SOME ASPECTS OF QUALITY OF CANCER MORTALITY AND INCIDENCE STATISTICS

D. Binder and A. Malhotra[1]

Statistics Canada, Canada's central statistical agency, has been compiling national mortality statistics, including those on cancer mortality since 1921. Also, cancer incidence data are available from 1969.

The data quality of these files may be assessed in a variety of ways. Ratios of cancer mortality to incidence give some information on coverage errors. Micro-data matches between incidence and mortality files give an indication of misclassifications. As well, multiple registrations for cancer incidence may be duplicates. Completeness and availability of data items are also important for special studies.

In this paper, the feasibility of using these measures of data quality and the implications of these measures are discussed.

1. INTRODUCTION

Population based cancer statistics are the basis of epidemiological research into the distribution and determinants of cancer and underlie health programmes for the prevention, diagnosis and treatment of cancer. Statistics Canada, Canada's central statistical agency, compiles two such types of data on cancer.

1. National mortality data which are based on reports from provincial vital statistics registration systems. These data date back to 1921.

2. National cancer incidence data which are based on notifications from provincial cancer registries. This data series was established in 1969.

---

[1] D. Binder, Institutional and Agriculture Survey Methods Division, Statistics Canada and A. Malhotra, Health Division, Statistics Canada.

Good statistics which provide reliable information on risk differentials depend on completeness and accuracy of cancer registration and comparability of the data between different registration areas and time periods. Cancer incidence and mortality data as reflections of the true cancer risk each have their own merits and limitations.

Cancer incidence data are particularly suitable for the epidemiological study of cancer because they provide information on all cancers, not only those that are fatal, because they can provide an early warning of emerging problems and because the diagnostic information is usually detailed and of a high quality. For example, the publication Cancer Incidence in Five Continents [1] emphasizes the role that international comparisons of cancer incidence play in yielding clues about the causes of cancer in spite of certain well known limitations of the data. These limitations include difficulty in achieving complete registration of new cancer cases and differences between registries in the extent to which this is achieved. Major factors influencing coverage are the number and types of data sources used, how active case finding is, the length of time a registry has been in operation and whether or not the reporting of cancer is a legal requirement in the registration area. Canadian cancer registries are quite heterogeneous in their data collection methodology but all attempt to follow international [2] as well as national [3] guidelines for the standardized recording of cancer incidence data. Differences in sources and techniques of registration not only influence coverage but also other aspects of data quality such as detail of socio-demographic and geographic information that is provided. Also, cancer incidence data are sensitive to such factors as mass screening programmes which result in the inclusion of previously undiagnosed prevalent cases.

The editors of Cancer Incidence in Five Continents use and discuss a number of indices which may be useful in assessing completeness of registration and reliability of the data [4]. These include cancer mortality-incidence ratios as indicators of completeness of registration (see Sections 2.1 and 3.1 of this paper).

The reporting of deaths is a legal requirement in most developed countries so that coverage error is assumed to be small. Known and suspected limitations

of mortality data for purposes of epidemiologic cancer research include a lack of information on non-fatal cancers, less precise diagnostic information and more frequent misclassification due to assignment and coding of underlying cause of death resulting in less accuracy compared with the diagnostic information in cancer incidence statistics.

A quality assessment of vital statistics [5] which was undertaken as a pilot study gives some indication of the quality of Canadian mortality data (all causes), particularly on error rates in the coding of underlying cause of death. This error rate was 7.2% in the data year 1976. About two-thirds of the errors involved the first or second digit of the 4-digit cause of death code. Variation in the error rate by specific causes of death was not investigated.

This paper is concerned with assessing the feasibility of measuring certain aspects of quality of the two cancer data files at Statistics Canada which are used in epidemiological studies, namely the cancer incidence file and subset of records from the mortality file with an underlying cause of death of cancer.

The aspects of data quality selected for investigation were:

1. Completeness of registration of new cancer cases through a comparison of cancer mortality with cancer incidence in the same period. This comparison is a crude but commonly used indicator of completeness of registration.

2. Consistency of assignment of diagnosis and cause of death codes through matching individual records on the two files.

3. Availability and completeness of data items through an analysis of how often valid values are present on the files.

4. Registration of multiple primary cancers on the incidence file through a matching of records within the file.

The period covered by the study is 1969-1978, the period for which cancer incidence data are available. Ontario was excluded from all investigations because the National Cancer Incidence Reporting system includes data for this province for 1969-1971 only[2].

A discussion of the approach taken in the investigations is contained in Section 2 of this report. In Section 3 the findings are discussed.

2. DESCRIPTION OF MEASURES

In this section we describe some methods for studying the data quality of the cancer mortality and incidence files.

2.1 Mortality-Incidence Ratios

In order to study relative rates of undercoverage among cancer incidence registrations, we consider the ratios of deaths to incidents of cancer. Since the Mortality System registers all deaths in Canada by cause of death and the concept of the National Cancer Incidence Reporting System (NCIR) is to register all new incidents of primary malignant neoplasms[3], if the two registration systems were of the same quality for all reporting registries, one would expect that the ratio of mortality rate to the incidence rate for a particular site would be fairly consistent within a population of given age and sex over sufficiently long periods of time. (We compute these ratios for deaths and incidents over 5-year and 10-year periods to reduce the effect of the time lag between the reporting of a cancer incident and death). Inconsistency of this ratio would arise if any of the following rates differ across reporting registries:

(a) rates of survival or sudden increase rate of incidence

(b) mortality rates from other competing risks,

(c) rates of error in coding of underlying cause of death,

---

[2] Ontario developed a passive registration system which makes use of reports on cancer patients made for other purposes. Data for recent years are currently being prepared by the province.

[3] By exception, metastatic cancers are registered when the primary site is unknown.

(d) rates of error in classifying cancer site for new cancers,
(e) rates of under-reporting or over-reporting of cancer incidents.

If the mortality-incidence ratios are grossly different for most sites, then differing rates in (a) and (b) can be discounted. With respect to the error rates in (c), studies on the coding error for underlying cause of death have yielded error rates of less than 10% [5]. In an unpublished report it was found that these errors rates could vary from 3% to 18% across reporting registries. However, the observed differences in the mortality-incidence ratios (see Tables 1 and 2) cannot be completely explained by these error rates. Also, since about 90% of the registrations of new cancers are confirmed histologically, the error rate in (d) would be small. Therefore, these mortality-incidence ratios do give an indication of the coverage error in the NCIR System.

Concentrating on those sites with leading diagnosis count (excluding skin cancer), based on the NCIR file, for each sex, we show in Tables 1 and 2 the national mortality-incidence ratios as well as provinces with largest and smallest ratios, for two five-year periods, broken down by age-groups. We omit Prince Edward Island from consideration because the number of observed events is too small for valid comparison.

2.2 Matching Mortality and Incidence Records

In order to assess the feasibility of evaluating errors in cause of death classification, or errors in cancer site classification, a sample or records can be selected from either the Mortality File or Cancer Incidence File and then the other file can be searched for matching records. The manual search does not guarantee that all true matches will be found, and, in fact, the rate of successfully matching may be different across reporting registries, because the level of detail of matching variables can vary from one registry to another. (See Section 2.3 for a study on availability of data).

Records with malignant neoplasms of the lung or bronchus (ICDA-8[4] is 162.1) for the years 1969-1978 were selected as starting points from both files. This choice was based on considerations of high incidence, high mortality and short survival times so that conditions were favourable for finding matching records on both files. In spite of this, because of the time difference between diagnosis and death, it is true that cancer deaths in the earlier years and newly diagnosed cancers in the later years are less likely to have a corresponding record on the matched file. The analysis of the results would be improved in future studies if the sample design controlled for year of death and year of diagnosis.

Two independent samples were selected: one from Mortality File and the other from the NCIR file.

2.2.1 Mortality to Incidence

All deaths from cancer between 1969 and 1978 should have, at least conceptually, a corresponding record on the NCIR system. Noteworthy exceptions to this rule are that the cancer was first diagnosed in Ontario or outside Canada or that the cancer was first diagnosed before 1969. Besides the exceptions, a lack of a corresponding record on the NCIR system is an indicationn of under-coverage. This, of course, assumes that the underlying cause of death in the Mortality File is error-free.

Therefore, if a sample of deaths from cancer are selected and matched to the NCIR system, we have a number of possible outcomes:
(a) no matching record is found,
(b) a match was found with a record having a different cancer site,
(c) a match was found with a record having the same cancer site.

If no match is found, this is an indication of under-coverage, or that the cancer was first diagnosed in Ontario or outside Canada or prior to 1969, or that the death was not really a death due to cancer. Alternatively, as

[4] International Classification of Diseases, Adapted for Use in the United States, Eighth Revision.

previously mentioned, the matching process itself is not perfect. If a match is found, but the records have different cancer sites, this may be an indication of an error on either the Mortality File or the NCIR System. As mentioned previously, it is generally believed that the NCIR system yields the more accurate disease classification, because of the high rate of histological confirmation.

A small scale study was undertaken to measure this phenomenon. A random sample of 56 records with underlying cause of death reported as a malignant neoplasm of the lung or bronchus (ICDA-8 is 162.1) was selected from each of the provinces except Ontario yielding a total sample of 504 records. Only deaths between 1969 and 1978 were selected.

The national rate of successful matches was 82.3%. The rates varied between 73.2% and 96.4% across the nine provinces. Among those with successful matches, 92.5% had the same 4-digit ICDA-8 classification. These rates varied from 74.4% to 100.0%. For those provinces with the lowest and highest rates of matches with the same ICDA-8 classification, we give in Table 3 the breakdown of the observed disease classifications.

2.2.2 Incidence to Mortality

A sample of records from the NCIR System was also selected and matched to the Mortality File. Fifty-six records with malignant neoplasms of the lung and bronchus (ICDA-8 is 162.1) from each of the nine reporting registries were randomly selected yielding a total sample of 504 records. The matched records were then checked for underlying cause of death on the complete Mortality File for 1969 to 1978. We did not check the cause of death on the original death certificate for this study, although this would be feasible for future studies.

The outcomes from this manual match may be classified as follows:

(a) no matching record is found,

(b) a match was found with a cause of death other than cancer,

(c) a match was found with a cause of death being cancer but not cancer of the lung or bronchus,

(d) a match was found with the same cause of death.

If no match is found, this is an indication of the inadequacy of the matching process, unless death occurred after 1978 or outside Canada or the person is still alive.

A match found with a different underlying cause of death is an indication of one of the following:

(a) a competing risk took precedence,

(b) cancer was a contributing cause of death but not the underlying cause of death or

(c) the underlying cause of death on the Mortality Data Base was incorrect, or the cancer site was incorrect on the NCIR system (the latter being assumed less likely).

The average rate of successful matches was 69.4%. The rates varied between 55.4% and 80.4% across the nine provinces. Among those with successful matches, 92.0% had the same 4-digit ICDA-8 classification. These rates varied from 85.3% to 100.0%. For those provinces with the lowest and highest rates of matches with the same ICDA-8 classification, we give in Table 4 the breakdown of the observed cause of death classifications.

2.3 Availability and Completeness of the Data

One simple measure of the quality of the data files is the relative frequency of valid data for specific items. For the National Cancer Incidence Reporting System and the cancer deaths on the Mortality File, we concentrate on the following items:

- date of birth (day, month, year)
- age
- place of birth
- county and subdivision of residence

We chose these items to exemplify how easy or difficult it would be to match records from other files (e.g. Section 2.2), or to create special tabulations, such as small area statistics. For each item we classify the data as being

valid or invalid. Besides blank values, invalid data would arise when alphabetic characters are found in a numeric variable or the numeric value is out of range. We have aggregated the relative frequencies into two five-year groupings (1969-1973 and 1974-1978) so that we can see whether the quality has changed significantly in the later years.

In Tables 5 and 6 we report the national averages for the two data bases as well as show the values for the provinces with largest deviation from the national average. For the Mortality File, we give the results only for cancer deaths in the nine provinves outside of Ontario so that the comparison with the NCIR system is more meaningful.

2.4 Multiple Registrations on the Cancer Incidence System

The concept of the National Cancer Incidence Reporting System is to register all new incidents of malignant neoplasms. An individual should be registered more than once when multiple malignant neoplasms develop. To avoid duplicate registration of the same incident or duplicate reporting of patients registered in more than one province, all provincial cancer registries follow routine procedures. In spite of this, duplicate reporting of the same cancer incident may occur. In order to evaluate the extent of the duplication, we searched for records which are likely duplicates. The search was nowhere near exhaustive, so that the number of potential duplicates found is an underestimate. Of the 457,158 records, we removed the records with invalid surnames or years of diagnosis. For those with missing birthyear, we calculated the birthyear from the age when available. We also removed skin cancer records (ICDA-8 is 173) since this is known to have multiple occurrences.
Of the remaining records, we found those cases where all the following occurred:

- birthyear or calculated birthyear matched exactly,
- surname matched exactly,
- first four letters of first given name agreed,
- the three digit ICDA-8 code agreed.

Of these records, we identified multiple registration as follows:

- year, month and day of birth was present and agreed, or
- day of birth was not present on at least one record but month of birth agreed.

We also manually verified all groups with at least 3 individuals where the month and day information did not agree and all groups of 2 individuals where the month or day information was missing on at least one record.

In all, this resulted in identifying 6113 records which where potential duplicates. These records correspond to 5947 individuals. (Note that some individuals were duplicated more than once.) We did not make the judgment as to whether these were legitimate multiple registrations or actual duplicates.

For each 3-digit ICDA-8 value, we show in Table 7 the breakdown of these potential duplicates according to whether the records came from the same reporting registry or different registries, as well as how many potential duplicates have the same fourth digit of the ICDA-8 classification.

3. DISCUSSION

3.1 Mortality - Incidence Ratios (Tables 1 and 2)

Ratios of cancer mortality to incidence can provide an indication of completeness of registration. The ratios will vary with cancer site (the highest ratios occur for sites with the lowest survival), age and sex for all registries. However, if a comparison of the ratios for different registries shows major differences within a given site, sex and age group, differences in completeness of registration of new cases of cancer must be suspected. A higher ratio, which means a higher proportion of deaths compared with newcases in the same period may indicate less complete registration of new cases.

In both time periods there were two registries which consistently had the highest ratios for all sites combined and for most of the major sites shown. There is little doubt that these high ratios do reflect underregistration of new cases - the registries are the only ones which do not use death

notifications as one of their sources of registration. In addition, one of these registries uses only a single data source, hospital reports, to register cancer cases. This registry had previously reported the results of a special study which showed that it was receiving notifications for only an estimated 70% of new cancer patients admitted to hospitals up to the end of 1976. Following this, major changes were made to improve the notification system. Since 1977 the registry has been reporting a higher number of cancer cases which is reflected in a marked reduction in mortality-incidence ratios.

All other Canadian cancer registries use multiple sources of registration which is considered essential to achieve good coverage and which could also have a positive impact on the completeness and quality of individual data items. The completeness of reporting of data items was examined in this study (see Subsection 2.3). However, it turns out that the one registry that uses only one data source actually ranks quite highly in terms of completeness of information for many data items.

A possible drawback associated with using multiple sources of registration is that duplicate registration may result. However an analysis of multiple registrations for the same individual and the same cancer site does not bear this out. In general, registries using a larger number of different sources of registration do not have more multiple registrations for the same site than registries using fewer sources of registration.

Cancer mortality-incidence ratios for the other six registries were more similar to each other. For these registries there was no consistent pattern of one registry always having higher or lower ratios for all sites and both time periods.

There are many factors that can influence variations in the observed ratios by cancer site. Factors which tend to result in less complete registration of new cases and therefore higher mortality-incidence ratios includes difficulty in diagnosing the cancer (e.g. in deep-seated organs) and lack of access to specific data sources (e.g. haematology reports confirming a diagnosis of leukaemia).

Factors which may lead to overregistration of new cases and lower ratios are mass screening programmes (which may lead to the inclusion of prevalent cases, especially, for slow-growing tumours), duplicate registration, inclusion of in-situ cases and inclusion of latent cancers discovered only at autopsy (this particularly affects cancer of the prostrate). In addition, differences in the accuracy of assignment of diagnosis or cause of death may lead to artefactual differences. For example, death certificates may state "cancer of the uterus, unspecified" or "leukaemia, unspecified" as the cause of death whereas a cancer registry will often have more precise information and will assign more precise codes [6]. An analysis of mortality-incidence ratios at the level of the more detailed diagnosis would therefore show gross discrepancies.

Of the leading cancer sites that were examined for males, cancer of the lung and stomach were associated with the highest mortality-incidence ratios for all registries but interprovincial variation in the ratios was greatest for cancer of the colon (excluding rectum) prostate and bladder. Use of cancer incidence data in studies designed to identify differences in cancer risk by geographic area (province) would therefore be more reliable for the former two cancer sites.

For females, of the leading sites examined, cancer of the colon and ovary had the highest ratios for all registries. Interprovincial variation in the ratios was greatest for cancer of the uterus and cervix uteri as well as for cancer of the colon. For purposes of interprovincial comparisons, incidence data for breast cancer and cancer of the ovary would therefore be more reliable.

In the case of the sites of cancer of the uterus (other than cervix) and cancer of the cervix, there are large interprovincial variations in the ratios if the sites are considered separately. This variation is greatly reduced if the two sites are combined, suggesting that there are differences in the accuracy of diagnosis and cause of death assignment for these sites.

The site-specific mortality-incidence ratios were examined for major age groups. The highest ratios consistently occur at older ages (65 and over) for

all registries and all sites shown. This is as expected, since the risk of death increases with age so that proportionately more death than new cases occur at older ages. It is also recognized that diagnosis and registration of cancers in older persons is generally more difficult. However, the relative increase in the ratios at older ages is much greater for the registries which have the highest average ratios to start with. This indicates that while all registries may have some difficulty in registering older patients, under-coverage of the older population is greater for registries which in general have less complete registration systems.

The Canadian data therefore lend support to recommendations made by the International Union against Cancer (1970) and the International Agency for Research on Cancer (1976) and the reiteration of this recommendation in a recent paper by Doll and Peto [7] that "reasonably reliable comparisons of cancer incidence are obtained only if comparisons are limited to men and women in middle life".

3.2 Matching Mortality and Incidence Records (Tables 3 and 4)

Since Statistics Canada is responsible for managing both the cancer incidence and the mortality data files, it is possible to compare reports for individuals who are listed in the two separate data files to verify the reported information.

In this part of the investigation of data quality, the accuracy of assignment of diagnosis and cause of death was of particular interest. Within the scope of the study it is only possible to describe the results - the reasons underlying the discrepancies found remain unknown. However, it is felt that the findings are revealing and do indicate, in the case of the particular cancer selected for analysis, lung cancer, that agreement on diagnosis between the two files is generally high, over 90%.

The study also indicates that a larger scale match would be feasible to assess the accuracy of diagnosis and cause of death codes. Of course, if a larger scale study were based on a sample, it would be preferable to stratify the sample by year of diagnosis or year of death. In theory, if computerized

matching techniques were used, this type of analysis is possible for all cancer sites. If such an undertaking were to be supplemented with, for example, studies on accuracy of coding of cause of death and diagnosis in the field, such as described in two U.S. reports [8] [9], interpretation of epidemiological research findings would be facilitated.

3.2.1 Mortality to Incidence File Search

Of the sample of 504 death records with an underlying cause of death of lung cancer from 1969-1978, 415 (82%) corresponding records on the cancer incidence file for the years of diagnosis 1969-1978 were found[5]. The rate of unsuccessful matches varied from 3.6% to 26.8% across the nine provinces. This rate is influenced by four main factors: (a) that the cancer was first diagnosed prior to 1969, (b) that it reflects underregistration of new cases, (c) that identifying information was not adequate to permit matching of records, or (d) that the cause of death code is incorrectly given as cancer. There were insufficient data to allow assessment of the relative contribution of each of these factors.

Of the 415 death records for which a corresponding record was found on the incidence file, there was agreement on the diagnosis, primary cancer of the lung, in 92.5% of cases. There was 95.2% agreement that a cancer of the respiratory system was present. The small number of remaining records, had diagnoses for sites other than respiratory cancer on the incidence file. It is generally accepted that the diagnostic information on cancer registry files is more accurate than the cause of death information on death certificates. However, given the scope of this study it is not possible to determine if misclassification on either of the files (or perhaps the fact that a lung cancer was first diagnosed prior to 1969 followed by a subsequent registration for another primary cancer) account for the disagreement. Across the provinces, the rate of agreement on diagnosis varied from 74.4% to 100%.

---

[5] In six cases more than one corresponding record for the same individual existed on the incidence file. Only one of these records was counted as a successful match.

Interestingly, for the province with total agreement on diagnosis there was also the highest success rate (96.4%) of locating corresponding records on the two files. This could possibly indicate close liaison between the provincial vital statistics office and the cancer registry. In the reverse match of a sample of cancer incidence records to mortality records (described in Section 3.2.2) it was the same province that had the highest rate of successful matches as well as complete agreement on diagnosis.

3.2.2 Incidence to Mortality File Search

The reverse search using the incidence file for the years 1969-1978 as a starting point and attempting to locate a corresponding record on the complete mortality file for the same period was successful for 69.4% of the selected sample of 504 records with a diagnosis of primary lung cancer. The rate of unsuccessful matches was higher than in the match from mortality to incidence records for all provinces and varied from 19.6% to 44.6%. Possible reasons for not finding a match include (a) that the patient was still alive at year end 1978, or (b) that identifying information was not adequate to permit a match. It is in general less likely that one will find a corresponding record in the search from incidence to mortality file since some persons diagnosed to have lung cancer do survive this whereas all persons who die from lung cancer should be registered as new cases either prior to death or at time of death.

For the records that were successfully matched there was agreement that the diagnosis was a primary lung cancer in 91.4% of cases, a rate very similar to that found in the reverse comparison. The samples for the two comparisons were chosen independently so the consistency of the findings concerning agreement on diagnosis is reassuring. Of the remaining cases, 4.3% had the underlying cause of death classified to cancer sites other than the respiratory system, and 3.7% had an underlying cause of death which was not cancer. For this latter group it is possible that cancer was mentioned on the death certificate as a contributing cause of death. This analysis is possible but was n-not carrier out.

3.2.3 Availability and Completeness of the Data (Tables 5 and 6)

One measure of quality and usefulness of the data files is the frequency of valid information for specific data items. This measure is crude because "valid" as defined here means valid according to computerized edit checks and does not preclude that imputation of missing information or errors in definition or classification of the data item render the information invalid.

Subject to the above caveats the measure may be useful in showing if and where there are improvements in reporting of data items over the years, and whether or not particular analyses of the data are feasible. For example, information on date of birth is important for purely statistical (age-specific) analysis of the data as well as for medical follow-up analyses which depend on good identifying information.

On the cancer incidence file, a complete birthdate (i.e., day, month and year) is on average present on only 68% of the records in the period 1969-1978. If the two time periods, 1969-1973 and 1974-1978 are considered separately some improvement in the more recent period becomes evident.

On cancer mortality records for the same time period a complete birthdate is present in over 95% of cases. However, at least part of this high rate is due to the fact that the mortality system imputes a date of birth from age and date of death when the birthdate is not reported. In 1976 the imputation rate was 11.5% [5]. No such imputations are carried out in the cancer incidence system[6].

Small area analyses of cancer occurrence require complete and detailed residence information. Cancer mortality data are much more useful for these purposes because census division (county) of residence codes are present on 99.8% of records and census subdivision (city, town, village) codes are present on 96.2% of records. In contrast, on the cancer incidence file, census division codes are present on 89.6% of records and census subdivision

[6] Imputations may be useful for statistical purposes but are actually mis-leading in medical follow-up studies unless it is made clear that the information is based on an imputation.

codes on only 25.2% of records. On the incidence file, there is improvement in the reporting of census subdivision information in the second time period.

3.2.4 Multiple Registrations on the Cancer Incidence System (Table 7)

Comparability of cancer incidence data is affected if there are differences in the reporting of multiple primary cancers in the same individual and in inadvertent duplicate registration.

The rules for reporting of multiple primary cancers are difficult to interpret, so some provincial differences in their application are expected.

Inadvertent duplicate registration of the same cancer incident may occur if a provincial registry cannot determine if the same case has been registered previously (perhaps because identifying information is inadequate) or if the same incident is reported by two different registries[7]. The search for multiple primary cancers was restricted to multiple entries for the same cancer site (at the 3-digit level of the ICDA code).

No attempt was made to separate duplicate registrations from multiple primaries, although it can be speculated that the majority of cases reported by two seperate registries may be duplicates whereas the cases reported by the same registry are more likely to be valid multiple primaries, particularly those that differ in the 4th digit of the diagnosis code.

Using very strict matching criteria and excluding skin cancers (other than melanoma of the skin), 1.7% (6113) of records on the 1969-1978 cancer incidence file were identified as multiple entries.

By province, this rate varied from 0.5% to 1.9%. Only 0.4% of multiple records were reported by two different registries. There was agreement down to the 4th digit level of the diagnosis code in 88.4% of cases.

---

[7] The national Cancer incidence System does not carry out routine checks on such duplication.

By cancer site, if only sites with more than fifty records are considered, the rate of multiple primaries varied from 1.0% for cancer of the stomach and pancreas to 3.6% for breast cancer. The high rate for breast cancer is not surprising since the current rules for reporting of multiple primary cancers require separate reports for cancers in both sides of (most) bilateral organs.

On the whole it is felt that while there is some inconsistency arising from multiple primary and duplicate reporting, this is very small compared with that arising from undercoverage.

4. SUMMARY

The techniques described in this paper have been successful at identifying differing levels of quality of cancer incidence and mortality data. It has been found that the mortality-incidence ratios, in particular, can be used to assess coverage errors, which are one of the major concerns of a high quality cancer incidence system. The data quality for those who are registered on the incidence system is sufficiently high that it is possible to assess the quality of the cause of death classification on the mortality system through a micro-data match. In fact a computerized micro-data match could be used to evaluate the undercoverage because the number of cancer deaths without previous registration on the NCIR system could be ascertained.

ACKNOWLEDGEMENTS

The authors are grateful to J. Gorman, D. Lawrence, K. McClean, S. Moore and P. Walsh for their assistance in preparing the data for this paper. We are also grateful to J. Silins for his comments. The referee's comments are also very much appreciated.

REFERENCES

[1] Waterhouse, J., Muir, C., et al., Editors, CANCER INCIDENCE IN FIVE CONTINENTS, VOLUME III, WHO, International Agency for Research on Cancer, Lyon, 1976, p.3.

[2] WHO HANDBOOK FOR STANDARDIZED CANCER REGISTRIES, World Health Organization, Geneva, 1976.

[3] MANUAL FOR CANCER RECORDS OFFICERS, National Cancer Institute of Canada.

[4] Op. cit., in reference 1 , pp. 45-51

[5] Nagnur, D.N., Currie, S.G., Heath, B., QUALITY ASSESSMENT OF VITAL STATISTICS (A pilot study), Statistics Canada, Health Division, Ottawa, 1981.

[6] King, H.S., Wigle, D.T., Hill, G.B., Silins, J., MORTALITY TRENDS FOR CANCERS OF THE CORPUS UTERI AND CERVIX UTERI, Alberta 1969-1978, CMAJ.

[7] Doll, R., Peto, R., THE CAUSES OF CANCER: Quantitative Estimates of Avoidable Risks of Cancer in the United States Today, JNCI, Vol. 66, No. 6, June 1981.

[8] Percy, C., COMPARISON OF THE CODING OF DEATH CERTIFICATES RELATED TO CANCER IN SEVEN COUNTRIES, Public Health Reports, Vol. 93, No. 4, July-August, 1978.

[9] Feigl, P., Breslow, N.E., Laszlo, J., Priore, R.L., Taylor, W.F., U.S. CENTRALIZED CANCER PATIENT DATA SYSTEM FOR UNIFORM COMMUNICATION AMONG CANCER CENTRES, JNCI, Vol. 67, No. 5, November, 1981, p. 1019.

Table 1

Cancer Mortality - Incidence Ratios

Deaths in Period (Mortality) as Percentage of New Cases Registered (Incidence)
Canada (excluding Ontario) and Provinces with the Highest and Lowest Ratios.

Males	Leading Sites by Age Group and Sex						
Cancer Site	Age	1969 - 1973			1974 - 1978		
		Canada	Highest Ratio	Lowest Ratio	Canada	Highest Ratio	Lowest Ratio
Lung (162)	**Total**	**96**	**110**	**83**	**95**	**101**	**80**
	0-24	27	37*	-	70	-	100*
	25-44	89	96	83	85	75	82
	45-64	94	106	81	88	86	78
	65+	100	118	83	101	118	82
Prostate (185)	**Total**	**44**	**55**	**35**	**39**	**54**	**33**
	0-24	71*	133	-	71*	-	100*
	25-44	39	54	-	22	-	-
	45-64	27	32	33	24	30	20
	65+	48	62	35	43	60	36
Colon (153)	**Total**	**76**	**101**	**54**	**67**	**86**	**56**
	0-24	56	60	-	40	-	-
	25-44	61	76	40	53	87	45
	45-64	66	86	41	58	61	56
	65+	84	115	61	73	100	58
Bladder (188)	**Total**	**35**	**42**	**26**	**28**	**35**	**23**
	0-24	20	25*	33*	3	-	-
	25-44	7	9	10	6	-	-
	45-64	23	30	16	18	22	13
	65+	44	53	32	36	45	29
Stomach (151)	**Total**	**104**	**124**	**82**	**90**	**112**	**80**
	0-24	29*	-	-	86*	-	-
	25-44	82	102	54	74	40	77
	45-64	93	102	79	80	89	80
	65+	112	144	86	96	130	80
All Cancers (140-209) excluding skin (173)	**Total**	**69**	**83**	**56**	**64**	**77**	**56**
	0-24	58	64	45	48	49	56
	25-44	51	62	39	45	55	37
	45-64	65	78	53	60	65	50
	65+	74	94	59	69	90	60

- Either mortality, incidence or both are zero.
* Ratios based on fewer than 10 cases for both mortality and incidence.

Table 2

Cancer Mortality - Incidence Ratios

Deaths in Period (Mortality) as Percentage of New Cases Registered (Incidence)
Canada (excluding Ontario) and Provinces with the Highest and Lowest Ratios.

Females	Leading Sites by Age Group and Sex						
Cancer Site	Age	1969 - 1973			1974 - 1978		
		Canada	Highest Ratio	Lowest Ratio	Canada	Highest Ratio	Lowest Ratio
Breast (174)	**Total**	**40**	**47**	**31**	**37**	**46**	**31**
	0-24	14	13	-	6	50*	-
	25-44	28	33	26	23	30	17
	45-64	37	44	31	35	43	29
	65+	50	62	35	46	55	38
Colon (153)	**Total**	**73**	**94**	**50**	**67**	**80**	**50**
	0-24	15	20	-	27	33*	-
	25-44	48	49	41	41	64	34
	45-64	63	77	43	55	57	42
	65+	83	114	55	74	95	55
Uterus (182)	**Total**	**27**	**38**	**17**	**20**	**26**	**15**
	0-24	33	40*	20*	-	-	-
	25-44	13	19	7	11	15	9
	45-64	16	25	9	12	15	10
	65+	49	64	33	35	43	26
Cervix Uteri (180)	**Total**	**39**	**45**	**29**	**32**	**48**	**27**
	0-24	12	20*	17*	5	-	4
	25-44	21	17	12	15	25	10
	45-64	39	54	32	34	49	33
	65+	66	64	51	53	65	50
Ovary (183)	**Total**	**69**	**79**	**54**	**69**	**78**	**63**
	0-24	20	33*	-	28	100*	-
	25-44	43	41	26	30	38	20
	45-64	68	74	52	64	72	55
	65+	87	105	77	95	98	106
All Cancers (140-209) excluding skin (173	**Total**	**57**	**67**	**46**	**53**	**64**	**45**
	0-24	46	49	43	37	44	28
	25-44	33	39	27	27	35	22
	45-64	48	56	38	44	51	39
	65+	75	94	58	69	85	58

- Either mortality, incidence or both are zero.
* Ratios based on fewer than 10 cases for both mortality and incidence.

Table 3

Match of Mortality with Incidence Records

Disease Classification on the Cancer Incidence File for a Sample of Lung Cancer Deaths (Percentage Distribution)

Canada (excluding Ontario) and
Provinces with the Highest and Lowest Rates of Lung Cancer Incidents among the Matches.

ICDA-8	CLASSIFICATION ON THE INCIDENCE FILE	CANADA	HIGHEST	LOWEST
	1. CANCER OF THE RESPIRATORY SYSTEM	95.2	100.0	83.7
162.1	(a) LUNG; Primary	92.5	100.0	74.4
160-163	(b) OTHER RESPIRATORY SYSTEM; Primary	1.0		4.7
197.0-197.3	(c) RESPIRATORY SYSTEM; Secondary	1.7		4.7
	2. OTHER CANCERS	4.8		16.3
174	(a) BREAST	0.7		2.3
200-209	(b) LYMPHATIC PHEMATOPOIETIC SYSTEM Primary	0.7		0.0
196	Secondary	1.0		4.6
	(c) OTHER SPECIFIED PRIMARY SITE	1.9		9.3
195,199	(d) ILL DEFINED OR UNDEFINED SITE	0.5		0.0
	TOTAL	**100.0**	**100.0**	**100.0**

SAMPLE SIZE FOR MATCHES	415	54	43
MATCH SUCCESS RATE (%)	82.3	96.4	76.8

Table 4

Match of Incidence with Mortality Records

Cause of Death Classification for a Sample of Lung Cancer Cases from the Cancer Incidence File (Percentage Distribution)

Canada (excluding Ontario) and
Provinces with the Highest and Lowest Rates of Lung Cancer Deaths among the Matches.

ICDA-8	CLASSIFICATION ON THE MORTALITY FILE	CANADA	HIGHEST	LOWEST
	1. CANCER OF THE RESPIRATORY SYSTEM	92.0	100.0	85.3
162.1	(a) LUNG; Primary	91.4	100.0	82.9
160-163	(b) OTHER RESPIRATORY SYSTEM; Primary	0.3		0.0
197.0-197.3	(c) RESPIRATORY SYSTEM; Secondary	0.3		2.4
	2. OTHER CANCERS	4.3		9.8
174	(a) BREAST	0.3		0.0
200-209	(b) LYMPHATIC PHEMATOPOIETIC SYSTEM Primary	0.3		0.0
196	Secondary	-		0.0
	(c) OTHER SPECIFIED PRIMARY SITE	3.1		9.8
195,199	(d) ILL DEFINED OR UNDEFINED SITE	0.6		0.0
	3. NOT CANCER	3.7		4.9
	TOTAL	**100.0**	**100.0**	**100.0**

SAMPLE SIZE FOR MATCHES	350	45	41
MATCH SUCCESS RATE (%)	69.4	80.4	73.2

Table 5

Cancer Incidence
Availability and Completeness of Data Items

Canada (excluding Ontario) and
Provinces with the Highest and Lowest Percentages of Data Completeness

DATA ITEM	YEAR OF DIAGNOSIS	CANADA	HIGHEST PERCENT	LOWEST PERCENT
DATE OF BIRTH				
Day	1969 - 1973	63.7	99.8	0.0
	1974 - 1978	71.6	99.8	4.0
	1969 - 1978	68.0	99.3	2.2
Month	1969 - 1973	65.6	98.8	0.0
	1974 - 1978	73.3	99.8	4.1
	1969 - 1978	69.8	99.4	2.2
Year	1969 - 1973	92.9	100.0	15.9
	1974 - 1978	96.3	100.0	21.9
	1969 - 1978	94.7	100.0	19.1
Complete Birthdate	1969 - 1973	63.6	98.7	0.0
	1974 - 1978	71.6	99.8	4.0
	1969 - 1978	68.0	99.3	2.2
AGE	1969 - 1973	99.4	100.0	98.6
	1974 - 1978	100.0	100.0	100.0
	1969 - 1978	99.7	100.0	99.4
BIRTHPLACE (Country or Province)	1969 - 1973	15.2	19.8	0.0
	1974 - 1978	24.4	71.1	0.0
	1969 - 1978	20.2	46.0	0.0
RESIDENCE Census Division	1969 - 1973	89.6	100.0	4.1
	1974 - 1978	89.6	100.0	82.8
	1969 - 1978	89.6	100.0	49.0
Census Subdivision	1969 - 1973	16.6	43.2	0.0
	1974 - 1978	32.3	76.4	0.0
	1969 - 1978	25.2	61.6	0.0

Table 6

Cancer Mortality
Availability and Completeness of Data Items

Canada (excluding Ontario) and
Provinces with the Highest and Lowest Percentages of Data Completeness

DATA ITEM	YEAR OF DEATH	CANADA	HIGHEST PERCENT	LOWEST PERCENT
DATE OF BIRTH				
Day	1969 - 1973	96.6	99.4	0.0
	1974 - 1978	97.8	99.7	43.9
	1969 - 1978	97.2	99.6	23.0
Month	1969 - 1973	97.0	99.7	0.0
	1974 - 1978	98.1	99.9	44.2
	1969 - 1978	97.6	99.8	23.2
Year	1969 - 1973	100.0	100.0	99.2
	1974 - 1978	99.9	100.0	99.6
	1969 - 1978	99.9	100.0	99.4
Complete Birthdate	1969 - 1973	96.6	99.4	0.0
	1974 - 1978	97.8	99.7	43.9
	1969 - 1978	97.2	99.6	23.0
AGE	1969 - 1973	100.0	100.0	100.0
	1974 - 1978	99.9	100.0	99.6
	1969 - 1978	100.0	100.0	99.8
BIRTH PLACE	1969 - 1973	98.3	100.0	98.9
(Country or Province)	1974 - 1978	52.2	99.9	0.0
	1969 - 1978	73.9	99.9	46.2
RESIDENCE	1969 - 1973	99.8	100.0	99.6
Census Division	1974 - 1978	99.7	100.0	99.8
	1969 - 1978	99.8	100.0	99.7
Census Subdivision	1969 - 1973	92.2	99.9	51.9
	1974 - 1978	99.7	99.5	99.4
	1969 - 1978	96.2	99.7	77.1

Table 7

Cancer Incidence
1969 - 1978
Multiple Primaries Within Each Site
(Canada excluding Ontario)

ICDA Cancer Site		Multiple Primaries			
		Number	Percent of Incidence	Percent Same Registry	Percent With Same 4αm Digit ICDA Code
Total All Sites (except skin, 173)		**6,113**	**1.7**	**76.8**	**88.4**
140	Lip	87	1.4	90.8	83.9
141	Tongue	30	1.6	63.3	63.3
142	Salivary Gland	8	0.6	12.5	75.0
143	Gum	8	1.5	100.0	62.5
144	Floor of Mouth	18	1.8	94.4	N.A.
145	Mouth, Other and Unspecified	16	1.4	75.0	62.5
146	Oropharynx	14	1.0	92.9	78.6
147	Nasopharynx	8	1.1	62.5	N.A.
148	Hypopharynx	3	0.5	100.0	33.3
149	Pharynx, Unspecified	3	1.3	66.7	N.A.
150	Oesophagus	37	1.1	75.7	N.A.
151	Stomach	178	1.0	71.9	74.2
152	Small Intestine	12	1.2	91.7	91.7
153	Lge. Intestine Excl. Rectum	623	1.9	82.3	44.0
154	Rectum	256	1.4	77.0	74.6
155	Liver	11	0.6	63.6	90.9
156	Gall Bladder	22	0.7	77.3	77.3
157	Pancreas	96	1.0	70.8	54.2
158	Peritoneum	6	0.7	100.0	100.0
159	Unspec. Digestive Organs	1	0.3	100.0	N.A.
160	Nose, Etc.	5	0.7	80.0	80.0
161	Larynx	94	2.0	77.7	57.4
162	Trachea, Bronchus, Lung	795	1.9	75.0	99.5
163	Resp. Organs, Other & NOS	7	0.7	85.7	85.7
170	Bone	28	2.0	78.6	82.1
171	Connective Tissue	32	1.3	81.3	75.0
172	Melanama of Skin	61	1.4	78.7	54.1
174	Breast	1,791	3.6	84.4	N.A.

Table 7 (concl'd)

Cancer Incidence
1969 - 1978
Multiple Primaries Within Each Site
(Canada excluding Ontario)

ICDA Cancer Site	Multiple Primaries			
	Number	Percent of Incidence	Percent Same Registry	Percent With Same 4αm Digit ICDA Code
180 Cervix Uteri	133	1.4	60.9	N.A.
181 Chorionepithelioma	1	1.0	0.0	N.A.
182 Other, of Uterus	149	1.1	73.8	91.3
183 Ovary, Etc.	126	1.5	80.2	97.6
184 F. Genital Organs, Other	31	1.5	71.0	83.9
185 Prostate	464	1.6	69.2	N.A.
186 Testis	27	1.5	44.4	N.A.
187 M. Genital Organs, Other	9	1.4	88.9	100.0
188 Bladder	274	1.6	77.7	N.A.
189 Urinary Org., Other & NOS	119	1.5	72.3	79.8
190 Eye	13	1.1	46.2	N.A.
191 Brain	93	1.6	46.2	N.A.
192 Other Nervous System	6	0.4	33.3	66.7
193 Thyroid Gland	49	1.5	55.1	N.A.
194 Other Endocrine Glands	5	0.6	80.0	100.0
195 Ill - Defined Sites	4	0.6	100.0	100.0
196 Sec. & Unspec. Lymph Nodes	5	0.3	100.0	100.0
197 Sec., Resp. & Digestive	9	0.3	88.9	77.8
198 Other Secondary	1	0.1	100.0	100.0
199 Without Spec. of Site	12	0.3	75.0	83.3
200 Lymphosarcoma, Etc.	68	1.1	60.3	97.1
201 Hodgkin's Disease	85	2.2	55.3	N.A.
202 Other of Lymphoid Tissue	22	0.8	90.9	72.7
203 Multiple Myeloma	45	1.3	64.4	N.A.
204 Lymphatic Leukaemia	64	1.5	62.5	84.4
205 Myeloid Leukaemia	32	1.0	53.1	90.6
206 Monocytic Leukaemia	4	1.0	75.0	100.0
207 Other & Unspec. Leukaemia	11	0.7	81.8	90.9
208 Polycythemia Vera	1	0.2	100.0	N.A.
209 Myelofibrosis	1	0.4	100.0	N.A.