
	Number of data sets	Latest study	Number of workers ^b	Period of employment		
Fernald, Ohio	1	1983	4,101	1952-72	1977	1 ^c
Hanford, Washington	4	1993	44,101 ^b	1944-85	1989	1
Los Alamos, New Mexico	3 ^d	1988	5,424 ^b	1943-88	1988	2 ^d
Linde, Missouri ^e	1	1987	995	1943-49	1979	1 ^c
Mallinckrodt, Missouri ^e	1	1994	2,542	1942-66	1988	1 ^c
Mound, Ohio	3	1991	4,697 ^b	1942-79	1984	1
Oak Ridge, Tennessee	11	1993	28,008 ^b	1943-82	1984	1 ^c
Pantex, Texas	1	1985	3,564	1951-78	1978	1
Rocky Flats, Colorado	1	1987	5,413	1951-79	1979	2
Savannah River, South Carolina	1	1988	9,860	1952-74	1980	1 ^c
Multiple sites	2	1993	59,995 ^b	1944-86	1986	none
Total	29					7

^aAll analytic data sets listed are from mortality studies except the Fernald data set, which is from a morbidity study. Two of the data sets added near the end of 1994 are also from morbidity studies.

^bThe number of workers studied excludes other workers at the site who were not subjects of the study. For sites with more than one study, the number shown is from the study covering the largest number of workers.

^cWorking data from the Oak Ridge Institute for Science and Education on approximately 420,000 people who worked at DOE's facilities between 1943 and 1991. Also included as part of this data set are working data for Fernald, Linde, Mallinckrodt, and Savannah River.

^dThe three analytic and two working data sets for Los Alamos include two data sets of workers of the Zia Company (a previous contractor) at Los Alamos, one analytical data set from an unpublished study, and one working data set that overlaps some of the working data set on Los Alamos in general.

^eThe Linde plant and the uranium facility at Mallinckrodt Chemical Works are no longer operational.

Source: Based on information from DOE.

We analyzed the contents of CEDR as of November 10, 1994. During our review, DOE was adding new data sets and updating others already in the system. For example, DOE added new analytic data sets from 1994 studies on workers at Fernald, Oak Ridge, Mallinckrodt, Savannah River, and other facilities and updated several working data sets, including data on workers at the Mound plant. In addition to the 36 data sets shown in table I.1, seven new analytical data sets, including two from multiple-site

Appendix I
Data Sets Included in CEDR

studies, were added. A total of 44 data sets were available through CEDR as of December 31, 1994. More additions and updates are planned for 1995.

DOE intends to make all the studies that it funds on exposures in or near DOE's facilities available through CEDR. DOE officials told us that during 1995 they plan to add new data sets to CEDR and update some of the existing data. Among the new data DOE plans to add are analytic data sets from additional studies of workers at several DOE facilities, a summary data set of epidemiological surveillance data for one or more sites, a data set on workers who painted radium dials, and data on exposures at DOE's Nevada Test Site. Updates are planned to the working data sets for at least two sites and the dosimetry data for several others.

Scope and Methodology

To determine how well CEDR meets its intended objective of being a comprehensive resource, we (1) reviewed recommendations from reports by the Secretarial Panel for the Evaluation of Epidemiologic Research Activities and National Academy of Sciences (NAS) on designing and implementing CEDR; (2) interviewed officials at DOE headquarters—including the Assistant Secretary for Environment, Safety, and Health; the Acting Director of the Office of Epidemiology and Health Surveillance; and the CEDR Program Coordinator—and contractor staff at the Lawrence Berkeley Laboratory concerning the current status of CEDR; (3) reviewed relevant DOE directives, program plans, progress reports, and documentation on CEDR; (4) interviewed eight of the nine members (attempts to contact the ninth member were unsuccessful) of the NAS committee responsible for monitoring and reporting on DOE's progress on CEDR; and (5) interviewed the officials from the National Institute for Occupational Safety and Health and the National Center for Environmental Health who were responsible for the studies conducted under the memorandum of understanding between DOE and the Department of Health and Human Services (HHS).

To determine how accessible and usable CEDR is for outside researchers we also (1) obtained authorization from DOE to become CEDR users and accessed and reviewed various files in the system and (2) interviewed CEDR users about their experiences with the system. We also discussed these issues with the officials on the NAS committee and at HHS mentioned above.

We performed our review between February 1994 and May 1995 in accordance with generally accepted government auditing standards. We discussed the facts presented in this report with CEDR program officials at the Lawrence Berkeley Laboratory and officials at DOE headquarters and incorporated their views where appropriate. As requested, we also provided a draft of this report to DOE for comment. Although DOE did not formally respond within the 15 days allowed, the views expressed by the Acting Director of the Office of Epidemiology and Health Surveillance and our evaluation of them are presented in the Agency Comments section of this report.

Major Contributors to This Report

Resources,
Community, and
Economic
Development
Division, Washington,
D.C.

Bernice Steinhardt, Associate Director
Alice G. Feldesman, Assistant Director
Carolyn K. McGowan, Assignment Manager

San Francisco
Regional Office

Margie K. Shields, Regional Management Representative
Randolph D. Jones, Evaluator-in-Charge
Daniel F. Alspaugh, Evaluator
Jonathan M. Silverman, Communications Analyst

Ordering Information

The first copy of each GAO report and testimony is free. Additional copies are \$2 each. Orders should be sent to the following address, accompanied by a check or money order made out to the Superintendent of Documents, when necessary. Orders for 100 or more copies to be mailed to a single address are discounted 25 percent.

Orders by mail:

U.S. General Accounting Office
P.O. Box 6015
Gaithersburg, MD 20884-6015

or visit:

Room 1100
700 4th St. NW (corner of 4th and G Sts. NW)
U.S. General Accounting Office
Washington, DC

Orders may also be placed by calling (202) 512-6000 or by using fax number (301) 258-4066, or TDD (301) 413-0006.

Each day, GAO issues a list of newly available reports and testimony. To receive facsimile copies of the daily list or any list from the past 30 days, please call (301) 258-4097 using a touchtone phone. A recorded menu will provide information on how to obtain these lists.

**United States
General Accounting Office
Washington, D.C. 20548-0001**

**Bulk Mail
Postage & Fees Paid
GAO
Permit No. G100**

**Official Business
Penalty for Private Use \$300**

Address Correction Requested



