RSP-0283

SUMMARY REPORT OF THE ONTARIO URANIUM MINERS COHORT MORTALITY AND CANCER INCIDENCE STUDY

MORTALITY LINKAGE: 1954 TO 2007
CANCER LINKAGE: 1969 TO 2005

Health Record Linkage Section
Health Statistics Division
Statistics Canada

July 2012

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This report describes the record linkage work carried out at Statistics Canada for an epidemiological mortality and cancer study of 30,914 Ontario Uranium Miners who worked for at least one week between the years 1954 to 2004.

The following record linkages have been carried out: (1) a linkage to the Historical Summary Tax File (HSTF) for the event years 1984 to 2009 inclusive; (2) a linkage to the Canadian Mortality Data Base (CMDB) for event years 1954 to 2007 inclusive; and (3) a linkage to the Canadian Cancer Data Base (CCDB) for event years 1969 to 2005 inclusive.

The cohort file was linked to the HSTF to determine the vital status of the individuals at the end of the follow-up period and to enhance the cohort file to facilitate the linkages to the mortality and cancer files. This was possible for 75 % of individuals who had a social insurance number (SIN) either supplied or found through the use of a composite key generated as part of the work process. On the original cohort file, 63% of the records had a valid SIN supplied, while the "exact key" match generated a SIN for another 12 % of individuals. Records which matched on social insurance number were then confirmed using other identifying information, such as names and birth dates.

Some 386 records were excluded from further processing due to insufficient identifiers or duplication. Therefore, 30,528 records were linked to the CMDB and the CCDB. Approximately 12,000 potential links to the 1954-2007 CMDB were reviewed. During the manual resolution phase, the team consulted detailed reports of common identifiers from the Ontario Uranium Miners Cohort and the CMDB.

In total, 8,794 (28.8) mortality links were found. There were 97 deaths identified on the HSTF and not in the mortality linkage. These could potentially be individuals who died outside of Canada. Some, 16,455 individuals (53.9 %) were determined to be alive at the end of the follow-up period and 5,182 (16.9 %) were lost to follow-up prior to the end of 2007.

Finally, a cancer linkage was carried out using the CCDB for the years 1969 to 2005. The pockets and rules used for this linkage were similar to those used for the mortality linkage. After manually resolving the potential links, there remained 4,320 accepted links to the CCDB.

1. INTRODUCTION

A national mortality and cancer linkage of the Ontario Uranium Miners Cohort was recently completed at Statistics Canada. The Ontario Uranium Miners Cohort consisted of 30,914 employees who worked at least one week from 1954 to 2004 in Ontario.

The purpose of this study is to further investigate the health outcomes of uranium miners in Ontario. The longer follow-up period for this study will increase the statistical power. The linkage to the cancer database will provide the study with greater details on cancer diagnosis information and the incidence of cancer.

1.1. OBJECTIVES

Specifically this study aims to:

1. Investigate the overall mortality and cancer incidence among uranium miners compared to the Canadian and Ontario populations;
2. Investigate the dose-response relationship of radon decay products (RDP) to health outcomes, taking into consideration factors such as the exposure rate, time since exposure, age at exposure and occupational code;
3. Identify associations between two forms of ionizing radiation found in uranium mines, radon decay products and gamma radiation, to cancer incidence and mortality;
4. Determine the relationship between gamma ray dose, and mortality and cancer incidence.

Under the Nuclear Safety Control Act (NSCA), the Canadian Nuclear Safety Commission (CNSC) is required to collect and monitor occupational radiation doses of nuclear energy workers in Canada in the National Dose Registry, and to use the registry for epidemiological research. The CNSC is required under this legislation to disseminate objective scientific information about the health effects on workers in the nuclear industry. The statistical information generated from this linked database will enable the CNSC to meet this obligation.

1.2. BACKGROUND

Underground uranium miners are occupationally exposed to radon and its radioactive radon decay products (RDP), which have been identified as the second most important cause of lung cancer after smoking. Past studies of underground uranium miners have found elevated risks of lung cancer.

Though the elevated health risk from exposure to radon and RDP is known, further understanding of the dose-response relationship is required. The United Nations Scientific

Committee on the Effects of Atomic Radiation (UNSCEAR, 2008) (6) and the Biological Effects of Ionizing Radiation Committee (BEIR)VI (4) have recommended the follow-up of existing miner cohort studies to fully understand the health effects of radon and its decay products in workplaces and homes, over time, and at different exposure levels.

Individuals in the Ontario Uranium Miners cohort have been linked to the Canadian Mortality Data Base in prior studies. The first study on Ontario miners was carried out by Statistics Canada in 1975, at the request of the Royal Commission on the Health and Safety of Workers in Mines, using a list of Ontario uranium miners. Their data were linked to the Statistics Canada mortality data base for 1955 to 1973. In 1979 and 1983, this linkage study was expanded to include Ontario uranium and non-uranium miners, and workers in other high-risk occupations. The mortality linkage was extended up to 1981.

The linkage methodology was improved in the 1980s and 1990s by including linkage to the Social Insurance Number master file and to income tax files to validate the mortality linkages. Data on the cohort of miners and workers in other high-risk occupations was linked for 1950 to 1993. These cohort studies have been the main sources of information on health risks from radon exposures and are internationally-recognized for their contribution to the radiation risk literature. (7, 8, 9) Due to the findings from these earlier studies, significant improvements were introduced to lower the exposure of miners to radiation. These changes included better ventilation and mining techniques, as well as the use of individual dosimetries and dose limits.

2. APPROVALS AND CONFIDENTIALITY

The study received ethical approvals from the Health Canada, Research Ethics Board on September 6, 2007 (REB-2007-0015).

The application to conduct mortality and cancer record linkages was approved by the Policy Committee at Statistics Canada.

The provincial and territorial vital statistics and cancer registrars approved the release of the de-identified linked mortality and cancer data to Cancer Care Ontario.

The confidentiality of individuals in the cohort has been protected through the creation of a de-identified mortality and cancer analysis files to be released to Cancer Care Ontario. The principal investigators at Cancer Care Ontario have signed a Statistics Canada Preservation of Confidentiality Statement which outlines the conditions under which they will receive the data, including the use of data limited to the objectives of the study and the secure storage and access of the data.

3. THE FILES

This section describes the various files that were used during the linkage activities of this study.

3.1.ONTARIO URANIUM MINERS COHORT FILE

The Ontario Uranium Miners Cohort consisted of 30,914 employees who worked at least one week from 1954 to 2004. This file is maintained by Cancer Care Ontario. The file structure contained the following identifiers for linkage: surname; first given name or initial only; second given name or initial only; sex code; birth date, country of birth; and last known date alive. The record layout is shown in section 7.1.

The National Dose Registry at Health Canada provided directly to Statistics Canada a file containing the social insurance numbers and miner's cohort study numbers for the Ontario Uranium Miners Cohort.

The Ontario Uranium Miners Cohort File was enhanced by adding the social insurance numbers (SIN) from the National Dose Registry File.

The distribution of the sex code and birth year variables from the cohort is shown in Table 1.

TABLE 1: DISTRIBUTION OF THE URANIUM MINERS COHORT, BY YEAR OF BIRTH AND SEX

Year of birth	Males		Females		Unknown		Total	
	Number	%	Number	%	Number	%	Number	%
Not stated	149	0.50	0	0.00	625	85.38	774	2.50
Prior to 1900	30	0.10	0	0.00	0	0.00	30	0.10
1900 to 1909	429	1.44	0	0.00	1	0.14	430	1.39
1910 to 1919	1,906	6.42	0	0.00	0	0.00	1,906	6.17
1920 to 1929	5,277	17.77	9	1.85	2	0.27	5,288	17.11
1930 to 1939	7,933	26.71	35	7.20	8	1.09	7,976	25.80
1940 to 1949	5,052	17.01	74	15.23	13	1.78	5,139	16.62
1950 to 1959	6,948	23.40	211	43.42	18	2.46	7,177	23.22
1960 to 1969	1,860	6.26	149	30.66	44	6.01	2,053	6.64
1970 to 1979	109	0.37	8	1.65	21	2.87	138	0.45
1980 to 1989	3	0.01	0	0.00	0	0.00	3	0.01
Total	29,696	100.00	486	100.00	732	100.00	30,914	100.00

3.2. HISTORIC SUMMARY TAX FILE (HSTF)

The Historic Summary Tax File (HSTF) contains approximately 33 million records and covers the years 1984 to 2009 of income tax returns. The file contains no financial data; rather, it only includes person identifiers and the minimum amount of data required to ascertain the vital status and location of individuals. The record layout is shown in section 7.2.

The validity of a cohort study depends fundamentally on complete ascertainment of the events of interest (e.g. cancer incidence, deaths) and the vital status of all cohort members. An evaluation of the extent to which follow-up losses have occurred is important, documentation of low loss rates adds to the credibility of the results.

The Historic Summary Tax File (HSTF) was used to:

- Enhance the Ontario Uranium Miners Cohort for linkage by filling in data gaps.
- Determine the vital status of the cohort members at the end of the follow-up period.
- Evaluate the results of the mortality linkage.

3.3. CANADIAN MORTALITY DATABASE (CMDB)

The Canadian Mortality Database (CMDB) contains all deaths registered in Canada since 1950. The size of the CMDB grows by approximately 200,000 death records each year. For the period 1954 to 2007, the database contains approximately 12.7 million death records.

The source of data for the CMDB is the vital statistics program at Statistics Canada that routinely collects demographic and medical (cause of death) information from all provincial and territorial vital statistics registries on all deaths in Canada. Some data are also collected on Canadian residents who die in some American states.

The CMDB is specially formatted to facilitate record linkage. Processing includes standardizing certain variables (e.g., names) and deriving variables required for record linkage such as the phonetic NYSIIS (New York State Individual Intelligence System) surname codes. Duplicate records, which contain data variations, such as alternate surnames, may be created during processing of the file.

The CMDB contains coded information on the underlying cause of death of the individual. This cause of death is coded using the version of the World Health Organization's *International Classification of Diseases* (ICD) that was in effect at the time of death.

ICD version	Death years
ICD-6	1950 to 1957
ICD-7	1958 to 1968
ICDA-8	1969 to 1978
ICD-9	1979 to 1999
ICD-10	2000 and onward

The record layout for the CMDB is shown in section 7.3.

3.4.CANADIAN CANCER DATABASE (CCDB)

At the national level, cancer incidence data for the years 1969 to 1991 were collected through the National Cancer Incidence Reporting System (NCIRS) and for the years 1992 to the present they are collected through the Canadian Cancer Registry (CCR). The provincial cancer registries and the health authorities in the two territories send their cancer incidence records to Statistics Canada where they are edited and standardized. Statistics Canada also transforms the data into a format suitable for record linkage, thereby creating the Canadian Cancer Database (CCDB). The CCDB is an historical electronic file, which at the time of this study contained approximately 5 million cancer events from 1969 to 2005.

The record layout for the CCDB is shown in section 7.4.

4. METHODOLOGY

4.1.OVERVIEW

Upon the receipt of the cohort files from Cancer Care Ontario and the National Dose Registry, Statistics Canada carried out the following steps:

- Assessment of the Ontario Uranium Miners Cohort;
- Linkage of the Ontario Uranium Miners Cohort to the 1984-2009 Historic Summary Tax File;
- Pre-processing of the Ontario Uranium Miners Cohort;
- Linkage of the Ontario Uranium Miners Cohort to the Canadian Mortality Data Base;
- Linkage of the Ontario Uranium Miners Cohort to the Canadian Cancer Data Base.

An overview of the process is shown in Figure 1.

FIGURE 1: FLOW DIAGRAM OF THE MORTALITY AND CANCER LINKAGE OF THE ONTARIO URANIUM MINERS COHORT

4.2.ASSESSMENT OF THE ONTARIO URANIUM MINERS COHORT

An initial quality assessment was carried out on the cohort to determine the completeness, and the validity of the data. This procedure serves to identify data errors (such as invalid dates) or omissions that may prevent the correct linkage of a record. The quality assessment also allows the staff at Statistics Canada to become more familiar with the data in the cohort file, which is required prior to developing and refining the linkage strategy.

The availability of linkage items is shown in Table 2.

TABLE 2: ONTARIO URANIUM MINERS COHORT: PERCENTAGE OF DATA AVAILABLE FOR VARIABLES.

Title	Available	
	Yes (%)	No (%)
Miners cohort study code	100.00	-
Miner study identification number	100.00	-
Surname (required for all linkages)	96.60	3.40
Surname initial only	0.00	-
Other surname used - Complete	6.10	93.90
Other surname initial only	0.00	-
First given name (required for all linkages)	91.10	8.90
First initial only	5.50	-
Second given name – Complete	51.60	48.40
Second initial only	10.80	-
Sex (required for all linkages)	97.60	2.40
Other sex code	0.05	99.95
Birth date (Century, Year, Month, Day) (required for all linkages)	96.20	3.80
Birth year only (Century, Year)	1.30	-
Other birth date – Complete	3.40	96.60
Other birth year	0.00	-
Birth province/country STC Code – Conversion made from the NDR record original entry	26.30	73.70
Birth province/country STC Code – Conversion made from the WHF record original entry	68.30	31.70
Work history earliest year	99.99	0.01
Work history latest year	99.99	0.01
Source: The Cancer Care Ontario document: Confirmation Report # 1, Contract 87055-08-0853, Preparation of Data for Linkage – Ontario Miner's Study Update, November 17, 2008. Appendix C: Availability of Information Form		

4.3.LINKAGE OF THE ONTARIO URANIUM MINERS COHORT TO THE HSTF

A social insurance number (SIN) was available for 63% of the cohort. These records were matched to the HSTF by SIN. In cases where the cohort records did not contain a SIN, a 19 character string of cohort information (first five characters of surname, first 4 bytes of the given name, year of birth, month of birth, day of birth) was used to match to the 1984 to 2009 HSTF.

Matches were then reviewed to verify that the records were, in fact, referring to the same individual. False matches were removed. For those cohort records with an accepted match, the following variables were appended to the cohort records: last tax year filed, filing province, date of death (if applicable), date of birth, surname, given names, postal code, standard geographical code and an historical tax filing flag vector. These variables were added to help with the record linkage process and also with the manual resolution of uncertain matches. As well, after completion of the mortality linkage, this data provided an indication of the number of individuals not linked to the CMDB that could be presumed "alive" as opposed to lost to follow-up.

After linking the cohort to the HSTF, a basic internal linkage was done to identify duplicate records. The manual resolution of duplicate records was conducted at Statistics Canada by deemed employees using detailed group reports of common items. One hundred and thirty-two records were reviewed and 96 records were classified as definite duplicates and excluded from the cohort. 36 records were deemed ‘possible’ duplicates and flagged.

An additional two hundred and ninety (n=290) records were deemed ineligible for linkage and excluded due to data gaps (i.e. incomplete dates of birth and name fields). Therefore the number of eligible records for linkage to the Canadian Mortality and Canadian Cancer Data Base was 30,528 records. Table 3 shows the distribution of birth year by sex of the final cohort file that was used in linkage.

TABLE 3: DISTRIBUTION OF THE FINAL URANIUM MINERS COHORT, BY YEAR OF BIRTH AND SEX

Year of birth	Males		Females		Unknown		Total	
	Number	%	Number	%	Number	%	Number	%
Not stated	116	0.39	0	0.00	326	75.99	442	1.45
Prior to 1900	29	0.10	0	0.00	0	0.00	29	0.09
1900 to 1909	428	1.45	0	0.00	1	0.23	429	1.41
1910 to 1919	1,903	6.43	0	0.00	0	0.00	1,903	6.23
1920 to 1929	5,273	17.81	9	1.85	1	0.23	5,283	17.31
1930 to 1939	7,925	26.76	35	7.20	8	1.86	7,968	26.10
1940 to 1949	5,042	17.03	74	15.23	13	3.03	5,129	16.80
1950 to 1959	6,928	23.40	211	43.42	17	3.96	7,156	23.44
1960 to 1969	1,857	6.27	149	30.66	43	10.02	2,049	6.71
1970 to 1979	109	0.37	8	1.65	20	4.66	137	0.45
1980 to 1989	3	0.01	0	0.00	0	0.00	3	0.01
Total	29,613	100.00	486	100.00	429	100.00	30,528	100.00

Once the cohort file was unduplicated and records with data gaps were omitted, the cohort file was preprocessed. Preprocessing involved four steps:

1. Edit and frequency checks to further identify potential errors;
2. Generate alternate records to improve the chances for a correct linkage;
3. Standardize common identifiers.
4. Expand the cohort by adding the results of the mortality linkage (i.e. date of death) to facilitate the cohort to cancer linkage.

During this phase all records were run through a sex code verification routine, which compared given names and sex codes on the cohort file to a Statistics Canada internal file that contained typically male and female names. An alternate version of the record was created for all records with doubtful given name/sex code combinations with the opposite sex code assigned. Alternate records were generated where the sex code was missing. The records on the cohort file were also put through a number of routines to verify and standardize the spelling of names. Where there was some question as to the accuracy of the data but confirmation was not available, an alternate or duplicate record was created which contained a variation of the data item in question. Expanding the data through the creation of alternate records increases the chances for a correct linkage.

5. BUILDING THE MORTALITY AND CANCER RECORD LINKAGE APPLICATIONS

5.1. PREPARING THE LINKAGE POCKET

As is the convention in probabilistic record linkage, it was necessary to identify a pocket within which records would be compared and pairs created. A pocket contains a group of records, which have a common value in the specified field. Using the defined pocket, as an example, records in File A, (cohort), would be compared only to records on File B, containing the same pocket value.

For this study, the NYSIIS code generated from the surname was used as the main pocket. The NYSIIS system is an encoding system that simplifies the spelling of surnames so that spelling variations from various source files might be eliminated when doing a comparison between records, thus improving the possibility of establishing a link. The NYSIIS code pocket available for this linkage included all records from the pre-processed cohort as well as the alternate records that were generated during the pre-processing of the file.

Two other variables were used as secondary pockets for this project, the birth date and birth year. Thus, any true matches that failed to satisfy the NYSIIS criteria on the first pocket pass were given a second and third chance to find a true link.

5.2. RULE DEFINITIONS AND OUTCOMES

The compare rules specified the relationships and outcomes resulting from each comparison. The two most obvious outcomes of a comparison are full agreement and disagreement. Where more than one relationship could be defined or where allowance was to be made for errors in the data, several levels of outcomes (or partial agreements) were defined.

Information required to define the rules and outcomes was obtained in part from the analysis and pre-processing of the files and in part from previous experience in mortality and cancer linkages of a similar nature.

The initial rules and outcomes were tested and refined using a sample of the data prior to running the complete mortality linkage. Some rules and relationships are obvious and are common to other projects. For example, one would expect that records which pertain to the same individual would probably agree on last name, given names and date of birth. Table 4 shows the availability of linkage items used in the comparison rules.

TABLE 4: PERCENTAGE OF DATA AVAILABLE FOR VARIABLES USED IN THE COMPARISON RULES

Variable	Ontario Uranium Miners cohort file % available	Canadian Mortality Database 1954-2007 % available
Last name	100.0	99.9
Alternate Last name	5.5	N/A
First given name	99.9	99.3
Second given name	64.0	51.7
Third given name (alternate)	2.9	3.0
Sex	100.0	99.9
Birth year	100.0	99.9
Birth month	100.0	95.3
Birth day	100.0	95.0
Country of birth	100.0	92.8

5.3. DEVELOPMENT OF WEIGHTS

The generalized record linkage system used at Statistics Canada is based on the mathematical theory of record linkage developed by Drs Ivan Fellegi and A.B. Sunter.[1]

Probabilistic record linkage methodology uses non-unique identifiers (e.g., name and birth date) to calculate the likelihood that records being matched refer to the same entity (e.g. person).

Comparisons between records are done field-by-field using comparison rules, with outcomes (for example, exact match, string proximity, missing information or fields disagree) generated by each rule based on the similarity of values in a pair of records. Each pair of records is assigned a comparison result pattern, and that pattern is evaluated to classify pairs as linked, possibly linked or not linked.

The theory of probabilistic record linkage is the premise that the results of certain comparison result patterns are characteristic of truly linked pairs, while other comparison result patterns are characteristic of truly unlinked pairs. Therefore, each rule outcome is assigned a weight based on the ratio of the estimated probability of the outcome occurring for true matches to the estimate probability of the outcome occurring for non-matches.

[1] Fellegi IP, Sunter AB. "A theory for record linkage", *Journal of the American Statistical Association*, American Statistical Association, December 1969, Vol. 64, No. 328, pp. 1183-1210.

The composition of the linked set is not known in advance, so the probabilities of result patterns for truly linked records are not known. Linked weight components are estimated from prior knowledge and early iterations of the linkage process, and refined by treating successive iterations of the linkage process on a test subset as generating better estimates of the truly linked set. The unlinked weight components were calculated based on the frequency with which the rule outcomes were observed among record pairs that do not belong together, which is approximately equal to the frequency with which the rule outcomes would be observed among randomly paired records. After repeated iteration of the linkage process, linked weight components stabilize and final weights are ready for use.

EXAMPLE

Suppose we have File A of N_A records and File B of N_B records. Then there are N_AN_B possible record pairs that could be created. If $N_A < N_B$ we expect to find at most N_A matches from File B, then of the N_AN_B possible record pairs, there would still be at least N_AN_B-N_A = $(N_B\text{-}1)N_A$ record pairs which are not true links.

5.4. COMPARISON PHASE

In this phase, the cohort file was compared to the Statistics Canada file (CMDB, CCDB) using the comparison rules, the weights were assigned based on rule outcomes, and the upper, lower and cut-off thresholds were set. The probabilistic record linkage process produces a total weight for a record pair which offers a quantitative measure of the likelihood two records refer to the same entity. Theoretically, record pairs with a weight above the upper threshold are accepted as “definite” links. Record pairs with a weight below a lower threshold are rejected as non-links. Record pairs with a weight that falls between the thresholds form the “grey zone”, which requires manual review and resolution.

The probabilistic record linkage process is often run iteratively, with refinements of weights and comparison rules between iterations. For this reason, the record pairs with a total weight below the lower threshold are still stored for further use. There is, however, a further cut-off threshold, below the lower threshold, below which the record pairs are not kept and are thus not available for re-evaluation in further iterations. This is done to save on computer resources.

Records were paired for comparison as a potential linkage only if they had the NYSIIS coding, or birth date agreement or birth year agreement, and these pairs were evaluated using the specified rules with their associated weights. Any pairs with a total weight falling below the cut-off threshold were not considered potential matches and were eliminated from further processing. The cut-off threshold for the mortality and cancer linkages was -150.

For both the mortality and cancer linkages, a large sample of record pairs over a range of total weight values was reviewed manually to decide whether each pair represented a true link or not. This process helped to refine the lower threshold settings, thereby raising its value to reflect where true links were discovered. After this process, the lower threshold was adjusted to 200 for the mortality and cancer linkages.

Record pairs were then grouped to bring together all pairs that referred to the same individual. In this way, all competing links could be evaluated collectively during manual resolution, and a decision made as to which, if any, was a good link.

5.5.MANUAL RESOLUTION

For both the mortality and the cancer linkages, manual resolution of potential links was completed on site at Statistics Canada by members of the study team who were sworn in as “deemed employees”.

The mortality manual resolution was carried out using detailed group reports containing the linked cohort to mortality records along with the results of the linked HSTF records. One member of the study team independently reviewed all links where the linkage weight was within the range of 250-985.

Upon completion of this review, links where the weight was less than or equal to 389, were classified as “grey” and subjected to an independent review by two other team members.

The following decision rules were followed during the review:

Links were accepted should the weight be equal to 303-385 and agreement was ascertained on five variables and partial agreement (± 2 years) on the date of birth or date of death variable.

Links were accepted should the weight be equal to 250-302 and agreement was ascertained on five variables and partial agreement (± 1 year) on the date of birth variable.

Once the separate reviews were complete two team members reviewed the resolution in order to verify decisions and resolve any discordance between the reviewers.

The cancer manual resolution was carried out using a detailed group report containing the linked cohort to cancer records along with the results of the alive follow-up and the mortality linkage. The decision rules were very similar to the mortality rules. Incidence manual resolution was carried out for links with a weight from 939 to 250. Two deemed employees did the manual resolution of cancer incidences linkage, blinded from the other person' work. These two employees are not involved otherwise in the study. The decision rules were similar to the

mortality manual resolution. Links were accepted should the total weight be equal to 250-302 and agreement was ascertained on five variables and partial agreement (± 1 year) on the date of birth variable. However, because of the added information (mortality and HSTF) the quality of the linkages was better than that of the mortality linkages so more information was available to base decisions on. Thus, decisions were easier to make.

5.6. QUALITY ASSURANCE

To assess the quality of the mortality linkage, records which did not link to the CMDB, but were deceased according to the results of the cohort to HSTF linkage, or to the cohort to CCDB linkage were subjected to a manual search. The online version of the CMDB and a secured Oracle application with limited-use access was used.

5.7. ANALYSIS FILES

The mortality and cancer analysis files were prepared upon completion of the manual resolution and finalization of the linkages. The mortality analysis file includes a Statistics Canada random sequence number, sex code, year and month of birth, birthplace code. For decedents: variables from the CMDB: the year and month of birth, date of death, province of death and the cause of death. The total linkage weight obtained during the linkage to the CMDB was also included. The record layout of the mortality analysis file for the years 1954-2007 is shown in section 7.5.

The cancer analysis file includes the Statistics Canada random sequence number and the following information from the CCDB for the period 1969-2005: ICD9 diagnosis, ICDO-T, ICDO-M, laterality, primary site, the date and province of diagnosis, province/country of residence and method of diagnosis. The linkage weight to the CCDB was also included. The record of the cohort cancer analysis file for the years, 1969-2005 is shown in section 7.6.

6. RESULTS

6.1. LINKAGE TO THE CANADIAN MORTALITY DATABASE

As shown in Table 4, of the 30,528 eligible individuals in the Ontario Uranium Miners cohort, a total of 8,794 were found in the Canadian Mortality Database for the death years 1954 to 2007. Table 5 provides a distribution of the year of death by sex and Table 6 shows the distribution by the province of death.

TABLE 5: MORTALITY IN THE ONTARIO URANIUM MINERS COHORT, BY YEAR OF DEATH AND SEX, 1954 TO 2007.

Year of death	Cohort links to the Canadian Mortality Database					
	Males		Females		Total cohort	
	Number of deaths	% male deaths	Number of deaths	% female deaths	Number of deaths	% total deaths
1954 to 1959	47	0.54	0	0.00	47	0.53
1960 to 1964	194	2.21	0	0.00	194	2.21
1965 to 1969	261	2.98	0	0.00	261	2.97
1970 to 1974	409	4.66	0	0.00	409	4.65
1975 to 1979	595	6.78	0	0.00	595	6.77
1980 to 1984	775	8.83	0	0.00	775	8.81
1985 to 1989	1,070	12.20	2	9.52	1,072	12.19
1990 to 1994	1,246	14.20	2	9.52	1,248	14.19
1995 to 1999	1,490	16.98	3	14.29	1,493	16.98
2000 to 2004	1,652	18.83	9	42.86	1,661	18.89
2005 to 2007	1,034	11.79	5	23.81	1,039	11.81
Total	8,773	100.00	21	100.00	8,794	100.00

TABLE 6: MORTALITY IN THE ONTARIO URANIUM MINERS COHORT, BY PROVINCE OF DEATH, 1954 TO 2007.

Province of Death	Total Cohort	
	Number of Deaths	% Deaths
United States	25	0.28
Newfoundland & Labrador	43	0.49
Prince Edward Island	9	0.10
Nova Scotia	152	1.73
New Brunswick	148	1.68
Quebec	1,118	12.71
Ontario	5,994	68.16
Manitoba	191	2.17
Saskatchewan	101	1.15
Alberta	235	2.67
British Columbia	740	8.41
Yukon	20	0.23
Northwest Territories	16	0.18
Nunavut	2	0.02
Total	8,794	100.00

6.2. LINKAGE TO THE HISTORICAL SUMMARY TAX FILES

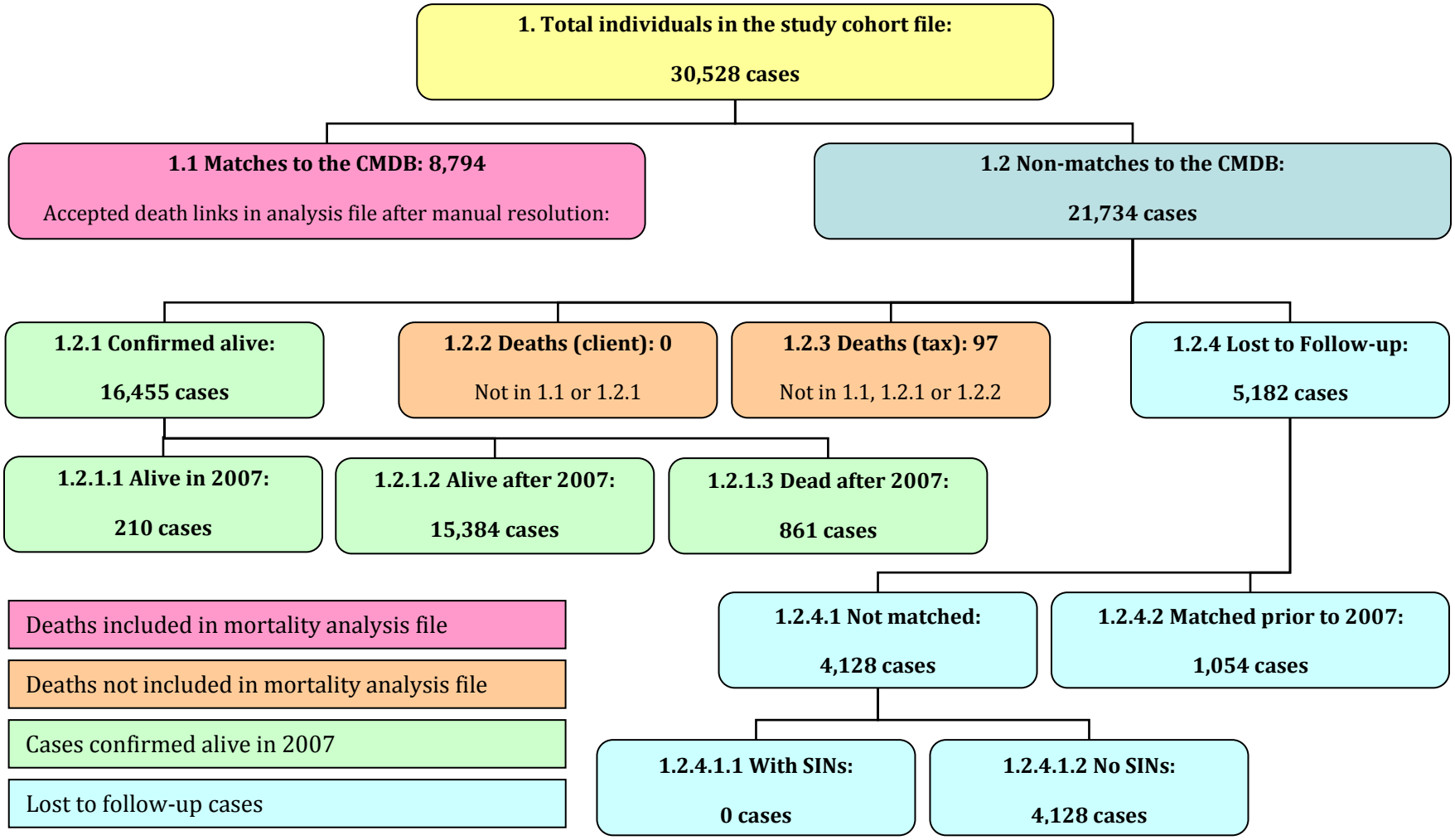
The primary purpose of the cohort to HSTF linkage is to confirm the vital status of cohort individuals at the end of the study period. For the Ontario Uranium Miners mortality linkage, the last year of the mortality linkage study was 2007.

If a death record for a cohort member was not found in the CMDB for the death years included in the study, a match in the tax file might provide information on whether the person was still alive at the end of the study period, if they died outside of Canada (and thus are not reported in the CMDB), or if they died in Canada during the study period but for some reason could not be found in the CMDB.

The study cohort was matched to the 1984 to 2009 Tax Summary file. If a cohort member was not found in either the CMDB (death years 1954 to 2007) or in the HSTF (tax years 2007 or later), that individual was considered to be 'lost to follow-up'.

No tax information was available prior to 1984, thus confirmation of vital status of cohort individuals for years prior to 1984 was not possible. As shown in Figure 2, 17 % of the cohort was deemed lost to follow-up.

FIGURE 2: VITAL STATUS OF ONTARIO URANIUM MINERS COHORT AS OF 2007



6.3. CANADIAN CANCER DATABASE LINKAGE

Of the 30,528 individuals in the Ontario Uranium Miners cohort, 3,906 individuals (3,874 males and 32 females) linked to the Canadian Cancer Database for the years 1969 to 2005. In total, 4,320 incidence records were extracted from the cancer file.

Table 7 shows the number of primary cancer incidence records per individual by sex, Table 8 shows the province of diagnosis (at the time of the cancer reporting). The distribution of year of diagnosis is shown in Table 9.

TABLE 7: ONTARIO URANIUM MINERS COHORT INDIVIDUALS LINKED TO THE CANADIAN CANCER DATABASE BY NUMBER OF INCIDENCES AND SEX

Number of primary cancer incidence records	Ontario Uranium Miners cohort individuals linked to the Canadian Cancer Database, 1969 to 2005		
	Males	Females	Total
1	3,509	32	3,541
2+	365	0	365
Total	3,874	32	3,906

TABLE 8: NUMBER OF CANCER INCIDENCE RECORDS BY REPORTING PROVINCE

Reporting Province	Total Cohort	
	Number of Cancer Incidence	% Cancer Incidence
Newfoundland & Labrador	22	0.51
Prince Edward Island	5	0.12
Nova Scotia	90	2.08
New Brunswick	83	1.92
Quebec	587	13.59
Ontario	2,901	67.15
Manitoba	79	1.83
Saskatchewan	46	1.06
Alberta	121	2.80
British Columbia	366	8.47
Yukon	13	0.30
Northwest Territories	6	0.14
Nunavut	1	0.02
Total	4,320	100.00

TABLE 9: NUMBER OF CANCER INCIDENCE RECORDS BY YEAR OF DIAGNOSIS

Year of Diagnosis	Number Incidence Records
1969	11
1970	18
1971	26
1972	28
1973	28
1974	31
1975	37
1976	30
1977	53
1978	52
1979	59
1980	68
1981	73
1982	76
1983	77
1984	103
1985	90
1986	116
1987	95
1988	112
1989	127
1990	115
1991	156
1992	146
1993	156
1994	185
1995	182
1996	170
1997	164
1998	182
1999	211
2000	226
2001	234
2002	214
2003	211
2004	210
2005	248
TOTAL	4,320

6.4.OTHER

In order to facilitate the analysis of the linked cohort, the Miner's Identification Number was replaced with the Random Statistics Canada Number in the following files:

1. The National Dose Registry File;
2. The Mining Master File.
3. The Analytical Variables File.

6.5.FUTURE STEPS

Subsequent requests to update the study will be made every five to ten years depending on the alive status and age distribution of the cohort.

7. RECORD LAYOUTS

7.1. RECORD LAYOUT OF THE ONTARIO URANIUM MINERS COHORT FILE

Source of record layout: Cancer Care Ontario [2]

Field	Size	Position	Type	Name	Description
1	1	1	A	MINER_TYPE	Miners cohort study code type
2	9	1-9	A/N	MINERID	Miner study identification number
3	15	11-25	A	SURNAME_STC	Surname
4	15	27-41	A	ALT_SURNAME_STC	Other surname
5	15	43-57	A	GNAME1_STC	First given name
6	15	59-73	A	GNAME2_STC	Second given name
7	1	75	N	SEX_STC	Sex
8	1	77	N	ALT_SEX_STC	Other sex code
9	8	79-86	N	DOB_STC	Birth date (century, year, month, day)
10	8	88-95	N	ALT_DOB_STC	Other birth date
11	3	97-99	N	BPLACE_NDR_STC	Birth province/country STC code – conversion made from the NDR record original entry
12	3	101-103	N	BPLACE_WHF_STC	Birth province/country STC code – conversion made from the WHF record original entry
13	4	105-108	N	FIRST_WORKYR_STC	Work history earliest year
14	4	110-113	N	LAST_WORKYR_STC	Work history latest year

[2] The record layout was taken from the Cancer Care document titled: Confirmation Report # 1, Contract 87055-08-0853, Preparation of Data for Linkage – Ontario Miner's Study Update, November 17, 2008. Cancer Care Ontario instructed Statistics Canada to use fields 1 to 14 in the cohort file for record linkage purposes.

7.2.RECORD LAYOUT OF THE HISTORIC TAX SUMMARY FILE

Field	Size	Type	Name	Description
1	9	N	SIN	Social Insurance Number
2	1	CH	DECEASED	Deceased Code (D=deceased)
3	4	BIN	DEATH	Date of Death (YYYYMMDD)
4	4	BIN	BIRTH	Date of Birth (YYYYMMDD)
5	1	CH	SEX	Gender Code
6	1	CH	MARSTAT	Marital Status
7	2	CH	SURNPOS	Surname position in the name field
8	30	CH	NAMES	Given names, Surnames
9	2	BIN	RESPROV	Province of Residence
10	8	BIN	LOCALITY	Locality Code
11	1	C	MIGRATN	Immigrant-emigrant code (0-3)
12	7	CH	POSTCODE	Postal Code at most recent filing
13	4	CH	SPOUGIV	Spouse's given name
14	9	N	SPOUSIN	Spouse's SIN
15	2	N	TAXYEAR	Last year in which tax return was filed
16	1	CH	LASTSTAT	Status of person on last tax filing
17	9	CH	STATUS	Filing history and status in each year . individual did not file a return A returned file – assumed alive D individual died during taxation year immigrated to Canada during taxation year emigrated from Canada during taxation year immigrated/emigrated in the same year

7.3.RECORD LAYOUT OF THE CANADIAN MORTALITY DATABASE SAS LINK FILE

Field	Size	Type	Name	Information contained in field
1	1	CH	CMDB_TYPE	Type of record
2	1	CH	CMDB_CANCEL	Cancel code
3	12	N	RECID	(CMDB_YEAR||CMDB_PROV||CMDB_REGNO)
4	2	N	DUPFLAG	Duplicate flag
5	4	CH	CMDB_YEAR	Year of registration
6	3	CH	CMDB_PROV	Province of registration
7	6	CH	CMDB_REGNO	Registration number
8	1	CH	CMDB_CCD	Control code digit (type of surname)
9	6	CH	CMDB_NYSIIS	NYSIIS code of surname
10	12	CH	CMDB_SURNAME	Surname
11	10	CH	CMDB_GIVEN1	First given name
12	10	CH	CMDB_GIVEN2	Second given name
13	10	CH	CMDB_GIVEN3	Third given name
14	3	CH	CMDB_TITLE	Title (Jr, Sr, Rev, Sis…)
15	1	CH	CMDB_SEX	Sex code (1=male, 2=female)
16	1	CH	CMDB_MARST	Marital status code (1-5)
17	8	CH	CMDB_BIRTHDATE	Date of birth (yyyymmdd)
17.1	4	N	CMDB_BIRTHYYYY	Year of birth (yyyy)
17.2	2	N	CMDB_BIRTHMM	Month of birth (mm)
17.3	2	N	CMDB_BIRTHDD	Day of birth (dd)
18	2	CH	CMDB_BIRTHPLAC2	Place of birth (2 digit- Vital Statistics code)
19	3	CH	CMDB_BIRTHPLACE	Place of birth (3 digit - ISO code)
20	2	CH	CMDB_RESPROV2	Residence – province or country (2 digit)
21	3	CH	CMDB_RESPROV	Province or country (3 digit)
22	2	CH	CMDB_RESCDIV	Census div. or country(2 digit)
23	3	CH	CMDB_RESCSUB	Census subdivision or locality (3 digit)
24	6	CH	CMDB_POSTAL	Residence – postal code
25	6	CH	CMDB_FATHNYS	Father – NYSIIS code of surname
26	12	CH	CMDB_FATHSURN	Father – surname
27.1	1	CH	CMDB_FATHINIT1	Father – initial1
27.2	1	CH	CMDB_FATHINIT2	Father – initial2
27.3	1	CH	CMDB_FATHINIT3	Father – initial3
28	2	CH	CMDB_FATHBPL2	Father – birth place (2 digit code)
29	3	CH	CMDB_FATHBPL	Father – birth place (3 digit code)
30	6	CH	CMDB_MOTHNYS	Mother – NYSIIS code of surname
31	12	CH	CMDB_MOTHSURN	Mother – maiden surname
32.1	1	CH	CMDB_MOTHINIT1	Mother – initial1
32.2	1	CH	CMDB_MOTHINIT2	Mother – initial2
32.3	1	CH	CMDB_MOTHINIT3	Mother – initial3
33	2	CH	CMDB_MOTHBPL2	Mother – birth place (2 digit code)

34	3	CH	CMDB_MOTHBPL	Mother – birth place (3 digit code)
35	6	CH	CMDB_SPOUNYS	Spouse – NYSIIS code of surname
36	12	CH	CMDB_SPOUSURN	Spouse – surname
37.1	1	CH	CMDB_SPOUINIT1	Spouse – initial1
37.2	1	CH	CMDB_SPOUINIT2	Spouse – initial2
37.3	1	CH	CMDB_SPOUINIT3	Spouse – initial3
38	8	CH	CMDB_DEATHDATE	Date of death (yyyymmdd)
38.1	4	N	CMDB_DEATHYYYY	Year of death (yyyy)
38.2	2	N	CMDB_DEATHMM	Month of death (mm)
38.3	2	N	CMDB_DEATHDD	Day of death (dd)
39	1	CH	CMDB_AGEP	Age period
39	3	N	CMDB_AGE	Age duration
40	3	CH	CMDB_DEATHPROV	Place of death - province or country
41-42	5	CH	CMDB_DEATHSGC	Place of death – census div/county code place of death – census subdivision/locality
43	15	CH	CMDB_DEATHCITY	Place of death – city or place name
44	4	CH	CMDB_CAUSE	ICD underlying cause of death
45	4	CH	CMDB_NATURE	ICD nature of injury (accidents)
46	1	CH	CMDB_AUTOPSY	Autopsy performed
47	1	CH	CMDB_AUTFIND	Autopsy findings used
48	1	CH	CMDB_AUTMORE	Autopsy further info expected
49	4	CH	CMDB_OCCUP	Kind of work done (occupation)
51	7	CH	CMDB_SEXNYSIIS	Concatenation of sex + NYSIIS variables
50	2	CH	CMDB_FILLER	Filler

7.4.RECORD LAYOUT OF THE CANADIAN CANCER DATABASE LINKAGE FILE

Field	Size	Type	Name	Description
1	1	CHAR	CCD	Control code digit
2	2	N	DUPFLAG	Duplicate Flag
3	9	N	SEQUENCE	Sequence number
4	8	N	RECID	Record identification number
5	2	CHAR	REPYEAR	Reporting year
6	9	CHAR	NOTID	Notification ID
7	2	CHAR	PROV	Reporting Province
8	9	CHAR	REGNO	Register Case Number
9	1	CHAR	D8	Eight Register Case Number Digit
10	7	CHAR	PREVNO	Previous Case Number
11	7	CHAR	PRPATNO	Province Patient Number
12	3	CHAR	WTCLASS	Weight Class Code
13	6	CHAR	NYSIIS	NYSIIS Code of Surname
14	12	CHAR	SURNAME	Surname
15	9	CHAR	GIVEN1	First Given Name
16	7	CHAR	GIVEN2	Second Given Name
17	1	AN	INIT3	Third Given Name Initial
18	3	CHAR	TITLE	Title
19	6	CHAR	ALTNYS	NYSIIS Code of Alternate Surname
20	10	CHAR	ALTSURN	Alternate Surname
21	9	CHAR	SPOUSE	Spouse
22	1	CHAR	SEX	Sex Code
23	1	CHAR	MARST	Marital Status
24	4	N	BIRTHYR	Year of Birth
25	2	N	BIRTHMN	Month of Birth
26	2	N	BIRTHDY	Day of Birth
27	8	CHAR	BTHDATE_C	Full birth date
28	7	CHAR	BIRTHSGC	Birthplace Standard Geographical Code (SGC) (PR, CD, CSD)
29	7	CHAR	RESSGC	Residence SGC
30	10	CHAR	RESPL	Place of Residence
31	6	CHAR	POSTCOD	Postal Code
32	3	N	AGEDIAG	Age at Diagnosis
33	1	CHAR	AGECONF	Age Confirmation
34	4	N	DIAGYR	Year of Diagnosis
35	2	N	DIAGMN	Month of Diagnosis
36	2	N	DIAGDY	Day of Diagnosis
37	8	CHAR	DIAGDATE_C	Full diagnosis date
38	4	CHAR	ICD9	ICD-9 Code
39	4	CHAR	ICD8	ICD-8 Code
40	4	CHAR	ICDTOP	ICDO-Topography
41	5	CHAR	ICDMOR	ICDO-Morphology

42	5	CHAR	MORCODE	Morphology –Code
43	1	CHAR	MORIND	Morphology-Indicator
44	1	CHAR	DIAGMTH	Method of Diagnosis (Most confirmed method)
45	1	CHAR	PRSITE	Primary Site Number
46	1	CHAR	LATERAL	Laterality
47	12	CHAR	HLTHNO	Health Insurance Number
48	1	CHAR	PATSTAT	Patient Status
49	1	CHAR	REGSRCE	Source of Registration
50	2	CHAR	DREGYR	Year of Death Registration
51	2	CHAR	DREGPR	Province of Death Registration
52	6	CHAR	DTHREGNO	Death Registration Number
53	4	N	DTHYEAR	Year of Death
54	2	N	DTHMON	Month of Death
55	2	N	DTHDAY	Day of Death
56	8	CHAR	DTHDATE_C	Full death date
57	9	N	FILLER	Filler
58	2	CHAR	AVAIL	Availability
59	1	CHAR	IDFLAG	I. D. Flag
60	2	N	AGEGRP	Age Group
61	2	CHAR	REPROV2	Vital Reporting Province
62	2	CHAR	RESPROV	Vital Residence Province
63	2	CHAR	DTHPROV	Vital Death Province
64	2	CHAR	BIRTHPL	Vital Birth Province
65	10	CHAR	POCKET	Pocket
66	7	CHAR	NYSIIS_SEX	NYSIIS/SEX combination

7.5.RECORD LAYOUT OF THE MORTALITY ANALYSIS FILE

Field	Position	Length	Type	Variable Name	Description
1	1	8	Char	STCNO	Statistics Canada random number
2	9	3	Char	Group_Duplicate	Identifies Duplicate Records
3	12	4	Char	TOWGHT	Mortality Linkage Weight
4	16	4	Char	CMDB_BIRTHYYYY	Mortality- Year of Birth
5	20	2	Char	CMDB_BIRTHMM	Mortality -Birth Month
6	22	3	Char	CMDB_RESPROV	Mortality -Province of Residence
7	25	8	Char	CMDB_DEATHDATE	Mortality -Date of Death (YYYY/MM/DD)
8	33	3	Char	CMDB_DEATHPROV	Mortality- Province of Death.
9	36	4	Char	CMDB_CAUSE	Mortality –Cause of Death
10	40	3	Char	COHO_Birthpl	Cohort – Birth Place
11	43	3	Char	COHO_Altbpl	Cohort – Alternate Birth Place Code
12	46	4	Char	COHO_First_workyr	Cohort – Work History Earliest Year
13	50	4	Char	Coho_Last_workyr	Cohort_- Work History Latest Year
14	54	1	Char	Cohort_sex	Cohort – Sex Code
15	55	1	Char	Cohort_oth_sex	Cohort - Other Sex Code
16	56	1	Char	Imp_sex	Sex Imputation Code
17	57	4	Char	Cohort_Birthyr	Cohort – Year of Birth
18	61	2	Char	Cohort_Birthmn	Cohort – Month of Birth
19	63	1	Char	Dup_indicator	Describes possible and definite duplicate records

7.6.RECORD LAYOUT OF THE CANCER ANALYSIS FILE

Field	Size	Position	Type	Name	Description
1	8	1 – 8	A	STUDYNO	Study number
2	1	9	N	SEX	Sex Code
3	2	10-11	A	REPROV2	Reporting Province
4	8	12-19	A	DIAGDATE_C	Diagnostic date from CCDB
5	4	20-23	CH	CCDB_birthyr	Birth year from Cancer File
6	2	24-25	CH	CCDB_birthmn	Birth Month from Cancer Fie
7	4	26-29	A	ICD9	ICD9 cause of death code
8	5	30-34	A	ICDMOR	ICDMOR code (see data dictionary)
9	4	35-38	A	ICDTOP	ICDTOP code (see data dictionary)
10	1	39	CH	MORIND	
11	1	40	CH	REGSRCE	
12	1	41	A	DIAGMTH	Diagnostic method (see data dictionary)
13	1	42	A	LATERAL	Laterality (see data dictionary)
14	1	43	A	PRSITE	Primary Site (see data dictionary)
15	1	44	A	PATSTAT	Patient Vital Status
16	2	45-46	CH	RESPROV	Province of Residence

17	8	47-54	CH	DTHDATE_C	Date of Death
18	2	55-56	CH	DTHPROV2	Province of Death
19	4	57-60	CH	TOTWGHT	Linkage Weight