

2

Overview of CEDR Data

The CEDR Program provides a repository of data that have been generated for epidemiologic or environmental health studies funded by DOE. Most of these studies have been published in peer-reviewed journals; however, there are also data that have been used in student dissertations, presentations, and technical reports.

The data contained in CEDR are collected from various sources and researchers. The data cover different time spans; the number, type, and format of data included in studies vary considerably; and the studies may represent different groups of people or cohorts. For example, some data files will be limited to specific locations, some to specific exposures, some to certain groups of people, and some to morbidity or mortality information.

A large portion of the data file sets offered in CEDR relate to specific individuals. This is because epidemiologic studies often require the collection of detailed information for each individual in the study. Therefore, the data files developed by the original researchers who performed DOE's Worker and Mortality Studies usually contained personal identifiers as well as many other variables that, when used in combination, might uniquely identify an individual. As a public agency, DOE has responsibilities not only to promote openness by sharing its health data, but also to maintain rights of privacy of its current and former workers. To meet these needs, it was necessary to establish a data release policy for CEDR. Details of the release policy can be found in Section 3. Briefly, the policy is that personal identifiers, such as name and social security number, are replaced with a unique identification number before the individual data are loaded into CEDR. Personal dates, including the starting and ending

Types of CEDR Holdings

- ❖ DOE Worker Health and Mortality Studies
- ❖ Classic Radiation-Effects Studies
- ❖ Off-Site Dose Reconstruction Data
- ❖ Other Workplace and Community Studies
- ❖ Epidemiologic Surveillance and Records Inventory Reports

dates of employment, date of birth, and date of death, are truncated leaving either year only or year and month. In addition, users of these data are required to complete access request and confidentiality forms stating that the data are to be used only for research purposes.

There are other data file sets and reports in CEDR that are not linked to particular individuals. Since there are no individual privacy concerns with these data, they are presented in their entirety, and no confidentiality statements are required to use the data. Currently, this type of CEDR holding includes general descriptions and documentation for the DOE Worker Health and Mortality Studies; data that support off-site dose reconstructions, such as that for the Nevada Test Site (NTS); epidemiologic surveillance summary reports; and reports describing inventories of various types of records at several major DOE sites.

The following subsections describe the types of data that currently comprise the CEDR holdings.

Data from DOE Worker Health and Mortality Studies

DOE's Worker Health and Mortality studies began in the early 1960s, when the feasibility of using existing employee and facility records for conducting long-term follow-up studies was explored. Based on the promising findings from the feasibility study, a 5-year pilot study was initiated in 1964. The primary DOE sites selected for this pilot study were the Hanford Site in the State of Washington and the Oak Ridge Site, which consists of three major facilities, in the State of Tennessee. Other facilities operating at several smaller sites were also included. To obtain the information needed for a retrospective cohort study, considerable effort was expended on locating, copying, and

Major Centers Providing Data

Historically, three DOE epidemiologic research centers conducted Worker Health and Mortality Studies:

- ❖ Los Alamos National Laboratory
- ❖ Oak Ridge Associated Universities (now the Oak Ridge Institute for Science and Education)
- ❖ Pacific Northwest Laboratory and the Hanford Environmental Health Foundation

abstracting data from various plant records located on- and off-site. When it was determined that further processing and validation of these data would be necessary to ensure their proper use in epidemiologic studies, the Hanford cohort became the primary focus of the Health and Mortality Studies.

In the 1970s, the study was expanded, and the responsibility for collecting and updating data for subsequent long-term worker studies was assigned to three epidemiologic research groups located at Oak Ridge Associated Universities (ORAU), at the Los Alamos National Laboratory (LANL), and at Hanford, where staff from Pacific Northwest Laboratory (PNL) and Hanford Environmental Health Foundation (HEHF) formed a team.

ORAU assumed responsibility for studies of worker populations at the Oak Ridge Site, including the K-25 Facility (also known as the Gaseous Diffusion Plant), the Y-12 Facility, and the X-10 Facility (also known as the Oak Ridge National Laboratory or ORNL). In addition, ORAU was assigned the studies of workers at two gaseous diffusion plants in Paducah, Kentucky, and Portsmouth, Ohio; at the Feed Materials Production Center in Fernald, Ohio; at two uranium processing plants located at the Mallinckrodt Chemical Works in St. Louis and Weldon Springs, Missouri; and at the nuclear fuel production plant, called the Savannah River Site, in South Carolina.

ORAU prepared a master roster of all contractor employees at each of these facilities and retrieved many of the death certificates for the three epidemiologic research groups. Each research group then extracted and coded the cause-of-death information from the death certificates for use in their respective studies. Each worker was assigned a unique identifier number, which allows information pertaining to that worker to be linked together, regardless of the

Sites Where Cohort Mortality Studies Have Been Performed

- ❖ LANL
 - LANL and Zia (at LANL)
 - Rocky Flats Plant
 - Mound Plant
 - Pantex Plant
- ❖ ORAU
 - Oak Ridge Site
 - Fernald Facility
 - Linde Plant
 - Mallinckrodt Facility
 - Savannah River Site
- ❖ PNL and HEHF
 - Hanford Site

Approximate size of CEDR data file sets pertaining to DOE Worker Health and Mortality Studies:

65 Data File Sets

225+ Data Files

Approximately 1,000 megabytes of Data

research center that collected the data or the file in which it appears. In total, the ORAU roster file contains this identification number and basic demographic data for about 420,000 workers. This includes rosters of contractor employees at the 10 sites that were the focus of long-term cohort mortality studies and partial listings of workers at about 40 other DOE facilities. It should be noted, however, that the most complete data for workers who were included in studies conducted by each epidemiologic research center are usually found in the data files that the particular research center generated and submitted to CEDR.¹

The epidemiologic research group at LANL assumed responsibility for studies of all workers monitored for exposure to plutonium by contractors conducting operations at the following sites: the LANL and Zia Company in New Mexico; the Rocky Flats Plant in Colorado; the Mound Laboratory in Ohio; and the Pantex Plant in Texas. The LANL study population also included ORNL and Savannah River Site (SRS) workers having potential exposures to plutonium. An additional, long-term clinical follow-up study that LANL performed was of a small group of workers exposed to plutonium at Los Alamos during the early days of the Manhattan Project.

The epidemiologic research team in Richland, Washington, with staff at HEHF and PNL, took primary responsibility for studying the operations workers who were employed by the major contractors at the Hanford Site in Washington. This research team also performed several related studies. One of the larger ones was a combined

¹ Recent publication providing history and overview of Worker Health and Mortality Studies: Fry, S. A., et al. 1994. *Health and mortality among contractor employees at U.S. Department of Energy facilities*. American Chemical Society Series: Advances in Chemistry, edited by J. P. Young and R. S. Yalow, vol. 243. Washington, D.C.: American Chemical Society.

mortality study that included workers at ORNL, Rocky Flats, and Hanford. The HEHF/PNL team, along with several researchers from the other centers, also participated in an international, combined mortality study, which was performed by the International Agency for Research on Cancer, that included workers from these three sites. Smaller studies conducted by the Hanford team included congenital malformations that occurred in offspring of individuals who lived in communities near the Hanford Site and lung cancers among monitored Hanford workers.

During the performance of the various studies, efforts were made to determine the vital status of each worker included in cohort mortality studies at the selected sites. Mortality was ascertained from employee rosters mainly through searches conducted by the Social Security Administration, the National Death Index, and other death-searching agencies, as appropriate. Death certificates, requested from state vital record offices, provide cause-of-death information, which was coded according to the International Classification of Diseases. Personal radiation-monitoring information available at the sites was examined or collected for most studies.

Until recently, these kinds of data were regularly collected by the three epidemiologic research centers. Each center used slightly different procedures and methodologies to merge, document, validate, and analyze the data. However, work performed by the three centers, as well as by a number of other researchers, during two decades has resulted in the collection of valuable information on a large number of workers at DOE sites and in the publication of numerous scientific papers and articles.

For the purpose of this CEDR catalog, the electronic data files generated in the course of worker studies are placed either in

working data file sets or analytic data file sets, which are briefly described below. Please refer to Section 3 for more detailed information on CEDR data definitions and collection procedures.

Working Data File Sets

The working data file sets mainly include the data collected and updated by the three epidemiologic research centers. A researcher selects or generates variables from these more dynamic data files in order to form analytic data files.

Individuals using these data file sets should note that the working data file sets may include variables not carried forth to the analytic data file or associated studies. There may be data entry and other errors in these files that have never been discovered or resolved; therefore, users should take appropriate precautions when working with these data. Because working data files may be updated during the course of a study, several versions of working data file sets may exist in CEDR. The most recent version is readily available in CEDR, while earlier versions are archived and made available upon request.

Detailed descriptions of working data file sets currently available are presented in **Appendix A**. For each working data file set, this information identifies the facility to which the data pertain, describes the data available, and lists the variables found in each working file of the data file set. Please see the summary table that alphabetically lists the short name for each data file set (page A-2).

Cause of Death	ICDA-8 Codes Included
Infectious & Parasitic Diseases	1 – 136
All Malignant Neoplasms	140 – 209
• Lip, Oral Cavity & Pharynx	140 – 149
• Digestive Organs & Peritoneum	150 – 159
• Respiratory System	160 – 163
• Bone & Connective Tissue	170 – 171
• Skin	172 – 173
• Breast	174 – 174
• Genitourinary System	180 – 189
• Brain/Central Nervous System (CNS)	191 – 192
• Other & Unspecified Sites, Except Brain/CNS	190, 193 – 199
• Lymphatic/Hematopoietic	200 – 209
All Benign Neoplasms	210 – 228
All Neoplasms, Unspecified	230 – 239
Endocrine, Nutritional & Metabolic Diseases	240 – 279
Diseases of Blood & Blood-Forming Organs	280 – 289
Mental Disorders	290 – 315
Diseases of Nervous System & Sense Organs	320 – 389
Diseases of Circulatory System	390 – 458
Diseases of Respiratory System	460 – 519
Diseases of Digestive System	520 – 577
Diseases of Genitourinary System	580 – 629
Complications of Pregnancy & Childbirth	630 – 678
Diseases of Skin & Subcutaneous Tissue	680 – 709
Diseases of Musculoskeletal System & Connective Tissue	710 – 738
Congenital Anomalies	740 – 759
Symptoms & Ill-Defined Conditions	780 – 796
Accidents, Poisoning & Violence (External Causes)	800 – 999

Figure 2-1. Groupings of ICDA-8 Codes Used in the Death Summary Tables

Analytic Data File Sets

Analytic data file sets contain more refined, static data files upon which publications of study findings are directly based. Analytic data files may include variables generated from, but not identical to, one or more variables found in the working data files. Also, depending on the purpose of the analysis, the analytic data file may contain additional variables, not found in the working files, that were collected, validated, or created by the individual investigators who were analyzing the data.

Detailed descriptions of each analytic data file set currently available are presented in **Appendix B**. This information includes the name of the site or facility to which the data file sets pertain, describes the files and data comprising the data file set, provides pertinent citations, and lists the variables found in each file of the data file set.

The International Classification of Disease (ICD) is used to organize cause-of-death data. Over time, various versions of ICD categories have been used by investigators whose studies are represented in CEDR. Earlier studies often utilized the *Eighth Revision, Adapted for Use in the United States* (ICDA-8). Later studies have used ICD-9 codes. ICD categories are not necessarily identical from one version to another. Additionally, ICD codes in the 800-900 range from the Oak Ridge Working Data File Set are intended to be read with an "E" prefix, rather than as standard ICD8-A codes.

For cohort mortality studies which include ICD-conforming coded data, CEDR provides a table that summarized the causes of death found in the analytic files. The table is clearly marked as to which ICD Classification is used, e.g., ICDA-8, ICD9 (Figure 2-1). This summary death table shows the number of deaths occurring in various groupings derived from the ICD classification used in the analysis. These tables, which appear in Appendix B along with the data file set descriptions,

help researchers quickly determine if that particular data file set is of interest to them. However, one should note that the number of deaths that appears in the table may not be exactly the same as the number given in the study description. This is because researchers sometimes retain individuals in their analytic files whose data are not used or discussed in their final analyses or published papers.

Pages B-2 and B-162 present summary tables that alphabetically list the short name for each analytic data file set and provides the name of the associated study and a brief description. These tables also indicate the order in which the analytic data file sets appear.

Classic Radiation-Effects Studies

Approximate size of CEDR files pertaining to radium dial painter studies:

1 Data File Set: RADPDW01
~ 30 Data Files
19.6 Megabytes

Radium Dial Painter Data

The opportunity to study a group exposed to a potentially toxic material is rare, and it is even more rare to collect sufficient data on the group to determine the potency and effects of the material. However, such an opportunity emerged in the 1920s when skeletal lesions became associated with workers, called luminizers or dial painters, who painted luminous dials on timepieces with paint containing radioactive radium as the energy source for luminosity.

Under the Internal Emitter Program conducted at DOE's Argonne National Laboratory for the last 25 years, researchers identified about 6,000 people with significant radium exposures, including about 500 individuals, mostly women, who were dial painters. Many were located and followed until death; in these cases, the cause of death and body content of radium were determined. The files available in CEDR contain much of the data, such as demographics,

measurements of radium body content, calculated doses, radium exposure history, and clinical examination results, that were generated during this research. However, as of the publication date of this CEDR catalog, loading of data and documentation for the approximately 30 files pertaining to this study has not been completed. Preliminary information is presented in **Appendix C**. As always, more complete information can be found in CEDRView (see page 26).

Approximate size of CEDR files pertaining to atomic bomb survivor studies:

2 Data File Set:	JALSSA02, JALSSA03
6 Data Files	
6.7 Megabytes	

See Appendix C for file and variable descriptions.

Radiation Effects Research Foundation Data

CEDR also contains one public-use data file set pertaining to the study of atomic bomb survivors, which was prepared and submitted by the Radiation Effects Research Foundation (RERF). RERF, established in April 1975, is a binational organization dedicated to the discovery, application, and dissemination of knowledge about health effects in the survivors of the atomic bombings of Hiroshima and Nagasaki and in their offspring. Prior to 1975, this important research was conducted by the Atomic Bomb Casualty Commission, an organization established in 1947 by the U.S. National Academy of Sciences.

The data submitted to CEDR support RERF's Life Span Study, which has been ongoing for more than 40 years. Unlike the Worker Health and Mortality Studies data, the RERF data does not directly present demographic and exposure information for each individual. Instead, the data consist of a cross-tabulation of death counts and person-years over five stratification variables: city, sex, age-at-exposure, dose estimated using the DS86 methodology, and calendar time. In addition, mean gamma and neutron transmissions stratified by city and age-at-exposure are presented.

A detailed description and list of variables contained in this data file set are presented in Appendix C.

Approximate size of CEDR files pertaining to U.S. Transuranium and Uranium Registries research:

1 Data File Set:	USTURW01
2 Data Files:	ADMIN; RADCHEM
1.1 Megabytes	

See Appendix C for file and variable descriptions.

U.S. Transuranium and Uranium Registries Data

The research carried out by the U.S. Transuranium and Uranium Registries (USTUR) involves the study of the distribution, biokinetics and possible biological effects of the actinide elements, including plutonium, uranium, americium, and thorium, in persons with documented exposure to these elements. This research helps in the refinement and validation of biokinetic models used to estimate doses due to deposition of these elements in humans.

The data file set in CEDR consists of two files for USTUR research that contain demographic data and results of more than 10,000 radiochemical analyses of various tissues. As more USTUR data containing exposure and work histories become available, they will be included in CEDR.

Dose Reconstruction Studies

In efforts to assess the health impacts of DOE operations on off-site communities, DOE has funded a number of lengthy studies commonly referred to as historical, off-site dose reconstructions. Dose reconstructions are usually complex, multidisciplinary studies that consist of several major steps, including cumulative estimates of the amounts of materials the site released to the environment, determination of the fate of those materials in the environment, and an estimate of the exposure of individuals to those materials.

One of the first comprehensive dose reconstructions to be performed was for populations downwind from the Nevada Test Site where atomic weapons were detonated. The single data file set that represents this study in CEDR contains some of the basic data upon which the results of this dose reconstruction depend.

Approximate size of CEDR files pertaining to NTS off-site dose reconstruction:

1 Data File Set:	NTORPW01
4 Data Files:	TOWNDB; CNTYDB; LFSDEMOG; LFSFARM
3.5 Megabytes	

See Appendix C for file and variable descriptions.

Other CEDR Data

Approximate size of CEDR files pertaining to the Complex Chemical Study:

1 Data File Set:	RFCHMW01
5 Data Files:	AGENT; DETSAMP; GENSAMP; EXPOSED; RESULTS

3.6 Megabytes

See Appendix C for file and variable descriptions.

Complex Chemical Feasibility Study

Some workers in the DOE nuclear weapons facilities are exposed to complex mixtures of chemicals and radioactive materials, such as plutonium. It is known that exposures to such mixtures of possible carcinogens may produce adverse health effects, but it is usually difficult to measure each individual's exposure to these agents. The purpose of this study was to examine the feasibility of using existing records and information to estimate a worker's exposure to such complex mixtures of chemicals, so that the estimates would be useful for epidemiologic studies. A methodology was developed that used data found in records about chemical purchases, work areas, and plant operations, and it was applied to a small pilot population of Rocky Flats workers. Data collected include information about worker locations, operations performed, and results of samples indicating workplace exposures to various chemicals.

Salmon Site Descriptive Study

Between 1964 and 1970, two nuclear and two non-nuclear bombs were detonated in a large underground salt dome in Lamar County, Mississippi, for the purpose of measuring the seismographic activity that these explosions generated in salt domes. Upon completion of the testing program, the site was deactivated and decommissioned in June 1972. Because of concerns that effluents from the tests might have increased cancer deaths among nearby residents, a mortality study of residents of Lamar County has been conducted. The purpose of the study was to determine if there was an association between living near the salt dome and death due to cancer.

Approximate size of CEDR files pertaining to the Salmon Site Study:

1 Data File Set:	TDBSRA01
1 Data File:	MORTDAT

See Appendix C for file and variable descriptions.

Epidemiologic Surveillance and Records Inventory Reports

Epidemiologic Surveillance Annual Reports

Epidemiologic surveillance consists of the regular and systematic collection, analysis, and interpretation of data on illness and injury in the DOE work force. The goals of epidemiologic surveillance are to:

- Determine incidence rates of disease among the current DOE work force;
- Monitor early indications of health problems that may be occurring among employees at DOE sites;
- Provide data in response to questions from workers, physicians, and others about possible occupational health effects;
- Identify areas for further investigation; and
- Provide a means to evaluate preventive and ameliorative programs.

Because of an increasing interest and awareness concerning the health trends of current workers at DOE facilities, the annual reports summarizing epidemiologic surveillance at participating sites are accessible via CEDR.

The first group of annual reports, published for the year 1992, pertain to four sites: Brookhaven National Laboratory (BNL), Hanford, Rocky Flats, and SRS. Since that time, four additional sites have been added: Fernald, the Idaho National Engineering Laboratory, Pantex, and Sandia National Laboratory. As additional reports are completed, they will be accessible via CEDR.

The 1992 annual reports for BNL and SRS contain information on illnesses and injuries from reports filed with the medical department

Category of Diagnoses	ICD9-CM Code	Number of Diagnoses†	Age-Adjusted Rate per 1,000*	Lower 95% Confidence Limit per 1,000	Upper 95% Confidence Limit per 1,000
Infections & parasitic diseases	001-139	14	3.1	1.9	5.3
Malignant neoplasms	140-208, 230-234	5	1.2	0.5	2.9
Benign neoplasms & other	210-229, 235-239	5	1.2	0.5	2.8
Endocrine & metabolic diseases	240-279	14	3.3	2.0	5.7
Blood & blood-forming organs	280-289	1	0.2	0.0	1.7
Mental disorders	290-319	4	1.5	0.5	5.0
Nervous system & sense organs	320-389	17	5.1	2.9	8.9
Circulatory system	390-459	35	9.1	6.4	13.0
Respiratory system	460-519	59	15.9	11.9	21.4
Digestive system	520-579	27	8.0	5.1	12.5
Genitourinary system	580-629	11	4.3	2.1	8.9
Pregnancy & childbirth	630-676	6	3.8	1.5	9.1
Skin & subcutaneous tissue	680-709	2	0.4	0.1	1.6
Musculoskeletal system	710-739	43	9.6	7.1	13.0
Other and ill-defined conditions	740-799	19	5.7	3.3	9.6
External causes of injury	E800-999	84	22.6	17.7	29.0
Total minus pregnancies		340	91.4	80.9	103.2
TOTAL		346	95.1	84.2	107.5

†Includes all diagnoses reported with an absence of 5 or more days.
*Standardized to age distribution of 1970 U.S. population.

Figure 2-2. Example of Table Showing Diseases and Injuries by Diagnostic Category

after an employee was absent for 21 or more consecutive work days. The 1992 annual reports for the Rocky Flats and Hanford sites contain information on illness and injuries from reports of absences of 5 or more consecutive workdays, as do all 1993 and future annual reports. Starting in 1994, data obtained from the Occupational Safety and Health Administration (OSHA) Log are also included in epidemiologic surveillance.

The summary information in the annual epidemiologic surveillance reports is arranged in five sets of tables that present:

- The distribution of the labor force by occupational category;
- The number of absences per person, number of diagnoses per absence, and diagnosis rates for the entire site;
- Diagnosis rates by type of disease or injury;
- Diagnosis rates by occupational category; and
- Relative risks for 16 major categories (with some additional breakdowns) of illness and injury for each occupational category.

One example of a table showing diagnosis rates by type of disease or injury is presented in **Figure 2-2**. Diagnosis rate is defined to be the number of new reported health events observed among DOE workers per 1,000 DOE workers at risk during a given time period. All rates are age-adjusted to the U.S. 1970 population. Confidence limits at the 95% level are also presented.

The reports can be viewed via CEDRView. Printed copies are available from the Office of Epidemiologic Studies (see page 3).

Epidemiologic Records Inventories

Site Name	Title
Oak Ridge Site	<ul style="list-style-type: none"> Volume 1: Y-12 Mercury Task Force Files: A Guide to Record Series of the Department of Energy and Its Contractors Volume 2: Records Relating to Cesium at the K-25 Plant: A Guide to Record Series of the Department of Energy and Its Contractors Volume 3: Records Relating to RaLa, Iodine-131, and Cesium-137 at the Oak Ridge National Laboratory and the Oak Ridge Operations Office: A Guide to Records Series of the Department of Energy and Its Contractors
Rocky Flats	<ul style="list-style-type: none"> The September 1957 Rocky Flats Fire: A Guide to Record Series of the Department of Energy and Its Contractors Guide to Record Series for Epidemiologic and Health-Related Records of the Department of Energy's Rocky Flats Site: <ul style="list-style-type: none"> Volume 1: Administrative and General Volume 2: Facilities Construction and Equipment Volume 3: Production and Materials Handling Volume 4: Waste Management Volume 5: Environmental and Workplace Monitoring Volume 6: Employee Health
Hanford Site - Epidemiologic Section of the Pacific National Laboratory	<ul style="list-style-type: none"> Hanford Site: A Guide to Record Series Supporting Epidemiologic Studies Conducted for the Department of Energy
Oak Ridge Institute for Science and Education	<ul style="list-style-type: none"> Oak Ridge Institute for Science and Education: A Guide to Record Series Supporting Epidemiologic Studies Conducted for the Department of Energy
Los Alamos National Laboratory Epidemiologic Section	<ul style="list-style-type: none"> Los Alamos National Laboratory: A Guide to Record Series Supporting Epidemiologic Studies Conducted for the Department of Energy

Figure 2-3. Records Guides Accessible via CEDR

In May 1992, each DOE site and DOE contractor was asked to prepare an inventory of all records pertinent to worker or community health-related studies. Each site was provided with guidelines that defined epidemiologic records, provided instruction for describing record series, and discussed the relationship of the epidemiologic inventory to DOE's comprehensive records inventory.

In August 1993, the Office of Health selected History Associates, Inc. (HAI), as its support services contractor to assist with the records inventory efforts. The work of DOE and HAI, a professional records management and historical research services firm, represents an ongoing endeavor to make epidemiologic and health-related records more accessible to researchers, and to address the complexities of handling modern, organizational records for archivists and records managers. Two of the tasks, verification of site records inventories and performance of epidemiologic records inventories, have led to the creation of a number of extensive guides to records that are potentially useful to investigators performing worker or community health-related studies.

For each series of records, the guide includes the following information: records series title and description, location, access restrictions, volume, and a point of contact for accessing the records.

Figure 2-3 lists the records guides that will be available in CEDR. Hard copies of the guides are available from the CEDR Program Manager (see Section 1 for address and phone number).

