



A Comprehensive Epidemiologic Data Resource

S. Jay Olshansky, Ph.D.,* R. Gary Williams, Ph.D.†

A Comprehensive Epidemiologic Data Resource (CEDR) of the health records of all present and former employees of the Department of Energy (DOE), its predecessor agencies, and their contractors is in the process of being established. Such a data base, as first proposed, would have enabled any researcher the opportunity to use DOE and contractor employee data to analyze, among other issues, the effects of low levels of radiation on the health of those exposed. However, insufficient funding and efforts to limit the definition of what data are to be included now threaten to reduce the breadth and usefulness of this data base. Vigilance and action by the concerned expert community and the general public are needed to ensure that CEDR is developed along the lines initially envisioned. [PSRQ 1991;1:145-156]

The Department of Energy (DOE) has been engaged in both legal and political battles since the Manhattan Project about access to health records on DOE employees and employees of DOE contractors. These records have never been available to researchers who did not have funding or approval from DOE. The DOE has argued until recently that these data should be maintained and analyzed by DOE and its contract epidemiologists because 1) the

privacy of the employees needed to be protected, 2) mortality data extracted from the death certificates could not be released to outside researchers without written permission from the individual states from which the certificates were purchased, 3) some contractors for DOE claim that records used for health studies are private corporate records that DOE and others have no right to use, and 4) DOE's contract epidemiologists claimed the right to publish from the data before releasing them to outside researchers. These four factors combined resulted in the preclusion of researchers outside the DOE system from analyzing the health effects of working at DOE, predecessor agencies, and at contractor facilities.

The Three Mile Island Public Health Fund brought the issue of access to DOE health records

0051-2438/1991/0103-0145\$03.00/0

© 1991 Physicians for Social Responsibility

* SJO is a Research Associate, Department of Medicine, The University of Chicago Pritzker School of Medicine, Chicago, Illinois, and a scientist, Argonne National Laboratory, Argonne, Illinois. Address all correspondence and requests for reprints to S. Jay Olshansky, Ph.D., Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439

† RGW is a scientist, Argonne National Laboratory, Argonne, Illinois

to the attention of the general public when it initiated a lawsuit against the DOE in pursuit of health and mortality data. In response to the lawsuit and the renewed public concerns about DOE's ability to police itself, the Secretary of the Department of Energy (Admiral Watkins) attempted to restore public confidence in DOE's epidemiology program. One of several actions taken by Admiral Watkins was a promise to create a Comprehensive Epidemiologic Data Resource (CEDR), a data base for public use that would contain health records on all DOE, predecessor agency, and contractor workers [1]. Although some within the DOE question the utility of such a data base, recently published studies on the relationship between exposure to low levels of radiation and health make the creation of this data base and its release to outside researchers even more important [2,3]. The CEDR initiative also represents a major change in the DOE's thinking regarding issues of health and safety, in that CEDR constitutes a mechanism to share data on worker health effects with researchers outside of the DOE.

The idea of centralizing DOE's epidemiologic data is not new. It was first suggested and attempted by Dr. Mancuso from the University of Pittsburgh in 1964 under a contract from the Atomic Energy Commission (AEC) [4]. From the vantage point of the 1990s, scientists and political officials are unanimous in the opinion that comprehensive health studies should have been a fundamental part of worker and off-site health and safety in the early years of the AEC. Data forms could have been developed and used consistently across sites, and longitudinal prospective health and mortality studies should have been carried out in a timely manner. Had this been done most of the current effort to establish a CEDR within the DOE would not be necessary.

Here we describe the latest effort by the DOE to put together, in one place, the data that are necessary to conduct retrospective and plan prospective epidemiologic studies of occupational safety and health of DOE and contractor employees in order to assess the more general issue of the relation between the exposure to low levels of radiation and hazardous chemicals and their long-term effects on human health. Included in this description is a review and summary of the relevant findings and recommendations of the Secretarial Panel for the Evaluation

of Epidemiologic Research Activities (SPEERA), the National Academy of Sciences (NAS) Committee on Radiation Epidemiological Research Programs, and the General Accounting Office Report to the Committee on Governmental Affairs. We also describe DOE's effort to standardize the collection of employee data that are routinely used for epidemiologic studies, and the mechanism that will allow access to these and other data by researchers outside of DOE. It should be made clear that the epidemiology program within the DOE is changing, and the methods for centralizing the relevant data are evolving as this article is being written. The description in this paper of the proposed content of CEDR, the methods for compiling these data into one central location, and data access issues are current as of July 23, 1991.

If completed in its originally proposed form, CEDR would become the most valuable data base available for assessing the more narrowly focused question of the health effects of the production of nuclear weapons and energy and the more general question of the health effects of long-term exposure to low levels of radiation. This data base would likely represent the basis on which research of this kind would be conducted internally at DOE, at the Centers for Disease Control (CDC), and at universities and research institutions throughout the world. Furthermore, other studies of interest could result from the creation of CEDR that would be of interest to the health care industry, the Occupational Safety and Health Administration, and to the employees and labor unions affected by the production of nuclear weapons and energy.

While the political will to create CEDR as originally planned may still exist in the DOE, developments within the past year indicate that CEDR, as originally planned, is in jeopardy. Specific problems that have arisen, to be discussed here, include 1) the lack of funding for CEDR, 2) the failure to meet deadlines for the installation of data into CEDR, 3) delays by the Office of Epidemiology and Health Surveillance (OEHS), the DOE office responsible for CEDR, in getting states to agree to allow cause-of-death data to be used, 4) the decision by OEHS to eliminate surveys of existing data, 5) the lack of a spirit of cooperation with the CDC, and 6) the lack of a presence of an established senior scientist with a long-standing track record of independence to

head the DOE's epidemiology program. As it currently stands, CEDR runs the risk of serving as nothing more than a repository for the analytical data files (and studies) that have already been conducted by DOE contractor epidemiologists, with little opportunity for outside researchers to test their own research hypotheses by modifying the data files on which such studies have been conducted in the past.

BACKGROUND

In the early 1960s the AEC set out to determine whether the records that were used to keep track of their employees could be used to conduct long-term follow-up mortality studies. Pilot studies to determine the feasibility of using worker records for health and mortality studies were conducted for the AEC by Dr. Mancuso and his assistants at the University of Pittsburgh from 1964 to 1969, and in a series of subsequently published internal reports it was demonstrated that records collected for routine purposes could be used to conduct longitudinal studies on employee health [4].

Data used in these and most of the subsequent studies conducted by the AEC, and later the DOE, were not originally collected for the purpose of performing epidemiologic research. Instead, they were collected as routine employer records for the purpose of following the job status of their employees and for addressing health and safety concerns that were of immediate interest to DOE and its contractors. Since most of the workers at weapons production facilities worked for individual nongovernment contractors, there were no standardized methods of collecting data on employees either within a given DOE facility or contractor site across time, or between facilities during any single time period. Thus, much of the detective work in epidemiology that is commonly associated with the search for confounding variables and the more elusive issue of causality was preceded in this case by the construction of usable analytical data files from the myriad of available information, much of which would be considered useless from an epidemiologist's perspective. Furthermore, the complexity of identifying data sources, verifying accuracy, tracking employees through time, linking employee records, and solving inconsistencies in the data have meant both slow progress in conducting research and the

requirement of numerous judgments about the data by the scientists who have been engaged in this work.

POLITICAL AND LOGISTIC ISSUES IN ESTABLISHING A CEDR

Consolidation of DOE Data

The most recent effort to consolidate the DOE data is a consequence of at least three factors: 1) the request from the Three Mile Island Public Health Fund for DOE data related to health and mortality, 2) DOE's inability (or unwillingness in the minds of some) to turn over the data to outside researchers, and 3) the environmental and organizational reforms initiated by Admiral Watkins [1]. Although the political will may exist within the DOE to carry out some of these reforms, wading through the departmental bureaucracy to see them to completion has been, and will continue to be, difficult. With respect to CEDR, examples of bureaucratic obstacles abound. Some DOE facilities allow contractor epidemiologists access to medical records but not personnel security questionnaires, while the reverse is true at other facilities. The DOE Systems of Records was revised in 1989 to allow data on current contractor employees to be used for epidemiologic studies. Funding for the creation of CEDR in fiscal year 1990 was drawn from existing programmatic funds with no new funding available from Congress, and Congress has yet to allocate more funds to the DOE explicitly for this purpose. Furthermore, funding for new review committees within the epidemiology program has been drawn from existing funds within the Office of Health and Environmental Research.

With regard to the data the DOE has been collecting for years, there are still lingering questions of ownership, problems associated with protecting confidentiality, and questions about access and release. An example of one of these problems is that the DOE has taken the position (on the basis of written agreements with some states) that information on cause of death from state death certificates belongs to the states and not to the DOE. This means that cause-of-death data cannot be released without specific permission from each individual state from whom the death certificate was purchased. Agreements with each of the 50 states will be necessary before information from death certificates may be included in the data base. To date, 43 states have

given their permission to include death certificate information in a CEDR. Without death certificate information from all 50 states, the DOE epidemiologic studies cannot be replicated and the findings independently verified. While as of May 1991, 10 researchers have requested access to the interim data system, those requesting data tapes have asked that the DOE wait to grant access until data from all states are available. It appears that those states that have not given permission to include death certificate data into CEDR are those with more complicated procedures for granting such permission. None of the states withholding permission are hosts for DOE weapons production facilities.

Oversight and Review of Research Program

In the past it has been suggested that the entire health and safety component of the DOE be detached from it and moved to another agency such as the Department of Health and Human Services (HHS) or the CDC. Bills have been presented in Congress to initiate such a transfer [5-8]. The response of Admiral Watkins to this suggestion has been that such a detachment would represent "a very serious violation of the principle of ownership of these important functions . . ." [9]. In spite of the Secretary's protests, the SPEERA panel recommended that portions of DOE's epidemiology program be transferred to HHS, a transfer that was just recently completed. The CDC will manage the program for HHS. Two advisory committees will be set up to negotiate and direct the DOE analytic epidemiology program. One will be a DOE Environmental Health and Safety Advisory Committee and the other will be a HHS Advisory Committee. The NAS has recommended that with regard to this transfer and CEDR "a formal, clear mechanism is required for coordination, cooperation, and exchange between DOE and HHS in the development of CEDR" [10]. The NAS went on to point out that the SPEERA recommendations to have HHS control the DOE's analytic epidemiology research "will almost certainly generate friction between the two agencies, the one (DOE) responsible for data generation, quality control, and descriptive epidemiology, and the other (HHS) responsible for analytic epidemiology" [10]. In fact, the NAS was prophetic in stating that "trouble lies ahead, unless preventive measures are taken immediately" [11]. No "preventive measures" have

been taken according to the CDC and, because of the unclear language in the Memorandum of Understanding between DOE and HHS, there are differing views about which organization is in charge and which organization has which responsibilities. One example of this is that OEHS currently does not plan to collect any data for CEDR because DOE views data collection as something tied to research directed by CDC.

While new advisory groups are being formed both inside and outside the DOE, it is important to note that the DOE-funded epidemiology programs at the existing contractor sites, as well as the overall epidemiology program, have had external scientific review panels that are composed largely of research scientists who work outside the DOE. As proper scientific protocol would dictate, the body of research based on the DOE's data and produced by DOE contractors has been published in the open scientific literature for critical review. A selected bibliography recently published by the DOE lists over 340 publications that present the research results obtained by scientists supported by the DOE's Epidemiologic Research Program [11]. The NAS committee "found the work by the DOE epidemiological research contractors at Hanford Environmental Health Foundation, Oak Ridge Associated Universities, and Los Alamos National Laboratories, to be highly credible" [10]. These contractors are, and have been, the core of the DOE efforts in epidemiology. Although the DOE has been charged with controlling the epidemiologic findings or the publication of these findings, the single DOE epidemiologist who directed the program had a standing informal policy not to review scientific papers produced by DOE contractors until they were accepted for publication in peer-reviewed journals. In the past year additional epidemiologists have been hired by the DOE and the OEHS has been created under the Office of Health. To reestablish public and scientific credibility in the DOE's epidemiology program, and to help ensure the independence of DOE and contractor researchers, such a hands-off policy will need to be formalized within this new organization. Even a formalized policy of independence between DOE and its contract researchers could leave the internal investigators vulnerable to censure and without an independent route of appeal. To the extent that the epidemiology program within the DOE is under the

control of established senior scientists with long-standing track records of independence, the scientific integrity of the program will be enhanced. To this end the NAS has recommended that an ombudsman office be created within the DOE to establish an independent environment—a position that should be filled by “a person with unquestionable scientific credentials and professional integrity, who would report directly to the secretary” [10]. To date such a person has not been appointed, nor has an ombudsman office been established within the epidemiology program at the DOE.

The key ingredients that have been missing from the DOE’s epidemiology program, however, are that outside researchers have not had the opportunity to 1) replicate the studies that have already been conducted by DOE contractors, 2) analyze the same data by using different methods of analysis, and 3) study different subcohorts and use different variables.

Access to DOE Data

The presence of CEDR, which will serve as the principal source of data for the DOE as well as outside researchers, was originally designed to eliminate these problems by allowing outsiders access to the same data as the DOE’s epidemiologists (assuming that all of the states will allow the inclusion of data from their death certificates in CEDR). This design would allow researchers the opportunity to make their own decisions with regard to the study populations and methods of analysis. The original plan regarding the content of, and access to, CEDR was that virtually all of the data (including the analytical data files, intermediate files, and raw hard-copy data) would be made available to outside researchers [12].

The OEHS recently concluded, however, that it will not conduct a comprehensive survey of the available data (a virtual requirement if all relevant data for epidemiologic studies are to be identified and made available) because “it would be more appropriate for the researchers managing the studies [e.g., HHS] to collect the pertinent data since they will determine the scope of the research” [13]. A working group composed of contractor epidemiologists, the authors of this paper, and outside researchers was unanimous in the recommendation that facility surveys be conducted at each DOE and DOE

predecessor agency facility [14]. This working group was disbanded by the OEHS at the same meeting at which the final recommendations were made. Furthermore, the Memorandum of Understanding between DOE and HHS states that “DOE will solicit input from HHS on the development and maintenance of . . . CEDR and the selection of data to include in CEDR” [13]. If the DOE does not survey its own facilities, outside researchers will never know what constitutes the universe of data that may be available for study. It may be that CDC will find it prudent to survey DOE facilities as a first step in overseeing DOE’s analytic epidemiology research.

An issue with CEDR has been defining how one actually goes about gaining access to the DOE data. Early in the planning process, DOE officials wanted the NAS to serve as a filter through which proposals from researchers outside of DOE would pass before these data will be made available. The NAS recommended a more open system, stating specifically that “data collected by DOE related to health and safety of citizens . . . whether DOE employees, DOE contractors’ employees, or members of the general public living near DOE facilities . . . should be widely available to interested parties.” [10]. The NAS also specifically stated that “there should be no test of qualification of persons seeking access to data on the health and safety of people working in or living near DOE facilities” [10]. The current policy, according to the OEHS representative in charge of CEDR, is that the researcher need only sign a disclosure form indicating that the data are to be used for scientific purposes, and that he or she would not try to identify individuals by combining information in the data set. Once the disclosure form is signed, researchers will be sent either data tapes or be provided with computer access to query the data base and download data.

Scope of Proposed Research Program

Finally, it should be emphasized that the research interests of the epidemiology program at DOE could be expanded considerably from that which has existed since its inception 25 years ago. The limits in the scope of the program in the past were the results of a combination of factors, including: 1) short-term health and safety concerns took precedence (in terms of funding priorities) over long-term epidemiologic studies, 2) some of the more important studies that

were begun early in the program required long-term follow up, 3) funding for the program diminished over time, and 4) the lack of a detailed long-term research plan. Although the DOE has agreed to expand the scope of its epidemiology program considerably beyond current limits, it appears at this time that the CDC will conduct most of the long-term epidemiologic studies using DOE data. As a result, the epidemiology program in the DOE may begin to focus most of its attention on the development of a health surveillance program.

Although previous outside reviews of DOE's epidemiology program recommended a considerable expansion of the program [15,16], these recommendations were not accompanied by the funding required to carry out such an expanded research agenda. The preliminary SPEERA report also recommended a considerable expansion of the program, both in terms of funding and personnel [17]. Admiral Watkins has pledged that "the conclusions from this report will be followed" [9]. This pledge will allow the scope of the program to be broadened to include an assessment of the more general issue of the relation between exposure to low levels of radiation, toxic substances, and their effects on human health. However, the scope of the external research program at HHS and the internal research program at the DOE will still depend on the content and availability of data. Questions continue to persist regarding the commitment of the DOE to identify and provide access to data that might be useful for future epidemiologic studies.

The data that are required to address these more general issues of the relation between exposure to low levels of radiation, toxic substances, and their effects on human health should be drawn not only from exposures that occurred among DOE employees and the employees of its contractors, but also from a wide range of outside sources. Included among the sources that at one time were being considered for inclusion into CEDR were data on workers at commercial and other kinds of nuclear power plants, naval shipyard workers, radium dial painters, military personnel involved in the testing of the atomic bomb, and uranium miners. In these early phases of the CEDR discussions, the DOE had proposed putting into one place the most comprehensive source of data in the world on the effects of nuclear weapons production and the energy indus-

try on human health, and it had proposed to expand the scope of the epidemiology program to address these issues. It now appears likely that a much more modest version of CEDR will be created, where outside researchers will be given access only to data already collected, and in most cases already analyzed by DOE contractors. What is different now, and what represents an improvement over earlier procedures, is that the CDC will be in a position to allow outside researchers the opportunity to compete for research funds to analyze these data. If, however, the data required for a research project were not already in CEDR, data collection will be the responsibility of the researcher. Specifically, this means that outside researchers will somehow have to obtain knowledge about the types and availability of data at sites for which such knowledge is not currently known, and they will then have to travel to these sites to collect the data in person. The formulation of research hypotheses without knowledge of the details of available data is not possible, nor is it likely that outsiders would be given access to DOE or contractor sites, even if they are aware of the presence of data that might be useful for epidemiologic studies.

EXISTING DATA SOURCES

In order to estimate the amount of time and effort required to establish CEDR, it is necessary to determine the location, form, and status of all of DOE's data (this term is meant to be inclusive of all data on DOE employees and its contractors' employees without regard to the question of ownership) that may be used for epidemiologic studies. The total number of people who are currently working, or who in the past have worked for the DOE or its contractors is estimated to be 600,000. There is, however, a great deal of uncertainty about this number because 1) DOE employees and its contractors' employees represent a very large currently employed population where some transition is expected, 2) there are some inaccuracies in the historical employment records, and 3) detailed information is not available on all employees. Moreover, this uncertainty may never be eliminated unless the DOE or CDC conducts a survey of the available sources of data to determine what might be useful for epidemiologic studies.

Of the estimated 600,000 people who were or are now associated with DOE and its contractors, 250,000 have some information on them in a computerized data base currently held by the DOE. Some of the data on these individuals are limited to name and social security number. The various kinds of data that are available on DOE employees have been defined by the Systems of Records notices published by the DOE in accordance with the Privacy Act of 1974 [18]. Among the main categories of records that have been used for epidemiologic studies are general employment records, personnel radiation exposure records, personnel security questionnaires, occupational and industrial accident records, and personnel medical records. Although these categories of records are required at all DOE facilities, not all DOE contractor facilities are required to maintain these same records [18]. Moreover, there has never been a standard format for DOE or its contractors regarding the collection of each of these kinds of records. As a result, there has been a great deal of variation in both the kinds of data that are available on DOE employees and its contractors, and the forms that have been used to collect each major category of data both between facilities at any single time period and within a given facility across time. In the case of dosimetry, for example, a number of different methods have been used both between and within facilities to transform dosimeter readings into a measure of exposure. This lack of consistency in the content and form of the data represents one of the major obstacles in conducting epidemiologic research on the employees of the DOE and its contractors.

AVAILABLE DATA AND DATA FILES

Because various types of data files exist on subcohorts of the 250,000 employees for whom some information is computerized, it is necessary to provide an explanation of the types of data files and the differences between raw data, intermediate data files, and analytical data files. The discussion that follows is based, in part, on an unpublished short summary of how these terms have been used by researchers at Oak Ridge Associated Universities [19]. It should be noted that the data are not consolidated or structured in exactly the same way at the other two facilities where the remainder of the DOE's epidemiology work is conducted.

Raw Data

Raw data are composed largely of hard copy forms that virtually everyone who has had a job has filled out, such as medical history forms, job history forms from personnel files, personnel security questionnaires for federal employees who require access to secured areas, etc. In addition to these examples, a relatively large number of other forms of raw data are available on DOE employees, such as death certificates, dosimeter readings, estimates of exposure, and others.

Although most of the raw data are in the form of hard copies such as the original paper on which the information was provided, photocopies, microfiche, microfilm, etc., some of the raw data may be computerized. For example, in the past film badges were read by a person who recorded the readings by hand on a piece of paper. Today the film badges are read automatically by a computer and the information is stored directly on computer files. In this case both the hard copy files that were used several decades ago and today's computer generated files are considered raw data. It should be emphasized that massive amounts of raw data are available on many individuals, most of which is of little or no use for epidemiologic studies, and there is a great deal of variation (across time and sites) in the degree to which the data are readily available, documented, and computerized.

When a researcher decides to conduct a study, an oversimplified version of what occurs is that the raw data are then transferred manually (in the case of hard copy) from their original form to computer files. It is at this point that decisions are made on exactly what data to include and exclude from the study. This inclusion/exclusion decision has been one of the points of contention between DOE epidemiologists and potential outside users. For example, Dr. Alice Stewart (a researcher whose studies are funded in part by the Three Mile Island Public Health Fund) argues that the frequency with which individuals are monitored for exposure to radiation may be used as a proxy for the level of danger associated with a given job. DOE epidemiologists have argued that such a variable is of little value. If DOE epidemiologists do not add the "frequency of monitoring" variable into their variable list, then it is not possible for Dr. Stewart to test her hypothesis, even if given the complete analytical

data files compiled by DOE epidemiologists. It is for this reason that outside researchers are being consulted now regarding the future content of CEDR.

It is also at this juncture that some of the raw data have to be interpreted, or accompanied by detailed documentation, in order for the data to carry intrinsic meaning to the researchers. For example, the transformation of a dosimeter reading to dose has been calculated differently both between and within sites. Typically, the methods used to calculate dose from exposure have varied because there have been one or two key individuals at each site who take on this responsibility, and these individuals seldom chose the same methods or assumptions. This issue about variation in the methods used to calculate dose from exposure has been, and will continue to be, a problem associated with these data. In the absence of first-hand knowledge as to how these data should be interpreted, or at least translated into a standard format that may be used for comparison purposes, they are of little use.

Intermediate Data Files

Once the researcher begins to process (e.g., alter) the raw data during the process of standardizing the variables, reformatting, or correcting for inconsistencies or any one of a number of other errors, the raw files are then referred to as intermediate data files (IDFs). Examples of a transformation from raw data to an intermediate form would be the standardization of the measure of exposure across time at a given facility, or the use of a nosologist to recode cause of death from the death certificates. The IDFs then undergo continuous processing to remove errors. It takes increasingly larger amounts of time and money to process the raw data into a form that approaches perfection, and it is safe to say that the attainment of perfect IDFs is not feasible given limited time and human resources. Besides, there is probably not much to be gained (in terms of accuracy) as the data are changed from, for example, 99% perfect to 100% perfect. It is important to remember, however, that the IDFs are constantly changing.

The processing of the raw data into its intermediate format is also a point of contention between DOE epidemiologists and outside researchers. For example, aggregating weekly or monthly dosimeter readings to yearly averages would make it impos-

sible for researchers to test some of their hypotheses. If exposures were provided only as yearly averages, for instance, it would not be possible to study the possible health effects of higher exposures over shorter time periods in contrast to lower exposures over longer time periods.

Analytical Data Files

Analytical data files (ADFs) are subsets of variables from the IDFs based on the specific research questions that are being addressed. Since ADFs are drawn from ephemeral IDFs, it is possible to replicate the ADF file exactly only if the IDF file is frozen (i.e., a copy is made) at the same time in which the ADF file is made. Since this has not been a standard practice for all DOE epidemiologists, it is not always possible to replicate exactly the ADFs that have already been created and analyzed by DOE epidemiologists. The reason for this is that data on more recently deceased workers are continuously added to the IDFs as they become available. However, the ADFs have been kept in their original form and are available on magnetic tape.

The ADFs are then processed to eliminate additional errors, and they are frequently supplemented with information from raw data that were not originally included in the IDFs. For example, an individual's vital status or worker identification number is verified with information from the personnel security questionnaires (a source of information that is not ordinarily used for the IDFs). Once the researcher is comfortable with the quality of the data in the file, the ADFs are analyzed. The methods used to analyze the ADFs represent the final point of contention between DOE epidemiologists and outside researchers [20]. A complete list of ADFs has been published [11].

SURVEY OF DOE DATA SOURCES AND FILES

According to the original plan, beginning in January, 1991 researchers at Oak Ridge Associated Universities, Los Alamos National Laboratories, Hanford, and Argonne National Laboratory were to have started conducting surveys at the currently active DOE facilities and contractor sites. These surveys were to identify the types of data that are available, the location of records, data tapes and forms, and the individuals (e.g., current and former employees) who are knowledgeable about these

data. Additionally, samples of all of the forms that have been used to collect information on employees (for epidemiological purposes) from the early 1940s to the present (inclusive of the Manhattan Engineer District, AEC, and DOE) were to have been collected during the site survey. The information from both the survey and the forms was to have been used to create a three-dimensional matrix of variables (i.e., fields on each form in which data should be encoded) by site and year. From this matrix it would have been possible to identify not only what variables are located on each form, for each site, for any given year, but also which forms may be the most useful for obtaining information on specific variables (under conditions when data on a single variable are present in more than one location). Additionally, the location and availability of the data for each site could have been determined from this survey. This identification of the universe of available data would have made it possible for researchers from both within and outside the DOE to develop research questions and hypotheses that have not been developed previously, and this in turn could have been used to determine which data should be placed into CEDR. Conversely, researchers with a specific research question would have been able to determine whether the data are available to address their question. It is therefore considered crucial at this time that researchers from both inside and outside the DOE provide input regarding what variables should be incorporated into CEDR. Although OEHS has made the decision not to proceed with the survey, CDC has expressed interest in making sure this work is initiated at a later date.

CEDR WORKING GROUPS

It was recognized early on in the planning for CEDR that the kinds of variables that would be placed into this data base would cross many disciplines. One way of addressing the complex problems associated with the data base was to create working groups along disciplinary lines. These working groups, composed of experts within the DOE and its contractor organizations, as well as experts from outside the DOE, would then identify the unique technical problems associated with the measurement, calibration, standardization, and interpretation of the variables within each discipline. It was also suggested that working groups be organized to

carry out the site surveys (described above), create the data base, and describe in detail the statistical tools that may be needed for researchers, both within and outside the Department, to evaluate the health effects of exposure to low levels of radiation.

Working groups on vital status, demography, site surveys, and information systems development were formed in early 1990 and have subsequently submitted their conclusions and recommendations to DOE. A fifth working group was to address issues on dosimetry, but this group has yet to meet. Other working groups may still be formed in the following areas: data reconstruction, nonradiologic exposures, statistical analysis requirements, and facilities and processes.

One of the major problems that is to be addressed by the dosimetry working group is the determination of whether it is possible to standardize the calculation of dose from an exposure to radiation across sites and time periods. The vital status working group has already provided detailed recommendations to the OEHS about what information should be extracted from death certificates and the format of the data bases (of varying detail) to be included in CEDR. The demography working group provided detailed suggestions about what demographic variables to include in CEDR and how OEHS should standardize the collection of these kinds of data in the future. For example, it was suggested that the OEHS conduct a quinquennial census of the entire currently active population of DOE and contractor employees for the purpose of standardizing data collection. Additionally, each of the working groups had, as one of its major goals, the creation of an accepted discipline-specific standardized format for collecting data in the future. The intent here is to alter data collection practices in the DOE and among its contractors in the future so that assessments of worker health and safety will become standardized and routine.

CURRENT TIME TABLE FOR INPUT OF DATA INTO CEDR

As of this writing, it has yet to be determined what data will be placed into CEDR, and the priorities that might be given to data on selected subcohorts of the population. It is anticipated that final decisions on these issues will be made by the National Academy of Sciences, the CDC, and the DOE.

A preliminary plan for the inclusion of data in CEDR was developed in 1989.

According to this preliminary plan (Fig), data on 70,000 employees were installed into CEDR as of February 1990. These data contain variables that conform to a protocol developed by the International Association for Research on Cancer (IARC). Unlike the data supplied in IARC, however, the data on the 70,000 employees now installed in CEDR do not contain cause-of-death information from all states. This issue about the inclusion of death certificate information in the CEDR files has yet to be resolved, although the vital status working group recommended that this problem be dealt with expeditiously. By September 1990, data on an additional 180,000 employees were to have been added by following an expanded version of the IARC protocol. It was anticipated that data on a total of 400,000 employees would be installed into CEDR by September 1992, and data on all 600,000 employees are expected to be available in CEDR by September 1995. It should be emphasized that this was an ambitious preliminary plan that depended on decisions being made in a timely manner about what variables to include into the data base other than those specified by the IARC protocol and the timing with which those data would be installed. As of this writing, the CEDR data base contains some analytical data files and data on the 70,000 employees

whose records were installed in February 1990. The original plan is now 18 months behind schedule.

INVOLVEMENT OF INDEPENDENT RESEARCHERS

One of the major changes associated with the reorientation of the DOE's epidemiology program is the attention given to potential outside researchers. This consideration of outside researchers is particularly important at this time because their input will help to shape the design of the data base, variable specifications, user interface, documentation, and the methods to be adopted by the DOE for providing updates and progress reports. There have been a number of steps taken to ensure that outside researchers have input into the current program.

1. *The establishment of SPEERA* This advisory group has actively sought comments from interested parties outside the DOE and has included union groups, the Three Mile Island Public Health Fund, and Physicians for Social Responsibility, among others. The final report from this group has been produced, and all documents were available in the DOE Public Reading Room (they have been removed). (These documents include the minutes from the CEDR Working Group Meetings and their recommendations.)
2. *The presence of the NAS Commission on Life Sciences Committee on DOE Radiation Effects Research Programs* This committee represents a broad external source of input into CEDR activities because it is made up mostly of university scientists. This standing committee is in a position to provide input into the design, operation, and future directions of CEDR.
3. *The encouragement of CEDR outreach activities* As part of the public involvement program being created specifically in regard to CEDR, individual scientists and research organizations that have expressed interest in the epidemiology program were to have been contacted by the spring of 1990 for suggestions and comments regarding the content and structure of the CEDR data base. They are also to be kept informed of decisions and progress regarding CEDR and the DOE's epidemiology program on a regular basis. The OEHHS recently hired an individual to conduct these activities although no work had been done as of May 1991.
4. *A role for professional associations and journals.* Presentations have been made at the American Public Health Association meetings held in Chicago in 1989, and future presentations were to have been made at appropriate professional meetings to inform scientists of the development of CEDR. In addition, articles describing the devel-

INSTALLATION OF CEDR

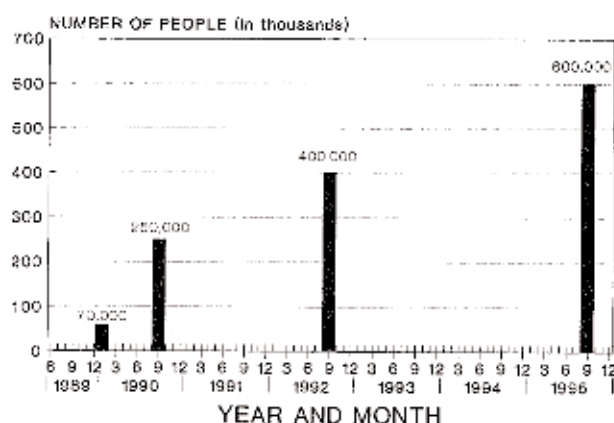


FIGURE Bar graph of original plan for installation of data into a Comprehensive Epidemiologic Data Resource (CEDR). As of June 1, 1991, only data on the first group of 70,000 workers have been installed, some without death certificate information.

opment of CEDR (similar to this one) are supposed to be sent to major journals and popular scientific magazines to inform potentially interested scientists of the program and to solicit input. Little progress, however, has been made here.

5. A fifth step, a CEDR quarterly newsletter was to be published and sent to all interested parties. The newsletter was intended to provide information on existing epidemiologic data, site survey work, CEDR variables, data base design, and general progress on the project. Letters from outsiders were to have been considered for publication in the newsletter. The first issue was never approved for publication by the OEHS.

CONCLUSIONS

The DOE originally proposed creation of the most comprehensive data base in the world to address issues associated with exposure to radiation and human health. This data base would have served as the focus of research by epidemiologists employed by the DOE, as well as for outside scientists. In addition, the data base had the potential to serve as the primary mechanism for evaluating data from the DOE's developing health surveillance program. The DOE's effort to create this data base has been supported by the NAS and a panel of outside experts appointed by the DOE to evaluate its epidemiology program (the SPEERA panel).

Since CEDR was originally planned, the OEHS has consistently chipped away at the breadth and detail of the data base so that the word comprehensive is no longer appropriate in the title of this enterprise. The most glaring example of this chipping away is the decision by the OEHS not to survey all of the available data to determine what might be useful for epidemiologic research—a decision based on the underlying premise by people in that office that everything that is worth studying has already been collected and analyzed. It also appears likely that if the DOE eventually develops a health surveillance program, the data resulting from such an effort may not be included in CEDR and therefore may not be available for outside researchers to analyze. Given that CEDR was originally designed to be a public use data base, the possible exclusion of health surveillance data from CEDR certainly brings into question the intentions of the newly formed office controlling CEDR and health surveillance. Furthermore, with respect to the creation of CEDR, there have been delays by the OEHS in fulfilling

the recommendations of SPEERA, the NAS, and the CEDR working groups—delays that could jeopardize the breadth and completeness of the data base.

The consequence of these problems is that CEDR, as it now stands, appears to be nothing more than a repository for the data files that have already been created by DOE contractor epidemiologists. This means that some of the key ingredients that have been missing from the DOE's epidemiology program in the past, ingredients that are necessary to re-establish credibility, such as the ability of outside researchers to choose their own variables and subcohorts of the population for study, are still missing. While it is not too late to save CEDR and make it truly comprehensive, such an effort will require that the political will and resources that exist at the top of the DOE filter down quickly to those responsible for the creation of the data base. To the extent that the concerned expert community external to the DOE and the general public can be alerted to the erosion that is taking place, pressure from these constituencies may also be helpful in restoring CEDR to a working status compatible with the initial vision of a world-class data base destined to allow all researchers the opportunity to study the health effects of the production of nuclear weapons and nuclear energy.

REFERENCES

1. Watkins JD. Setting the new DOE course. Secretary of Energy Notice. Washington, DC: U.S. Department of Energy, 1989. Department of Energy publication SEN-11-89.
2. Gardner MJ, Snee MP, Hall AJ, Powell CA, Downes S, Tenell JD. Results of case-control study of leukaemia and lymphoma among young people near Sellafield nuclear plant in West Cumbria. *Br Med J* 1990;300(6722):423-429.
3. Wing S, Shy CM, Wood JL, Wolf S, Cragle DL, Frome EL. Mortality among workers at Oak Ridge National Laboratory: evidence of radiation effects in follow-up through 1984. *JAMA* 1991;265(11):1397-1402.
4. Mancuso TP, Sanders BS, Brodsky A. Feasibility study of the correlation of lifetime health and mortality experience of AEC and AEC contractor employees with occupational radiation exposure. Progress Report No. 1. Pittsburgh, PA: Department of Occupational Health, Graduate School of Public Health, University of Pittsburgh, AEC Contract No. AT(30-1)-3394, 1965.
5. U.S. Congress. House. Committees on Armed Services and Energy and Commerce. Nuclear Weapons Production Health and Safety Act. Hearings on House bill 1643. March 23, 1989.
6. U.S. Congress. Senate. Committee on Armed Services. Radiation and Toxic Substances Research Reorganization Act. Hearings on Senate bill 3212. August 4, 1989.
7. U.S. Congress. Senate. Committee on Energy and Natural Resources. Radiation Research Reorganization Act. Hearings on Senate bill 972. May 11, 1989.

- 8 U.S. Congress. Senate. Committee on Energy and Natural Resources. DOE Nuclear Safety and Environment Act Hearings on Senate bill 1304. July 12, 1989.
- 9 Watkins JD. Testimony of Secretary of Energy Watkins before the Secretarial Panel for the Evaluation of Epidemiologic Research Activities. Washington, DC: September 12, 1989.
- 10 Providing access to epidemiological data. Washington, DC: National Academy of Sciences, 1990.
- 11 DOE Epidemiologic Research Program. Selected bibliography. Washington, DC: Epidemiologic Research Task Force, Office of Energy Research, 1990. Department of Energy publication DOE/ER-0437.
- 12 Goldsmith R. Presentation before the Secretarial Panel for the Evaluation of Epidemiologic Research Activities. Chicago, IL: October 27, 1989.
- 13 Nuclear health and safety: efforts to strengthen DOE's health and epidemiology program. Washington, DC: General Accounting Office 1991. General Accounting Office publication GAO/RCED-91-57.
- 14 Williams RC, Dupree B, Marsh G, et al. Report from the site survey working group to the Comprehensive Data Resource (CFDR) steering committee. Chicago, IL: Argonne National Laboratory, August 29, 1990.
- 15 Review of Department of Energy research on human health effects of low doses of ionizing radiation. A report to the Department of Energy from the committee on DOE research on health effects of ionizing radiation, Division of Medical Sciences, National Research Council, Washington, DC: National Academy of Sciences 1980.
- 16 Review of the Office of Health and Environmental Research Program. Epidemiology. Prepared by a Subcommittee of the Health and Environmental Research Advisory Committee (HERAC), Department of Energy. Washington, DC: Office of Energy Research, 1984.
- 17 Gebbie C. Final report to the Secretary of Energy. Secretarial Panel for the Evaluation of Epidemiologic Research Activities for the Department of Energy. Washington, DC: 1990.
- 18 Department of Energy. Privacy Act of 1974; Annual Publication of System Notices. Part II. Volume 47, no. 64, April 2, 1982. Washington, DC: Federal Register, 1982.
- 19 Cragle D. Definitions of data terms. Unpublished manuscript Oak Ridge Associated Universities, Medical Sciences Division, Center for Epidemiological Research, October 13, 1989.
- 20 Alvarez R. Occupational radiation health risks: folklore and fact. An assessment of the health and mortality studies of federal nuclear workers in the United States. Washington, DC: Environmental Policy Institute, 1985.