Statistics Canada / Statistique Canada

Methodology Branch

Social Survey Methods Division

Direction de la méthodologie

Division des méthodes d'enquêtes sociales

Canada

WORKING PAPER NO. SSMD-87-008

METHODOLOGY BRANCH

CAHIER DE TRAVAIL NO. SSMD-87-008

MÉTHODOLOGIE

METHODOLOGY REPORT: ESTIMATING 1987 CANCER MORTALITY and INCIDENCE IN CANADA

by

D. Dolson
K. McClean

SSMD

April, 1987

* This is a preliminary version. Do not quote without author's permission. Comments are welcome.

Abstract

In March 1987, "Canadian Cancer Statistics - 1987" was published by the Canadian Cancer Society, Toronto, Canada. This publication included projections of cancer incidence and mortality counts and age-standardized rates for 1987. The present paper describes the regression method used to compute these projections. Its application, using incidence data from 1970-1982 and mortality data from 1970-1985, is discussed.

Résumé

En mars 1987, la Société canadienne du Cancer de Toronto a publié "Statistiques canadiennes sur le cancer, 1987". Cette publication contenait des projections du nombre de cas et de décès et les taux normalisés par âge pour 1987. Le présent article décrit la méthode de régression utilisée pour leur calcul. Il contient une discussion de leur application qui emploie les données sur le nombre de cas pour 1970-1982 et sur le nombre de décès pour 1970-1985.

1. Introduction

The publication "Canadian Cancer Statistics 1987" [1] includes projections of cancer incidence and mortality counts and age-standardized rates for 1987. This report describes the methodology and data sources that were used to derive those projections.

2. Preparation of the Original Data Files

The data used to estimate 1987 cancer incidence and mortality were obtained from mortality data files and the National Cancer Incidence Reporting System maintained by the Vital Statistics and Health Status Section at Statistics Canada[2,3]. Records for all cancers except non-melanomatic skin cancer(ICD=173) were extracted from these data bases for the ten provinces. Aggregate data on cancer incidence by five year age group and sex for selected cancer sites were provided by the Ontario Cancer Treatment and Research Foundation for years prior to 1981; these were integrated into the cancer incidence data files.

Ten major cancer sites were chosen for each sex based primarily on a review of the ten most common sites of cancer incidence and mortality. For five sites the ICD-9 codes were grouped as follows: oral, 140-149; colorectal, 153-154; uterine, 179-182; lymphoid, 200-203; and leukemia, 204-208. The remaining sites were: stomach, 151; pancreas, 157; lung, 162; breast, 174; ovary, 183; prostate, 185; bladder(male only), 188; kidney(male only), 189. In addition, all cancers together(except non-melanomatic skin cancer, 173), 140-208, was considered.

Canada totals by sex for cancer incidence and mortality were found as the sum of the number of cases in the ten provinces. If no records were found for any sex, site, year, province combination it was assumed that zero events had occurred. This action was taken for a small number of records for Prince Edward Island and Newfoundland. The mortality data were available for 1970-85 and the incidence data were available for 1970-82.

Population data were obtained from the Demography Division of Statistics Canada as follows. Census counts and inter-censal estimates[4,5] for 1970-81, post-censal estimates[6] for 1982-85, and population projections[7] for 1986 and 1987 were extracted for the provinces for both sexes. The populations of the two territories were excluded from the Canadian population records.

All analyses were performed in SAS (Statistical Analysis System).

3. Notation

The regression models described below will be specified in terms of the following notation. Let

y	indicate year(1970-1987)
p	indicate province(Nfld., ... , B.C., Can.)
c	indicate cancer site
s	indicate sex
$n(p,y)_s$	be the population in year y, province p, and sex s
$i(p,y)_{sc}$	be the incidence of cancer c in year y, province p, and sex s
$m(p,y)_{sc}$	be the mortality due to cancer c in year y, province p, and sex s
$t=(y-1977)/10$	

4. Estimating the Number of Cancer Deaths for 1987

The number of cancer deaths for 1987 was estimated using the mortality and population data sets as created above. The mortality data set included the following variables: province, sex, year, an ICD9 code identifying the cancer site, and the corresponding number of deaths. The population data set used the following variables: province, sex, year, and the corresponding population count.

The estimation was done using the SAS regression procedure, PROC REG. To prepare the data for this procedure, the two files were merged together. Crude cancer mortality rates were calculated as the number of deaths divided by the population, for each province, sex, site and year combination.

$$r(p,y)_{cs}=m(p,y)_{cs}/n(p,y)_s$$

A separate regression was performed for Canada and each province for each sex and cancer site. Since the regressions were of the same form for all cancer sites and both sexes, these parameters will be omitted in the following descriptions. The dependent variable was the crude cancer mortality rate and the independent variables were the created time variable, t, and its squared value t^2. This squared term was included to account for the possible curvature of these series over time.

$$\hat{r}(p,y)=\hat{\beta}_0(p)+\hat{\beta}_1(p)t+\hat{\beta}_2(p)t^2 \qquad \text{for each } p$$

Three options of the SAS regression procedure were requested. They provided the following statistics for each year: the predicted rate, the 95% prediction interval for an individual prediction and the standard error of the individual predicted rate.

Each predicted crude mortality rate for 1987 was then multiplied by the corresponding estimated population size to arrive at a predicted number of cancer deaths.

$$\tilde{m}(p,1987)=\hat{r}(p,1987)n(p,1987) \qquad \text{for each } p$$

These predicted numbers of deaths were adjusted so that for each sex and cancer site, the sum of the predicted number of deaths in the provinces added to the predicted number of deaths in Canada. In particular, the difference between the predicted number of deaths in Canada and the sum of the predicted numbers for the provinces was calculated by

$$d=\tilde{m}(Can,1987) - \sum_{p \neq Can} \tilde{m}(p,1987)$$

This difference was allocated to each province proportional to its predicted crude cancer mortality rate to arrive at the final predicted number of cancer deaths.

$$\hat{m}(p,1987)=\begin{cases} \tilde{m}(p,1987) + \dfrac{(d\hat{r}(p,1987))}{\sum_{p \neq Can} \hat{r}(p,1987)} & p \neq \text{Canada} \\ \bar{m}(p,1987) & p=\text{Canada} \end{cases}$$

The 95% prediction interval was calculated for each m(p,1987). Finally, coefficients of variation were calculated for each projection by dividing each standard error by the corresponding predicted value. It was assumed estimated populations had zero variance.

An investigation was done to judge whether the t^2 term should be included in the regressions for mortality. The 1984 actual cancer mortality counts had just become available at that time. Using mortality data for 1970-83, predictions for 1984 and their 95% prediction intervals were found by regressions with and without the t^2 term. When the actual numbers of deaths for 1984 were compared with these 95% prediction intervals, it was found that prediction intervals with the squared time variable more often included the actual number of deaths (96% of sex, site, province combinations) than did the prediction intervals without the squared time variable(91% of sex, site, province combinations). A large number of the cases where the regressions with the t^2 term performed better were for lung cancer and all cancers. Since it had been decided, for simplicity, that the same model form should be used for all number of death regressions, and even though the regressions without the t^2 term usually performed adequately, the above evidence was considered sufficient to retain the t^2 term.

5. Estimating the Number of Cancer Incidence Cases for 1987

The number of cancer incidence cases was estimated using the incidence and population data sets created above. With a few exceptions the projection method was similar to that used for estimating the number of deaths. Again, since the regressions were of the same form for both sexes and all cancer sites these parameters are omitted for simplicity. Crude incidence rates were computed for each province, sex, site, and year combination by

$$r(p,y)=i(p,y)/n(p,y)$$

In the regression procedures, the squared time variable was not included as an independent variable. It was felt that regressions including the t^2 term based upon only 13 years of data were too likely to yield models with excessive(and inappropriate) curvature especially when projections 5 years into the future were to be made. A graphical example illustrating this effect is included as Appendix 1. This is in comparison to the mortality models where 16 years of data were available to project only 2 years into the future. Finally, cross-validation and other diagnostic methods are being used to further study the appropriateness of these models. These studies are now being developed for both the incidence and mortality series.

The model for crude incidence rates used was of the form

$$\hat{r}(p,y)=\hat{\beta}_0(p)+\hat{\beta}_1(p)t \qquad p\neq\text{Quebec}$$

It is known that changes in reporting procedures in Quebec between 1976 and 1977 resulted in improved coverage. If it is assumed that a constant fraction, (1-u), of incidence was not covered prior to 1977 then the observed crude incidence rate (r(Que,y)) can described by

$$r(Que,y)=\begin{cases} r^*(Que,y) & y=1970\text{-}1976 \\ r^*(Que,y)/u & y>1976 \end{cases}$$

Thus if the model for $r^*(Que,y)$ is given by

$$\hat{r}^*(Que,y)=\hat{\beta}_0(Que)+\hat{\beta}_1(Que)t$$

then the resulting segmented model for r(Que,y) is given by

$$\hat{r}(Que,y)=\begin{cases} \hat{\beta}_0(Que)+\hat{\beta}_1(Que)t & y=1970\text{-}1976 \\ (\hat{\beta}_0(Que)/\hat{u})+(\hat{\beta}_1(Que)/\hat{u})t & y>1976 \end{cases}$$

This nonlinear model was fitted using SAS PROC NLIN. Note that the value for u was allowed to change for different site/sex combinations.

This necessitated other changes. Regressions were done for Canada minus Quebec rather than for Canada. Projections for Canada were thus arrived at by the sum of that for Canada minus Quebec plus that for Quebec. The adjustment procedure was also changed. Projections for the other nine provinces were adjusted to add up to the Canada minus Quebec projection. Thus

$$\tilde{i}(p,1987)=\hat{r}(p,1987)n(p,1987) \qquad \text{for each } p$$

$$d=\tilde{i}(\text{Can-Que},1987) - \sum_{p\neq \text{Can,Que}} \tilde{i}(p,1987)$$

and final predictions of cancer incidence are given by

$$\hat{i}(p,1987)=\begin{cases}\tilde{i}(p,1987)+\dfrac{d\hat{r}(p,1987)}{\sum_{p\neq \text{Can,Que}}\hat{r}(p,1987)} & p\neq \text{Canada, Quebec}\\ \tilde{i}(\text{Que},1987) & p=\text{Quebec}\\ \tilde{i}(\text{Can-Que},1987)+\tilde{i}(\text{Que},1987) & p=\text{Canada}\end{cases}$$

6. Age-standardization of Cancer Mortality and Incidence Rates

The second phase of the study was to estimate the 1987 age-standardized cancer mortality and incidence rates for the major cancer sites for each province and sex. The first step was to create time series of age-standardized mortality rates (ASMR) and age-standardized incidence rates (ASIR).

The actual numbers of mortality and incidence cases were available for each cancer site, province and sex in 18 five-year age groups: 0-4, 5-9, ..., 85 and over. Matching population data were also available. Mortality rates for 1970-85 and incidence rates for 1970-82 were calculated for each age group, cancer site, province and sex.

The World population[8], shown below, was used as the basis for age-standardization. This population was used for calculation of both male and female rates.

World Population

Age	Population	Age	Population
0-4	12,000	45-49	6,000
5-9	10,000	50-54	5,000
10-14	9,000	55-59	4,000
15-19	9,000	60-64	4,000
20-24	8,000	65-69	3,000
25-29	8,000	70-74	2,000
30-34	6,000	75-79	1,000
35-39	6,000	80-84	500
40-44	6,000	85+	500
Total			100,000

Thus age-standardized rates were calculated by

$$s(p,y) = \sum_j w_j r(p,y)_j \;/100{,}000$$

where

$s(p,y)$ is the age-standardized rate for province p and year y

w_j is the world population in age group j, j=1,...18

$r(p,y)_j$ is the cancer incidence or mortality rate in province p, year y, and age group j

The series created were ASMR for Canada and each province by sex for the major cancer sites for 1970-85, and ASIR for Canada and each province by sex for the major cancer sites for 1970-82.

ASIRs for Canada for 1970-1976 were adjusted to account for the estimated undercoverage in Quebec in those years. Prior to calculation of standardized rates, the Canada incidence numbers were calculated as the sum of the incidence in the nine provinces other than Quebec plus an adjusted incidence count for Quebec. The Quebec count was adjusted according to the estimate of undercoverage resulting from the regression described in Section 5.

$$i(Can.,y)(adjusted) = \sum_{p \neq Can,Que} i(p,y) + i(Que.,y)/\hat{u} \qquad y=1970\text{-}1976$$

7. Estimation of Cancer ASMR and ASIR for 1987

With the exception of Quebec ASIRs, the estimation of cancer ASMR and ASIR for 1987 was done using the SAS regression procedure, PROC REG. Separate regressions were run for ASMR and ASIR for Canada and each province, for both sexes and for each cancer site of interest. The dependent variable was either ASMR or ASIR. In both situations, the independent variable used was the time variable t. The squared time term was not included for these regressions.

$$\hat{s}(p,y)=\hat{\beta}_0(p)+\hat{\beta}_1(p)t \qquad p\neq\text{Quebec}$$

For Quebec ASIRs segmented regressions similar to those described in Section 5 were fitted using PROC NLIN.

$$\hat{s}(Que,y)=\begin{cases}\hat{\beta}_0(Que)+\hat{\beta}_1(Que)t & y=1970\text{-}1976\\ (\hat{\beta}_0(Que)/\hat{u})+(\hat{\beta}_1(Que)/\hat{u})t & y>1976\end{cases}$$

Using the same options described in estimating the number of cases, the results obtained from these regressions were the predicted age-standardized rate, its 95% prediction interval and the standard error for an individual prediction.

8. Rounding of Estimates for Presentation in Tables

A rounding procedure was developed to present the estimates in each table, such that most estimates would include just two significant digits. For estimates of numbers of cancers or numbers of deaths, the following procedure was used:

≥1000	round to nearest 100
[50,1000)	round to nearest 10
[5,50)	round to nearest 5
[0,5)	suppress

Age-standardized rates were presented as rates per 100,000 population. These rates were rounded to the nearest integer.

References

[1] Canadian Cancer Society "Canadian Cancer Statistics 1987", Toronto, Canada, 1987

[2] Statistics Canada Catalogue 82-207 "Cancer in Canada".

[3] Statistics Canada Catalogue 84-206 "Mortality - Summary List of Causes, Vital Statistics Volume III".

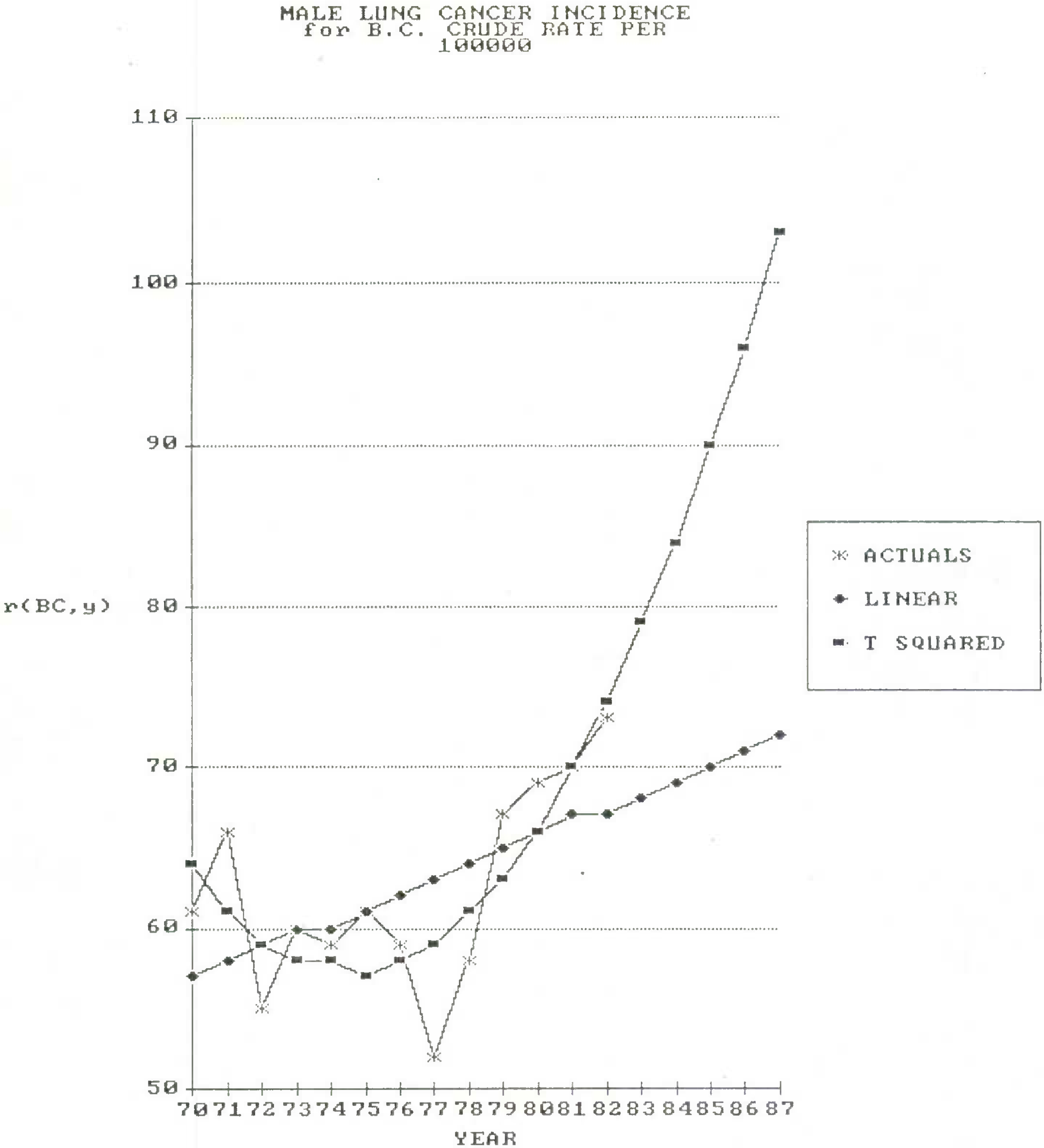
[4] Statistics Canada Catalogue 91-518 "POPULATION Intercensal Annual Estimates of Population, by Sex and Age for Canada and the Provinces, 1971-1976, and 1976-1981".

[5] Statistics Canada Catalogue 91-512 "POPULATION 1921-1971 Revised Annual Estimates of Population, by Sex and Age for Canada and the Provinces".

[6] Statistics Canada Catalogue 91-210 "Post Censal Annual Estimates of Population by Marital Status, Age, Sex and Components of Growth for Canada, Provinces and Territories, 1982, 1983, 1984, 1985".

[7] Statistics Canada Catalogue 91-520 "Population Projections for Canada, Provinces and Territorries 1984-2006".

[8] Waterhouse et al (Eds) "Cancer Incidence in Five Continents, vol. IV", CIRC, Scientific Publication no 42, International Agency for Research on Cancer, Lyon 1982.


MALE LUNG CANCER INCIDENCE
for B.C. CRUDE RATE PER
100000
110
100
90
80
70
60
50
r(BC,y)
70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87
YEAR
ACTUALS
LINEAR
T SQUARED