

Case-Control Selection Program Documentation

**National Institute for Occupational
Safety and Health**

March 2003

S. Nowlin, K. Steenland,
R. Rinsky, J. Beaumont



Table of Contents

I.	Introduction	2
II.	Method of Control Selection	4
III.	Program Library and System Specifications	8
IV.	Input and Output Files	9
V.	Run Options and Requirements	10
VI.	System Process Outline	14
VII.	“Installing” the System	20
VIII.	Submitting the Case-Control Selection Program	21
IX.	Reports	24

Appendices

A.	PVC Input File Format	25
B.	DEDS Input File Format	27
C.	Dose File Format	32
D.	Case-Control Selection Output Files	33
E.	Request Form	37
F.	Future Modification requests	39
G.	Glossary of Terms	40
H.	Source Code	42

I. Introduction

A case control study is a method of analysis which looks for a causal relationship between a certain disease and a given exposure. The first step in a case control study is to select the workers who have the disease being investigated. These workers are called the Acases@. The next step is to assign Acontrols@ to the cases. There are various methods of selecting controls for particular cases. This document describes a method of selection used by the National Institute for Occupational Safety and Health (NIOSH).

This program selects controls for nested case-control studies which are derived from retrospective occupational cohort studies. Cases typically occur over an extended time period. The specific diseases of interest are usually chronic diseases, frequently with long latent periods. In addition to an exposed group, there should be a non-exposed or less exposed group. Nested case-control studies select cases and controls from among a cohort of exposed and non-exposed individuals who have already been followed over a long period of time to determine disease status. Typically, disease rates (or mortality rates) for the exposed and non-exposed have already been compared. The nested case-control study attempts to examine in more detail the relationship between exposure and disease by examining specific jobs or work areas which might be associated with an excess disease rate for the exposed. Nested case-control studies also allow more multi-variate analyses than are typically conducted in the overall cohort analysis

The basic purpose of this program is to select controls from a pool of eligible controls (the Ariskset@) who are at risk of failure at the time the case fails. In this program the variable used for Atime@ is age. All members of the population are eligible to serve as a control for the index case if they have lived to an age greater than or equal to the age that the index case failed.

The program gives the option to either include the entire riskset for a given case in the final case control analysis file (Full Matched Set) or to randomly select a limited number of controls per case from the riskset (Sample Matched Set). All time-related covariates for a control (e.g. time since

first employment, duration employed, duration exposed, etc.) are calculated at the time the control attained the age of the index case. For example, if a case in a study of lung cancer died at age 38, all other cohort members who lived to an age greater than or equal to 38 are eligible to serve as controls for the index case. The time-related covariates for the chosen controls are then calculated for all controls when they were 38.

Options in the program allow the User to not only match eligible controls to the cases on age but additionally on year of birth, duration of employment, race and/or sex. Other options enable the User to restrict both cases and controls to those who have obtained a minimum time-since-first-employed and/or a minimum length of employment.

II. Method of control selection

Each worker that is read into the control selection program from a study cohort file will eventually fall into one of three categories. One, he can be rejected due to error(s) in his data or user specified restrictions; two, he can be a case; or three, he can be a control. A worker can also be both a case and a control for another case.

A worker will be rejected from the study if his date of birth, date first employed or death date is invalid. If the User chooses to use the Date Last Observed option or the Eligibility Date option and these dates are invalid, the worker will also be rejected. A date is considered invalid if:

1. It contains non-numeric characters;
2. The month portion is ≥ 1 and ≤ 12 ; or
3. The day portion is ≥ 1 and ≤ 31

A worker will also be rejected for any of the following reasons:

- Date first employed or date of birth is missing
- Date first employed is greater than the study end date

- Date last employed is less than the study begin date
- Worker has no valid work histories

If the worker passes these data requirements then subsequent data checks are performed on additional data points depending on options selected by the User. Options which will trigger these additional data verifications are:

- If the Date Last Observed (DLO) option is chosen then the worker must have a valid date last observed.
- If the Eligibility Date option is chosen then the worker must have a valid eligibility date.
- If a Minimum Length of Employment is specified then the worker had to work as long as or longer than the minimum length of employment requirement specified.
- If a Time Since First Employed is specified then the worker had to work as long as or longer than the time since first employed requirement specified.

If the worker passes all required and optional data checks then he continues to be analyzed to see if he is a case, a prospective control or both. Each worker is processed to determine whether he is a case, and each worker is also put into a prospective control file (cases can also be controls for other workers, depending on their respective ages at failure).

In order to be a case, a worker must have a failure code (ICD number); this is one of the ICD=s chosen by the User. Also the year of the failure must be within the range specified by the User to be used with the ICD code in question. If the worker meets the ICD criteria, his age at failure is computed and he is written to the case file.

Regardless of whether or not the worker has been designated as a case, he/she is then processed as a prospective control. The processing calculates the worker=s age at study end date (or date last observed if optionally chosen), and also his date at entry into the study, from which his age at entry date is calculated. A worker=s entry date equals the worker=s date first employed plus the minimum length of employment or the study begin date, which ever date is later.. However, if the eligibility date option is selected then the entry date equals the greater date value between the study

begin date and the worker's eligibility date.

The next step is the selection of controls for each case. A pool of controls for each case is created from all eligible cohort members. A cohort member is eligible to be a control if his/her age at study end date (or optionally, date last observed) is greater than or equal to the case's age at failure, and the control's age as of the entry date is less than or equal to the case's age at failure (i.e. the control had to have entered the study before the control attained the age at which the index case failed, so as to be at risk). Assignment to the pool of controls means that the prospective control is now eligible to be selected as a control for the index case.

Once prospective controls have been identified based on the age criteria mentioned above, additional data checks are performed based on User selected options for matching controls to cases. Options which may further limit the number of controls in the cases riskset are:

- Match Controls to Cases by Sex
- Match Controls to Cases by Race
- Match Controls to Cases by Date of Birth
- Match Controls to Cases by Date First Employed

Should the control meet all of these possible matching options, he becomes a control for the case.

The control selection process repeats itself until all eligible controls are selected for the case. This is the riskset for the index case.

If the User chooses to create a Sampled Matched Set as the control selection criteria, then a random number is assigned to each control in the pool. The random number assignments are then sorted into ascending order. Controls for the case are then selected by taking the first X number of controls; where X equals the maximum number of controls to select per case that the user specified in the program options.

A cut-off date for the control is computed to be the case's age plus the control's date of birth. In other words, the cut-off date is that date at which the control's age equals the case's age. This is

the date beyond which exposure and covariate data should be ignored in the analysis.

Any work histories whose begin date occurs after the cut-off date is removed. If a specific job in the work history has a begin date less than the cut-off date, but an end date occurring after the cut-off date, the history end date is changed to be the cut-off date. If the User has chosen to calculate exposures for each case and control, then the time exposed is also limited according to each control=s cut-off date.

Cut-Off Date Illustration

A sample case lived to be 50 years old. All cohort members whose age at last observed was greater than or equal to 50 and whose age at date first employed is less than or equal to 50 would be eligible to be in the riskset for the index case.

A sample control in the risk set for the case was born on 1/1/1905. The control=s cut-off date, the date at which his age equaled the case=s age, is 50 years after the control=s birth date of 1/1/1905, namely 1/1/1955.

If the eligible control=s date last observed (DLO) and/or his date last employed (DLE) occur after the cut-off date of the 1/1/1955, then these dates are replaced by the cut-off date 1/1/1955.

Any work history that began after the cut-off date is removed from the file. If the eligible control=s last work history began on 10/1/1960, which is after 1/1/1955, then this last work history is removed. A work history having a begin date prior to the cut-off date and an end date later than the cut-off date has the end date replaced by the cut-off date. If our sample control=s second to the last work history began on 12/18/1952, but ended after the cut-off date on 9/30/1960, then 1/1/1955 would replace 9/30/1960 as the history end date of the record.

III. Program Library and System Specifications

The NIOSH Case-Control Selection Program is written using Version 8.2 Base SAS® software. A SAS® license is required to run the application. See SAS® Software documentation for complete list of compatible operating environments. Only one program module (JCL.SAS) needs to be modified to conform to the users operating environment. The default operating environment used by NIOSH is Windows 2000.

The application is made up of a set of SAS program modules. These modules are called by the main program module, CASECNTL.SAS, as needed based on the user options specified.

The user only needs to modify the JCL.SAS program to change values for options each time the program is submitted. This program will submit the CASECNTL.SAS program which will call the other program modules as needed.

Following is the list of the Case-Control Selection program modules and their descriptions:

Program Module	Description
JCL.SAS	Program submitted by the user containing macro variables to supply values for the case-control selection options.
CASECNTL.SAS	Main Case-Control Selection Program module. This module calls other program modules as needed based on user selected options.
CCINPVC.SAS	Input program for PVC file format.
CCINDEDS.SAS	Input program for DEDS file format.
CCINDOSE.SAS	Input program for the Dose File format (Exposure Data File)
CCDOSEWH.SAS	Assigns a dose factor using the dose file to each work history for each case and control.
CCEXPOS.R.SAS	Calculates exposure variables for each case and control.
CCEXPAUC.SAS	Calculates area under the curve for each case and control.

IV. Input and Output Formats

The Case-Control Selection Program currently accepts one of two input file formats for worker demographic, cause of death/disease and work history data. The first format is known as the NIOSH PVC file format. This input structure is a set of two flat ASCII files (Cohort file, Multiple Cause of Death file). The second format is known as the NIOSH/HERB DEDS file format. This input structure is a set of related SAS datasets. The user must ensure the data is in one of these two file formats prior to submitting a case-control selection job.

The data dictionary for the NIOSH PVC file format can be found in Appendix A.

The data dictionary for the NIOSH/HERB DEDS file format can be found in Appendix B.

If the user has exposure data that can be related to the worker either by area worked (Plant, Department, Operation) or personal exposure data, then an exposure factor can be assigned to each case and control's work history record. To do this, the exposure data needs to be formatted in the NIOSH Dose file format, a flat ASCII file format. The data dictionary for the NIOSH Dose file format can be found in Appendix C.

The Case-Control Selection Program will produce 4 output files all in SAS format. The data dictionary for the Case-Control Selection output files can be found in Appendix D.

Descriptions of the output SAS dataset files are:

Dataset	Description
CASECNTL	Case and Control Demographics (one record per each case and each of their matched controls)
CC_WH	Case and Control Work History Data (one or more detail work history records per each case and each of their matched controls).
WARNINGS	Identifying information about any records that received warning messages during the data quality checks within the program.
RUN_OPTS	Options chosen by the user in running the Case-Control Selection

	Program.
--	----------

V. Run Options and Requirements

Run Options and Requirements			
Option	Option Description	Option Choices	Choice Description
Type of Case-Control Selection		S	Sample Matched Set
		F	Full Matched Set
Maximum number of controls to select per case	Sample Matched Set Only	Enter #	Number should be the maximum number of controls to select from the riskset per case.
Cause of death (or incidence) criteria	Used to identify cases. Indicates whether only underlying cause of death (or incidence) is to be used, or underlying and contributory causes.	UCOD	Program will only look at Underlying Cause of Death (UCOD) to identify cases.
		ALL	Program will look at Underlying Cause of Death (UCOD) and all contributing causes of death (or incidence) to identify cases.
Study Begin Date	This date is the earliest date at which an individual may enter the study. Workers with final employment dates occurring before this date will be rejected for selection as a case or control.	Enter a date in mmddyy format	
Study End Date	Used to calculate age for alive workers (default).	Enter a date in mmddyy format	
Date Last Observed option	This option enables the use of Date Last Observed (DLO) to be used in the calculation of AGE instead of using the Study End date.	Y	Yes, use Date Last Observed (DLO) to calculate AGE.
		N	No, do NOT use DLO to calculate AGE. (Study End date will be used).
Minimum length of employment	Workers without this minimum length of employment will be rejected.	0	No minimum length of employment restriction.
		Enter #	Enter the number of days a worker must have worked in order to be considered as a case or control.

Run Options and Requirements			
Option	Option Description	Option Choices	Choice Description
Time Since First Employed (TSFE) restriction	<p>This restriction gives the option of requiring a worker to have been employed a given amount of time before he can be eligible to be a case or a control. This restriction applies differently to cases and controls.</p> <p>NOTE: This option can NOT be chosen if the Eligibility Date option is used.</p>	0	No TSFE restriction
		Enter # of days	<p><u>CASES:</u> this time span is determined as the difference between the case's failure date and the date first employed (DFE). If this calculated time span is not equal to or greater than the TSFE value supplied by the User, then the worker cannot be a case, nor can the worker ever be a prospective control for other cases.</p> <p><u>CONTROLS:</u> calculated as the difference between the date of first employment and the cut-off date (i.e., the date at which the control reaches the age the matched case failed).</p>
Eligibility Date option	<p>If this option = >Y= then the individuals eligibility date, hard coded in the workers record, is used as the Entry Date. Exception: if Study Begin date is > the eligibility date then Study Begin date is used for calculating the Entry Date.</p> <p>NOTE: This option can NOT be chosen if the TSFE option is used.</p>	Y	Yes, use the individual eligibility date to calculate Entry Date.
		N	No, do not use the individual eligibility date to calculate the Entry Date.
Match Controls to Cases by Sex	<p>If this option = >Y= then controls for a given case will be restricted to those cohort members of the same sex.</p>	Y	Yes, use this restriction in selecting controls
		N	No, do not use this restriction to select controls
Match Controls to Cases by Race	<p>If this option = >Y= then controls for a given case will be restricted to those cohort members of the same race.</p>	Y	Yes, use this restriction in selecting controls
		N	No, do not use this restriction to select controls.
Race Recode option	<p>This option will enable the collapsing of Race code values to White and Non-white.</p>	Y	Yes, recode unknown and Hispanic race values to White and Asian and other race values to Non-white.
		N	No, do not recode the race values.

Run Options and Requirements			
Option	Option Description	Option Choices	Choice Description
Match Controls to Case by Date of Birth	The Date of Birth (DOB) restriction gives the User the option of restricting an eligible control=s date of birth to be within a desired period of time of the matched case=s date of birth. For instance, a DOB restriction value of 5 years means that an eligible control must have been born within 5 years (plus or minus) of the date of birth of the case in order to be a control for the case.	0	No matching on DOB is performed
		# of days	Enter the maximum number of days permitted difference in Dates of Birth for a control to be considered a match for the given case.
Match Controls to Case by Date First Employed	This option restricts the controls to those who began work at a similar time as the case. The Date First Employed (DFE) restriction gives the user the option of restricting an eligible control=s DFE to be within a desired period of time of the case=s DFE. For example, a DFE restriction value of 10 years means that an eligible control must have started work within 10 years (plus or minus) of the time the case began working in order to be a control for the case.	0	No matching on DFE is performed
		# of days	Enter the maximum number of days permitted difference in Date First Employed for a control to be considered a match for the give case.
Exposure Option	This option will assign a dose per day value to each work history record for all cases and controls using a dose exposure file supplied by the User. If this option is chosen then additional variables will be added to the final case control file (such as, duration exposed, cumulative exposure, peak exposure, time since first exposed, etc. - see case control output documentation)	A	Exposure factor per day will be assigned to each work history based on the area worked (plant, dept, oper) and dates worked/exposed.
		P	Exposure factor per day will be assigned to each work history based on personal dose exposure values and dates worked/exposed.
		Blank	The exposure option will not be run
Area Under the Curve Option	If this option = >Y= then the program to calculate area under the curve will be called.	Y	Yes, calculate area under the curve.
		N	No, do not calculate area under the curve.
Half life units	If the area under the curve option is selected, then enter the exposure	#	Enter the exposure half life in years.

Run Options and Requirements			
Option	Option Description	Option Choices	Choice Description
	half life (units in years)		
Input file format	Specifies which format the input cohort is in.	PVC	PVC file format
		DEDS	DEDS file format
Exposure File Format	Specifies which format the exposure file is in	P	PVC exposure file format
		L	PC LTAS exposure file format (NOT AVAILABLE YET)
Random Number Generator Initial Value (Seed)	Leave blank or enter a numeric string having 9 digits. A blank causes the program to generate its own initial >seed= value based on a computer date/time function. Otherwise, the initial >seed= value used by the random number generator (used in Sample Matched Set only) is the 9 digit numeric string supplied by the User taken from a previous run of the case control selection program. If a number is supplied from a previous run then the same controls will be selected per case.	Blank	Program will generate its own initial seed value
		9 digit number	Program will use this seed value for generating the list of random numbers. - used to select same controls per case as a previous run in a Sample Matched Set.

VI. System Process Outline

Outline of Program Logic (Business Rules) - as of 1/17/2000

Program Module: JCL.SAS – used to select options and submit the Case-Control Selection Program

(* This is the only program needing modification by the user***)**

- Identify input files
 - Study Cohort
 - Dose (optional)
 - Multiple Cause of Death (optional – only for PVC input file format and if selecting cases based on all causes of death/disease)
- Identify ICD codes used to select Cases
- User selection of program run options
- Identify temporary hold libraries
- Identify program library

Program Module: CASECNTL.SAS - Main program module)

- Initialize System Macro variables
- Verify valid entries in program run parameters (options)
- If invalid values entered for any of the program options then abort program
- Create Error codes format library
- Setup system macros
- Set Master Seed Value
- Define library for final Case Control output
- Read Study Cohort File
 - If input file format selection is >PVC= then include the **CCINPVC.SAS** program module
 - Perform data error checks
 - Place records in Reject temp file if invalid data found
 - Recode data where necessary and/or requested

- Create DEMOGR, WORKHIST, REJECT, WARN, REJECTWH, REJECTME temporary datasets
- If input file format selection is >DEDS= then include the **CCINDEDS.SAS** program module
 - Perform data error checks
 - Place records in Reject temp file if invalid data found
 - Recode data where necessary and/or requested
 - Create DEMOGR, WORKHIST, REJECT, WARN, REJECTWH, REJECTME temporary datasets
- Check Time Since First Employed (TSFEMP) Restriction
- If option is selected (i.e., TSFEMP > 0) then
 - Calculate TSFEMP:
 - if using the DLO option (DLO_OPT==Y=) then $TSFEMP = DLO - DFEMP$
 - else $TSFEMP = \text{Study End Date} - DFEMP$
 - If $TSFEMP < \text{number of days restriction specified by the user}$ then reject the record.
- Keep track of records rejected with reason(s) rejected for rejection report

- Calculate Age, Entry Date and Age at Entry for each individual not rejected
 - If Date of Death (DOD) not missing and DOD less than or equal to the Study End date then $AGE = DOD - DOB$ (Date of Birth)

 - If DOD not missing but $DOD > \text{Study End date}$ then $AGE = \text{Study End Date} - DOB$

 - If DOD missing and Date Last Observed (DLO) option = >Y= (Yes) then
 - $AGE = DLO - DOB$

 - If DOD missing and DLO Option = >N= (No) then
 - $AGE = \text{Study End Date} - DOB$

Calculate Entry data

- If $DFEMP + \text{Minimum Length of Employment} < \text{Studybeg date}$ then Entry Date = Studybeg Date
 - Else Entry Date = $DFEMP + \text{Minimum Length of Employment}$
 - If the Eligibility Date option = >Y= and the individuals Eligibility Date (position 1570-1575 in PVC) is greater than or equal to the Study Begin date then the Entry Date = the individuals Eligibility date.
- Calculate age at Entry Date (ENTRYDTE)
 - $AGE_ENTR = ENTRYDTE - DOB$

Identify Cases

- Read the ICD codes supplied by the user to be used to identify cases
 - Compare each cohort member=s cause(s) of death/disease to the ICD codes read.
 - If the Cause of Death/Disease option (COD) = >UCOD= then only compare the underlying cause of death (UCOD)
 - If the Cause of Death/Disease option (COD) = >ALL= then compare the UCOD and all contributing causes
 - If a cohort member=s cause(s) of death/disease equals one of the ICD codes and the date of death/disease of the individual falls within the year range supplied by the user, then that individual is identified as a Case.
- Set case flag=1
 - Set CASE_ID = NIOSH_ID
 - Output record to the CASES temporary dataset
-
- Generate Seed Value for Each Case

- Remove contributing causes of death if not used to identify cases.
 - Assign Exposure factor to each work history record (if Exposure option chosen).
 - Read the exposure file, include program module: **CCDOSEIN.SAS**
 - Include program module: **CCDOSEWH.SAS** to assign exposure factors to work histories.

 - Create Risk Sets
 - For each case, compare the case=s data with all other individuals in the cohort to identify eligible non-cases
 - Non-Case=s age must be greater than or equal to the case=s age at death/disease
 - and
 - the non-case=s age at Entry Date (AGE_ENTR) must be less than or equal to the case=s age at death/disease.
 - Calculate each eligible control=s cutoff date
 - $CUTOFF = DOB + CASE_AGE$
-
- Check control matching criteria supplied by the user
 - SEX match
 - RACE match
 - Year of Birth match (plus or minus # days)
 - Date First Employed match (plus or minus # days)
 - Inclusion/exclusion options
 - Time Since First Employed restriction (# days)
 - Entry date restriction
 - Entry date must be less than or equal to CUTOFF

 - If Sample Match set option chosen, perform random selection of maximum number of controls per case from the pool of eligible controls (risk set).

- For each control in each resulting matched set, cutoff work histories to the case=s age at death/disease (CUTOFF).
- Combine all matched sets together into one dataset
- If exposure option chosen, calculate the exposure variables:
 - CCEXPOSR.SAS
 - CCEXPAUC.SAS
- Create the final output files
 - CASECNTL - Demographic data on the combined set of matched sets
 - CC_WKHST - Work History data on the combined set of matched sets
 - WARNINGS - File containing all the records that had warning messages
 - RUN_OPTS - File containing the options selected in creating the CASECNTL and CC_WH file (including the Task number, Seed value (if Sample matched set), Project Officer=s and programmer=s names)
- Produce Reports and Sample Listings
 - Run Parameters Report (indicates which options were used)
 - Summary Report (summarizes total number of cases and controls selected and rejected)
 - Rejection & Warnings Summary Table
 - Rejection Report
 - Sample Warnings Report
 - Control Audit Report
 - Sample Listings of the final case control files
 - Proc Contents of the final case control output SAS library

Program Module: CCINPVC.SAS – reads PVC input file and performs data QA

- Read in the demographic information, create DEMOGR temp dataset
- Assign a unique NIOSH_ID to each record

- If no detail work histories exist then use the general DFEMP and DLE dates (positions 246 - 257) as the only work history record, create WORKHIST temp dataset.
- Else read in the detail work histories, create WORKHIST temp dataset
- If the COD option equals >ALL= then read the PVC Multiple Cause of Death file and merge with the DEMOGR temp dataset

Program Module: CCINDEDS.SAS – reads DEDS input file and performs data QA

- Create a DEMOGR temp dataset from the DEDS DEMOGR dataset
- Create a WORKHIST temp dataset from the DEDS WORKHIST dataset
- Merge the COD DEDS dataset with the DEMOGR temp dataset

... continued processing for CCINPVC.SAS or CCINDEDS.SAS:

- Recode data where necessary and/or requested
 - If SEX code = unknown then SEX = male
 - If RACE RECODE option = >Y= (YES) then recode:
 - RACE code of unknown or Hispanic to RACE = white
 - RACE code of Asian or other to RACE = non-white
 - Else RACE value stays as is in cohort study file.
 - If DAY of Birth, Death or DLO is missing then set day to >15'
 - If MONTH of Birth Death or DLO is missing then set month to >7'
- Keep track of all fields recoded for warnings report
- Output identifying information (SSN, Last Name, NIOSH_ID) and any fields recoded to a WARNINGS file
- Reject records if invalid data found
 - If invalid DOB then reject record
 - If invalid DOD and UCOD exists then reject record
 - If invalid DLO and the DLO option for calculating age is chosen the reject record.
 - If the Eligibility Date option (ELIGDATE==Y=) is chosen and the

individuals Eligibility Date, which is hard coded into the individuals record, is invalid then reject record

- If all work histories occur prior to the study begin date then reject record
 - If invalid work history dates then reject only that work history record from the WORKHIST file.
 - If minimum length of employment requirement is not met then reject record
 - If individual has no work history records then reject record
-
- **Program Module: CCDOSEWH.SAS** – assign dose factor to each work history.
 - Reformat the exposure file so that for each unique exposure key (area=plant/dept/oper or personal=SSN) a dynamic array is created containing all dose exposure levels and time period ranges. Essentially creating one record per unique exposure key value.
 - Match the reformatted exposure file with the Work Histories file by the exposure key variable
 - Compare the begin and end dates of each work history record to each begin and end date elements of the exposure array in order to assign the appropriate dose exposure factor to each work history.

VII. “Installing” the System

1. Create a directory to store the set of Case-Control Selection programs.
2. Copy the 9 programs to the new directory
3. Open the program JCL.SAS
4. Edit the filename statement near the bottom of the program to reference the directory created in step 1. `filename pgm 'enter new directory full path name here';`

VIII. Submitting the Case-Control Selection Program

For NIOSH/DSHEFS/HERB users, copy the program JCL.SAS from the o:\app_dev\source\case_control directory and modify as necessary (see instructions within the JCL.SAS program) to submit a case-control job.

For Users outside NIOSH\DSHEFS\HERB, once the steps in section VII. Have been completed, copy and modify the JCL.SAS program as necessary to submit a case-control job.

A sample job follows. Items highlighted in yellow MUST be modified by the user. Those in turquoise are optional items to be modified.

```
*****;
* FILENAME JCL.SAS ;
* NIOSH/DSHEFS - CASE CONTROL SELECTION PROGRAM ;
*
* Version 1.0 1/1/2000
*
* 7/99 S. Nowlin ;
* This SAS Program is the program used to create a Case-Control Analysis file. ;
* ;
* The user must modify the filename statements ;
* for the 'IN','DOSE' and 'COD' files to point to the location of their ;
* 'master' cohort file, exposure file (IF requesting exposure variables to be ;
* calculated), and multiple cause of death file (IF requesting cases be ;
* identified using all causes of death on the death certificate). ;
* ;
* Additionally the user must enter the ICD codes in the 'ICD Codes to Identify ;
* Cases' section for the appropriate selection of cases. ;
* ;
* Lastly, the user must modify the 'RUN OPTIONS' macro variables to ;
* reflect the options required for the selection of the controls ;
* NOTE: ALL MODIFICATIONS TO THE RUN OPTIONS MUST BE IN UPPER CASE! ;
*****;

**-----**
** 1. Users Must Do !!! **
** You must specify the cohort file name. **
**-----**

*filename in 'c:\casecntl\sample_pvc.fil' lrecl=1600; * IF PVC FORMAT;
LIBNAME IN 'L:\GROUP\HERB\TRANSFER\CASECNTL'; * IF DEDS FORMAT;
```

NIOSH Case-Control Selection Program Documentation

```
**-----**;  
** 2. Users Optional. **;  
** Specify the dose exposure file name if **;  
** requesting exposure variables to be calculated. **;  
** If not using dose, comment this statement out **;  
** by placing a '*' infront of the word 'filename'.**;  
**-----**;  
*filename dose 'c:\casecntl\sample_dose.fil' lrecl=80;  
  
**-----**;  
** 3. Users Optional. **;  
** Specify file name for cause of death, only **;  
** need to do this if below option COD=ALL and **;  
** INPUT=PVC. **;  
** If not using this option, comment this statement**;  
** out by placing a '*' infront of the word **;  
** 'filename'. **;  
**-----**;  
*filename cod 'd:\abc\rawdata\your filename';  
  
**-----**;  
** 4. Users MUST do!!! **;  
** Enter the ICD Codes and Year of Death ranges **;  
** for indentification of Cases. **;  
**-----**;  
  
*--- ICD Codes to Identify Cases ---*;  
  
data ICD;  
  INPUT @1  LOW_YEAR  4.  
        @6  HIGHYEAR  4.  
        @11 LOW_ICD   $4.  
        @16 HIGH_ICD  $4. ;  
  
cards;  
1950 1967 204  2049  
1968 1978 204  2079  
1979 1998 204  2089  
;  
*****;  
* 5. Users Must enter/select RUN OPTIONS: *;  
*****;  
%LET TASK=TEST; * TASK REQUEST NUMBER;  
%LET PROGRAMR=NOWLIN; * PROGRAMMER'S NAME WHO RAN JOB;  
%LET PROJOFF=SCHUBAUER-BERIGAN; * STUDY PROJECT OFFICER'S NAME;  
%LET STUDYID=TES; * 3 CHAR. STUDY ID;  
%LET OUTDIR=c:\casecntl; * THE PC DIRECTORY TO STORE OUTPUT;  
* MUST USE ONLY ONE LEVEL SUBIDRECTORY;  
  
%LET TITLE=TEST RUN OF THE CASE/CONTROL PROGRAM;  
  
%LET INPUT=DEDS; * (PVC OR DEDS);  
%LET CCTYPE=F; * (S=SAMPLE MATCHED SET, F=FULL MATCHED
```

NIOSH Case-Control Selection Program Documentation

```

                                                    SET);
%LET MAXCNTLS=0;          * MAX # CONTROLS PER CASE
                          (SAMPLE MATCHED SET ONLY);
%LET COD=UCOD;          * CAUSES OF DEATH (UCOD OR ALL);
%LET STUDYBEG=MDY(01,01,01); * STUDY BEGIN DATE;
%LET STUDYEND=MDY(12,31,96); * STUDY END DATE;
%LET DLO_OPT=N;        * USE DLO TO CALC AGE (Y OR N);
%LET MIN_EMP=0;        * MIN LENGTH EMPLOYMENT (DAYS);
%LET TSFEMP=0;        * TIME SINCE FIRST EMPLOYED (DAYS);
%LET ELIG_OPT=N;      * USE ELIGIBILITY DATE? (Y OR N);
%LET SEX=N;           * MATCH ON SEX;
%LET RACE=N;         * MATCH ON RACE;
%LET RACE_REC=N;     * RECODE RACE TO WHITE AND NON-WHITE;
%LET DOBDAYS=0;     * +- # OF DAYS TO MATCH ON DOB;
%LET DFEDAYS=0;     * +- # OF DAYS TO MATCH ON DFE;
%LET DOSEXP=A;      * (P=PERSONAL, A=AREA, BLANK=NO DOSE);
%LET DOSEFRMT=P;    * (P=PVC FORMAT, L=PC LTAS FORMAT);
%LET SEED=;         * SAMPLE MATCHED SET ONLY. BLANK IF
                    1ST RUN, ENTER SEED # FROM PREVIOUS
                    RUN IF RE-RUNNING TO SELECT SAME;
%LET AUC=N;         * CALCULATE AREA UNDER CURVE (Y or N);
%LET HALFLIFE=;     * EXPOSURE HALF LIFE (UNIT IN YEAR);
*****;
```

```
filename pgm 'k:\group\casecntl\pgmlib';
```

```

**-----**
** Submit the Case Control Selection Program **
**-----**
```

```
%INCLUDE PGM(CASECNTL);
```

```
OPTIONS NOTES;
```

```
footnote;
```

```
title;
```

```
RUN;
```


IX. Reports

The following reports are produced with each run of the Case-Control Selection Program:

- Run Parameters Report (indicates which options were used)
- Summary Report (summarizes total number of cases and controls selected and rejected)
- Rejection & Warnings Summary Table
- Rejection Report
- Sample Warnings Report
- Control Audit Report
- Sample Listings of the final case control files
- Proc Contents of the final case control output SAS library

APPENDIX A. PVC Input File Format

Cohort File

<u>FIELD DESCRIPTION</u>	<u>FORMAT</u>	<u>POSITION(S)</u>
Employee social security#	num 9	1-9
Employee's Last Name	char 14	10-23
Sex code 1 = male 2 = female 3 = unknown	num 1	33
Race code 1 = white 2 = black 3 = oriental 4 = other 5 = unknown 6 = hispanic	num 1	34
Employee's birth information		
birth month (MM)	num 2	35-36
birth day (DD)	num 2	37-38
birth year (CYY)	num 3	39-41
Vital status 1 = alive 3 = unknown 5 = dead	num 1	116
Employee's death information		
death month (MM)	num 2	117-118
death day (DD)	num 2	119-120
death year (CYY)	num 3	121-123
underlying cause of death (left justified)	alphanum 4	124-127
contributory cause of death # 1 (left justified)	alphanum 4	128-131
contributory cause of death # 2 (left justified)	alphanum 4	132-135
General employment information		
date first employed (MMDDYY)	num 6	246-251
date last employed (MMDDYY)	num 6	252-257
plant number	alphanum 2	260-261
Detail employment histories (Repeats up to 80 times)		
department	alphanum 2	262-263,etc
operation	alphanum 2	264-265;etc
beginning date (MMDDYY)	num 6	266-271,etc
ending date (MMDDYY)	num 6	272-277,etc
Number of detail employment histories	num 2	1542-1543
Date last observed (MMDDYY)	num 6	1546-1551
Person Years Begin Date (MMDDYY)	num 6	1570-1575

APPENDIX A. PVC Input File Format (cont'd)

Multiple Cause of Death File

<u>FIELD DESCRIPTION</u>	<u>FORMAT</u>	<u>POSITION(S)</u>
Social Security Number	num 9	1-9
First Four characters of last name	char 4	10-13
Number of causes of death in array	num 2	14-15
Array of 20 elements: ICD1-ICD20, each element is 5 characters long, the value of the first element is the underlying cause of death.	char5 (occurs 20 times)	16-115

APPENDIX B. DEDS Input File Format

Dataset Name: DEMOGR				
Variable Name	Variable Description	Type	Length	Valid Values
NIOSH_ID	Assigned ID number used by NIOSH's DEDES to manage and link relational data files. The value of this field is set per study and is a unique number per individual in the cohort.	\$	9	
SSN	Social Security Number	N	9	
LAST	Last Name	\$	22	
FIRST	First Name	\$	15	
MIDDLE	Middle Name	\$	15	
PREFIX	Name Prefix	\$	4	MR MS MRS MISS DR REV
SUFFIX	Name Suffix	\$	3	JR SR I II III IV V
OTH_IDNO	Other Major Individual ID Number	\$	10	
DOB	Date of Birth	\$	8	00 coded in Month or Day indicates an unknown value. Year must be a valid century & year value based on study specifications/estimates for earliest and latest birth year. Month range: 1 - 12 Day range 1 - 31
BRTHCITY	Birth City	\$	23	
BRTHST	Birth State	\$	2	Post Office State abbreviation or If territory: AS=American Samoa CZ=Canal Zone FM=Federated States of Micronesia

NIOSH Case-Control Selection Program Documentation

Dataset Name: DEMOGR				
Variable Name	Variable Description	Type	Length	Valid Values
				GU=Guam MH=Marshall Islands MP=Northern Mariana Islands PR=Puerto Rico PW=Palau TT=Trust Territories VI=Virgin Islands or UK=Unknown
BRTHCNTY	Birth Country	\$	10	
DOD	Date of Death	\$	8	See DOB.
DTHSTATE	The verified state in which the worker died.	\$	2	See BRTHST field. Additional valid values for Death State only: FN=Foreign Death WD=War Death
DCR	Death Certificate Received	\$	1	Y=Yes N=No U=Unknown
DTHSRCE	Source used to obtain Death Information.	\$	2	See SOURCE CODE values
VS	Vital Status	\$	1	A=Alive D=Dead
DLO	Date Last Observed	N	8	Valid date (SAS Date)
DLOSRC	Source used to obtain Date Last Observed information	\$	2	See SOURCE CODE values
SEX	Sex Code	\$	1	M=Male F=Female U=Unknown
RACE	Race Code	\$	1	W=White B=Black O=Other H=Hispanic I=American Indian S=Asian U=Unknown
HEIGHT	Height in inches	N	3	
WEIGHT	Weight in pounds	N	3	
MARITAL	Marital Status Code	\$	1	S=Single Never Married M=Married D=Divorced P=Separated

NIOSH Case-Control Selection Program Documentation

Dataset Name: DEMOGR				
Variable Name	Variable Description	Type	Length	Valid Values
				W=Widowed U=Unknown
MILITARY	Military Status Code	\$	1	N=None U=Unknown 1=Active 2=Reservist 3=4F 4=Vet (not Vietnam) 5=Disabled Vet 6=Vietnam Vet 7=Disabled Vietnam Vet
EDUCATN	Education Code indicating highest level of education attained	\$	1	N=None 1=<9 years 2=9 - 12 years 3=High school Graduate 4=< 2 years College, No Degree 5=Associate Degree 6=>2 years College, No Degree 7=Bachelor Degree 8=Advanced Degree U=Unknown
EMPSTAT	Employment Status as of Study End Date (Active/Inactive)	\$	1	A=Active I=Inactive U=Unknown
PYRSBEG	Person Years Begin Date	n	8	Valid Date (SAS Date)

Dataset Name: WORKHIST				
Variable Name	Variable Description	Type	Length	Valid Values
NIOSH_ID	Assigned Id number used by NIOSH's DEDS to manage and link relational data files. The value of this field is set per study and is a unique number per individual in the cohort.	\$	9	
AREA	Area Name Abbreviation	\$	4	
DEPT	Department Abbreviation	\$	4	
OPER	Operation/Jobcode Abbreviation	\$	4	
BEGIN	Begin Date	\$	8	See DOB of Worker Demographic file.

NIOSH Case-Control Selection Program Documentation

Dataset Name: WORKHIST				
Variable Name	Variable Description	Type	Length	Valid Values
END	End Date	\$	8	See DOB of Worker Demographic file.
LOA_REAS	Leave of Absence Reason Code	\$	1	A=Maternity D=Disability/Medical E=Education/Research/Training F=Family Leave J=Jury Duty L=Laid off M=Military P=Personal S=Sick Leave T=Terminated U=Unknown V=Vacation W=Work Place Injury X=Shut Down/Strike
SHIFT	Shift Code	\$	1	1=First Shift 2=Second Shift 3=Third Shift 4=Fourth Shift R=Rotating O=Other U=Unknown
PAYCODE	Pay Code/Status	\$	1	E=Salary - Exempt H=Hourly N=Salary - Non-exempt S=Salary - Unknown O=Other U=Unknown

Dataset Name: COD (Cause of Death)				
Variable Name	Variable Description	Type	Length	Valid Values
NIOSH_ID	Assigned ID number used by DEDS to manage and link relational datasets.	N	6	
UCOD	Underlying Cause of Death	\$	8	
WPID	Work Place Injury Death	\$	1	Y=Yes N=No U=Unknown
CAUSE1	Contributing Cause of Death 1	\$	5	
CAUSE2	Contributing Cause of Death 2			
CAUSE3	Contributing Cause of Death 3			

Dataset Name: COD (Cause of Death)				
Variable Name	Variable Description	Type	Length	Valid Values
CAUSE4	Contributing Cause of Death 4			
CAUSE5	Contributing Cause of Death 5			
CAUSE6	Contributing Cause of Death 6			
CAUSE7	Contributing Cause of Death 7			
CAUSE8	Contributing Cause of Death 8			
CAUSE9	Contributing Cause of Death 9			
CAUSE10	Contributing Cause of Death 10			
CAUSE11	Contributing Cause of Death 11			
CAUSE12	Contributing Cause of Death 12			
CAUSE13	Contributing Cause of Death 13			
CAUSE14	Contributing Cause of Death 14			
CAUSE15	Contributing Cause of Death 15			
CAUSE16	Contributing Cause of Death 16			
CAUSE17	Contributing Cause of Death 17			
CAUSE18	Contributing Cause of Death 18			
CAUSE19	Contributing Cause of Death 19			
CAUSE20	Contributing Cause of Death 20			

APPENDIX C. Dose File Format

<u>Field Description</u>	<u>Format</u>	<u>Positions</u>
Study ID	Char 3	1-3
Chemical ID	Char 4	4-7
Dosage Type P=Personal A=Area	Char 1	8
Dosage Key Personal = SSN (zero filled) Area = Plant (Char 3), Dept (Char 3), Operation (Char 3)	Char 9	9-17
Dose Period Begin Date (mmddy format)	Num 6	18-23
Dose Period End Date (mmddy format)	Num 6	24-29
Dose Factor (<i>implied decimal after 3 positions</i>)	Num 6	30-35

NOTE:

This dose file format is taken from the NIOSH LTAS system. Enhancements/modifications to the date fields to expand the year value have been listed in the 'Future Modification Requests', Appendix F.

APPENDIX D. Case-Control Selection Output Files

Four SAS files are created:

- 1) **CASECNTL:** Case & Control Demographic Data (one record per Case/Control match)
- 2) **CC_WH:** Case & Control Work History Data (potentially multiple detailed work history records per Case/Control match)
- 3) **WARNINGS:** Identifying information about any records that received warning messages
- 4) **RUN_OPTS:** Options chosen by the user in running the Case Control Selection program to select cases and their matched controls.

FILE 1: CASECNTL: Case & Control Demographic Data

<u>Field Name</u>	<u>Field Description</u>
NIOSH_ID	Unique record identifier. (Assigned by the Case Control Selection program if using the PVC input option)
CASE	Case identifier (1=case, 0=control)
SSN	Social Security Number of the individual.
CASE_ID	NIOSH_ID of the case in a riskset. This number links the riskset together.
LAST	Last Name of the individual.
SEX	Sex Code
RACE	Race Code
AGE	Age of the individual at death or Study End date, which ever is earlier. If the DLO option is chosen then the age is at Date Last Observed if less than the Study End date. (YEARS)
AGE_DFE	Age at Date First Employed (YEARS)
CASE_AGE	Age of the case at death. (YEARS)
DOB	Date of birth.
DOD	Date of death.
UCOD	Underlying Cause of Death.
CCOD2-CCOD20	Contributing Cause(s) of Death (only included in the final file if the Cause of Death/Disease option equals >ALL=)
DFE	Date First Employed (Missing if DFE > CUTOFF).
DLE	Date Last Employed up to the age the case failed (Missing if DFE >CUTOFF).

APPENDIX D. Case-Control Selection Output Files (cont'd)

NIOSH Case-Control Selection Program Documentation

DLO	Date Last Observed.
ENTRYDTE	Entry date.
PYRS	Person Years Begin date.
CUTOFF	Date the individual reached the age of the case. This date is used to limit ("cut off") the length of employment once the individual reached the age of the case. Also used to cut off the calculation of cumulative exposure and duration exposed.
TSFE	Time Since First Employed
DFEXP	Date First Exposed (Missing if first exposure date > CUTOFF)
DLEXP	Date Last Exposed up to the age the case failed (Missing if first exposure date > CUTOFF)
AGE_FEXP	Age at date first exposed. (YEARS)
AGE_LEXP	Age at date last exposed. (YEARS)
DUREXP	Duration exposed up to the age the case failed.
CUMEXP	Cumulative exposure up to the age the case failed.
EXPRATE	Exposure Rate. (CUMEXP / DUREXP)
PEAKEXP	Peak exposure. Highest exposure received in any work history as of the age the case failed.
LATENCY	Time since first exposed. (YEARS)
TSLEXP variable]	Time since last exposed. (YEARS) [See TSLE_FLG for explanation of calculation of this
TSLE_FLG	1=Time since last exposed as of the age the case failed: (CASE_AGE - AGE_LEXP) 2=Not exposed as of the age the case failed, used DFE: (CUTOFF - DFE) / 365.25 3=Not exposed as of the age the case failed and DFE > CUTOFF, therefore, TSLEXP=missing.
(LAG variables)	
LAG5	Date 5 years prior to the date the individual reached the age of the case. (CUTOFF - 5 YEARS)
DUREXP5	Duration exposed lagged 5 years.
CUMEXP5	Cumulative exposure lagged 5 years.
EXPRAT5	Exposure rate lagged 5 years.
<u>APPENDIX D.</u> Case-Control Selection Output Files (cont'd)	
PEAKEX5	Peak exposure lagged 5 years.

NIOSH Case-Control Selection Program Documentation

LAG10	Date 10 years prior to the date the individual reached the age of the case. (CUTOFF - 10 YEARS)
DUREXP10	Duration exposed lagged 10 years.
CUMEXP10	Cumulative exposure lagged 10 years.
EXPRAT10	Exposure rate lagged 10 years.
PEAKEX10	Peak exposure lagged 10 years.
LAG15	Date 15 years prior to the date the individual reached the age of the case. (CUTOFF - 15 YEARS)
DUREXP15	Duration exposed lagged 15 years.
CUMEXP15	Cumulative exposure lagged 15 years.
EXPRAT15	Exposure rate lagged 15 years.
PEAKEX15	Peak exposure lagged 15 years.
LAG20	Date 20 years prior to the date the individual reached the age of the case. (CUTOFF - 20 YEARS)
DUREXP20	Duration exposed lagged 20 years.
CUMEXP20	Cumulative exposure lagged 20 years.
EXPRAT20	Exposure rate lagged 20 years.
PEAKEX20	Peak exposure lagged 20 years.

FILE 2: CC_WH: Case & Control Detailed Work History Data

<u>Field</u>	<u>Field Description</u>
NIOSH_ID	Unique record identifier. (Assigned by the Case Control Selection program if using the PVC input option)
CASE_ID	NIOSH_ID of the case in a riskset. This number links the riskset together.
SSN	Social Security Number of the individual.
CUTOFF	Date the individual reached the age of the case. This date is used to limit ("cut off") the length of employment once the individual reached the age of the case. Also used to cut off the calculation of cumulative exposure and duration exposed.
AREA	Work Area Code
DEPT	Department Code

NIOSH Case-Control Selection Program Documentation

OPER	Operation Code
BEG	Work History Begin Date
END	Work History End Date
DOSE	Dose per day factor.

NIOSH Case-Control Selection Program Documentation

If **Yes**, what type of dose file will be used?

P = Personal dose

A = Area dose

12) Calculate Area Under the Curve?

Yes No

If **Yes**, enter _ = _____

Control Criteria:

12) Match Controls to Cases by Sex?

Yes

No

13) Match Controls to Cases by Race?

Yes

No

14) Do you want the Race code to be recoded to WHITE and NON-WHITE only?

Yes

No

15) Match Controls to Cases by Date of Birth?

Yes

No

If >YES=, what is the maximum number of **days** difference permitted to be considered a match? _____

16) Match Controls to Cases by Date First Employed?

Yes

No

If >YES=, what is the maximum number of **days** difference permitted to be considered a match? _____

Seed Value

Enter the seed value if rerunning a previously created Sample Matched Set Case Control File (in order to obtain the same controls selected per case): _____

APPENDIX F. Program Modification Requests (as of 3/2003)

Modifications/Enhancements to be made:

1. Enable the dynamic creation of exposure LAG variables. Give the option to select the number of LAG break points and the values (length of time) of each LAG break point.
2. Enable the pool of eligible controls (or non-cases) to be determined in two different ways:
 - i. (Current default) Non-cases are eligible to be in the pool of controls if they reached the case=s age at death/disease and their age at entry to the study (Entry Date) is less than or equal to the case=s age at death/disease.
 - ii. (Cut off date = DOB - Case=s age at death/disease)
3. (new option) Non-cases are matched to cases by Date of Birth (DOB), they are then eligible to be in the pool of controls if they reached the case=s date of death/disease.
 - i. (Cut off date - Case=s Date of Death/Disease)
4. Enable the option to use a global eligibility date to establish entry date (instead of only being able to specify the eligibility date within each record of the cohort).
5. Redesign the Dose file format to expand data variables to include century. Modify programs to incorporate.
6. Add counter matching capabilities.
7. Add HEDS input file structure.

Epi Issues:

If there are not enough people in the pool of eligible controls to meet the maximum number of controls to select in a Sample Matched Set, should the case be rejected?

(For example, if the user requested that 5 controls be selected per case but only 4 people meet the requirements to be in the pool of eligible controls, should the matched set for that case contain those 4 controls while other cases contain 5 controls per case?)

APPENDIX G. Glossary of Terms

Glossary of Terms	
Term	Definition
Case	A persons having the disease being investigated in a case-control study.
Case-Control Study	A scientific study that compares a group of people with disease, such as leukemia, to a similar group of people without that disease. This type of study compares the levelof exposure (radiation or chemical) each group had prior to the appearance of disease.
Contributing Cause of Death/Disease	Diseases or other factors which contributed to the underlying cause of death for an individual.
Control	Individuals selected as a comparison group for cases in a case-control study. <ol style="list-style-type: none"> 1) Controls should be free of the study disease and should be similar to the cases in regard to past potential for exposure during the time period of risk under consideration. 2) (NIOSH) Members of the population eligible to serve as control for the index case must have lived to an age greater than the age that the index case failed.
Cut Off Date	Date the Control reached the age of the Case at death/disease
DEDS file format	NIOSH DSHEFS Epidemiological Database System. A relational file structure developed by DSHEFS/HERB IT specialists in 1991 using SAS Software to manage HERB epi research data.
DFEMP	Date First Employed
DFEXP	Date First Exposed
DLE	Date Last Exposed
DLO	Date Last Observed Alive
DOB	Date of Birth
DOD	Date of Death
Duration Exposed	The accumulated time a worker was exposed to any particular agent while working at a site/plant.
Eligibility Date	The date at which a cohort member is eligible to be a control in a case-control study (i.e., the control had to have entered the study before the control attained the age at which the index case failed, so as to be at risk)
Entry Date	The earliest date at which a worker can enter the study.
Full Matched Set	The full set of matched controls for a given case, i.e., the pool of eligible controls for a case.
ICD	International Classification of Disease

Glossary of Terms	
Term	Definition
Incidence Density Sampling	A sampling design used in nested case-control studies.
Lag	The time between the initial exposure and the effect of that exposure.
LTAS	NIOSH Life Table Analysis System
NIOSH ID	Unique ID used in the NIOSH DEEDS databases as the primary key to link related tables.
Non-Case	Persons who lived longer than the case's age at death/disease
PVC File Format	A legacy data structure developed in the late 1970's to manage NIOSH/DSHEFS epidemiological study data..
Risk Set	The pool of eligible controls for a case.
Sample Matched Set	A randomly selected sample of controls for a given case taken from the entire pool of eligible controls.
SAS Dataset	A collection of information stored in a form that the SAS System can use.
SAS Library	A shortcut name or pointer to a storage location on a specific operating environment where SAS files are stored. The storage location (directory) contains one or more SAS files, each file being a member of the library.
Seed Value	Numeric value used to initialize the random number generator function which supplies random numbers to each control in a case's riskset, thereby enabling the random selection of a sample set of controls for a case. By default, this value is automated using the date and time the case control selection program was run.
Study Begin Date	The date workers had to have been employed during or after in order to be part of the study.
Study End Date	The date workers had to have been employed before in order to be part of the study.
TSFEMP	Time Since First Employed
TSFEXP	Time Since First Exposed
UCOD	Underlying Cause of Death/Disease

APPENDIX H. **Source Code**