DEPARTMENT OF HEALTH AND HUMAN SERVICES

Public Health Service
Centers for Disease Control
and Prevention (CDC)

Memorandum

Date: October 4, 2001

From: Michael J. Galvin, Jr., Ph.D., Research Grants Program Officer
Office of Extramural Programs, NIOSH, D30

Subject: Final Report Submitted for Entry into NTIS for Grant 5 K01 OH000161-03.

To: William D. Bennett
Data Systems Team, Information Resources Branch, EID, NIOSH, P03/C18

The attached final report has been received from the principal investigator on the subject NIOSH grant. If this document is forwarded to the National Technical Information Service, please let us know when a document number is known so that we can inform anyone who inquires about this final report.

Any publications that are included with this report are highlighted on the list below.

Attachment
cc: Sherri Diana, EID, P03/C13

List of Publications missing page 62

Title: Occupational Cancer Surveillance Through Record Linkage
Investigator: Debora Boyle, Ph.D.
Affiliation: Minnesota Department of Health
City & State: Minneapolis, MN
Telephone: (612) 623-5765
Award Number: 5 K01 OH000161-03
Start & End Date: 6/1/1997–5/31/2001
Total Project Cost: $162,000
Program Area: NORA
Key Words:

Abstract:
Objectives: Statewide cancer registries offer researchers a source of high quality, detailed information on cancer diagnosis. They also create an opportunity to apply new methods to use this information in the evaluation of cancer risk in occupational cohorts. The aims of this study were to: (1) determine the feasibility of utilizing statewide cancer surveillance systems in the
evaluation of cancer incidence within occupational cohorts; (2) compare and contrast the relative merits of standardized incidence ratios with standardized mortality ratios as determined from cancer surveillance incidence data and death certificate mortality data, respectively; and (3) provide recommendations concerning how and under what circumstances statewide cancer registries should be utilized in the evaluation of occupational cohorts.

Methods: Two occupational cohorts, the Highway Worker cohort (N=3497) and Mineral Board Worker cohort (N=5086), were used to conduct standardized mortality ratio (SMR) and standardized incidence ratio (SIR) analyses for 1988-1996. Cancer mortality data were obtained from state mortality records and the National Death Index. Cancer morbidity data were obtained from the Minnesota Cancer Surveillance System, a statewide cancer registry. Because the cancer registry only collects information for Minnesota residents, it was necessary to conduct record-linkages with a variety of information sources to determine Minnesota residency status for each worker for each year from 1988 to 1996. Three models were tested to determine the effect of different residency status assumptions on the SIR estimates (i.e., follow-up bias). SMR and SIR estimates and their 95 percent confidence intervals were calculated.

Results: Over the nine years of follow-up (1988-1996), for male mineral board workers, there were 209 cancer diagnoses and 104 cancer deaths. For male highway workers, there were 297 cancer diagnoses and 153 cancer deaths. For female mineral board workers there were 27 cancer diagnoses and 11 cancer deaths.

Difficulties in validating the residency history occurred more frequently among older workers, females, and those with earlier dates of employment. Incomplete information concerning full name, date of birth, and social security number increased the difficulties in determining residency. Of the three models tested, use of out-of-state migration rates to complete unknown

residency status was the best method for completing the residency history.

For cancers with long-term survival, such as prostate cancer, melanoma of the skin, and urinary bladder cancer the SIR was higher than the SMR. Among male highway workers, however, the SIR, for bladder cancer was much lower than the SMR. For cancers with very short relative survival such as lung, pancreatic and stomach, the SMR and SIR were similar. For male mineral board workers, however, lung cancer SIRs were higher than the SMRs. Differences were noted between the SMR and SIR for which cancers reached the usual statistical significance ($p < 0.05$) including colorectal, prostate, lung, urinary bladder, non-Hodgkin's lymphoma and all cancers.

Conclusions: Standardized incidence ratio analyses are a potentially useful tool for examining the rate of cancer within an occupational cohort. Cancer mortality and incidence data provide very different perspectives about potential cancer issues within occupational cohorts. The potential problem of follow-up bias is substantial however, and such analyses should only be
undertaken after careful consideration. If both SMR and SIR analyses are conducted, differences or similarities between their estimates can be evaluated to determine if a priori study expectations are met regarding cancer diagnosis and mortality.

Publications
No publications to date.

Occupational Cancer Surveillance through Record Linkage

Grant # 5 K01 OH00161

Final Report

August 31, 2001

Principal Investigator:
Debora Boyle, DVM, PhD

Co-Investigators:
Alan Bender, DVM, PhD
Allan Williams, PhD

Project Director:
Wendy Brunner, MS

Sponsors:
Centers for Disease Control and Prevention
National Institute for Occupational Safety and Health

Chronic Disease and Environmental Epidemiology
Minnesota Department of Health
Minneapolis, MN 55440-9441

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

HCFA	Health Care Financing Administration
MDH	Minnesota Department of Health
MCSS	Minnesota Cancer Surveillance System
NCOA	National Change of Address
NDI	National Death Index
SIR	Standardized Incidence Ratio
SMR	Standardized Mortality Ratio

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Objectives: Statewide cancer registries offer researchers a source of high quality, detailed information on cancer diagnosis. They also create an opportunity to apply new methods to use this information in the evaluation of cancer risk in occupational cohorts. The aims of this study were to: (1) determine the feasibility of utilizing statewide cancer surveillance systems in the evaluation of cancer incidence within occupational cohorts; (2) compare and contrast the relative merits of standardized incidence ratios with standardized mortality ratios as determined from cancer surveillance incidence data and death certificate mortality data, respectively; (3) provide recommendations concerning how and under what circumstances statewide cancer registries should be utilized in the evaluation of occupational cohorts.

Methods: Two occupational cohorts, the Highway Worker cohort (N=3497) and Mineral Board Worker cohort (N=5086), were used to conduct standardized mortality ratio (SMR) and standardized incidence ratio (SIR) analyses for 1988-1996. Cancer mortality data were obtained from state mortality records and the National Death Index. Cancer morbidity data were obtained from the Minnesota Cancer Surveillance System, a statewide cancer registry. Because the cancer registry only collects information for Minnesota residents, it was necessary to conduct record-linkages with a variety of information sources to determine Minnesota residency status for each worker for each year from 1988 to 1996. Three models were tested to determine the effect of different residency status assumptions on the SIR estimates (i.e., follow-up bias). SMR and SIR estimates and their 95 percent confidence intervals were calculated.

Results: Over the nine years of follow-up (1988-1996), for male mineral board workers, there were 209 cancer diagnoses and 104 cancer deaths. For male highway workers, there were 297 cancer diagnoses and 153 cancer deaths. For female mineral board workers there were 27 cancer diagnoses and 11 cancer deaths.

Difficulties in validating the residency history occurred more frequently among older workers, females, and those with earlier dates of employment. Incomplete information concerning full name, date of birth, and social security number increased the difficulties in determining residency. Of the three models tested, use of out-of-state migration rates to complete unknown residency status was the best method for completing the residency history.

For cancers with long-term survival, such as prostate cancer, melanoma of the skin, and urinary bladder cancer the SIR was higher than the SMR. Among male highway workers, however, the SIR, for bladder cancer was much lower than the SMR. For cancers with very short relative survival such as lung, pancreatic and stomach, the SMR and SIR were similar. For male mineral board workers, however, lung cancer SIRs were higher than the SMRs. Differences were noted between the SMR and SIR for which cancers reached the usual statistical significance ($p < 0.05$) including colorectal, prostate, lung, urinary bladder, non-Hodgkin's lymphoma and all cancers.

Conclusions: Standardized incidence ratio analyses are a potentially useful tool for examining the rate of cancer within an occupational cohort. Cancer mortality and incidence data provide very different perspectives about potential cancer issues within occupational cohorts. The potential problem of follow-up bias is substantial however, and such analyses should only be undertaken after careful consideration. If both SMR and SIR analyses are conducted, differences or similarities between their estimates can be evaluated to determine if *a priori* study expectations are met regarding cancer diagnosis and mortality.

SIGNIFICANT FINDINGS

Background and Purpose:

Statewide cancer registries offer researchers a readily available source of high quality, detailed information on cancer diagnosis. Use of this information to conduct standardized incidence ratio (SIR) analyses for occupational cohorts is appealing. However, unlike mortality data, this information is not available on a national basis in the United States. Cancer diagnosis information is only available on state residents, since that is a usual criterion for eligibility into a state registry. Difficulties in determining the residency history for cohort members may result in follow-up bias.

Two different occupational cohorts (Highway Worker cohort and Mineral Board Worker cohort) were used to evaluate the potential for using standardized incidence ratios as a method for evaluating cancer incidence. (Note: Because this is a methodologic analysis, resulting standardized mortality ratios or standardized incidence ratios from the two cohorts are not in themselves useful nor should they be viewed as the end result of this research.) The Minnesota Cancer Surveillance System, a statewide cancer registry, was used to identify all incident cancers for both cohorts. Death certificates provided all cancer mortality data. The follow-up period for the mortality and incidence study was 1988 to 1996.

Cohort members were only "eligible" for diagnosis with cancer during the years that they resided in Minnesota. A series of computerized record-linkage activities were conducted to determine the residency history for all cohort members. Three models about completing unknown residency were developed and tested to determine their impact on the standardized incidence ratios (i.e., follow-up bias). The total person-years for each model also were compared against a "gold standard."

Cancer mortality and incidence information were used to calculate standardized mortality ratios and standardized incidence ratios for ten cancers (pancreas, prostate, lung and bronchus, kidney and renal pelvis, urinary bladder, non-Hodgkin's lymphomas, leukemias, stomach, colorectal, and all cancers). The accompanying 95 percent confidence intervals also were calculated.

1. **Determine the feasibility of utilizing statewide cancer surveillance systems in the evaluation of cancer incidence within occupational cohorts.**

   Through this study, it was demonstrated that it was feasible to use statewide cancer registry data to evaluate cancer incidence for two occupational cohorts through the use of standardized incidence ratios. A standardized mortality ratio also was calculated for the two occupational cohorts.

   The amount of time required to complete this study, and the level of complexity in conducting the record-linkages, increased with each additional information source that was utilized. In particular, missing data (i.e., name, date of birth, social security number,

address, employment dates) for cohort members or within the information source created considerably more work when adjudicating record-linkage results.

The inclusion of females in the historical occupational cohort created a uniquely difficult problem for conducting record-linkages because of potential name changes.

SIRs were calculated using the person-years at risk life table analyses and indirect standardization for age, gender, and year. To account for residency status, person-years were accumulated between 1988 and 1996 using three different residency models Difficulties in validating the residency history occurred more frequently among older workers, females, and the "historical" Mineral Board Worker cohort. The person-years matrix for each model was multiplied by the year-, sex-, and age-specific rate using Minnesota's statewide cancer rates. The 95 percent confidence intervals were calculated based on a Poisson distribution.

2. Compare and contrast the relative merits of standardized incidence ratios with standardized mortality ratios as determined from cancer surveillance incidence data and death certificate mortality data, respectively.

Over the nine years of follow-up (1988-1996), for male mineral board workers, there were 209 cancer diagnoses and 104 cancer deaths. For male highway workers, there were 297 cancer diagnoses and 153 cancer deaths. For female mineral board workers there were 27 cancer diagnoses and 11 cancer deaths.

Because of the increased number of cancer diagnoses, when compared with cancer deaths for the same time period, the confidence intervals for the standardized incidence ratios were consistently narrower and thus more precise than the standardized mortality ratios.

Standardized incidence ratio estimates were most reliable for the Highway Worker cohort followed by the male Mineral Board Worker cohort and were least reliable for the female Mineral Board Worker cohort. The SIR estimates for females were questionable because of the limited residence history and the resulting potential for follow-up bias.

In general, the same differences and similarities were seen between the standardized mortality ratio (SMR) and standardized incidence ratio (SIR) estimates in both cohorts. For cancers with long-term survival, such as prostate cancer, melanoma of the skin, and urinary bladder cancer the SIR was higher than the SMR. Among male highway workers, however, the SIR for bladder cancer was much lower than the SMR. For cancers with very short relative survival such as lung, pancreatic and stomach cancer the SMR and SIR were similar. Lung cancer SIRs, however, were clearly higher than the SMRs for male mineral board workers.

Use of out-of-state migration rates (Model C) to complete unknown residency status seemed to be a valid method for completing the residency history for both cohorts when compared to our "gold standard."

Use of capture-recapture methods were not effective in calculating the Minnesota residency history due to very wide confidence intervals around the population estimates.

Overall, the SIR facilitated our overall knowledge about cancer occurrence for these two occupational cohorts. Differences were noted between the SMR and SIR for which cancers reached the usual statistical significance ($p < 0.05$) including colorectal, prostate, lung, urinary bladder, non-Hodgkin's lymphoma and all cancers. It was useful to have the SMR as a tool for making comparisons. Cancer mortality and incidence data provide very different perspectives about potential cancer issues within occupational cohorts.

3. **Provide recommendations concerning how and under what circumstances statewide cancer surveillance systems should be utilized in the evaluation of occupational cohorts.**

Standardized incidence ratio analyses are a potentially useful tool for examining the rate of cancer within an occupational cohort. The potential problem of follow-up bias is substantial however, and such analyses should only be undertaken after careful consideration.

Difficulties in validating the residency history occurred more frequently among older workers, females, and the "historical" Mineral Board Worker cohort. Incomplete information concerning full name, date of birth, and social security number increased the difficulties in determining residency and also increased the staff time needed to complete the study. Increased levels of in- and out-state migration also compounded the potential for follow-up bias.

Although we did not examine the impact of cancer diagnoses missed due to out-of-state residency in detail within this study, its potential impact should be evaluated. Occupational cohorts experiencing higher levels of migration in- and out-of-state will be at higher risk of having these "missed" diagnoses. Because of the interplay between the observed and expected cancer diagnoses (i.e., potential for follow-up bias), it is difficult to predict whether this error would increase or decrease the SIR.

If both SMR and SIR analyses are conducted, differences or similarities between their estimates can be evaluated to determine if *a priori* study expectations are met regarding cancer diagnosis and mortality. Very little additional staff time and costs were directed to the SMR analyses, outside of what was already needed for the SIR analyses, since the only additional information required was the cause of death.

USEFULNESS OF FINDINGS

Background and Purpose:

One of the standard methodologies used to examine the risks for cancer and other causes of death among occupational cohorts has been the standardized mortality ratio (SMR). This method is well understood and the causes of death are available nationally through state vital records offices or through the National Death Index. However, there is always the potential for "missed" cancer diagnoses or misclassification of cancer on death certificates, especially for cancers such as mesothelioma.

Statewide cancer registries offer researchers an alternate source of high quality, detailed information on cancer diagnosis. Use of this information to conduct standardized incidence ratio (SIR) analyses is appealing. However, unlike mortality data, cancer registry data are not available on a national basis in the United States. Cancer diagnosis information is only available on state residents, since that is a usual criterion for eligibility into a state registry. Difficulties in determining the residency history for cohort members would result in follow-up bias. For example, if the total person-years at risk were overestimated (by assuming, for example, that all cohort members have remained in the state), the standardized incidence ratio would decrease. The impact of follow-up bias would depend on the validity of the person-years at risk estimate.

To evaluate the potential for using standardized incidence ratios as a method for evaluating cancer incidence within occupational cohorts, we used two different existing occupational cohorts (Highway Worker cohort and Mineral Board Worker cohort). (Note: Because this is a methodologic analysis, resulting SMRs or SIRs from the two cohorts are not in themselves useful nor should they be viewed as the end result of this research.)

The Highway Worker cohort is an example of a more "current" cohort. It was comprised of 3,497 males who were employed as highway workers by the state of Minnesota for at least one year between 1945 and 1989. These workers were located throughout Minnesota. As a result of prior studies, virtually complete information was available for record-linkage activities (i.e., name, date of birth, and social security number).

The Mineral Board Worker cohort is an example of a more "historic" cohort. It was comprised of 5,086 males and females who were employed in a single mineral board manufacturing plant in northeastern Minnesota. These workers were employed at any time between 1958 and 1974, the time period in which large quantities of raw asbestos were used in the manufacturing of ceiling tiles and other products. Because workers were at high risk of asbestos-related diseases, a risk notification program was launched in which company employment records and Social Security Administration payroll records were used to identify all people employed at this facility. Due to the fact that Social Security Administration data, which provided only name and social security number, was a major source of information used to determine cohort eligibility, a large percentage of workers were missing date of birth or gender information. Name and social security number information were available for the complete cohort.

In this study, we used the Minnesota Cancer Surveillance System, a statewide cancer registry, to identify all incident cancers for both cohorts. Because this registry was implemented on January 1, 1988, our follow-up period for this study began on that date and continued until December 31, 1996. Cancer mortality data were obtained from state mortality records and the National Death Index. The follow-up period for the mortality study was the same as for the incidence study (1988-1996).

Cohort members were only "eligible" for diagnosis with cancer during the years that they resided in Minnesota. A series of computerized record-linkage activities were conducted to determine the residency history for all members of both cohorts. Information sources that provided residency data included credit bureau records, drivers license records, voter registration records, death certificates, National Death Index records, and the statewide cancer registry. Minnesota Department of Revenue information was only used as a "gold standard" for comparisons of total person-years at risk; it was not used to create the residency histories. Although these sources did help create a residency history, only a small percentage of the workers had complete residency information from 1988 to 1996. Three models were developed to test different assumptions about completing unknown residency. Model A assumed all Minnesota residency until 1996 or date of death. Model B utilized all of the residency record-linkages, but assumed Minnesota residency for all unknown years. Model C utilized all of the residency information from the record-linkages, but incorporated out-of-state migration information for all unknown years.

Cancer mortality and cancer incidence data were used to calculate standardized mortality ratios and standardized incidence ratios for ten cancers (pancreas, prostate, lung and bronchus, kidney and renal pelvis, urinary bladder, non-Hodgkin's lymphomas, leukemias, stomach, colorectal, and all cancers). The accompanying 95 percent confidence intervals also were calculated. The total person-years for each model were then compared against a gold standard and Model C had the closest estimates.

1. Determine the feasibility of utilizing statewide cancer surveillance systems in the evaluation of cancer incidence within occupational cohorts.

Information from statewide cancer registries provides a number of benefits when compared to cancer information obtained from death certificates. (1) The expected numbers of incident cancers per year are approximately twice that of cancer deaths. (2) For cancers with longer survival, exposure information may be obtained by interviewing workers rather than through spouses or other surrogates. (3) Some statewide cancer registries have existed long enough to contribute years or decades of information about cancer in their states, although this is still less than is available through death certificates. (4) The latency period from exposure to cancer diagnosis is often much shorter than for exposure to death. (5) There is a breadth of validated, accurate, information concerning cancer diagnosis, multiple primaries, cell type, stage and grade potentially available from cancer registries (i.e., less misclassification). This information is not available from death certificates. (6) Confounding may be limited to factors affecting cancer diagnosis, whereas confounding could affect both cancer diagnosis and survival for mortality studies.

A primary consideration – and limitation – to using registry data to conduct standardized incidence ratio analyses is the potential for follow-up bias. Because the United States does not have a national data repository on cancer diagnosis, cohort members are only "eligible" for diagnosis with cancer during the years that they resided in the state with the cancer registry. Follow-up bias will affect the standardized incidence ratio if the residency history has not been established accurately. Invalid residency estimates will lead to an invalid estimate for the number of "expected" cancers, due to errors in calculating total person-years at risk.

Information sources used in this study to provide residency information included credit bureau records, voter registration records, drivers license records, Minnesota death certificates, National Death Index records, and cancer registry records. Minnesota death certificates and National Death Index records provided date of death and cause of death information for the SMR analyses. The cancer registry provided date of diagnosis and type of cancer diagnosis data for the SIR analyses.

Sensitivity analyses were conducted using three separate assumptions about residency status (i.e., Models A, B, and C). Model A was the most liberal model in terms of determining Minnesota residency (i.e., it maximized the number of person-years at risk in the denominator of the SIR estimate). In this model, none of the residency information from any of the information sources was used. Each worker was assumed to have been a Minnesota resident from 1988 until 1996 or until the date of his or her death. Although this model was basically unsound in its assumptions, it formed the basis for the most extreme example possible concerning Minnesota residency and resulted in the lowest SIR.

Model B used all of the residency information provided by the information sources, but then assumed Minnesota residency for all unknown residency years (i.e., all unknown residency years were "filled" as Minnesota residency). This model was also quite liberal in terms of assuming Minnesota residency, but it did utilize all information that was known concerning out-of-state (non-Minnesota) residency.

Model C was the closest estimation to "truth" about residency status. Initially, it used all of the information about residency status from each of the information sources, but then supplemented the data in two ways. First, a set of assumptions was used to "fill" state of residency information between gaps found in one of the information sources. For example, if a cohort member was found to have voted in Minnesota during 1994 and 1996, then they were assumed to be a Minnesota resident in 1995.

Second, the remaining unknown residency years were filled using information about Minnesota's out-of-state migration rate (3%). For example, using the 3 percent migration rate, if we knew that a worker lived in Minnesota in 1988, but then had no information on residency status until 1996, the succeeding years were filled in the following manner: 1989=0.97; 1990=0.97^2; 1991=0.97^3; 1992=0.97^4; 1993=0.97^5; 1994=0.97^6; and 1995= 0.97^7. If there was no residency information for 1988, then the same pattern was used to fill succeeding years beginning with the last known date of employment.

Although information about out-of-state migration was available, there was no available information on the number or percent of individuals who moved out of Minnesota and subsequently returned. For this reason, the 3 percent figure for out-of-state migration also was used for in-state migration. For example, if we knew that a worker was not a Minnesota resident in 1988 and had no further residency status until 1996, the succeeding years were filled in the following manner: 1989=(1 - 0.97); 1990=(1 - 0.97^2); 1991=(1 - 0.97^3); 1992=(1 - 0.97^4); 1993=(1 - 0.97^5); 1994=(1 - 0.97^6); and 1995= (1 - 0.97^7). All unknown residency years were filled in this manner.

A comparison of Model A (i.e., assumed all Minnesota residence with no information sources used) to Model B (i.e., all residency information sources were utilized, but all unknown residency was assumed to be Minnesota) showed no significant differences in the standardized incidence ratio or the resulting confidence intervals. Unless additional information sources were used or additional employment information was available, Models A and B would not provide a valid standardized incidence ratio for cohorts that were known to have any significant levels of out-of-state migration.

In Model C, the standardized incidence ratio decreased as the number of information sources used to determine/confirm residency status increased. The lowest standardized incidence ratios were calculated when all residency information sources were utilized; the loss of even one source increased the SIR. Credit bureau records contributed significantly, in this study, to the residency status because they included data about in-state and out-of-state residency for a long time period. It is likely that the use of Health Care Financing Administration data, or other national information sources, would also be very useful in establishing out-of-state residency. Where migration probabilities were not utilized (Model B), the number of information sources had little impact on the SIRs.

The total years of Minnesota residency, calculated from Model C, were similar to residency information obtained from the Minnesota Department of Revenue, our "gold standard." This comparison was closer for the Highway Worker cohort than for the Mineral Board Worker cohort however. This difference was probably due to difficulties in conducting record-linkage activities for the Mineral Board Worker cohort because of missing or incomplete data, unavailability of employment information within the follow-up period, and the inclusion of females.

Model C performed better in estimating person-years at risk for both the more "current" cohort of highway workers as well as for the "historic" cohort of mineral board workers. The benefit to using this approach is that it allows researchers to supplement known residency information in a way that is appropriate for their state and consistent with their knowledge about available information sources. State population migration statistics are available nationally and can be validated, to a limited degree, by comparison with known, out-of-state residency data gained from the record-linkage activities with the information sources.

Based on the results of these three models, it is clear that there is great potential for bias in the standardized incidence ratio. Every percentage difference in the total number of residency years for the cohort resulted in a corresponding change in the opposite direction for the standardized

incidence ratio. For example, a 5 percent decrease in residency years resulted in a 5 percent increase in the SIR estimate. All assumptions concerning residency status need to be examined carefully for the standardized incidence ratios to be valid.

The percentage of females within the Mineral Board Worker cohort with no initial residency data from any of the information sources was much higher (61%) than for males (27%). The best residency data was available for highway workers; only five percent of the cohort was missing residency information for the entire follow-up period (1988-1996). Based on this information alone, it is likely that the most valid SIR estimates were for the highway workers and the least valid estimates were for the female mineral board workers. The validity of SIR estimates derived from very "thin" residency history data, as demonstrated by the female mineral board workers, would be questionable.

Researchers should consider identifying data sources that could act as a "gold standard" for their occupational cohorts. Previous tracing efforts in this state for a high-risk worker notification program had demonstrated that the Department of Revenue was able to locate state residents not found by any other source or combination of sources. Although state data privacy regulations precluded using Minnesota revenue address information on an individual level for this study, the Department of Revenue was able to provide annual counts of the number of Minnesota residents in each cohort. This information was critical in evaluating the performance of our three models.

Capture-recapture analysis was conducted to determine if this method could be used to determine the number of years of Minnesota residency for each cohort. It was clear from our analyses that the point estimates and extremely wide confidence intervals produced using this methodology were inadequate to determine person-years at risk (i.e., Minnesota residency).

Probabilistic record-linkage resulted in very few "false" matches when both the cohort and the information source provided complete and accurate information. When complete and accurate information was not available, however, the number of matches to be evaluated, and the staff time required to review the data, increased significantly. The amount of time that was required to complete this study, and the level of complexity in conducting the record-linkages, increased with each additional information source utilized. In particular, missing data (i.e., name, date of birth, social security number, address, employment dates) for cohort members or within the information source created considerably more work when adjudicating record-linkage results. Because of missing or incorrect data, the time required to conduct and adjudicate the record-linkage activities for the males and females in the Mineral Board Worker cohort (N=5086) was at least twice that needed for the Highway Worker cohort (N=3497).

The inclusion of females in the historical occupational cohort created a uniquely difficult problem for conducting record-linkages because of potential name changes. Women may change their middle and/or last names when they marry or divorce. If the study investigators are aware of name changes they can create "duplicate" records with different name combinations to ensure that the record-linkage process finds all possible matches. However, conflicting name information, when found for women, does not automatically equate to be a non-match, especially if the date of birth or city of residence are the same. Manual adjudication of these potential matches was always required.

Overall, this study demonstrated that it was feasible to use statewide cancer registry data to evaluate cancer incidence for two occupational cohorts through the use of standardized incidence ratios. Adequate resources for staffing, computers, programming skills, and sources of residency information need to be available for the development of valid residency histories. All assumptions regarding completing unknown residency information need to be carefully evaluated.

2. Compare and contrast the relative merits of standardized incidence ratios with standardized mortality ratios as determined from cancer surveillance incidence data and death certificate mortality data, respectively.

For all cancers combined, the ratio of the number of male workers diagnosed with a new pathology-based cancer compared to the number of cancer deaths was consistent with known population-based cancer statistics. Approximately twice as many cancers were identified from 1988-1996 through the cancer registry as were found on death certificates (i.e., National Death Index records, Minnesota death certificates). For male mineral board workers, there were 209 people diagnosed with cancer and 104 cancer deaths. For male highway workers, there were 297 people diagnosed with cancer and 153 cancer deaths. For female mineral board workers there were 27 cancer diagnoses and 11 cancer deaths, which is also proportionate to state and national data.

Differences between the standardized mortality ratio (SMR) and standardized incidence ratio (SIR) estimates could be due to several factors acting alone or in combination with each other. First, some differences might be expected solely from random variation, especially when the number of observations (diagnoses or deaths) was small and the estimate had a wide confidence interval. Second, there are large differences in the survival rate for different cancers. A rapidly fatal cancer is much more likely to be found in mortality records. Third, assumptions about Minnesota state residency status directly affected the SIR. Difficulties in validating the residency history occurred more frequently among older workers, females, and the "historical" Mineral Board Worker cohort. Fourth, cancer registry data on cancer diagnosis is more accurate than cancer diagnoses identified on death certificates (i.e., diagnosis misclassification).

Because of the increased number of cancer diagnoses, when compared with cancer deaths for the same time period, the confidence intervals for the standardized incidence ratios were consistently narrower and thus more precise than the standardized mortality ratios.

It was assumed that the occurrences, confidence intervals, and SMR, SIR estimates would be similar for cancers with very short relative survival rates. Among the ten cancers we evaluated, pancreas, lung, and stomach cancer had among the worst survival rates of 3.9, 13.3, and 19.1 percent, respectively (5-year relative survival rate for white males, 1992-1997) (Ries, 2001). Lung cancer estimates were nearly equivalent for highway workers and female mineral board workers, but SIRs were clearly higher than SMRs for male mineral board workers (discussed below). Pancreatic cancer was relatively rare, so the confidence intervals for both the SMR and SIR were quite broad. Although the actual point estimates were different, the confidence

intervals overlapped. Stomach cancer estimates and confidence intervals were very similar in both cohorts, but the number of stomach cancer deaths among male mineral board workers, however, was very low (N=2) and resulted in a very wide confidence interval.

It was assumed that occurrences, confidence intervals, and SMR, SIR estimates would be quite different for cancers with good, long-term survival such as prostate cancer, melanoma of the skin, urinary bladder cancer, and female breast cancer. Their respective 5-year relative survival rates for white males, 1992-1997 are 97, 86.6, and 83.9 percent. The 5-year relative survival rate for breast cancer in white females is 86.8 percent (Ries, 2001). As expected, the SIR for prostate cancer was higher than the SMR. Likewise, the SIR for melanomas of the skin was higher, although there were no deaths reported among male highway workers and only one death reported among male mineral board workers. The SIR for bladder cancer was higher than the SMR for male mineral board workers, but the confidence interval fell within the confidence interval for the SMR. The opposite situation occurred among male highway workers with the SIR being much lower than the SMR. In addition, there was only a minimal degree of overlap between the SIR and SMR confidence intervals.

For female breast cancer, the SMR was higher than the SIR was, but the confidence intervals were wide and largely overlapped.

Differences were noted for which cancers the SMR and SIR reached the usual criteria of statistical significance ($p < 0.05$). Among male highway workers, colorectal, prostate, and "all" cancer SIRs were significantly elevated, while their corresponding SMRs were not significantly elevated. The opposite situation occurred for lung cancer, urinary bladder cancer, and non-Hodgkin's lymphoma. Among male mineral board workers, lung cancer, kidney and renal pelvis cancer, and "all" cancer SIRs were significantly elevated, while their corresponding SMRs were not significantly elevated. There were no instances where the SMR was elevated significantly. Among female mineral board workers, none of the cancers were significantly different than their expectation.

Although unexpected, the SIR and SMR estimates for lung cancer were different for male mineral board workers. Similarities and differences between death certificate diagnosis of lung cancer and cancer registry diagnosis of lung cancer were evaluated to determine why these estimates were different. There were 55 men identified with lung cancer from either source. Eighteen men were identified with lung cancer by both sources. Twelve men did not have any cancer diagnosis in the cancer registry but were identified by death certificates as dying from lung cancer; 4 of these 12 men died out-of-state and of the remaining 8, the date of death was 1988 (2), 1989 (1), 1991 (2), 1992 (2), and 1995 (1). Fifteen men had a registry diagnosis of lung cancer and had no date of death; 8 of these 15 men were known to have died after 1996 (from a partial Minnesota death certificate search) and the others were diagnosed in 1988 (1), 1989 (1), 1993 (1), 1995 (1) and 1996 (3). Seven men had a diagnosis of lung cancer in the registry but the death certificate had a different cause of death (6 died from other endocrine and 1 died from non-cancer causes). Three men were in the registry with other cancers but the death certificate listed lung cancer as the cause of death.

It is clear from the example above concerning lung cancer that there are a variety of factors that result in a different standardized incidence ratio and standardized mortality ratio for even a very lethal cancer. The same workers are not necessarily identified by both death certificates and cancer registries. Here, over a nine-year period, only 18 individuals were included in both the cancer registry and in the death records as a lung cancer, while 37 individuals were counted as a lung cancer in one source but not the other. Certainly, the out-of-state cancer diagnoses and deaths should not have been identified by the cancer registry. Also, the cancer diagnosis might have been made out-of-state and the person may have moved back into the state to receive treatment or to die. In a third scenario, the person may have been diagnosed with cancer, but died from a cause other than cancer.

In the early years of follow-up (e.g., 1988, 1989, or 1990) the cancer registry may not have identified a cancer that was recorded on a death certificate. Intuitively, these cancers may have been diagnosed before the follow-up period began. In the later years of follow-up (e.g., 1994, 1995, or 1996), the cancer registry may have recorded a cancer diagnosis, but the resulting cancer death may not have occurred until outside the follow-up period. Short periods of follow-up, such as when a cancer registry has only existed for a few years, might affect the resulting SIR. Although this occurred to a limited degree with this example of lung cancer, the effect would probably be more prevalent among cancers with longer survival rates.

Any problems with identifying matches within either the cancer registry or the death certificates would also cause changes to the SMR or SIR. An addition or subtraction of even one "case" may have an effect on the estimate; especially for rare cancers. It is likely that any "false" matches would tend to occur with death certificate data rather than cancer registry data because of differences in data quality for name, date of birth or social security number.

In addition, misclassification of cancer on the death certificate would cause differences between the SMR and SIR. Cancer registries are frequently subjected to quality control studies, but death certificates are not held to the same standard for data quality or completion. As an example, only one cancer death among the Mineral Board Worker cohort was coded as a "pleural" cancer or any other code consistent with mesothelioma. The cancer registry, however, revealed 11 malignant mesotheliomas in this cohort. Malignant mesothelioma, which has a life expectancy of less than one year, is frequently miscoded on death certificates.

In this study the standardized incidence ratio was a useful adjunct to the standardized mortality ratio. The SIR and SMR estimates were, for the most part, similar for lethal cancers and different for cancers with good five-year survival rates. In the one example where this was not true (i.e., lung cancer among male mineral board workers), it was evident that there were several logical reasons for this difference. By having both estimates, it allowed us to postulate the reasons for the level of consistency and/or inconsistency for each of the cancers.

Use of out-of-state migration rates to complete unknown residency status seemed to be a valid method for completing the residency history for both cohorts. The SIRs for the Highway Worker cohort were probably slightly closer to "truth" than those for the Mineral Board Worker cohort due to differences in our ability to confirm residency status for the cohorts. Significant elevations

in the SIR occurred for cancers that did not have elevated SMRs. Overall, the SIR did expand our overall knowledge about cancer occurrence for these two occupational cohorts.

3. Provide recommendations concerning how and under what circumstances statewide cancer surveillance systems should be utilized in the evaluation of occupational cohorts.

Statewide cancer registries offer a newly emerging opportunity for obtaining information about incident cancers diagnosed within an occupational cohort. Many statewide cancer registries have now been in existence for enough years that sufficient incidence data exists to conduct specific analyses. The use of a standardized incidence ratio (SIR) analysis, however, requires that information about residency history be collected for all the years of follow-up. Because the United States does not have a national data repository on cancer diagnosis, cohort members would only be "eligible" for diagnosis with cancer during the years that they resided in their own state.

An accurate and precise standardized incidence ratio estimate, by cancer type, can only be estimated if the cancer registry has been collecting data for a number of years. In addition, the state of residency must be determined for each cohort member on an annual basis for the entire follow-up period. A number of factors must be considered before deciding if this residency status history can be compiled in a manner that is accurate and complete enough to provide valid SIR estimates. These factors include:

- A complete roster of cohort members with a minimum of 20 years of latency until cancer diagnosis.
- Identifying information needed for record-linkage activities and "match" resolution with information sources including complete and accurate information for first, middle and last name, date of birth, gender, and social security number. Historical and/or current address information would be extremely helpful, as would a listing of all name changes for females.
- Sources of information for determining date of death and cause of death information (i.e., state death certificates, National Death Index, Social Security Administration vital status).
- Sources of information for determining date of cancer diagnosis and type of cancer diagnosis (i.e., population-based state cancer registry).
- Potential sources of information for determining a residency status history for each worker in the cohort (i.e., employment records, voter registration records, drivers license records, credit bureau records, death records, cancer registry records, Health Care Financing Administration records, pension records, state tax records). Multiple sources of information are required to create a residency status history. Validation of out-of-state residency requires the use of a minimum of one national information source such as credit bureau records. The validity of the SIR estimate increases as additional sources of residency information are utilized.
- A method for completing unknown residency status must be determined and tested (e.g., using the rate of out-of-state migration).

- Additional money, study time, computer programming resources, and staffing are needed to conduct an SIR analysis as compared with the resources needed for an SMR analysis.
- Because date of death information is required to build the residency status history, it is recommended that a standardized mortality (SMR) analysis be conducted to coincide with the SIR analysis. Differences or similarities between SMR and SIR estimates can be evaluated to determine if *a priori* expectations are met.

Unless each of these factors is addressed, an SIR analysis should not be conducted in conjunction with an SMR for occupational cohorts since the results may not be interpretable. Invalid residency histories may result in follow-up bias that increases or decreases the SIR. All SMR and SIR estimates need to be carefully evaluated using all available information about work histories, possible exposures, and the potential movement of workers in- and out-of-state.

SCIENTIFIC REPORT

The increased number and enhancement of statewide cancer surveillance systems in recent decades has created an opportunity to apply new methods to use this information in the evaluation of cancer risk in occupational cohorts. Cancer research studies that utilize incidence data will be particularly valuable because: (1) cancer registries provide high quality cancer incidence data, (2) a number of occupationally-related cancers have good survival, (3) cancer morbidity requires a shorter latency period than mortality, and (4) cancer incidence studies require fewer subjects than mortality studies.

The goal of this study was to determine if and when cancer risks could be estimated by establishing record-linkages between a statewide cancer registry and occupational cohorts. The primary issue to be resolved in this evaluation was the potential impact that follow-up bias might have on the resulting standardized incidence ratios. The potential for follow-up bias was directly related to the requirement that cohort members be a resident of the state with the cancer registry before they would be counted as an eligible person for that system. In essence, cohort members were only "at risk" for being diagnosed with an incident cancer during the years that they were a state resident. While it may be assumed that a state population-based cancer registry will identify state residents who develop cancer, it cannot be assumed that all cohort members are state residents during follow-up. In- and out-of-state migration would impact the person-years of the occupational cohort, and therefore the expected cases and corresponding standardized incidence ratios.

The aims of this study were to:

- Determine the feasibility of utilizing statewide cancer surveillance systems in the evaluation of cancer incidence within occupational cohorts.
- Compare and contrast the relative merits of standardized incidence ratios with standardized mortality ratios as determined from cancer surveillance incidence data and death certificate mortality data, respectively.
- Provide recommendations concerning how and under what circumstances statewide cancer surveillance systems should be utilized in the evaluation of occupational cohorts.

1. Description of Occupational Cohorts

Note: Because this is a methodologic analysis, resulting SMRs or SIRs for the following cohorts are not in themselves useful nor should they be viewed as the end result of this research.

Mineral Board Worker Cohort

In 1988, the Minnesota Department of Health conducted a screening program of 1552 former employees of the mineral board manufacturer and their spouses (MDH, 1989; Williams, 1994). (These individuals were a selected subset of a much larger cohort of former employees for whom follow-up information was available.) Asbestos-containing ceiling tile and wallboard was

manufactured at the Cloquet, Minnesota, facility from 1958 to 1974. This was the largest single mineral board plant in the world and employed about 1,200 people at any given time.

This screening program was conducted to: confirm and expand the findings of a union study (Robins and Green, 1988); inform individuals whether they had signs of asbestos-related diseases; identify those categories of people with higher risk by virtue of their work histories; determine to what extent family members have been indirectly exposed; and identify previous workers and inform them and their families of their risks. Screening data showed clear evidence that the workers at this facility had significant exposure to asbestos. Because workers were at high risk of asbestos-related diseases, a risk notification program was launched in which company employment records and Social Security Administration payroll records were used to identify all people employed at this facility at any time between 1958 and 1974 (Bender et al., 1993).

For purposes of this study, the Mineral Board Worker cohort consisted of 5,086 former employees who worked some time between 1958 and 1974 and were alive as of January 1, 1988. Of the 1552 individuals that participated in the 1988 medical screening, a subcohort of 874 male employees were available for a stratified analysis about potential confounding. A description of the study cohort prior to any record-linkage activities is included in Table 1. Record-linkage activities were used to complete missing data.

The Mineral Board Worker cohort was utilized in this effort to conduct standardized mortality ratio, standardized incidence ratios, and capture-recapture analyses. The subcohort of screened workers was utilized, through a stratified analysis, to examine the possible impact of confounding.

Highway Worker Cohort

This cohort was a subset of the Minnesota Department of Health's Highway Worker cohort made up of 5,253 male highway maintenance workers employed by the Minnesota Department of Transportation for at least one year between 1945 and 1989 (MDH, 1987; Bender, Parker et al., 1989; Parker, Bender et al., 1989; MDH, 1990; MDH, 1993). Work sites for these workers were located throughout Minnesota. For purposes of this study, the criteria for inclusion required that workers were alive as of January 1, 1988. Employment information also was available for those employed in 1988 and 1989. A description of the study cohort prior to any record-linkage activities is included in Table 1. Record-linkage activities were used to complete missing data.

The Highway Worker cohort was utilized in this effort to conduct standardized mortality ratio, standardized incidence ratio, and capture-recapture analyses.

Table 1. Mineral Board Worker Cohort (Males and Females) and Highway Worker Cohort (Males): Cohort Characteristics and Data Available for Evaluating Record-Linkage Matches

Cohort Characteristics	Highway Worker Cohort Number (Percent)		Mineral Board Worker Cohort Number (Percent)	
Number of Eligible Workers	3,497		5,086	
Gender				
Male	3,497	(100)	1,061	(20.9)
Female	0	(0)	129	(2.5)
Unknown	0	(0)	3,896	(76.6)
Date of Birth				
Complete Information	3,497	(100)	3,060	(60.2%)
Range	1892 - 1969		1888 - 1959	
Median	1937		1943	
Years Employed (Range)	1945 - 1989		1958 - 1974	
Work Location	Throughout Minnesota		One plant in northeastern Minnesota	
Exposure Information	Unavailable		Medical screening information available for a subcohort	
Data Available for Evaluating Record-Linkage Matches				
Social Security Number	3,434	(98)	5,086	(100)
Month, Day, & Year of Birth	3,497	(100)	3,060	(60.2)
First Name Information				
Complete Information	3,493	(99.9)	4,803	(94.4)
Initial Only	4	(0.1)	281	(5.5)
No Information	0	(0)	2	(0.04)
Middle Name Information				
Complete Information	1,699	(48.6)	2,549	(50.1)
Initial Only	1,602	(45.8)	2,254	(44.3)
No Information	196	(5.6)	301	(5.9)
Last Name Information				
Complete Information	3,497	(100)	5,083	(99.9)
Initial Only	0	(0)	2	(<0.1)
No Information	0	(0)	1	(<0.1)
Multiple Last Names	0	(0)	146	(2.9)
Address Information	193	(5.5)	1,243	(24.4)

3M Cohort

This cohort was made up of male and female employees of 3M's St. Paul or Maplewood facilities, who worked at least one year between 1947 and 1992. For the purposes of this study, the criteria for inclusion required that workers were alive as of January 1, 1988. Work histories also were available for those employed between 1988 and 1996.

The 3M cohort was to be utilized in this effort to conduct capture-recapture analyses. Standardized mortality ratio and standardized incidence ratio analyses were to be conducted independently by 3M as a separate initiative.

2. Description of Minnesota Cancer Surveillance System (MCSS)

The Minnesota Cancer Surveillance System (MCSS) is a population-based cancer registry for the state of Minnesota and collects only microscopically confirmed cancers. MCSS was implemented in Minnesota on January 1, 1988 and remains an ongoing program within the Section of Chronic Disease and Environmental Epidemiology at the Minnesota Department of Health.

MCSS data have been shown to meet high standards of data quality (Minnesota Cancer Surveillance System, 1997; Chen, Howe et al., 2000). An external audit (Minnesota Cancer Surveillance System, 1997) performed in June 1996 by the Cancer Surveillance and Control Program, in collaboration with the North American Association of Central Cancer Registries (NAACCR), estimated that MCSS hospital-based case finding was 99.6 percent complete for microscopically confirmed cancers, and 99.1 percent complete for all cancers. (The audit included colorectal, prostate, lung, and female breast cancers.) The most recent adjusted NAACCR estimate of completeness for MCSS data was 101.1 percent (Chen, Howe et al., 2000). The NAACCR estimate is based on a standard of a rate ratio for SEER incidence and U.S. mortality, applied to the state mortality rates for the registry's geographic coverage area and is adjusted based on an assessment of the number of duplicates within the registry's data. The MCSS estimates that ascertainment is between 95 and 97 percent complete considering all cancer sites and all methods of diagnosis.

3. Description of Standardized Mortality Ratio (SMR) and Standardized Morbidity Ratio (SIR) Methods

Standardized mortality ratios (SMRs) make use of age-, sex-, and race-specific mortality rates to compare health outcomes between populations. Indirect standardization (adjustment) involves applying mortality rates from some selected reference population, adjusted for age, gender, and year, to the study population (Fleiss, 1981). This procedure generates the number of deaths that would be "expected" if the study population had experienced the same disease mortality as the reference population. In this study Minnesota served as the reference population.The SMR is obtained by dividing the observed number of deaths by the expected number.

In this study, SMRs were calculated using the person-years at risk life table analyses (Waxweiler, 1983; Hill, 1972). Person-years were accumulated by five year age groups (0-4, 5-9, ...85+), calendar year (1988-1996), and sex, from 1988 until death or the end of follow-up (1996).

The expected number of deaths was calculated according to the methods described by Hill. The person-years matrix for each cohort was multiplied by the year-, sex-, and age-specific rate using Minnesota state-wide mortality rates.

The observed number of cancer deaths was obtained through a search of the National Death Index records and Minnesota death certificates. For the period of this study, ICD 9 cause-of-death codes were used. Only the underlying cause of death was analyzed.

The SMR was calculated by dividing the observed number of deaths by the expected number of deaths. The 95% confidence intervals were calculated based on a Poisson distribution (Bailar and Ederer, 1964; Breslow and Day, 1987).

Standardized incidence ratios (SIRs), in the context of this study, were calculated in a similar fashion as standardized mortality ratios. Rather than comparing the ratio of observed and expected deaths, SIRs compare the ratio of observed and expected incident (newly-diagnosed) cases. Statewide cancer registry data were used to enumerate incident cancer cases within the cohort and for Minnesota-specific population-based incidence rates.

Because individuals were eligible to be counted by the state cancer registry (Minnesota Cancer Surveillance System) only when they were Minnesota residents, it was necessary to determine residency status before the denominator could be calculated. An individual's person-years accumulated only while they were alive and a Minnesota resident.

SIRs were calculated using the person-years at risk life table analyses (Waxweiler, 1983; Hill, 1972). Person-years were accumulated by five year age groups (0-4, 5-9, ...85+), calendar year (1988-1996), and sex. To account for residency status, person-years were accumulated between 1988 and 1996 using different assumptions (residency models) about annual Minnesota residency status. For all residency models, person-years stopped accumulating at death. For residency models A and B, the person-years were adjusted by whole years; for model C (which incorporated migration rates), partial years were also used. (Residency models A, B, and C are described in detail in Section 11.)

The expected number of cancers was calculated according to Hill. The person-years matrix for each residency model was multiplied by the year-, sex-, and age-specific rate using Minnesota state-wide cancer incidence rates from the Minnesota Cancer Surveillance System. Minnesota rates were used rather than national or SEER rates because of known differences in rates for several major cancers in Minnesota.

The observed number of newly-diagnosed cancers was obtained through record-linkage with the Minnesota Cancer Surveillance System, which became fully implemented in 1988. Cancer

incidence codes (ICD-O) were converted to corresponding ICD 9 codes to allow comparisons of incidence and mortality.

The SIR was calculated by dividing the observed number of incident cancers by the expected number of incident cancers. The 95% confidence intervals were calculated based on a Poisson distribution (Bailar and Ederer, 1964; Breslow and Day, 1987).

4. Description of Record-Linkage Methods

A number of different sources of information were used to determine the residency status of the members of the two cohorts, including voter registration records, drivers license records, credit bureau records, cancer registry records, Minnesota death certificate records, National Death Index records, and Minnesota Revenue records. (These information sources are described in detail in Section 5.) Record-linkage processes were required to determine whether an individual in the Highway Worker or Mineral Board Worker cohort was potentially the same individual listed within an information source. For example, was John J. Johnson in the Highway Worker cohort the same John J. Johnson listed in the voter registration records (i.e., a true match)? Because of the large numbers of potential "matches" to be evaluated for each information source, this process needed to be computerized.

Two types of computerized record-linkage methods utilized in this study. Probabilistic record-linkage was used for each information source evaluated at the Minnesota Department of Health (MDH). A variation of an exact match requirement record-linkage was used for each information source that was evaluated outside of the MDH. These two methods are described in detail below.

Probabilistic Record-Linkage: The Minnesota Department of Health has developed its own set of record-linkage FORTRAN programs. This is a probabilistic system used to identify which cases from an external study file match with Minnesota Cancer Surveillance System (MCSS) cancer cases (Punyko, Bushhouse et al., 1995). This system, developed by Alan Bender, D.V.M., Ph.D., is based on Bayesian analysis and decision theory (Fellegi and Sunter, 1969) and is modeled with corrections after the approach developed in the *SEER Data Management System II Linkage System Statistical Users Guide*, 5/13/82 (Nelson, 1982).

The MCSS record-linkage program is powerful and fast, capable of comparing more than 1 billion records per day. The efficiency of the program is measured in terms of sensitivity and specificity. Sensitivity is the proportion of cases who match that were correctly identified as matching by the program, while specificity is the proportion of cases who do not match that were correctly identified as non-matches by the program. The performance of the MCSS record-linkage system has been measured for high quality data sets in which all of the key identifiers have valid data. Based on several hundred million comparisons, sensitivity was found to be 98.2 percent and specificity was greater than 99.9 percent (Punyko, Bushhouse et al., 1995).

The MCSS record-linkage program performs pairwise comparisons between every record in an external or study file and every record in a specially compiled set of MCSS data. The comparisons are performed separately for each sex and a selected set of personal identifiers from the two data sets

are compared: last name, first name, middle initial, social security number, sex, and date of birth. For each pair of records, the program calculates a total weight that reflects the overall agreement in these key personal identifiers. Using this total weight and a set of decision rules, the study cases are grouped into one of three categories: (1) cases that match, (2) cases that do not match, and (3) cases that need adjudication. This latter set of cases are said to fall into the gray zone. Decisions about the gray zone cases are based on a manual review of all available information.

Upper and lower thresholds establish the decision rules for determining which group a case will fall into. The choice of particular threshold values is based on a preliminary run of the program, in which the sensitivity and specificity of the linkage process are evaluated using a subset of the study data. There are three decision rules for categorizing the study cases. (1) If the total weight exceeds the upper threshold, the cases are considered a match. (2) If the total weight is less than the lower threshold, the cases are considered a non-match. (3) If the total weight is greater than or equal to the lower threshold and less than or equal to the upper threshold, the cases fall in gray zone and a review of case information is needed. MCSS algorithms minimize the size of the gray zone to its theoretical limit, often less than 5 percent of the records. The decision of a match also depends on weights for individual personal identifiers and on calibrating the thresholds. These aspects of the record-linkage system are discussed below.

The total weight measures the likelihood that the two records being compared represent the same person. It is calculated as the sum of seven individual weights, one for each of the key personal identifiers used in linkage, with date of birth separated into the components of month, day, and year. An individual weight represents the likelihood of a match based on a comparison of a single pair of identifiers. For example, the weight for last name indicates the likelihood of a match based only on a comparison of the last names. Individual weights are calculated as the logarithm of the ratio of two conditional probabilities: the probability of the observed pattern of agreement between the two identifiers compared, given that the two records represent the same person, divided by the probability of the observed pattern of agreement between the two identifiers compared, given that the two records represent different people. This ratio has the structure of the logarithm of an odds ratio.[1]

In record-linkage, three rules are used to indicate the direction and strength of the evidence for a match. (1) Positive individual weights suggest the case is a match. The larger the positive value, the more likely it is that the two records represent the same person. (2) Negative individual weights suggest the case is a non-match. The smaller the negative value, the more likely it is that the two records represent different persons. (3) Zero-valued individual weights are assigned when an identifier has a missing value (i.e., when there is no information). Although the algorithms perform optimally when all data fields are available, they are not all required. The primary impact of an imperfect data set is the increased size of the gray zone. Details about calculating and assigning individual weights and calibrating threshold values are documented in a 1995 MCSS Technical Report (Punyko, Bushhouse et al., 1995).

---

[1] log (odds ratio) = log(P[observed agreement pattern | same person]/P[observed agreement pattern | not the same person]).

Exact Match Record-Linkage: The record-linkages with credit bureau, National Death Index and Minnesota Revenue records were conducted using a variation of an exact match. Cohort records were submitted to the sources in the format requested. Each information source used different criteria for determining possible matches to the cohort records. In the case of Experian, the credit bureau evaluation, the matching criteria used was not available. The National Death Index program used a series of 12 algorithms; a match on any one of the algorithms was treated as a possible match and the death certificate information was provided. The Minnesota Revenue records utilized an exact match based only on social security number.

5. Description of Information Sources

A variety of information sources were used to gather data on the state of residency, date of death, and cancer diagnosis including: voter registration records, drivers license records, credit bureau records, cancer registry records, Minnesota death certificate records, National Death Index records, and Minnesota Revenue records. Data was available on an individual level for all sources except for the Minnesota Revenue records. Characteristics of the different sources, plus details about performing record-linkages with the sources is presented in Table 2. Table 3 lists data items (e.g. last name, date of birth) provided by each information source.

Conducting standardized mortality ratio analyses required information to identify all causes of death in the cohort. Cause of death information was obtained from Minnesota death certificates and National Death Index records.

Conducting standardized incidence ratio analyses required information to: (1) identify all cancer diagnoses for individuals in the cohort; and (2) identify residency status (i.e., whether each person lived in Minnesota or in another state) for each year from 1988 through 1996. Cancer diagnosis information was obtained entirely from Minnesota Cancer Surveillance System records. Individual residency status information was obtained from voter registration records, drivers license records, credit bureau records, cancer registry records, Minnesota death certificate records, and National Death Index records. Only summary counts by year (i.e., 1988 through 1996) of Minnesota residency were available from Minnesota Revenue records.

Although residency status information was available from all of the sources, they differed in the amount of information they provided.

- Voter registration records (i.e., voted or did not vote) were available on a year by year basis for 1993 through 1996. Because these records also included the original registration date, they often provided information on residency prior to 1993.
- Drivers license records were limited to one year of residency validation for each individual matched in the record-linkage process. The reason for this was that the drivers license database was continually overwritten; that is, each time a person's drivers license was renewed, the new date overwrote the previous renewal date.
- Credit bureau records from Experian provided start and end dates of residence for each address that was listed for an individual. This information was not always complete,

however, and occasionally different addresses were reported for the same year for the same person.

- National Change of Address information was obtained for all people who "matched" within the credit bureau records. This information was relatively uninformative, however, because there were very few people for whom there was a date associated with an address change. Only 111 individuals in the Mineral Board Worker cohort and 61 individuals in the Highway Worker cohort actually had a date associated with an address change. In addition, the move date had to be 1996 or later, therefore, the residency information was only useful for 1996.
- The cancer registry records were used to validate Minnesota residency for each year that an individual had an incident diagnosis of cancer. A total of 33 workers in both cohorts were diagnosed with multiple primary incident cancers.
- Death certificate and National Death Index records were used to validate the state of residency at the time of death.

Probabilistic record-linkage was conducted at the Minnesota Department of Health for voter registration, drivers license, cancer registry, and Minnesota death certificate records. Probabilistic record-linkage methods potentially identified a larger number of potential matches than record-linkage methods that relied strictly on an exact match (i.e., social security number) or a series of algorithms. Exact match record-linkage was conducted outside of MDH for the National Death Index, credit bureau (Experian), and Minnesota Revenue records.

It had been anticipated that Health Care Financing Administration (HCFA) records could be used as a source of residency information. However, use of HCFA records was not possible due to conflicts between federal law and Minnesota statutes regarding data retention schedules.

Table 2. Information Sources: Characteristics and Record-Linkage Information

Information Source	Years Searched	Scope of Information	Fields Required & (Desired) for Record-Linkage	Type of Linkage Conducted	Site Initial Record-Linkage Performed	Information Used for:
Voter Registration Records	Registration Date & 1993-1996	Minnesota	Name Birth Year (Address)	Probabilistic	MDH*	State of Residency
Drivers License Records	1988-1996	Minnesota	Name Birth Date Gender (Address)	Probabilistic	MDH	State of Residency
Credit Bureau Records**	1988-1996	National	Name SSN***	Unspecified	Experian	Date of Death State of Residency
Cancer Surveillance Records	1988-1996	Minnesota	Name Birth Date Gender SSN	Probabilistic	MDH	Date of Diagnosis Cancer Diagnosis State of Residency
Minnesota Death Certificate Records	1988-1996	Minnesota	Name Birth Date Gender SSN	Probabilistic	MDH	Date of Death Cause of Death State of Residency
National Death Index Records	1988-1996	National	Name Birth Date Gender SSN	12 Matching Algorithms Utilized	National Death Index	Date of Death Cause of Death State of Residency
Minnesota Revenue Records****	1988-1996	Minnesota	SSN	Exact Match	Department of Revenue	State of Residency

*MDH = Minnesota Department of Health

**Credit Bureau Records included Address Information, National Change of Address Information, and Date of Death Information

***SSN = Social Security Number

****Revenue Records were only used to obtain total residency counts by year; information was not available for any individual.

Table 3. Information Sources: Data Available

Information Sources	First Name	Middle Name	Middle Initial	Last Name	Date of Birth	Date of Death	Gender	Social Security Number	Address
Voter Registration Records	X	X		X	Birth Year*				X
Drivers License Records	X	X		X	X		X		X
Credit Bureau Records	X		X	X	X	X	X	X	X
Cancer Surveillance Records	X	X		X	X		X	X	
Minnesota Death Certificate Records	X	X		X	X	X	X	X	X
National Death Index Records	X		X	X	X	X	X	X	State**

*Voter Registration Records have "Birth Year" listed rather than a birth date.

**National Death Index Records listed the state of death.

6. Evaluation of Potential Record-Linkage Matches

Although the initial program for determining potential matches varied by the information source (i.e., probabilistic or exact matches), both methods provided a list of potential matches for each of the information sources. These lists varied in length and quality of matches, ranging from those that contained almost entirely true matches (e.g., Experian) to those that provided thousands of potential matches (e.g., drivers license records).

The number of potential matches that needed to be evaluated between the Highway Worker cohort and the Mineral Worker cohort was very different due to the quality of the original information available for the cohorts (Table 1). The Highway Worker cohort had virtually complete data on all cohort members, while the Mineral Board cohort had a large percentage of missing or invalid information, especially for the date of birth (40%). In addition, some females included in the Mineral Board cohort had multiple last names that required evaluating potential matches for each possible last name.

The number of potential matches ranged from a single record to several hundred records for an individual in the cohort. Very large numbers of potential matches per cohort record frequently occurred when a cohort member had a common surname and/or a missing date or birth or social security number.

The first step in evaluating potential matches from the record-linkage output involved ruling out incorrect matches. A computerized program was used to evaluate whether contradictory information was present between the cohort and the information source (i.e., two different dates of birth listed for a person with the same name). All database management/programming at the Minnesota Department of Health was conducted using Microsoft FoxPro 2.6. Potential matches with clearly contradictory information (i.e., a year of birth that was $\pm$ 2 years or more) were classified as non-matches. The program also checked for information inconsistencies such as name reversals. This computerized program was quite conservative in determining a non-match (i.e., very few true matches would have been evaluated as non-matches at this step).

The second step in evaluating potential matches involved using all of the information sources together as a group to adjudicate unresolved matches. A data screen was created to display all of the linkage results from each of the information sources; this also included multiple screens to facilitate looking at multiple possible matches from an individual information source. Names, dates of birth, social security numbers, and addresses were manually adjudicated to determine which specific records actually represented the same person. For example, there may have been ten possible matches to individuals within the vehicle license records. Using the screen that displayed all of the available information on the person it could be determined that the John J. Johnson who lived in St. Cloud, Minnesota was the true match from the vehicle license records. Frequently, a verified match within one information source assisted in validating the match for another information source. Specific written criteria were utilized in making all decisions about matches or non-matches. Figure 1 depicts the process used to evaluate the potential record-linkages.

Although most possible matches were resolvable as matches or non-matches, there were a few individuals for whom there was no conflicting information, but also not enough information to determine whether or not it was the same individual. This situation occurred primarily with females. For example, the first and middle names, and date of birth from the cohort matched a drivers license record, but the last name was different. It was impossible to decide in this instance whether it was a true match or not. Based on data from the probabilistic record-linkage program, the weights could be theoretically utilized to determine a match. Because of these records, a sensitivity analysis was conducted to estimate the difference in the total number of residency years that would occur if these records were considered matches or non-matches. As a result of this sensitivity analysis, a decision was made to count all of these potential matches as non-matches; there were not enough of these matches to significantly impact the residency history for the cohort.

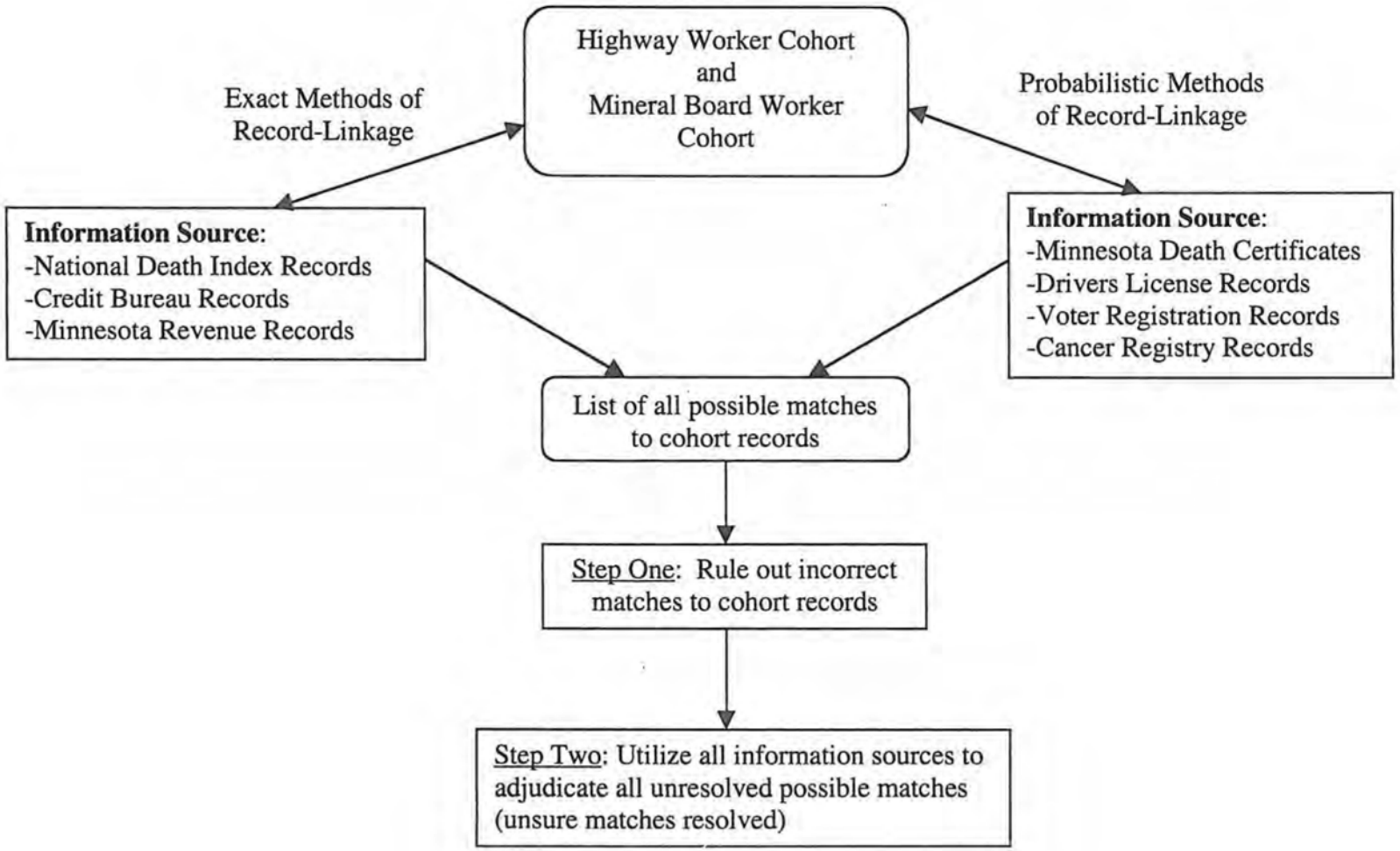
Work Time and Cost Estimates for Completing Record-Linkage Activities

A large amount of staff time was utilized to: (1) obtain access to all of the information sources (i.e., contracts and purchase agreements); (2) prepare the required information sources for record-linkage activities; (3) conduct the record-linkage activities either at the Minnesota Department of Health or elsewhere; (4) evaluate the initial record-linkages for each information source separately; and (5) evaluate the final matches and non-matches using all of the available information sources. In essence, it took considerably more staff time to create the needed residency history for the standardized incidence ratio analyses than would have been required for the standardized mortality ratio analyses.

The ease of access and direct cost associated with each information source would vary from state to state and also, potentially, for the type of institution (i.e., university versus state agency) conducting a standardized incidence ratio and/or standardized mortality ratio study. For instance, we were unable to utilize Health Care Financing Administration records as an information source due to conflicting state law and federal statutes. The largest direct costs were associated with the National Death Index search as well as with the credit bureau records. Both of these information sources conducted their own initial record-linkage evaluations, however, so this did decrease the amount of staff time in doing that task. The National Death Index results were easily read into data management software. In contrast, the credit bureau results were received in two different formats and took a considerable number of hours to write and run special computer programs to combine and place them into a database for evaluating residence history.

The direct costs associated with drivers license records, voter registration records and Minnesota death certificate records were minimal, but there was often a large amount of work required to create databases suitable for linkage from these records. For example, name fields were often combined rather than separated into first, middle, and last names. The drivers license database also was extremely large and required additional time to do queries and routine database management. The Minnesota Cancer Surveillance System (MCSS) had by far the best quality data available, but there were strict regulations about how this record-linkage was completed. This work was completed by MCSS staff.

Figure 1. Record-Linkage Evaluation Process



By far, the greatest of amount of computer and staff time spent in developing the residency histories involved determining which of the many possible matches provided by an information source for a particular cohort member was most likely to be the true match. This was referred to as the "gray" zone in the discussion of probabilistic record-linkage. The more missing data there was in the cohort, the more possible matches an information source provided. There were also more matches to be evaluated when the information source was missing data on date of birth or social security number.

Although the same record-linkages were completed with the same information sources for the Highway Worker and Mineral Board Worker cohorts, the amount of work resulting from these activities was quite different. When probabilistic record-linkage activities were completed with voter registration records, the Highway Worker cohort averaged 4.1 possible matches per cohort member as compared to 6.2 and 8.3 possible matches for male and female mineral board workers, respectively. This increase in the number of possible matches was due primarily to missing or incomplete data within the cohort (e.g., only 60 percent of mineral board workers had a known date of birth).

Although the number of hours it took to evaluate each possible match varied slightly between information sources, the total number of possible matches to be evaluated from the initial probabilistic record-linkage activities was indicative of the resulting staff time needed to complete the computerized and manual adjudication. For the Highway Worker cohort (N=3,497) there were 14,461 and 5,290 possible matches to be evaluated for voter registration records and drivers license records, respectively. For the male Mineral Board Worker cohort (N=4,350) there were 27,081 and 21,403 possible matches to be evaluated for voter registration records and drivers license records, respectively. For the female Mineral Board Worker cohort (N=736) there were 6,120 and 5,305 possible matches to be evaluated for voter registration records and drivers license records, respectively. For females, the average number of possible matches per cohort member was much larger due to the inclusion of possible alternate last names (i.e., maiden name versus married name); a different last name did not necessarily indicate that it was or was not a match.

Time estimates were computed for equivalent record-linkage activities across a number of computing platforms. This test was conducted to determine the effect that different processors, operating systems, speed, and memory had on the number of record-linkage comparisons that could be completed each day. As shown in Table 4, the SUN Ultra SparcII (300MHz, 252 MB) conducted the greatest number of record-linkage comparisons per day of activity at 16.72 billion. Of the platforms tested, the SUN Micro SparcII (70MHz, 32MB) completed the lowest number of comparisons per day at 2.74 billion. Use of a "slower" computer could potentially add significantly to the time required to conduct a series of comprehensive record-linkage activities.

Table 4. Record-Linkage Comparisons by Platform

Processor	Operating System	Speed	Memory	Billions of Record-Linkage Comparisons per Day
SUN Sparc61	UNIX (SOLARIS)	100 MHz	128 MB	4.18
SUN Micro SparcII	UNIX (SOLARIS)	70 MHz	32 MB	2.74
SUN Ultra Sparc	UNIX (SOLARIS)	200 MHz	892 MB	11.52
SUN Ultra SparcII	UNIX (SOLARIS)	300 MHz	252 MB	16.72
Pentium III	DOS	600 MHz	256 MB	2.75

7. Determining the Residency Status for 1988 to 1996

The reason for conducting the record-linkages and making decisions about which record-linkages were matches or non-matches was to create a residency history for each worker in the cohorts. This was important for the standardized incidence ratio analyses because a worker was only eligible for being diagnosed with cancer in the Minnesota Cancer Surveillance System if they were a resident of Minnesota. In essence, only the years when they resided in Minnesota could be counted as years when they were at "risk" of being diagnosed with cancer in Minnesota.

The first step in creating this residency history was to create a database that chronicled all "known" residency years from each of the information sources. Initially, each source was evaluated separately before having the information combined. In the example below, worker number one was found to have lived in Minnesota during 1988, 1989, and 1994 and lived in another (other) state in 1991 and 1996. Worker number two lived in Minnesota in 1989 and 1992 and died in 1992. Because he could no longer be at "risk" for developing cancer at all from 1993-1996 (i.e., he was dead for the entire time period) he was recorded as dead for those years.

Table 5. Example: Creation of Residency Status for 1988 to 1996 in Database Format

Example	1988	1989	1990	1991	1992	1993	1994	1995	1996
Worker #1	MN	MN		O			MN		O
Worker #2		MN			MN	Dead	Dead	Dead	Dead
Worker #3	MN		O			O			

MN = Minnesota resident
O = Out of State (Non-Minnesota) resident
Dead = Deceased for the entire year per Minnesota death records and/or National Death Index—contributes no person-years of Minnesota residency

Tables 6 to 9 show the number of true matches identified for the Highway Worker cohort (males) and Mineral Board Worker cohort (males and females) for each of the information sources separately. Employment records were only available for the Highway Worker cohort for 1988

and 1989. However, this source contributed significantly to Minnesota residency information because almost half of the cohort was still employed during that time period. Overall, for each of the individual sources, more people were found from the Highway Worker cohort than for the Mineral Board Worker cohort. This was primarily due to the difference in cohort information on name, date of birth, and social security number (i.e., the Mineral Board Worker cohort contained more missing information). Also, the inclusion of females in the Mineral Board Worker cohort made it more difficult to identify matches or non-matches based on possible or known last name changes. Males always had a greater number of matches to information sources than females for the Mineral Board Worker cohort.

Although matches were found for voter registration records from 1989 to 1996, there was a large variation in the number of matches that were made for each of the follow-up years. This variation was due to two factors: (1) all linkages prior to 1993 were based on information concerning the original date the person registered to vote; and (2) the number of voters increased dramatically during national elections, such as in 1996. Overall, about 50 percent of the highway workers and 35 percent of the mineral board workers were found in voter registration records.

Drivers license records did not contain social security numbers. Therefore, only name, date of birth, and address could be compared with cohort records to determine matches. Since 40 percent of the Mineral Board Worker cohort was missing date of birth information (compared to 0 percent for the Highway Worker cohort), fewer workers were identified in this source. Although a large percentage of individuals were found in drivers license records, it only contributed information one year of Minnesota residency (e.g., the year they renewed their drivers license) for each person that was matched. Drivers license records proved to be a good source of residency information for both cohorts, but a significantly higher percentage of highway workers were found (77%) than mineral board workers (47%).

Approximately 45 percent of both cohorts were found within the credit bureau records, but a higher percentage of Mineral Board Workers were identified as living outside of Minnesota (10%) as compared to the Highway Workers (3%). Although there were a few more matches found for 1992-1996 than for 1988-1991, this information source contributed a large number of years of residency status for each of the workers that were found. Most of the addresses identified within the credit bureau records had a start and end date associated with them. Because this information source provided a range of dates (e.g., 1988-1992 as a Minnesota address), it contributed a large share of the total residency status information.

The three remaining sources, cancer registry records, Minnesota death certificates, and National Death Index records, were critical information sources for identifying date and type of cancer diagnosis, and date and cause of death information, respectively. Although the total number of workers found in each source was small, they also contributed to information about residency status. Death information, especially, facilitated identifying each of the succeeding residency years as ineligible for being diagnosed with cancer.

Table 10 depicts the total number of matches, i.e., the total number of confirmed years of Minnesota or non-Minnesota (other state) residency, for each of the information sources. The percent listed was derived by dividing the total number of matches by the total possible number

of matches for each cohort. Overall, credit bureau records confirmed 34 percent of the possible residency years for males in the Highway and Mineral Board Worker cohorts. This was by far the greatest percentage of confirmed residency years when compared to the other information sources. Overall, 63 percent of the total possible matches were confirmed for highway workers as compared to 45 and 20 percent of male and female mineral board workers, respectively.

The number and percent of workers identified by different combinations of the information sources is shown in Table 11 and in Figure 2. Employment information for the Highway Worker cohort was ignored as an information source for the purposes of this table.

For the Highway Worker cohort, the combinations that, by far, resulted in the highest percentage of identified workers were that of voter registration/drivers license (24%), and voter registration/drivers license/credit bureau (22%). These two combinations of sources also identified the greatest percentage of female mineral board workers at 7 and 9 percent, respectively. For male mineral board workers, the voter registration/drivers license/credit bureau combination resulted in the greatest percentage of matches (18%), followed by credit bureau records (12%) and drivers license/voter registration (11%).

Table 6. Highway Worker Cohort (Males): Residency Status by Information Source and Year of Follow-Up

	Employee Records		Voter Registration Records		Drivers License Records		Credit Bureau Records		Cancer Surveillance Records		Minnesota Death Certificates		National Death Index		Number of People Found	
Year of Follow-Up	*MN*[*]	*O*[**]	*MN*	*O*	*MN*	*O*	*MN*	*O*	*MN*	*O*	*MN*	*O*	*MN*	*O*	*MN*	*O*
1988	1712 (49)[#]	-	-	-	100 (3)	-	466 (13)	28 (1)	30 (1)	-	67 (2)	-	49 (1)	8 (<1)	2058 (59)	145 (4)
1989	1639 (47)	-	109 (3)	-	58 (2)	-	1023 (29)	50 (1)	25 (1)	-	60 (2)	-	49 (1)	10 (<1)	2197 (63)	291 (8)
1990	-	-	109 (3)	-	43 (1)	-	1144 (33)	59 (2)	29 (1)	-	59 (2)	-	5 (<1)	4 (<1)	1300 (37)	335 (10)
1991	-	-	32 (1)	-	48 (1)	-	1185 (34)	70 (2)	35 (1)	-	52 (1)	-	4 (<1)	4 (<1)	1279 (37)	348 (10)
1992	-	-	116 (3)	-	114 (3)	-	1250 (36)	76 (2)	38 (1)	-	58 (2)	-	8 (<1)	4 (<1)	1425 (41)	399 (11)
1993	-	-	649 (19)	-	230 (7)	-	1273 (36)	77 (2)	38 (1)	-	56 (2)	-	8 (<1)	11 (<1)	1785 (51)	405 (12)
1994	-	-	1759 (50)	-	438 (13)	-	1263 (36)	85 (2)	39 (1)	-	51 (1)	-	6 (<1)	7 (<1)	2354 (67)	418 (12)
1995	-	-	670 (19)	-	425 (12)	-	1290 (37)	85 (2)	46 (1)	-	47 (2)	-	6 (<1)	3 (<1)	1843 (53)	420 (12)
1996	-	-	1899 (54)	-	1226 (35)	-	1285 (37)	84 (2)	28 (1)	-	57 (2)	-	5 (<1)	5 (<1)	2517 (72)	425 (12)
Number of People Found	1716 (49)	-	2001 (57)	-	2682 (77)	-	1467 (42)	114 (3)	297 (8)	-	507 (14)	-	140 (4)	56 (2)		

[*] MN = Denotes Minnesota state residency

[**] O = Denotes a Non-Minnesota (other) state residency

[#] Percent = Number of People Found / Number of Cohort members (N=3497) rounded to the nearest percent

Table 7. Mineral Board Worker Cohort (Males and Females): Residency Status by Information Source and Year of Follow-Up

	Employee Records		Voter Registration Records		Drivers License Records		Credit Bureau Records		Cancer Surveillance Records		Minnesota Death Certificates		National Death Index		Number of People Found	
Year of Follow-Up	*MN*[*]	*O*[**]	*MN*	*O*	*MN*	*O*	*MN*	*O*	*MN*	*O*	*MN*	*O*	*MN*	*O*	*MN*	*O*
1988	-	-	180	-	55	-	405	139	15	-	39	-	0	10	648	145
	-		(4)[#]		(1)		(8)	(3)	(<1)		(1)		(0)	(<1)	(13)	(3)
1989	-	-	27	-	36	-	1010	285	24	-	28	-	0	7	1091	291
	-		(1)		(1)		(20)	(6)	(<1)		(1)		(0)	(<1)	(21)	(6)
1990	-	-	104	-	50	-	1195	335	26	-	38	-	1	5	1338	335
			(2)		(1)		(23)	(7)	(1)		(1)		(<1)	(<1)	(26)	(7)
1991	-	-	19	-	50	-	1255	347	24	-	43	-	1	5	1340	348
			(<1)		(1)		(25)	(7)	(<1)		(1)		(<1)	(<1)	(26)	(7)
1992	-	-	143	-	67	-	1400	391	26	-	32	-	0	13	1537	399
			(3)		(1)		(28)	(8)	(1)		(1)		(0)	(<1)	(30)	(8)
1993	-	-	515	-	264	-	1418	401	34	-	42	-	1	9	1796	405
			(10)		(5)		(28)	(8)	(1)		(1)		(<1)	(<1)	(35)	(8)
1994	-	-	1468	-	456	-	1418	417	27	-	33	-	0	11	2197	418
			(29)		(9)		(28)	(8)	(1)		(1)		(0)	(<1)	(43)	(8)
1995	-	-	472	-	571	-	1445	424	31	-	42	-	1	8	1871	420
			(9)		(11)		(28)	(8)	(1)		(1)		(<1)	(<1)	(37)	(8)
1996	-	-	1702	-	833	-	1464	431	37	-	53	-	2	6	2356	425
			(33)		(16)		(29)	(8)	(1)		(1)		(<1)	(<1)	(46)	(8)
Number of People Found	-	-	1762	-	2382	-	1696	528	236	-	350	-	6	74		
			(35)		(47)		(33)	(10)	(5)		(7)		(<1)	(1)		

[*] MN = Denotes Minnesota state residency [**] O = Denotes a Non-Minnesota (other) state residency

[#] Percent = Number of People Found / Number of Cohort members (N=5086) rounded to the nearest percent

Table 8. Mineral Board Worker Cohort (Males): Residency Status by Information Source and Year of Follow-Up

	Employee Records		Voter Registration Records		Drivers License Records		Credit Bureau Records		Cancer Surveillance Records		Minnesota Death Certificates		National Death Index		Number of People Found	
Year of Follow-Up	*MN*[*]	*O*[**]	*MN*	*O*	*MN*	*O*	*MN*	*O*	*MN*	*O*	*MN*	*O*	*MN*	*O*	*MN*	*O*
1988	-	-	168	-	54	-	375	129	14	-	34	-	-	9	601	134
	-		(4)[#]		(1)		(9)	(3)	(<1)		(1)		-	(<1)	(14)	(3)
1989	-	-	27	-	35	-	935	270	19	-	27	-	-	5	1010	275
	-		(1)		(1)		(21)	(6)	(<1)		(1)		-	(<1)	(23)	(6)
1990	-	-	98	-	43	-	1108	321	21	-	37	-	1	5	1236	322
			(2)		(1)		(25)	(7)	(<1)		(1)		(<1)	(<1)	(28)	(7)
1991	-	-	18	-	47	-	1169	333	22	-	40	-	1	5	1247	335
			(<1)		(1)		(27)	(8)	(<1)		(1)		(<1)	(<1)	(29)	(8)
1992	-	-	132	-	64	-	1309	376	22	-	29	-	-	13	1433	385
			(3)		(1)		(30)	(9)	(<1)		(1)		-	(<1)	(33)	(9)
1993	-	-	478	-	238	-	1323	387	31	-	39	-	1	9	1664	391
			(11)		(5)		(30)	(9)	(1)		(1)		(<1)	(<1)	(38)	(9)
1994	-	-	1347	-	417	-	1321	402	25	-	29	-	-	11	2015	403
			(31)		(10)		(30)	(9)	(1)		(1)		-	(<1)	(46)	(9)
1995	-	-	435	-	519	-	1349	408	28	-	38	-	1	8	1726	404
			(10)		(12)		(31)	(9)	(1)		(1)		(<1)	(<1)	(40)	(9)
1996	-	-	1558	-	756	-	1364	414	34	-	51	-	2	6	2156	408
			(36)		(17)		(31)	(10)	(1)		(1)		(<1)	(<1)	(50)	(9)
Number of People Found	-	-	1615	-	2173	-	1582	507	209	-	324	-	6	71		
			(37)		(50)		(36)	(12)	(5)		(7)		(<1)	(2)		

[*] MN = Denotes Minnesota state residency

[**] O = Denotes a Non-Minnesota (other) state residency

[#] Percent = Number of People Found / Number of Cohort members (N=4350) rounded to the nearest percent

Table 9. Mineral Board Worker Cohort (Females): Residency Status by Information Source and Year of Follow-Up

	Employee Records		Voter Registration Records		Drivers License Records		Credit Bureau Records		Cancer Surveillance Records		Minnesota Death Certificates		National Death Index		Number of People Found	
Year of Follow-Up	*MN*[*]	*O*[**]	*MN*	*O*	*MN*	*O*	*MN*	*O*	*MN*	*O*	*MN*	*O*	*MN*	*O*	*MN*	*O*
1988	-	-	12 (2)[#]	-	1 (<1)	-	30 (4)	10 (1)	1 (<1)	-	5 (1)	-	-	1 (<1)	47 (6)	11 (1)
1989	-	-	-	-	1 (<1)	-	75 (10)	15 (2)	5 (1)	-	1 (<1)	-	-	2 (<1)	81 (11)	16 (2)
1990	-	-	6 (1)	-	7 (1)	-	87 (12)	14 (2)	5 (1)	-	1 (<1)	-	-	-	102 (14)	13 (2)
1991	-	-	1 (<1)	-	3 (<1)	-	86 (12)	14 (2)	2 (<1)	-	3 (<1)	-	-	-	93 (13)	13 (2)
1992	-	-	11 (1)	-	3 (<1)	-	91 (12)	15 (2)	4 (<1)	-	3 (<1)	-	-	-	104 (14)	14 (2)
1993	-	-	37 (5)	-	26 (4)	-	95 (13)	14 (2)	3 (<1)	-	3 (<1)	-	-	-	132 (18)	14 (2)
1994	-	-	121 (16)	-	39 (5)	-	97 (13)	15 (2)	2 (<1)	-	4 (<1)	-	-	-	182 (25)	15 (2)
1995	-	-	37 (5)	-	52 (7)	-	96 (13)	16 (2)	3 (<1)	-	4 (<1)	-	-	-	145 (20)	16 (2)
1996	-	-	144 (20)	-	77 (10)	-	100 (14)	17 (2)	3 (<1)	-	2 (<1)	-	-	-	200 (27)	17 (2)
Number of People Found	-	-	147 (20)	-	209 (28)	-	114 (15)	21 (4)	27 (4)	-	26 (4)	-	-	3 (<1)		

[*] MN = Denotes Minnesota state residency

[**] O = Denotes a Non-Minnesota (other) state residency

[#] Percent = Number of People Found / Number of Cohort members (N=736) rounded to the nearest percent

Table 10. Mineral Board Worker Cohort (Males and Females) and Highway Worker Cohort (Males): Total Residency Status for 1988-1996 by Information Source

	Voter Registration Records	Drivers License Records	Credit Bureau Records	Cancer Surveillance Records	Death Certificates	All Sources Combined
Highway Worker Cohort (Males; N=3497)						
Total Number of Links	5,343 (17)[#]	2,682 (9)	10,793 (34)	308 (1)	2,859 (9)	19,671 (63)
Mineral Board Worker Cohort (Males; N=4350)						
Total Number of Links	4,261 (11)	2,173 (6)	13,293 (34)	216 (1)	1,876 (5)	17,611 (45)
Mineral Board Worker Cohort (Females; N=736)						
Total Number of Links	369 (6)	209 (3)	887 (13)	28 (<1)	515 (8)	1338 (20)

[#]Percent = Number of Matches / Total Possible Number of Matches (N=31,473 for male Highway Workers; N=39,150 for male Mineral Board Workers; N=6,624 for female Mineral Board Workers) rounded to the nearest percent. The Total Possible Number of Matches = 9 years x total number of cohort members.

Table 11. Mineral Board Worker Cohort (Males & Females) and Highway Worker Cohort (Males): Residency Status by Unique Linkage Source Combinations

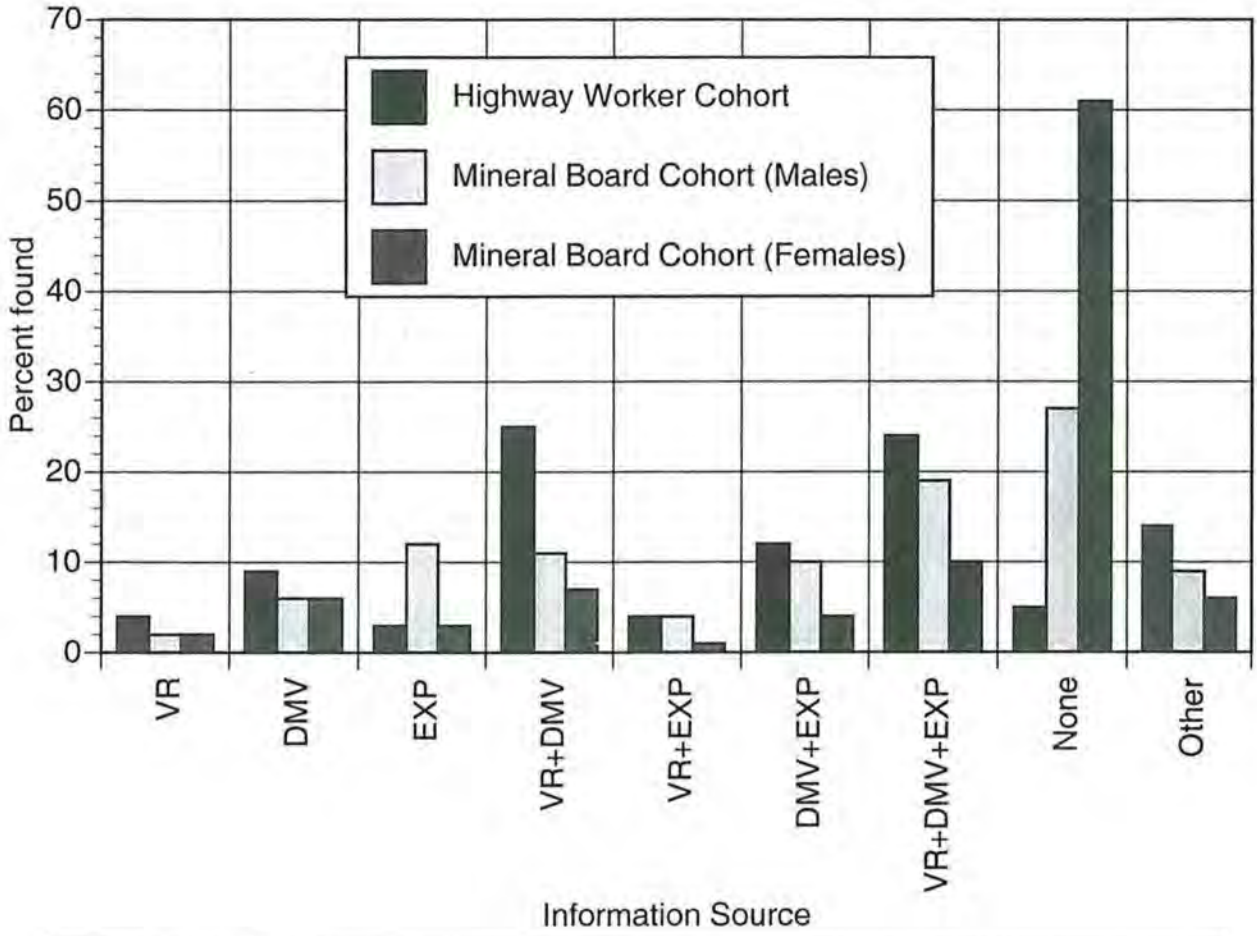
Information Source					Total Number of Information Combinations					
Voter Registration Records	Drivers License Records	Credit Bureau Records	Cancer Surveillance Records	Death Certificates	Mineral Board Worker Males		Mineral Board Worker Females		Highway Worker Males	
					Number	*Percent*	*Number*	*Percent*	*Number*	*Percent*
	x				254	6[#]	45	6	311	9
	x	x			365	8	24	3	291	8
	x	x		x	47	1	3	<1	59	2
	x		x		21	<1	6	1	32	1
	x	x	x		19	<1	1	<1	21	1
	x	x	x	x	23	1	-	-	51	1
	x		x	x	27	1	2	<1	63	2
	x			x	102	2	3	<1	149	4
		x			501	12	25	3	109	3
		x		x	20	<1	1	<1	21	1
			x		14	<1	9	1	2	<1
		x	x		5	<1	-	-	8	<1
		x	x	x	-	-	-	-	1	<1

Table 11 (Continued). Mineral Board Worker Cohort (Males & Females) and Highway Worker Cohort (Males): Residency Status by Unique Linkage Source Combinations

Information Source					Total Number of Information Combinations					
Voter Registration Records	Drivers License Records	Credit Bureau Records	Cancer Surveillance Records	Death Certificates	Mineral Board Worker Males		Mineral Board Worker Females		Highway Worker Males	
					Number	*Percent*	*Number*	*Percent*	*Number*	*Percent*
			x	x	22	1	3	<1	12	<1
				x	150	3	17	2	198	6
x					108	2	15	2	138	4
x	x				460	11	52	7	834	24
x	x			x	1	<1	-	-	-	-
x	x	x			790	18	68	9	778	22
x	x	x		x	-	-	-	-	1	<1
x	x		x		29	1	3	<1	49	1
x	x	x	x		35	1	2	<1	43	1
x		x			178	<1	6	1	143	4
x			x		5	<1	1	<1	5	<1
x		x	x		9	<1	-	-	10	<1
No Sources					1165	27	450	61	168	5

[#] Percent = Number of People Found / Number of Cohort members (N=3497 for male Highway Workers; N=4350 for male Mineral Board Workers; N=736 for female Mineral Board Workers) rounded to the nearest percent

Figure 2. Mineral Board Worker Cohort (Males & Females) and Highway Worker Cohort (Males): Percentage of Workers Identified by Unique Linkage Source Combinations



Information Sources:
DL=Drivers license records; VR=Voter registration records; CB=Credit bureau records
None=No source
Other=Cancer registry records and/or death records only

8. Determining the Effect of Birth Cohort and Last Date of Employment on Being Found by Information Sources

Tables 12 to 14 display the number of people found by three different information sources (voter registration, drivers license, and credit bureau records) for each birth cohort. Little difference was seen for the birth cohorts between 1988-1992 because of the paucity of voter registration matches. Overall, equal percentages of people were found within the voter registration records, by birth cohort, in 1993 and 1995 as well as 1994 and 1996. Similar percentages of voting were found among highway workers with the exception of people born before 1920, who voted at a much lower percentage than other birth cohorts. Male mineral board workers had similar patterns of voting, but lower percentages occurred for workers born before 1930. Female mineral board workers had the highest percentage of voters among those born before 1920 and 1930-1939.

Fewer differences were found between birth cohorts among those found in drivers license records. Somewhat fewer highway workers were found among those born before 1920 (57%) than for those born in 1920 or later (78-82%). Among male mineral board workers, the percentage ranged from 60, 37, 60, 43, to 65 percent, respectively for those born $\leq$1919, 1920-1929, 1930-1939, 1940-1949, and $\geq$1950. There were no real differences in birth cohort seen among female mineral board workers; the numbers of women found in each birth cohort were very low compared to the men.

The percentage of male workers found by credit bureau records within each birth cohort was greater for younger workers. For instance, male highway workers ranged in the percentage found from 14 percent ($\leq$1919) to 50 percent ($\geq$1950). Among female mineral board workers the lowest percentages of workers found within credit bureau records were among the 1920-1929 and 1940-1949 birth cohorts.

Table 12. Highway Worker Cohort (Males): Residency Status by Birth Cohort, Information Source, and Year of Follow-Up

Information Source / Birth Cohort	1988	1989	1990	1991	1992	1993	1994	1995	1996
Voter Registration Records									
≤1919	-	7	5	2	5	47	113	50	125
(N=629)		(1)[#]	(1)	(<1)	(1)	(7)	(18)	(8)	(20)
1920-1929	-	21	10	5	13	146	331	144	365
(N=652)		(3)	(2)	(1)	(2)	(22)	(51)	(22)	(56)
1930-1939	-	21	18	6	27	141	379	142	402
(N=670)		(3)	(3)	(1)	(4)	(21)	(57)	(21)	(60)
1940-1949	-	37	33	12	30	187	533	200	571
(N=878)		(4)	(4)	(1)	(3)	(21)	(61)	(23)	(65)
≥1950	-	23	43	7	41	128	403	134	436
(N=668)		(3)	(6)	(1)	(6)	(19)	(60)	(20)	(65)
Drivers License Records									
≤1919	61	28	17	23	54	40	19	14	113
(N=629)	(10)	(4)	(3)	(4)	(9)	(6)	(3)	(2)	(18)
1920-1929	19	21	8	10	33	66	49	58	246
(N=652)	(3)	(3)	(1)	(2)	(5)	(10)	(8)	(9)	(38)
1930-1939	12	1	7	11	14	49	112	85	246
(N=670)	(2)	(<1)	(1)	(2)	(2)	(7)	(17)	(13)	(37)
1940-1949	4	3	9	1	10	48	125	139	377
(N=878)	(<1)	(<1)	(1)	(<1)	(1)	(5)	(14)	(16)	(43)
≥1950	4	5	2	3	3	27	133	129	244
(N=668)	(1)	(1)	(<1)	(<1)	(<1)	(4)	(20)	(19)	(37)

Table 12 (Continued). Highway Worker Cohort (Males): Residency Status by Birth Cohort, Information Source, and Year of Follow-Up

Information Source Birth Cohort	1988	1989	1990	1991	1992	1993	1994	1995	1996
Credit Bureau Records									
≤1919	55	103	107	107	106	109	103	104	94
(N=629)	(9)	(16)	(17)	(17)	(17)	(17)	(16)	(17)	(15)
1920-1929	95	190	209	203	220	227	228	225	223
(N=652)	(15)	(29)	(32)	(31)	(34)	(35)	(35)	(33)	(34)
1930-1939	117	253	280	284	298	297	295	300	299
(N=670)	(17)	(38)	(42)	(42)	(44)	(44)	(44)	(45)	(45)
1940-1949	134	314	354	377	398	401	400	412	410
(N=878)	(15)	(36)	(40)	(43)	(45)	(46)	(46)	(47)	(47)
≥1950	93	213	253	284	304	316	322	334	343
(N=668)	(14)	(32)	(38)	(43)	(46)	(47)	(48)	(50)	(51)

[#] Percent = Number of People Found / Total Number of People in each Birth Cohort rounded to the nearest percent

Table 13. Mineral Board Worker Cohort (Males): Residency Status by Birth Cohort, Information Source, and Year of Follow-Up

Information Source Birth Cohort	1988	1989	1990	1991	1992	1993	1994	1995	1996
Voter Registration Records									
≤1919	5	-	5	-	4	12	58	16	61
(N=273)	(2)[#]		(2)		(1)	(4)	(21)	(6)	(22)
1920-1929	15	2	7	-	10	48	143	42	157
(N=617)	(2)	(<1)	(1)		(2)	(8)	(23)	(7)	(25)
1930-1939	30	2	16	5	19	94	268	91	299
(N=673)	(4)	(<1)	(2)	(1)	(3)	(14)	(40)	(14)	(44)
1940-1949	70	12	40	7	53	209	557	200	643
(N=1982)	(3)	(1)	(2)	(<1)	(3)	(11)	(28)	(10)	(32)
≥1950	48	11	30	6	46	115	321	86	398
(N=805)	(6)	(1)	(4)	(1)	(6)	(14)	(40)	(11)	(49)
Drivers License Records									
≤1919	16	9	15	7	14	21	26	22	33
(N=273)	(6)	(3)	(5)	(3)	(5)	(8)	(10)	(8)	(12)
1920-1929	13	6	11	10	10	30	46	29	73
(N=617)	(2)	(1)	(2)	(2)	(2)	(5)	(7)	(5)	(12)
1930-1939	6	8	5	8	5	47	68	113	142
(N=673)	(1)	(1)	(1)	(1)	(1)	(7)	(10)	(17)	(21)
1940-1949	15	5	7	11	25	94	179	203	319
(N=1982)	(1)	(<1)	(<1)	(1)	(1)	(5)	(9)	(10)	(16)
≥1950	4	7	5	11	8	46	98	152	189
(N=805)	(<1)	(1)	(1)	(1)	(1)	(6)	(12)	(19)	(23)

Table 13 (Continued). Mineral Board Worker Cohort (Males): Residency Status by Birth Cohort, Information Source, and Year of Follow-Up

Information Source Birth Cohort	1988	1989	1990	1991	1992	1993	1994	1995	1996
Credit Bureau Records									
≤1919	17	45	54	51	53	49	48	50	45
(N=273)	(6)	(16)	(20)	(19)	(19)	(18)	(18)	(18)	(16)
1920-1929	45	106	118	116	125	129	127	123	123
(N=617)	(7)	(17)	(19)	(19)	(20)	(21)	(21)	(20)	(20)
1930-1939	93	220	265	283	311	313	306	310	319
(N=673)	(14)	(33)	(39)	(42)	(46)	(47)	(45)	(46)	(47)
1940-1949	244	584	679	713	801	814	836	855	865
(N=1982)	(12)	(29)	(34)	(36)	(40)	(41)	(42)	(43)	(44)
≥1950	105	250	313	339	395	405	406	419	426
(N=805)	(13)	(31)	(39)	(42)	(49)	(50)	(50)	(52)	(53)

[#] Percent = Number of People Found / Total Number of People in each Birth Cohort rounded to the nearest percent

Table 14. Mineral Board Worker Cohort (Females): State of Residency by Year of Birth, Linkage Source, and Year of Follow-Up

Information Source Birth Cohort	1988	1989	1990	1991	1992	1993	1994	1995	1996
Voter Registration Records									
≤1919	-	-	-	-	-	-	8	1	8
(N=24)							(33)[#]	(4)	(33)
1920-1929	-	-	-	-	1	2	10	6	11
(N=71)					(1)	(3)	(14)	(8)	(15)
1930-1939	3	-	2	-	3	10	32	8	39
(N=103)	(3)		(2)		(3)	(10)	(31)	(8)	(38)
1940-1949	4	-	3	-	3	17	52	15	64
(N=411)	(1)		(1)		(1)	(4)	(13)	(4)	(16)
≥1950	5	-	1	1	4	8	19	7	22
(N=127)	(4)		(1)	(1)	(3)	(6)	(15)	(6)	(17)
Drivers License Records									
≤1919	1	-	1	-	-	3	1	3	2
(N=24)	(4)		(4)			(13)	(4)	(13)	(8)
1920-1929	-	-	1	1	-	4	2	2	3
(N=71)			(1)	(1)		(6)	(3)	(3)	(4)
1930-1939	-	-	-	-	1	6	11	10	24
(N=103)					(1)	(6)	(11)	(10)	(23)
1940-1949	-	1	3	2	2	5	20	26	31
(N=411)		(<1)	(1)	(<1)	(<1)	(1)	(5)	(6)	(8)
≥1950	-	-	2	-	-	8	5	11	17
(N=127)			(2)			(6)	(4)	(9)	(13)

Table 14 (Continued). Mineral Board Worker Cohort (Females): State of Residency by Year of Birth, Linkage Source, and Year of Follow-Up

Information Source Birth Cohort	**1988**	**1989**	**1990**	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**
Credit Bureau Records									
≤1919	-	2	2	2	3	3	4	5	5
(N=24)		(8)	(8)	(8)	(13)	(13)	(17)	(21)	(21)
1920-1929	4	8	8	8	8	7	8	8	8
(N=71)	(6)	(11)	(11)	(11)	(11)	(10)	(11)	(11)	(11)
1930-1939	11	23	25	23	23	26	24	26	27
(N=103)	(11)	(22)	(24)	(22)	(22)	(25)	(23)	(25)	(26)
1940-1949	20	38	44	45	48	49	50	48	50
(N=411)	(5)	(9)	(11)	(11)	(12)	(12)	(12)	(12)	(12)
≥1950	5	19	22	22	24	24	26	25	27
(N=127)	(4)	(15)	(17)	(17)	(19)	(19)	(20)	(20)	(21)

[#] Percent = Number of People Found / Total Number of People in each Birth Cohort rounded to the nearest percent

Tables 15 to 17 show the number of people found by three different information sources (voter registration, drivers license, and credit bureau sources) for groups of workers ending their employment in succeeding decades (≤1959, 1960-1969, 1970-1979, and ≥1980). Employment information for the Highway Worker cohort was last updated in December 31, 1989, but members of the Mineral Board cohort had all retired, quit, or were laid off by 1978 due to the shutdown of the manufacturing plant.

There were few differences in the voter registration record matches for the different employment dates for the years 1988-1992 because of the paucity of matches. Overall, more voters were found among highway workers leaving employment in 1980 or later compared to the other years. For example, in 1996, 64 percent of voters ended their employment in 1980 or later as compared to those ending their employment earlier (30-38%). This was similar to the pattern found for female mineral board workers. The percentage of male mineral board workers found in the voter registration records increased as the last date of employment increased.

There were similar patterns among male highway workers, male mineral board workers, and female mineral board workers concerning the percentage of matches within the drivers license records. In general, the percentage of matches from drivers license records increased as the last date of employment increased.

Male and female workers among both the Highway Worker and Mineral Board Worker cohorts were found within credit bureau records. Workers leaving employment on or before had the fewest matches while those leaving in 1960-1969 and 1970-1979 had similar, but larger numbers of matches. The highest percentages of male highway workers found by credit bureau records were for those ending their employment in 1980 or later.

Table 15. Highway Worker Cohort (Males): Residency Status by Last Year of Employment, Information Source, and Year of Follow-Up

Information Source Last Year Employed	1988	1989	1990	1991	1992	1993	1994	1995	1996
Voter Registration Records									
≤1959	-	3	1	2	3	26	53	23	54
(N=183)		(2)	(1)	(1)	(2)	(14)	(29)	(13)	(30)
1960-1969	-	7	6	4	7	55	139	69	159
(N=419)		(2)	(1)	(1)	(2)	(13)	(33)	(16)	(38)
1970-1979	-	11	10	3	11	84	218	86	248
(N=659)		(2)	(2)	(<1)	(2)	(13)	(33)	(13)	(38)
≥1980	-	88	92	23	95	484	1349	492	1438
(N=2236)		(4)	(4)	(1)	(4)	(22)	(60)	(22)	(64)
Drivers License Records									
≤1959	6	8	5	6	9	12	9	15	22
(N=183)	(3)	(4)	(3)	(3)	(5)	(7)	(5)	(8)	(12)
1960-1969	20	4	11	13	12	40	34	45	77
(N=419)	(5)	(1)	(3)	(3)	(3)	(10)	(8)	(11)	(18)
1970-1979	47	21	8	8	39	38	53	70	162
(N=659)	(7)	(3)	(1)	(1)	(6)	(6)	(8)	(11)	(25)
≥1980	27	25	19	21	54	140	342	295	965
(N=2236)	(1)	(1)	(1)	(1)	(2)	(6)	(15)	(13)	(43)

Table 15 (Continued). Highway Worker Cohort (Males): Residency Status by Last Year of Employment, Information Source, and Year of Follow-Up

Information Source Last Year Employed	**1988**	**1989**	**1990**	**1991**	**1992**	**1993**	**1994**	**1995**	**1996**
Credit Bureau Records									
≤1959	13	37	36	35	38	38	38	36	33
(N=183)	(7)	(20)	(20)	(19)	(21)	(21)	(21)	(20)	(18)
1960-1969	47	100	113	124	134	132	130	133	132
(N=419)	(11)	(24)	(27)	(30)	(32)	(32)	(31)	(32)	(32)
1970-1979	88	161	178	187	208	212	206	213	209
(N=659)	(13)	(24)	(27)	(28)	(32)	(32)	(31)	(32)	(32)
≥1980	346	775	876	909	946	968	974	993	995
(N=2236)	(15)	(35)	(39)	(41)	(42)	(43)	(44)	(44)	(44)

[#] Percent = Number of People Found / Total Number of People in Last Employment Year rounded to the nearest percent

Table 16. Mineral Board Worker Cohort (Males): Residency Status by Last Year of Employment, Information Source, and Year of Follow-Up

Information Source Last Year Employed	1988	1989	1990	1991	1992	1993	1994	1995	1996
Voter Registration Records									
≤1959	5	-	5	1	1	14	45	13	52
(N=374)	(1)[#]		(1)	(<1)	(<1)	(4)	(12)	(3)	(14)
1960-1969	74	15	36	9	50	218	534	212	618
(N=2028	(4)	(1)	(2)	(<1)	(2)	(11)	(26)	(10)	(30)
1970-1979	89	12	57	8	81	246	768	210	888
(N=1948)	(5)	(1)	(3)	(<1)	(4)	(13)	(39)	(11)	(46)
Drivers License Records									
≤1959	3	4	-	1	2	11	9	22	25
(N=374)	(1)	(1)		(<1)	(1)	(3)	(2)	(6)	(7)
1960-1969	18	7	9	12	25	89	152	206	316
(N=2028	(1)	(<1)	(<1)	(1)	(1)	(4)	(7)	(10)	(16)
1970-1979	33	24	34	34	37	138	256	291	415
(N=1948)	(2)	(1)	(2)	(2)	(2)	(7)	(13)	(15)	(21)
Credit Bureau Records									
≤1959	32	63	73	77	82	84	86	88	89
(N=374)	(9)	(17)	(20)	(21)	(22)	(22)	(23)	(24)	(24)
1960-1969	242	569	680	717	806	823	827	845	862
(N=2028	(12)	(28)	(34)	(35)	(40)	(41)	(41)	(42)	(43)
1970-1979	230	573	676	708	797	803	810	824	827
(N=1948)	(12)	(29)	(35)	(36)	(41)	(41)	(42)	(42)	(42)

[#] Percent = Number of People Found / Total Number of People in Last Employment Year rounded to the nearest percent

Table 17. Mineral Board Worker Cohort (Females): Residency Status by Last Year of Employment, Information Source, and Year of Follow-Up

Information Source Last Year Employed	1988	1989	1990	1991	1992	1993	1994	1995	1996
Voter Registration Records									
≤1959	1	-	-	-	-	3	8	4	8
(N=79)	(1)[#]					(4)	(10)	(5)	(10)
1960-1969	4	-	1	-	3	15	41	15	52
(N=402)	(1)		(<1)		(1)	(4)	(10)	(4)	(13)
1970-1979	7	-	5	1	8	19	72	18	84
(N=255)	(3)		(2)	(<1)	(3)	(7)	(28)	(7)	(33)
Drivers License Records									
≤1959	-	-	1	-	-	1	2	4	6
(N=79)			(1)			(1)	(3)	(5)	(8)
1960-1969	-	-	3	1	1	7	16	22	26
(N=402)			(1)	(<1)	(<1)	(2)	(4)	(5)	(6)
1970-1979	1	1	3	2	2	18	21	26	45
(N=255)	(<1)	(<1)	(1)	(1)	(1)	(7)	(8)	(10)	(18)
Credit Bureau Records									
≤1959	1	4	5	5	5	5	5	6	6
(N=79)	(1)	(5)	(6)	(6)	(6)	(6)	(6)	(8)	(8)
1960-1969	20	35	40	40	40	44	44	43	44
(N=402)	(5)	(9)	(10)	(10)	(10)	(11)	(11)	(11)	(11)
1970-1979	19	51	56	55	61	60	63	63	67
(N=255)	(7)	(20)	(22)	(22)	(24)	(24)	(25)	(25)	(26)

[#]Percent = Number of Matches / Total Number of People in Last Employment Year rounded to the nearest percent

9. Unknown Information: Methods for Determining Data for Mineral Board Worker Cohort

After linkage with all of the available information sources, 21 percent of the 5,086 mineral board cohort workers had an unknown date of birth. Date of birth information was required for both the standardized mortality ratio and standardized incidence ratio analyses. The method for determining the date of birth for these individuals was as follows. First, the median year of birth for employees beginning their first year of employment from 1958 to 1974 was determined. For instance, for workers beginning employment at the mineral board plant in 1958 the median year of birth was 1929. For 1959 to 1974, the median years of birth were 1940, 1940, 1941, 1943, 1943, 1945, 1946, 1947, 1948, 1948, 1949, 1951, 1952, 1953, 1953, and 1954, respectively. These years (with the month and day of July 1) were assigned to the workers who also started in that year, but for whom the date of birth was unknown (N=1,063).

The twelve individuals that had no date of birth or start date were assigned July 1, 1940 for date of birth. This date was based on the average year of birth for the entire cohort. In addition, they were assigned the average start and end dates for the whole cohort, which were 1964 and 1968, respectively.

For individuals with a last date of employment, but no starting date of employment (N=3), the start date was calculated by subtracting four years (average length of employment for the cohort) from the known last date of employment. For individuals with a starting date, but no last date of employment (N=14), the ending date was calculated by adding four years (average length of employment for the cohort) to their starting date of employment.

After linkage with all of the available information, five percent of the 5,086 mineral board workers had an unknown gender. For purposes of this study, these workers were assumed to be male.

10. Estimating Minnesota State Residency through Capture-Recapture Techniques

Capture-recapture methodology was evaluated to determine its feasibility for estimating the number of people with Minnesota residency. This analysis was based on information from credit bureau records, cancer registry records, drivers license records and voter registration records. These methods rest on numerous assumptions, especially concerning the relationships between the various sources. The usual log-linear models for example require that there were no highest order interactions between the sources.

Table 18, below, delineates the estimated number of workers with Minnesota residency status for each residency year as well as the accompanying 95 percent confidence interval. The total being estimated was the total number of people living in Minnesota for that year, not including those known to have died or moved out of state. The capture-recapture model used all four sources and included all 2-way interactions. In the years with small capture rates by many of the sources, the confidence intervals were extremely wide, indicating little ability to pin down the population total. Although the confidence intervals were narrower for later years, this amount of variability would have a large impact on any standardized incidence ratios that would be based upon it.

Table 18. Mineral Board Worker Cohort (Males & Females) and Highway Worker Cohort (Males): Capture-Recapture Estimates of Minnesota Residency Status by Year of Follow-Up

Follow-Up Year	Highway Worker Cohort			Mineral Board Worker Cohort		
	Number of Workers Alive for Any Portion of the Residency Year	Capture-Recapture Estimate of Minnesota Residency*	95 Percent Confidence Interval	Number of Workers Alive for Any Portion of the Residency Year	Capture-Recapture Estimate of Minnesota Residency*	95 Percent Confidence Interval
1988	3497	568	568 - 22,343	5086	3798	609 - Infinity
1989	3425	5277	3296 - Infinity	5039	3133	1063 - Infinity
1990	3360	5356	1527 - Infinity	5004	3520	1300 - Infinity
1991	3297	5358	1406 - Infinity	4961	5482	1714 - Infinity
1992	3241	1370	1368 - 10517	4913	3410	1549 - Infinity
1993	3179	2798	2275 - 3649	4868	3692	2723 - 5407
1994	3112	2713	2570 - 2917	4816	2659	2491 - 2893
1995	3054	3164	2651 - 3926	4774	2883	2476 - 3485
1996	3004	2728	2650 - 2828	4724	2617	2523 - 2744

*The total being estimated does not include people who are known to have died, or were known to have been living out-of-state. The model used includes all 2-way interactions.

From this analysis of the mineral board and highway worker cohorts, it was clear that that capture-recapture estimates of denominator data would not be useful in calculating standardized incidence ratios. Because this analysis was not deemed useful, it was not repeated for the 3M cohort after consultation with study staff and investigators.

11. Sensitivity Analyses and Follow-Up Bias: Three Models for Completing Unknown Residency Status Information

To calculate a standardized incidence ratio (SIR), two basic numbers are required: the observed number of events and the expected number of events. The expected number of events, in turn, is dependent on two quantities: the cancer rates used in the reference or comparison population and the person-years to which those rates are applied. In this study, the state cancer registry was used to identify all cancers diagnosed in Minnesota in the cohort and also to provide the reference cancer rates. It was assumed that these measures were reasonably complete and accurate. However, the person-years for Minnesota residency was clearly incomplete. Therefore, the expected number of cancers and corresponding SIR would be strongly affected by the assumptions and information used to fill in missing residency years. This represents a significant potential for follow-up bias, a major source of concern in all cohort studies. Sensitivity analyses were conducted to analyze residency assumptions and the potential for follow-up bias.

Although record-linkage activities had been completed with all of our available information sources there were still a large percentage of workers with unknown residency status information. Overall, 5 percent of male highway workers, 27 percent of male mineral board workers, and 61 percent of female mineral board workers had no residency status information at all (i.e., they were not found within any of the information sources). Also, many individuals had unknown residency data for one to eight of the possible nine year residency history (1988-1996).

It was clear that the method used to "fill" or complete the unknown residency years as Minnesota or non-Minnesota residency would potentially have a major impact on the standardized incidence ratios. Therefore, three sets of different assumptions (Models A, B, and C) were tested to determine their influence on the total years of Minnesota residency for each cohort as well as its influence on the SIR itself for a number of selected cancers. In addition, we used residency information from the Minnesota Department of Revenue (Minnesota Revenue records) as the "gold standard" to determine how each of these models compared to their estimation of Minnesota residency (See Section 12). Previous tracing efforts in Minnesota demonstrated that Minnesota Revenue was able to locate state residents not traceable by any other source.

Model A: Model A was the most liberal model in terms of determining Minnesota residency (i.e., it maximizes the number of person-years at risk in the denominator of the SIR estimate). In this model, none of the residency information from any of the information sources was used. Each worker was assumed to have been a Minnesota resident from 1988 until the end of follow-up. Although this model is basically unsound

in its assumptions, it formed the basis for the most extreme example possible and would result in the lowest possible SIR estimate because of the increased years of Minnesota residency.

Model B: Model B used all of the residency information provided by the information sources, but then assumed Minnesota residency for all unknown residency years (all unknown residency years were "filled" as Minnesota residency). This model was also quite liberal in terms of assuming Minnesota residency, but it did utilize all information that was known concerning out-of-state (non-Minnesota) residency.

Model C: Model C was the closest estimation to "truth" about residency status. Initially, it used all of the information about residency status from each of the information sources, but then supplemented the data in two ways. First, a set of assumptions was used to "fill" state of residency information between gaps for each person found by one of the information sources as delineated below.

- Death certificate or national death index records: If the residency status for the year preceding death was unknown, it was assigned the same residency as the state of residency at the time of death. For example: If a worker died in Minnesota in 1995 and there was no residency status for 1994, then 1994 was assigned as Minnesota residency.
- Drivers license records: If the preceding residency year was unknown then it was assigned as Minnesota residency. If the next succeeding two years were unknown then they were assigned Minnesota residency. If only one succeeding year was unknown then it was assigned Minnesota residency.
- Voter registration records: A check was conducted for unknown residency status between the years when the worker was known to have voted. If a "gap" of one or two years was identified, then that gap was filled with Minnesota residency. If there was only one year when the individual voted, the preceding and succeeding years were assigned as Minnesota residency if they were unknown.
- Credit bureau records: A check was conducted for unknown residency status between the years when the worker was known to have lived in one state. If a "gap" of one or two years was identified, then that gap was filled with Minnesota residency status when the preceding year or succeeding year were Minnesota residency. If a "gap" of one or two years was identified, then that gap was filled with "Other" residency status when the preceding year or succeeding year were non-Minnesota residency. If there was only one year when the individual had a known residency, then the preceding and succeeding year were assigned with the same residency status (Minnesota or Other State) if they were unknown.

Second, the remaining unknown residency years were filled using information about Minnesota's out-of-state migration rate (3%). For example, if we knew that a worker lived in Minnesota in 1988, but then had no information on residency status until 1996, the succeeding years would be filled in the following manner: 1989=0.97; 1990=0.97^2;

1991=0.97^3; 1992=0.97^4; 1993=0.97^5; 1994=0.97^6; and 1995= 0.97^7. If there was no residency information for 1988, then the same pattern was used to fill succeeding years beginning with the last known date of employment.

Although we had information about out-of-state migration, we had no information on the number or percent of individuals who moved out of Minnesota and subsequently returned. For this reason, we also decided to use the 3 percent figure for in-state migration. For example, if we knew that a worker was not a Minnesota resident in 1988 and had no further residency status until 1996, the succeeding years would be filled in the following manner: 1989=(1 - 0.97); 1990=(1 - 0.97^2); 1991=(1 - 0.97^3); 1992=(1 - 0.97^4); 1993=(1 - 0.97^5); 1994=(1 - 0.97^6); and 1995= (1 - 0.97^7).

Although we had incomplete information concerning the number of workers in both cohorts that moved in or out of Minnesota between 1988-1996, some information was available based on the confirmed matches with the different information sources. Overall, 149 (4%) of the highway workers lived out-of-state for at least one year between 1988 and 1996; 56 of the 149 workers (38%) also listed a Minnesota residency for a minimum of one year. Of the 56 workers with both a Minnesota and out-of-state residence, 27 lived in Minnesota and then moved out-of-state, 11 lived out-of-state and then returned to Minnesota, and 18 moved between Minnesota and another state several times between 1988 and 1996.

Of the 4350 male mineral board workers, 554 (13%) lived out-of-state for at least one year between 1988 and 1996; 109 of the 554 workers (20%) also listed a Minnesota residency for a minimum of one year. Of the 109 workers with both a Minnesota and out-of-state residence, 38 lived in Minnesota and then moved out-of-state, 23 lived out-of-state and then returned to Minnesota, and 48 moved between Minnesota and another state several times between 1988 and 1996. Of the 736 female mineral board workers, 23 (3%) lived out-of-state for at least one year; 5 of the 23 workers (22%) also listed a Minnesota residency for a minimum of one year. Three workers lived in Minnesota and then moved out-of-state and two workers lived out-of-state and then returned to Minnesota.

In addition to creating three separate models, we also evaluated the effect of seven different combinations of information sources on the SIRs for Models B and C. These seven different combinations were studied because study investigators elsewhere may not have access to all of the information sources used in this study and would potentially want to use a subset of the information sources. To examine the effect that this might have within each model type, the SIR was analyzed by different combinations of information sources. Assuming that all investigators would have access to death data and cancer diagnosis information to conduct an SIR study, the residency information provided from those sources was evaluated in seven different combinations with:

(1) drivers license (DL) records, voter registration (VR) records, and credit bureau (CR) records;
(2) drivers license (DL) records;
(3) voter registration (VR) records;

(4) credit bureau (CB) records;
(5) drivers license (DL) records and credit bureau (CB) records;
(6) voter registration (VR) records and credit bureau (CB) records; and
(7) drivers license (DL) records and voter registration (VR) records.

12. Standardized Incidence Ratio (SIR) Sensitivity Analyses: Summaries by Cancer

The three sensitivity models, (A, B, and C) described previously, are displayed in a graphical format on a log scale from 0.1 to 10.0 with the standardized incidence ratio (SIR) and their 95 percent confidence intervals shown (Figures 3 through 24). For males, ten individual cancers were evaluated as well as all cancers combined. The individual cancers are listed below.

Pancreas	Stomach
Colon & Rectum	Lung and Bronchus
Prostate	Kidney and Renal Pelvis
Urinary Bladder	Melanomas of the Skin
Non-Hodgkin's Lymphomas	Leukemias

Fifteen different SIRs were calculated for each of the ten cancers. Model A had only one SIR estimate because Minnesota residency status was assumed until 1996 or the workers date of death. Models B and C each included seven different SIR estimates based on the different combinations of information sources used to generate the residency status. Assuming that all investigators would have access to death data and cancer diagnosis information to conduct an SIR study, the residency information provided from those sources was evaluated in seven different combinations with:

(1) drivers license (DL) records, voter registration (VR) records, and credit bureau (CR) records;
(2) drivers license (DL) records;
(3) voter registration (VR) records;
(4) credit bureau (CB) records;
(5) drivers license (DL) records and credit bureau (CB) records;
(6) voter registration (VR) records and credit bureau (CB) records; and
(7) drivers license (DL) records and voter registration (VR) records.

Because of the low number of females in the Mineral Board Worker cohort (N=736), only two cancers, breast and lung, were evaluated. The SIRs also were calculated for all cancers combined. Model A only had one SIR estimate because Minnesota residency status was assumed until 1996 or the workers date of death. Models B and C were calculated using all available residency information (i.e., death data, cancer diagnosis, drivers license records, voter registration records, and credit bureau records). Figures 25

through 27 display the SIR estimates and corresponding 95 percent confidence intervals for females.

Tables 19 and 20 provide information concerning the number of observed cancers, the number of expected cancers, and the ratio of observed to expected cancers by cancer for both cohorts (males and females) for Model A, Model B, and Model C. Every percentage difference in the total number of residency years for the cohort resulted in a corresponding change in the opposite direction for the standardized incidence ratio. For example, a five percent decrease in residency years resulted in a five percent increase in the SIR estimate. This occurred because the numerator (observed number of cancers) was based on information from cancer registry records and did not vary between the different models for the cancers; the only information that changed between the models was the calculation of the denominator (number of years of residency in Minnesota).

An examination of the figures and tables for both cohorts shows a similar pattern. For the most part, very little variation is seen in the seven SIRs for Model B. Figures 3-27 do not portray significant differences between the SIRs or their 95 percent confidence intervals.

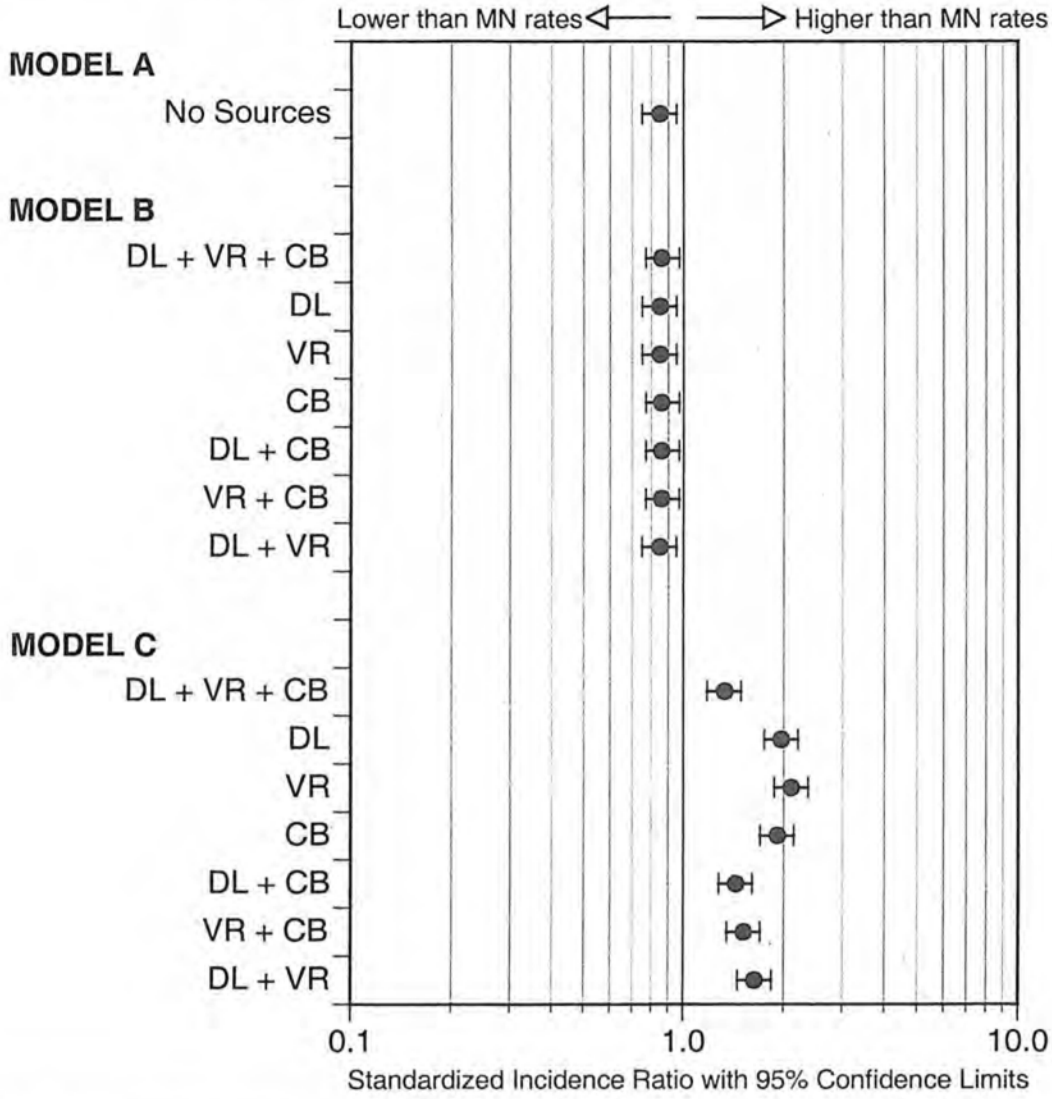
The difference between the ratio of observed cancers/expected cancers for Model A (assumed all Minnesota residency) and Model B (used residency status from all information sources) ranged from 0.02 to 0.07 for male mineral board. The difference between the ratio of observed cancers/expected cancers for Model A and Model B (all information sources) ranged from 0.01 to 0.03 for male highway workers. The difference between Model A and Model B SIR estimates was negligible for both cohorts, a somewhat unexpected finding since Model B did take advantage of any known residency information.

Model C with its seven SIRs was quite different than Model A or B. There also were significant differences between the SIRs and their confidence intervals when the different information source combinations were compared. The information sources providing the lowest to highest SIR estimates are listed below for Model C. This pattern was seen with every cancer.

(1) drivers license (DL) records + voter registration (VR) records + credit bureau (CR) records
(2) drivers license (DL) records + credit bureau (CR) records;
(3) voter registration records (VR) + credit bureau (CR) records;
(4) drivers license (DL) records + voter registration (VR) records;
(5) credit bureau (CB) records;
(6) drivers license (DL) records; and
(7) voter registration (VR) records.

In essence, the lowest SIR estimate resulted from using all sources, the next lowest estimates came from using at least two sources, and the highest estimates were from residency histories derived from only one information source.

Figure 3. Highway Worker Cohort Standardized Incidence Ratio (SIR) by Residency Model: Males, All Cancers

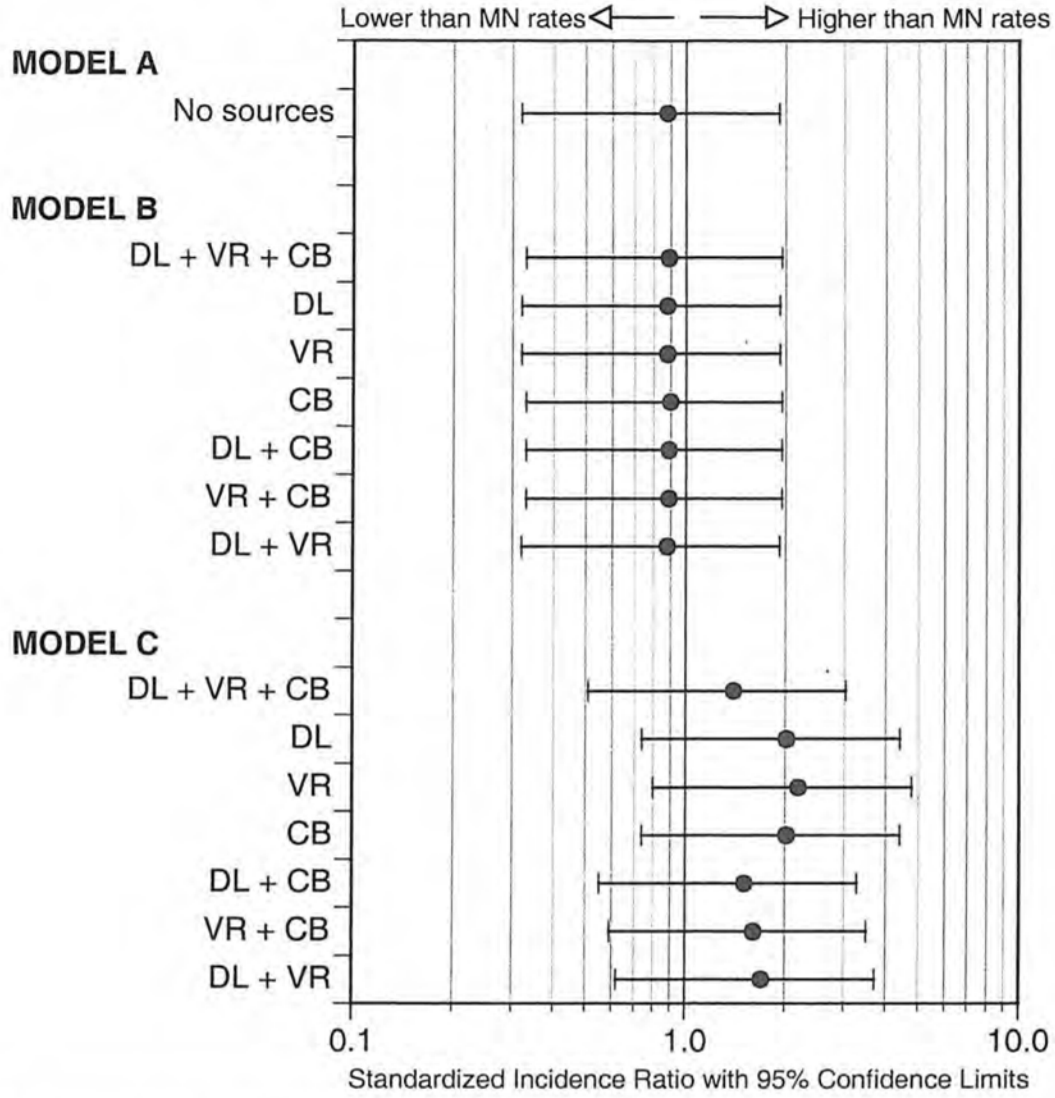


Information Sources:
DL=Drivers license records; VR=Voter registration records; CB=Credit bureau records

Residency Models:

- Model A assumes all were MN residents until the end of follow-up. No other sources were used.
- Model B assumes all were MN residents unless there was information to indicate otherwise.
- Model C assumes a 3% in- or out-of-state migration rate per year. (See text for explanation.)
- Models B and C also include residency information from cancer registry records and death records.

Figure 4. Highway Worker Cohort Standardized Incidence Ratio (SIR) by Residency Model: Males, Stomach

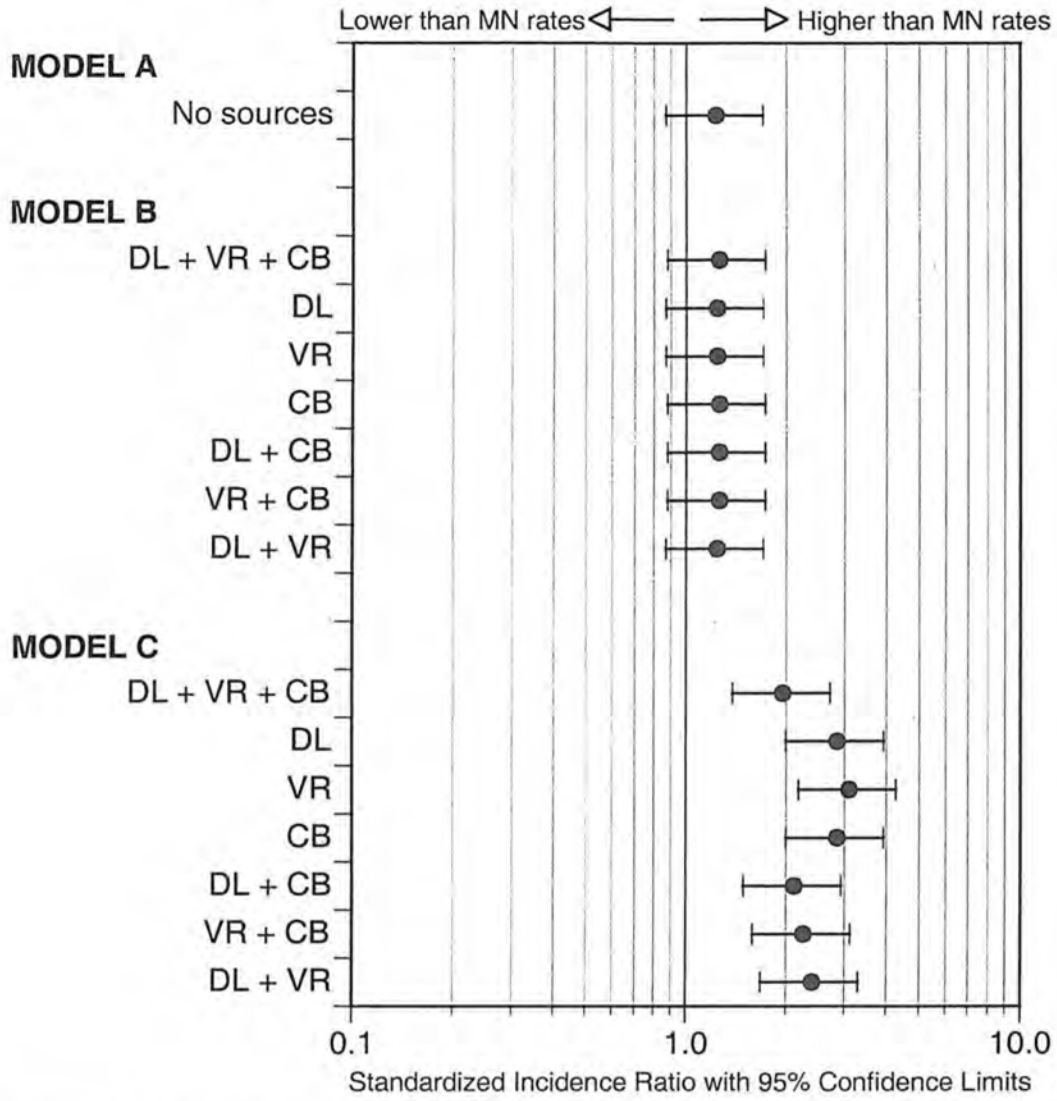


Information Sources:
DL=Drivers license records; VR=Voter registration records; CB=Credit bureau records

Residency Models:

- Model A assumes all were MN residents until the end of follow-up. No other sources were used.
- Model B assumes all were MN residents unless there was information to indicate otherwise.
- Model C assumes a 3% in- or out-of-state migration rate per year. (See text for explanation.)
- Models B and C also include residency information from cancer registry records and death records.

Figure 5. Highway Worker Cohort Standardized Incidence Ratio (SIR) by Residency Model: Males, Colon (Excluding Rectum)

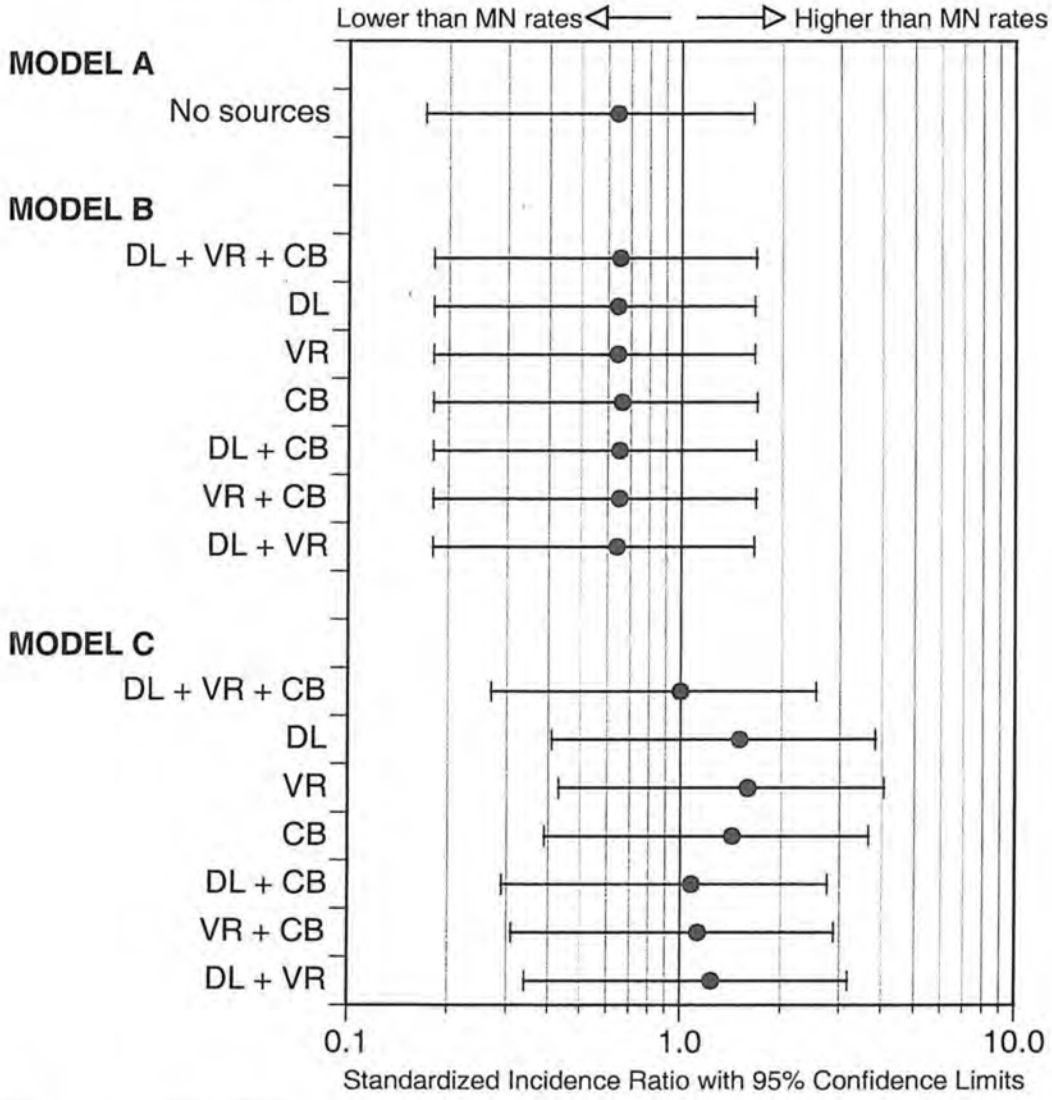


Information Sources:
DL=Drivers license records; VR=Voter registration records; CB=Credit bureau records

Residency Models:

- Model A assumes all were MN residents until the end of follow-up. No other sources were used.
- Model B assumes all were MN residents unless there was information to indicate otherwise.
- Model C assumes a 3% in- or out-of-state migration rate per year. (See text for explanation.)
- Models B and C also include residency information from cancer registry records and death records.

Figure 7. Highway Worker Cohort Standardized Incidence Ratio (SIR) by Residency Model: Males, Pancreas

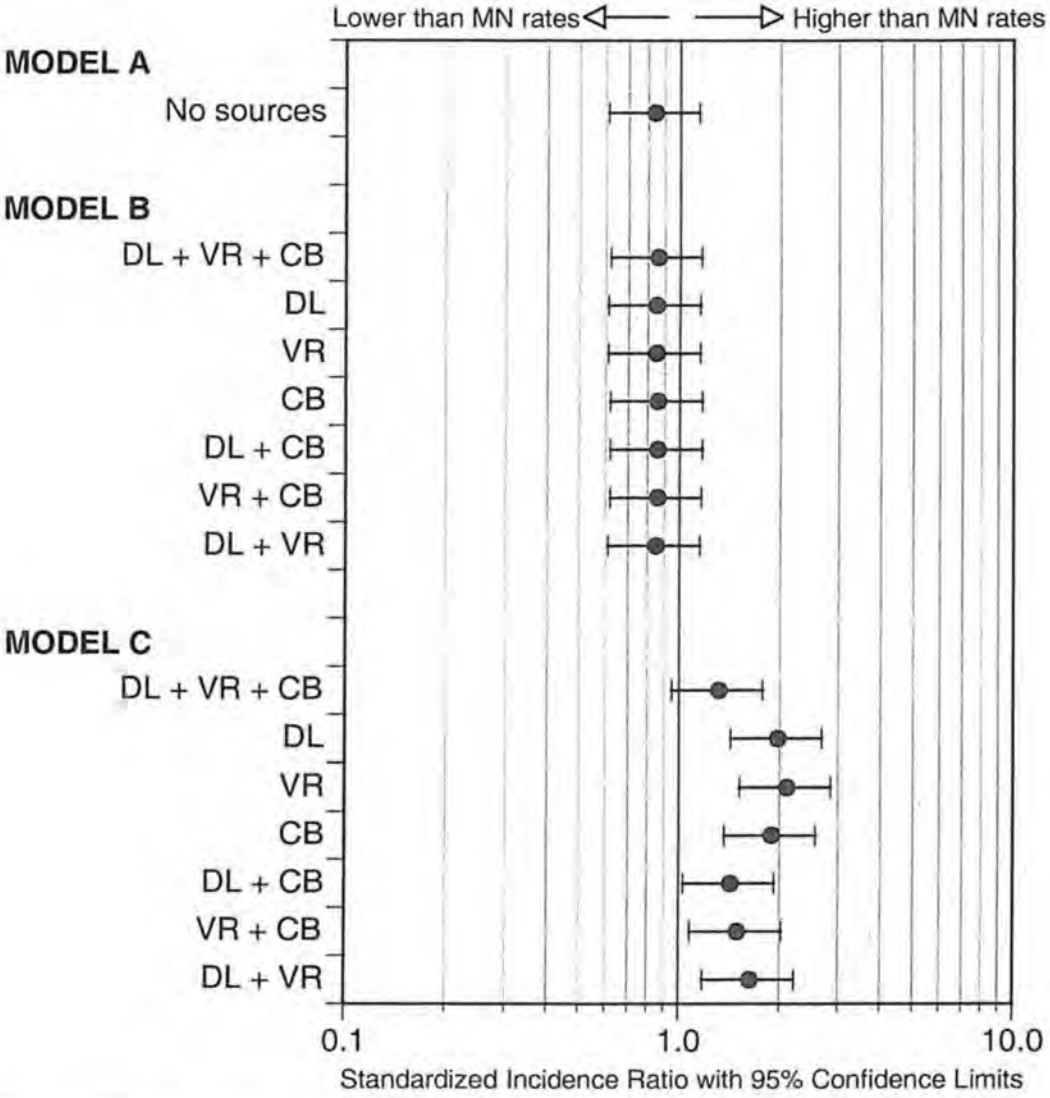


Information Sources:
DL=Drivers license records; VR=Voter registration records; CB=Credit bureau records

Residency Models:

- Model A assumes all were MN residents until the end of follow-up. No other sources were used.
- Model B assumes all were MN residents unless there was information to indicate otherwise.
- Model C assumes a 3% in- or out-of-state migration rate per year. (See text for explanation.)
- Models B and C also include residency information from cancer registry records and death records.

Figure 8. Highway Worker Cohort Standardized Incidence Ratio (SIR) by Residency Model: Males, Lung and Bronchus

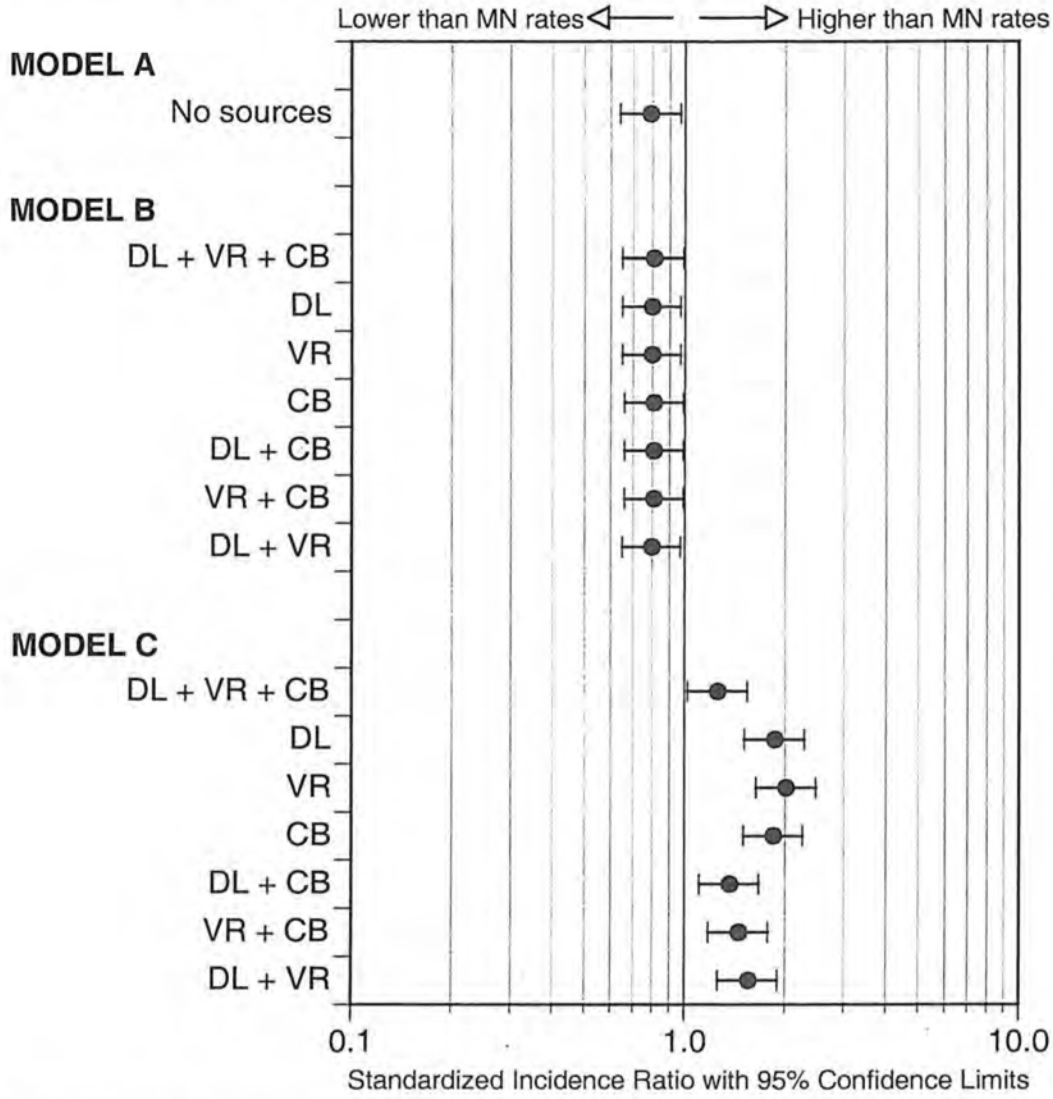


Information Sources:
DL=Drivers license records; VR=Voter registration records; CB=Credit bureau records

Residency Models:

- Model A assumes all were MN residents until the end of follow-up. No other sources were used.
- Model B assumes all were MN residents unless there was information to indicate otherwise.
- Model C assumes a 3% in- or out-of-state migration rate per year. (See text for explanation.)
- Models B and C also include residency information from cancer registry records and death records.

Figure 9. Highway Worker Cohort Standardized Incidence Ratio (SIR) by Residency Model: Males, Prostate

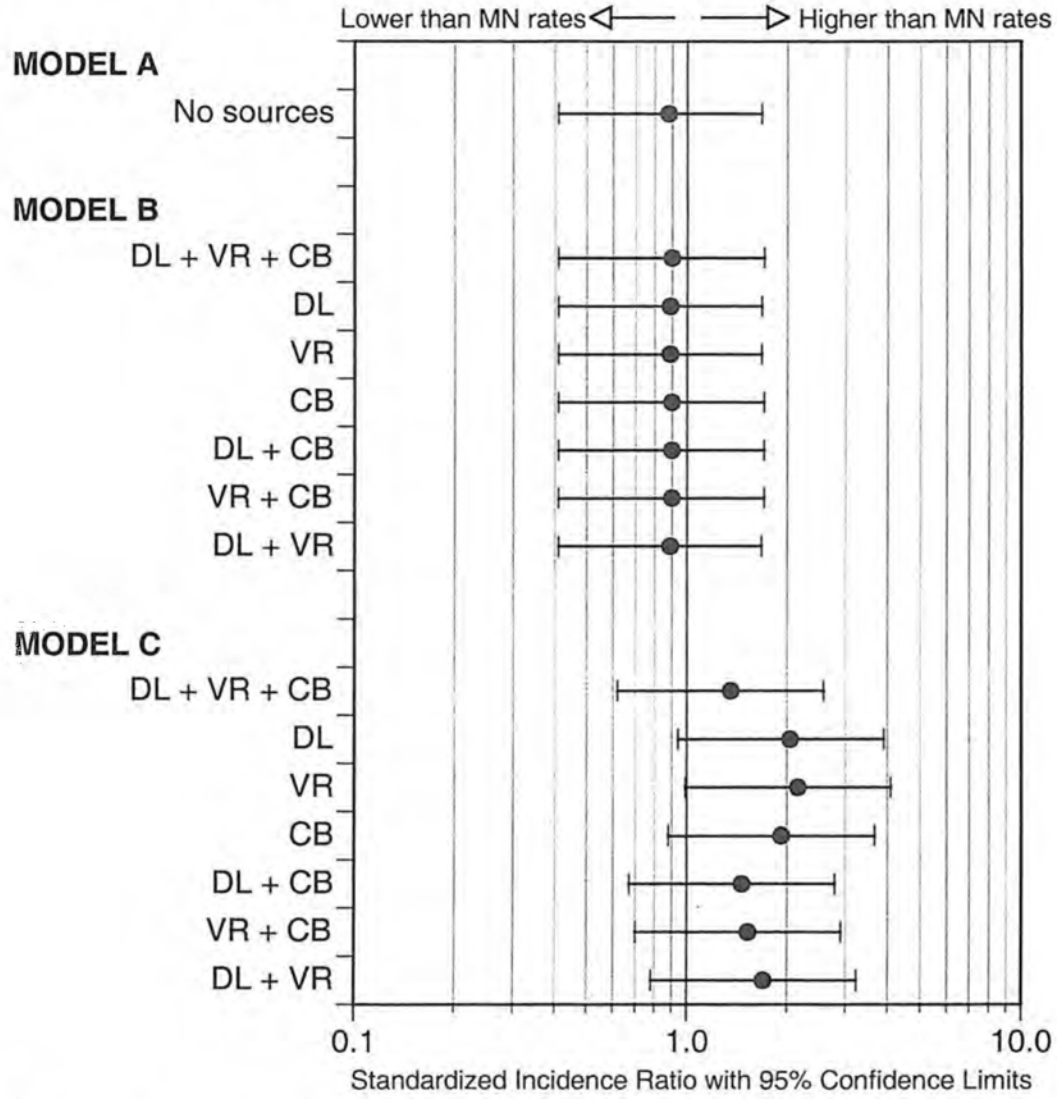


Information Sources:
DL=Drivers license records; VR=Voter registration records; CB=Credit bureau records

Residency Models:

- Model A assumes all were MN residents until the end of follow-up. No other sources were used.
- Model B assumes all were MN residents unless there was information to indicate otherwise.
- Model C assumes a 3% in- or out-of-state migration rate per year. (See text for explanation.)
- Models B and C also include residency information from cancer registry records and death records.

Figure 10. Highway Worker Cohort Standardized Incidence Ratio (SIR) by Residency Model: Males, Kidney and Renal Pelvis

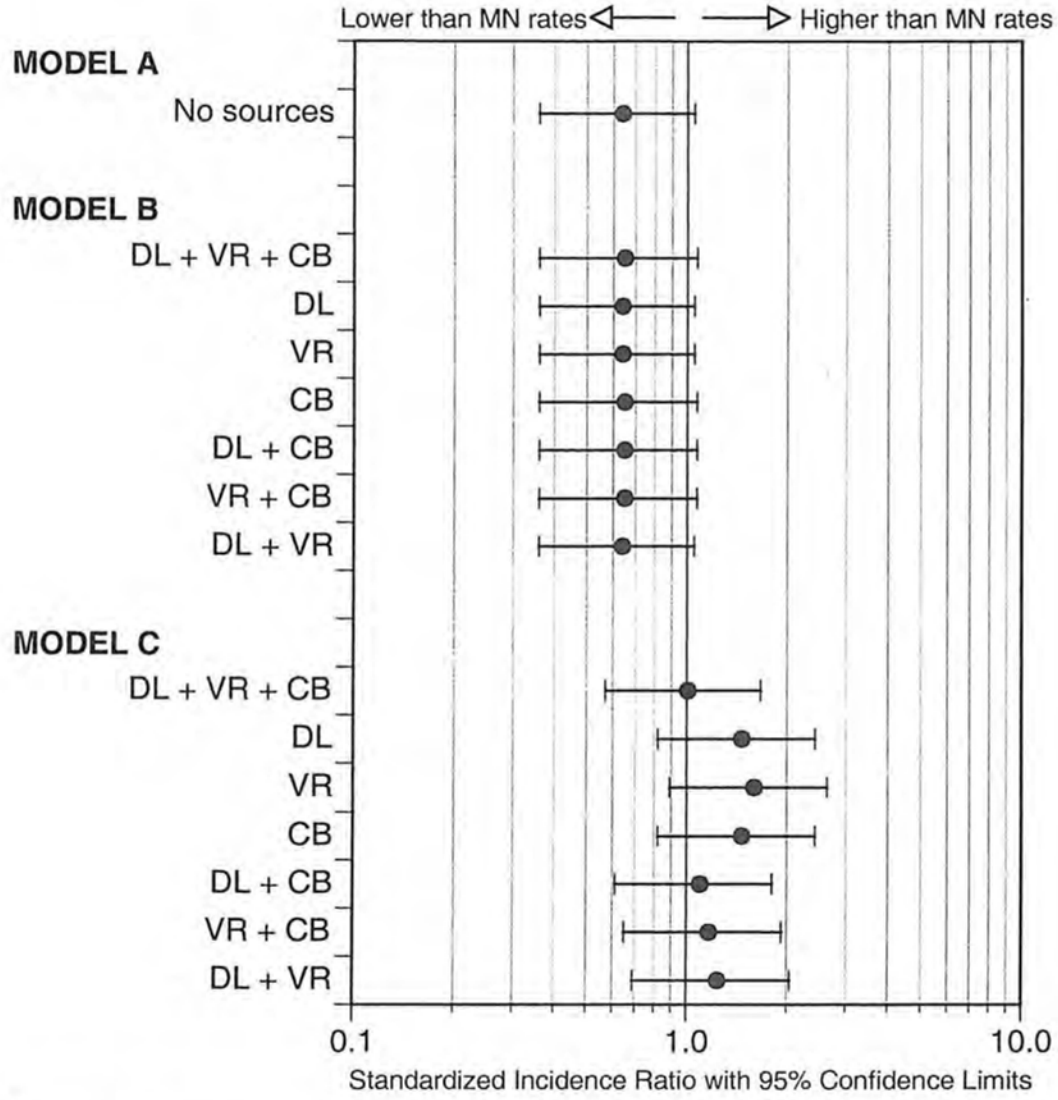


Information Sources:
DL=Drivers license records; VR=Voter registration records; CB=Credit bureau records

Residency Models:

- Model A assumes all were MN residents until the end of follow-up. No other sources were used.
- Model B assumes all were MN residents unless there was information to indicate otherwise.
- Model C assumes a 3% in- or out-of-state migration rate per year. (See text for explanation.)
- Models B and C also include residency information from cancer registry records and death records.

Figure 11. Highway Worker Cohort Standardized Incidence Ratio (SIR) by Residency Model: Males, Urinary Bladder

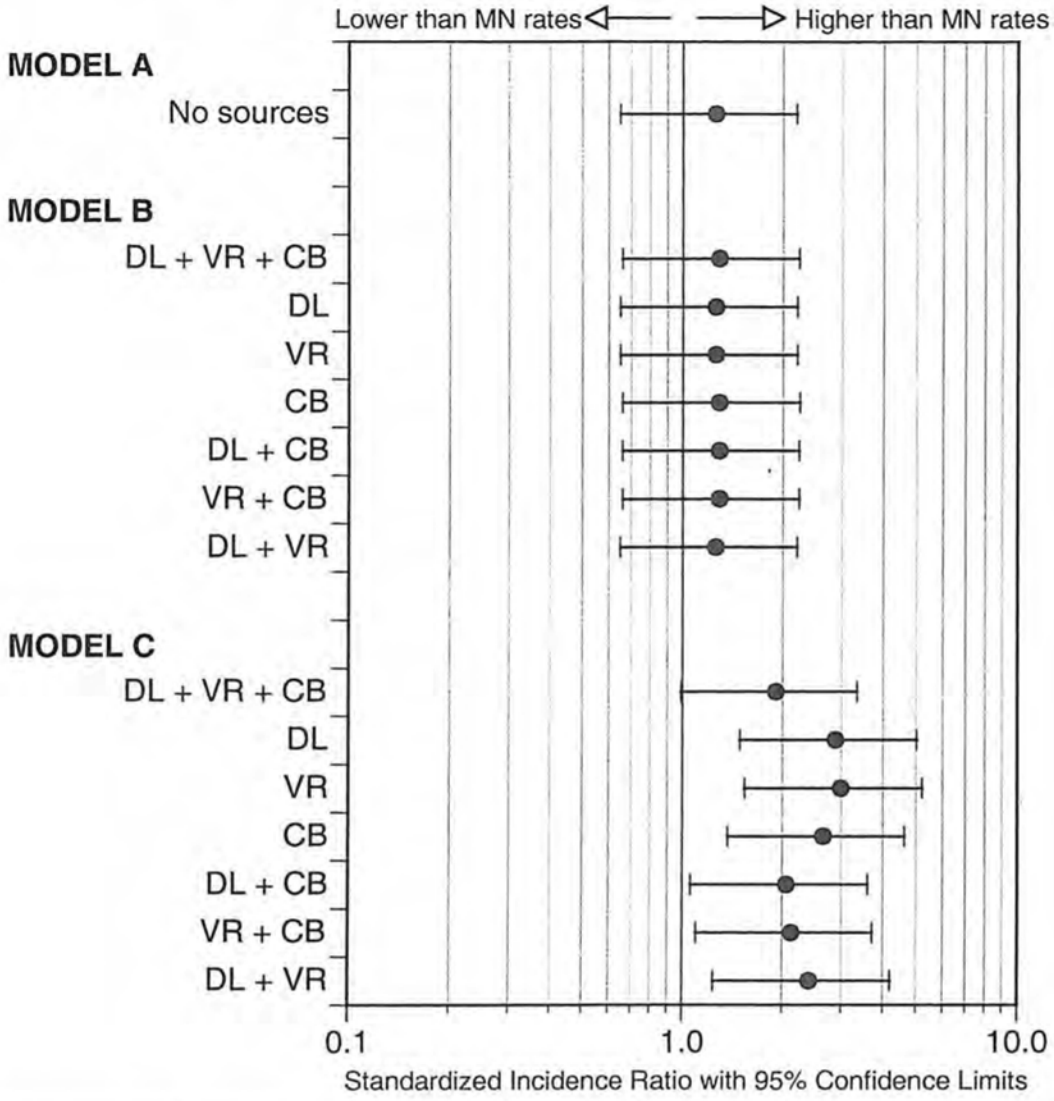


Information Sources:
DL=Drivers license records; VR=Voter registration records; CB=Credit bureau records

Residency Models:

- Model A assumes all were MN residents until the end of follow-up. No other sources were used.
- Model B assumes all were MN residents unless there was information to indicate otherwise.
- Model C assumes a 3% in- or out-of-state migration rate per year. (See text for explanation.)
- Models B and C also include residency information from cancer registry records and death records.

Figure 12. Highway Worker Cohort Standardized Incidence Ratio (SIR) by Residency Model: Males, Melanomas of the Skin

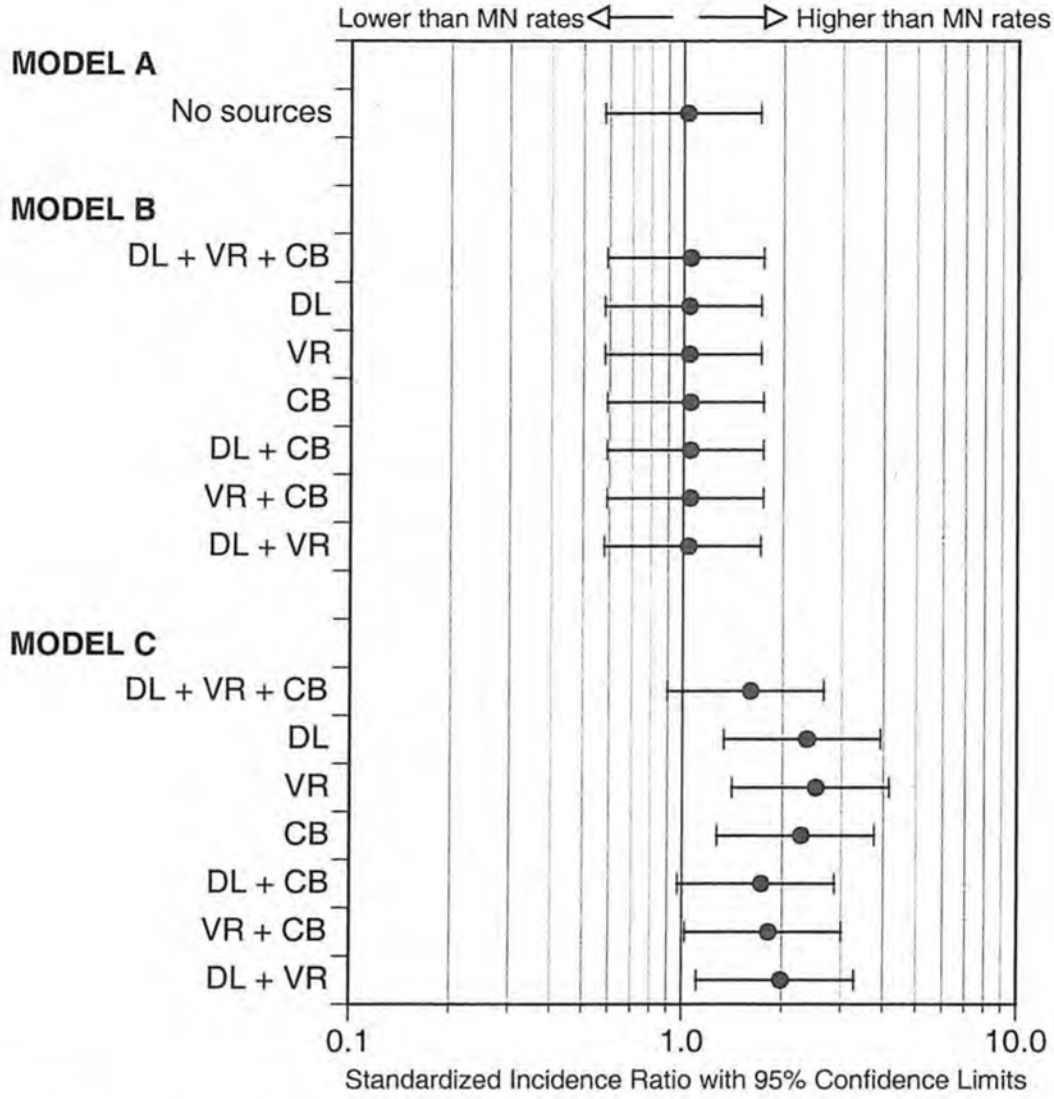


Information Sources:
DL=Drivers license records; VR=Voter registration records; CB=Credit bureau records

Residency Models:

- Model A assumes all were MN residents until the end of follow-up. No other sources were used.
- Model B assumes all were MN residents unless there was information to indicate otherwise.
- Model C assumes a 3% in- or out-of-state migration rate per year. (See text for explanation.)
- Models B and C also include residency information from cancer registry records and death records.

Figure 13. Highway Worker Cohort Standardized Incidence Ratio (SIR) by Residency Model: Males, Non-Hodgkin's Lymphomas

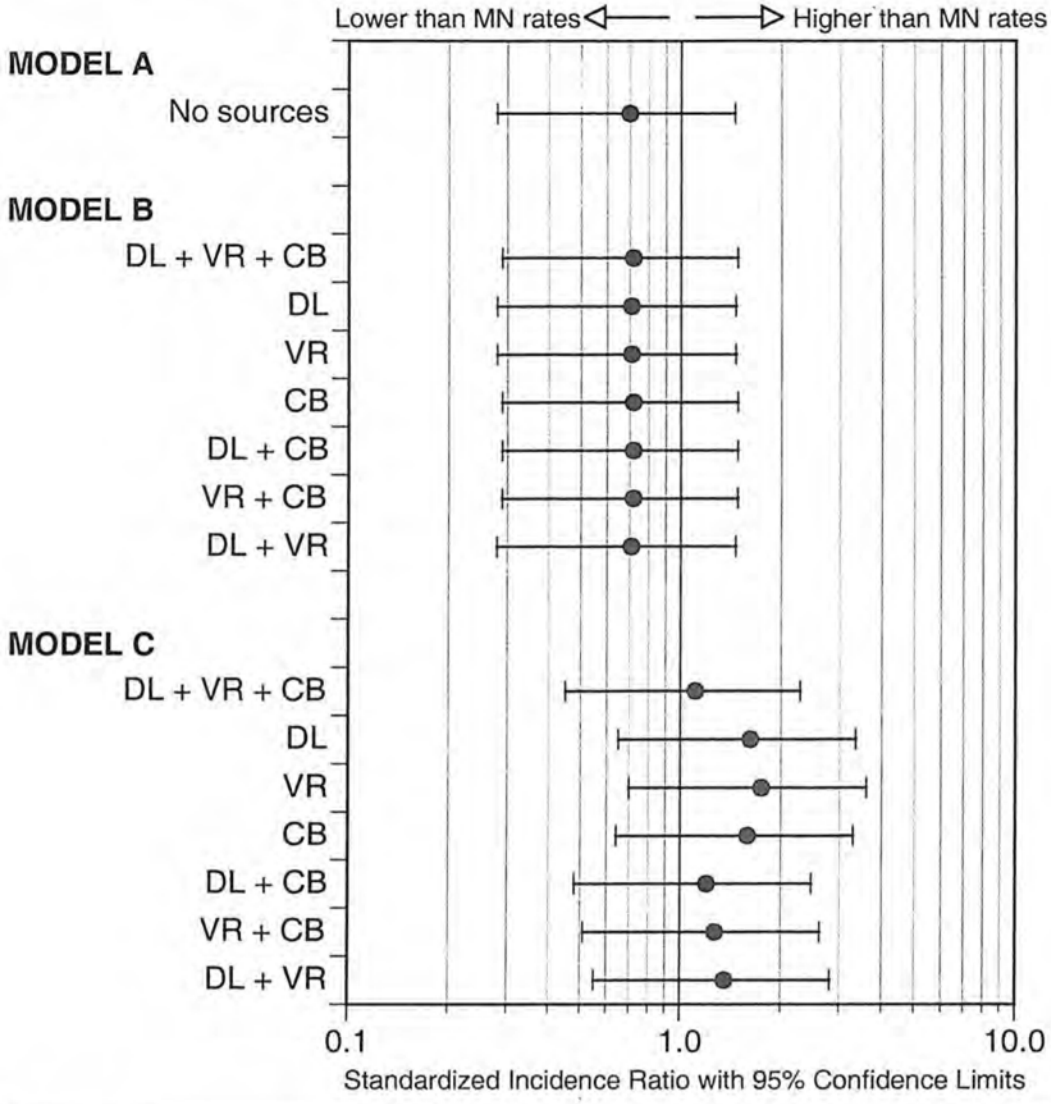


Information Sources:
DL=Drivers license records; VR=Voter registration records; CB=Credit bureau records

Residency Models:

- Model A assumes all were MN residents until the end of follow-up. No other sources were used.
- Model B assumes all were MN residents unless there was information to indicate otherwise.
- Model C assumes a 3% in- or out-of-state migration rate per year. (See text for explanation.)
- Models B and C also include residency information from cancer registry records and death records.

Figure 14. Highway Worker Cohort Standardized Incidence Ratio (SIR) by Residency Model: Males, Leukemias

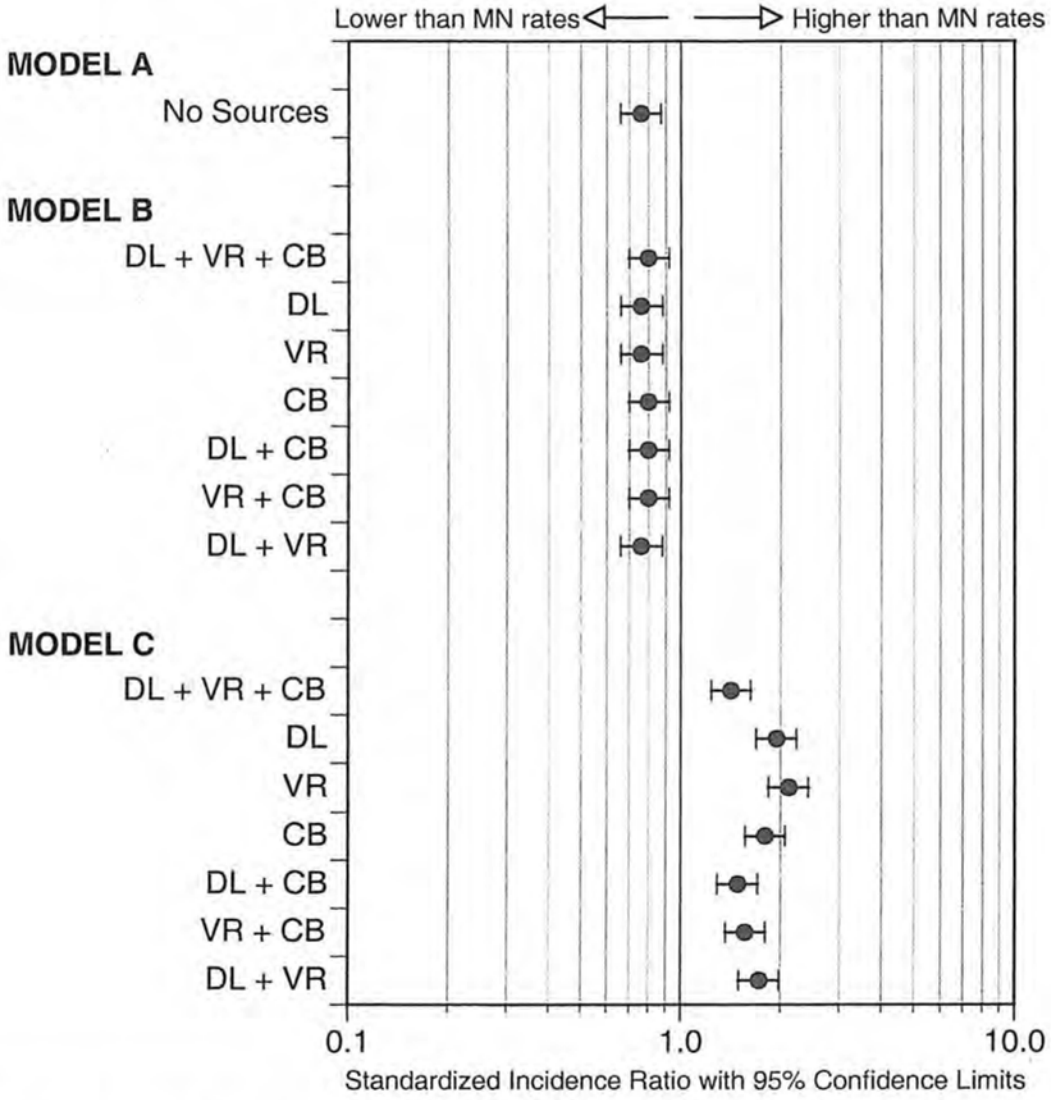


Information Sources:
DL=Drivers license records; VR=Voter registration records; CB=Credit bureau records

Residency Models:

- Model A assumes all were MN residents until the end of follow-up. No other sources were used.
- Model B assumes all were MN residents unless there was information to indicate otherwise.
- Model C assumes a 3% in- or out-of-state migration rate per year. (See text for explanation.)
- Models B and C also include residency information from cancer registry records and death records.

Figure 15. Mineral Board Worker Cohort Standardized Incidence Ratio (SIR) by Residency Model: Males, All Cancers

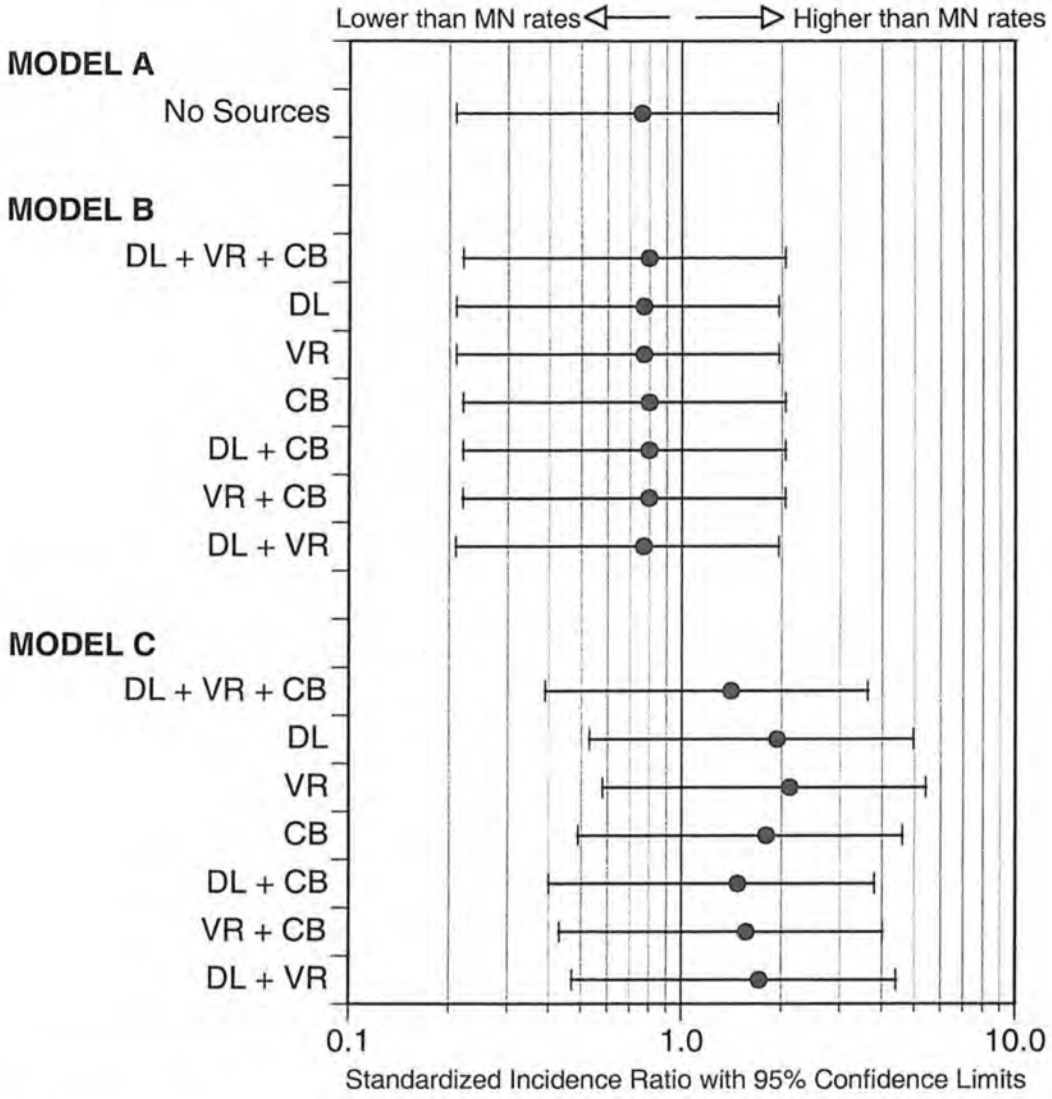


Information Sources:
DL=Drivers license records; VR=Voter registration records; CB=Credit bureau records

Residency Models:

- Model A assumes all were MN residents until the end of follow-up. No other sources were used.
- Model B assumes all were MN residents unless there was information to indicate otherwise.
- Model C assumes a 3% in- or out-of-state migration rate per year. (See text for explanation.)
- Models B and C also include residency information from cancer registry records and death records.

Figure 16. Mineral Board Worker Cohort Standardized Incidence Ratio (SIR) by Residency Model: Males, Stomach

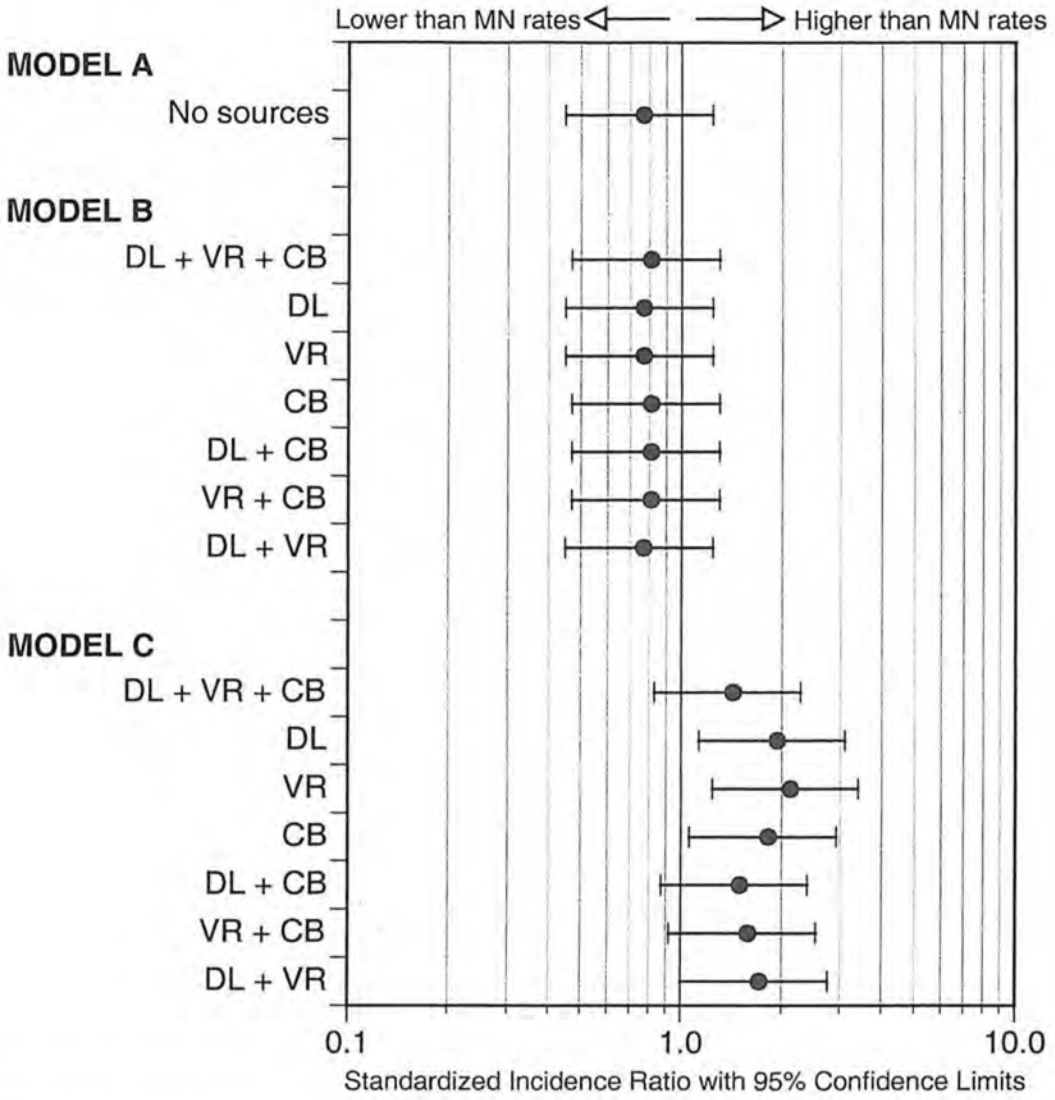


Information Sources:
DL=Drivers license records; VR=Voter registration records; CB=Credit bureau records

Residency Models:

- Model A assumes all were MN residents until the end of follow-up. No other sources were used.
- Model B assumes all were MN residents unless there was information to indicate otherwise.
- Model C assumes a 3% in- or out-of-state migration rate per year. (See text for explanation.)
- Models B and C also include residency information from cancer registry records and death records.

Figure 17. Mineral Board Worker Cohort Standardized Incidence Ratio (SIR) by Residency Model: Males, Colon (Excluding Rectum)

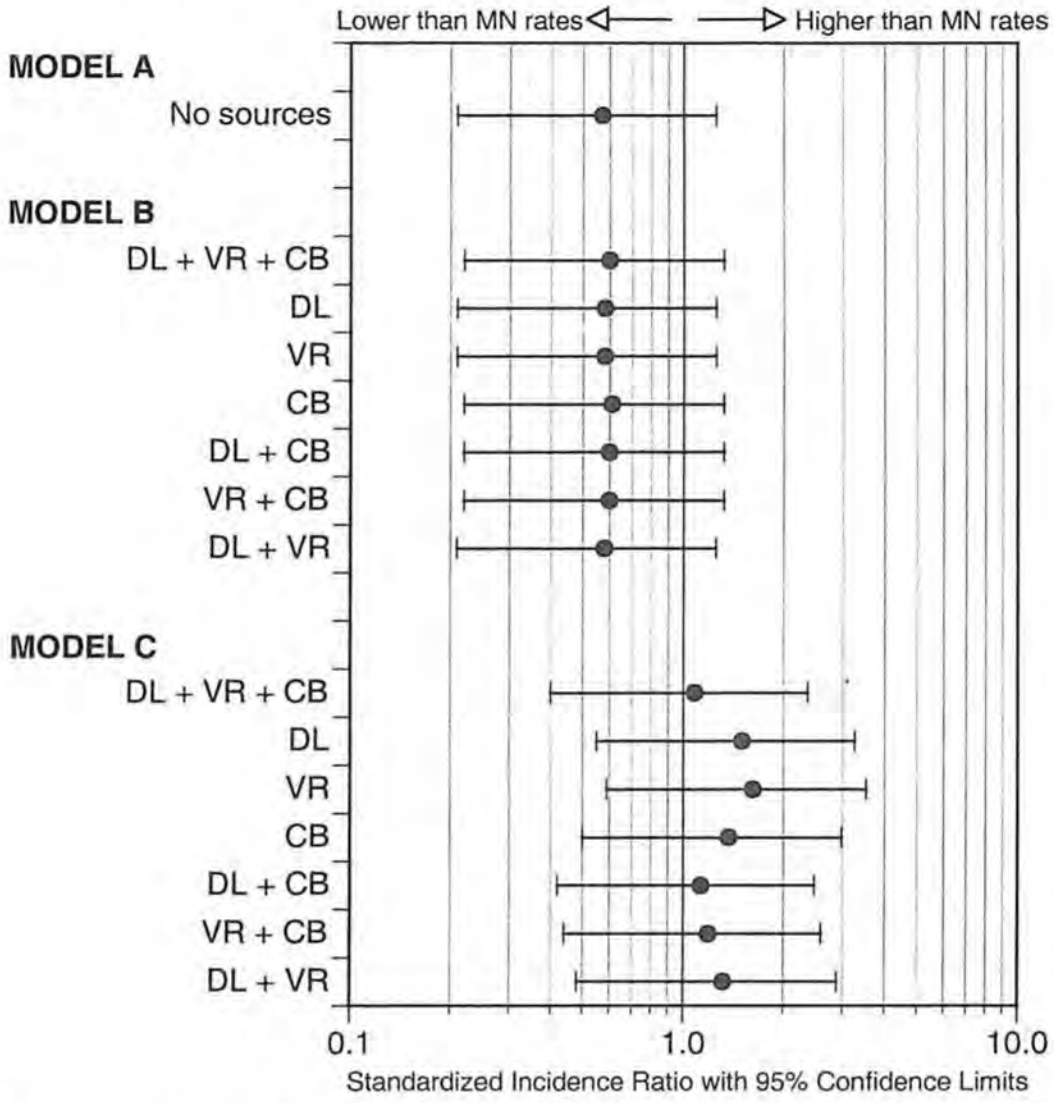


Information Sources:
DL=Drivers license records; VR=Voter registration records; CB=Credit bureau records

Residency Models:

- Model A assumes all were MN residents until the end of follow-up. No other sources were used.
- Model B assumes all were MN residents unless there was information to indicate otherwise.
- Model C assumes a 3% in- or out-of-state migration rate per year. (See text for explanation.)
- Models B and C also include residency information from cancer registry records and death records.

Figure 18. Mineral Board Worker Cohort Standardized Incidence Ratio (SIR) by Residency Model: Males, Rectum and Rectosigmoid

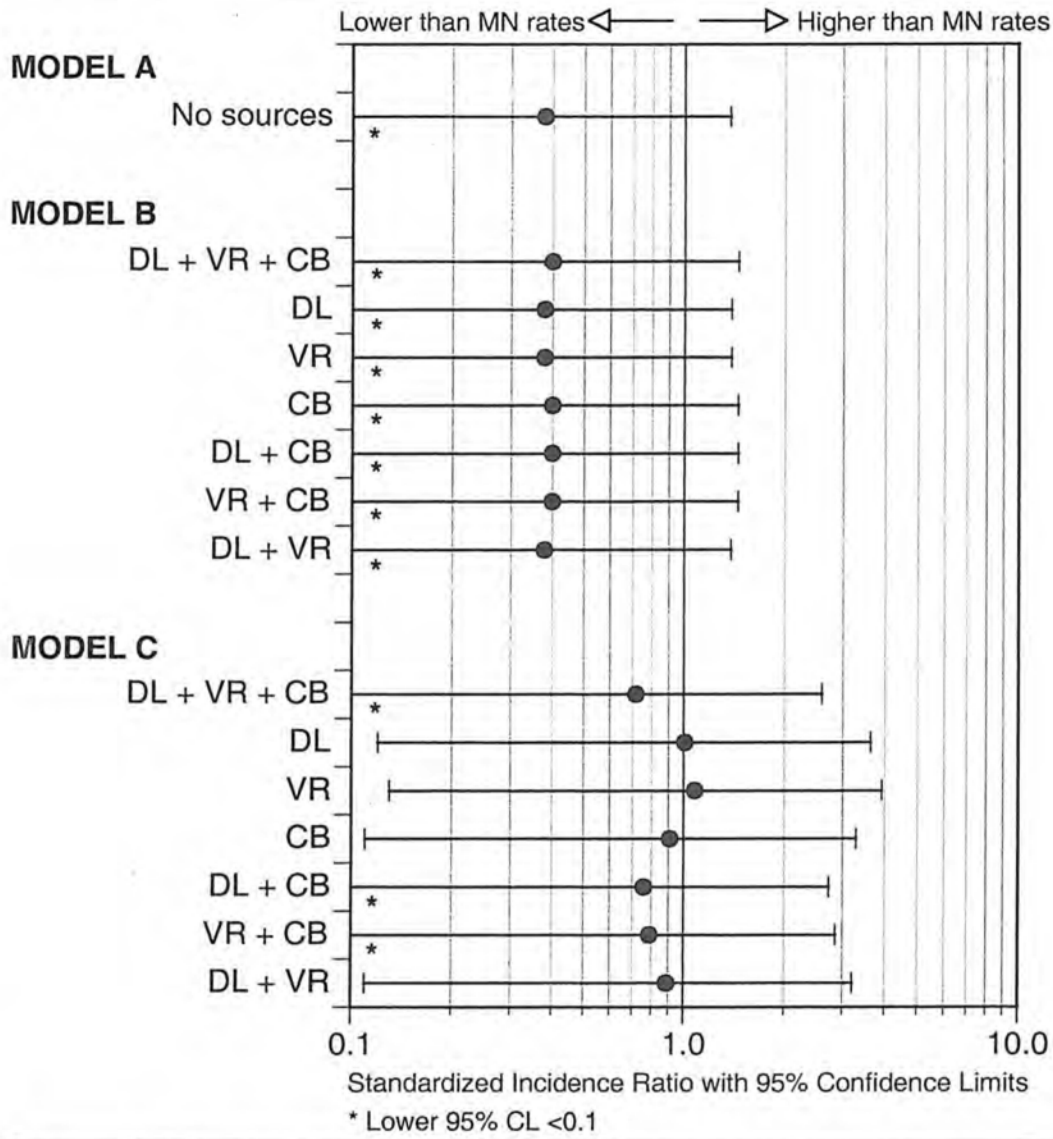


Information Sources:
DL=Drivers license records; VR=Voter registration records; CB=Credit bureau records

Residency Models:

- Model A assumes all were MN residents until the end of follow-up. No other sources were used.
- Model B assumes all were MN residents unless there was information to indicate otherwise.
- Model C assumes a 3% in- or out-of-state migration rate per year. (See text for explanation.)
- Models B and C also include residency information from cancer registry records and death records.

Figure 19. Mineral Board Worker Cohort Standardized Incidence Ratio (SIR) by Residency Model: Males, Pancreas

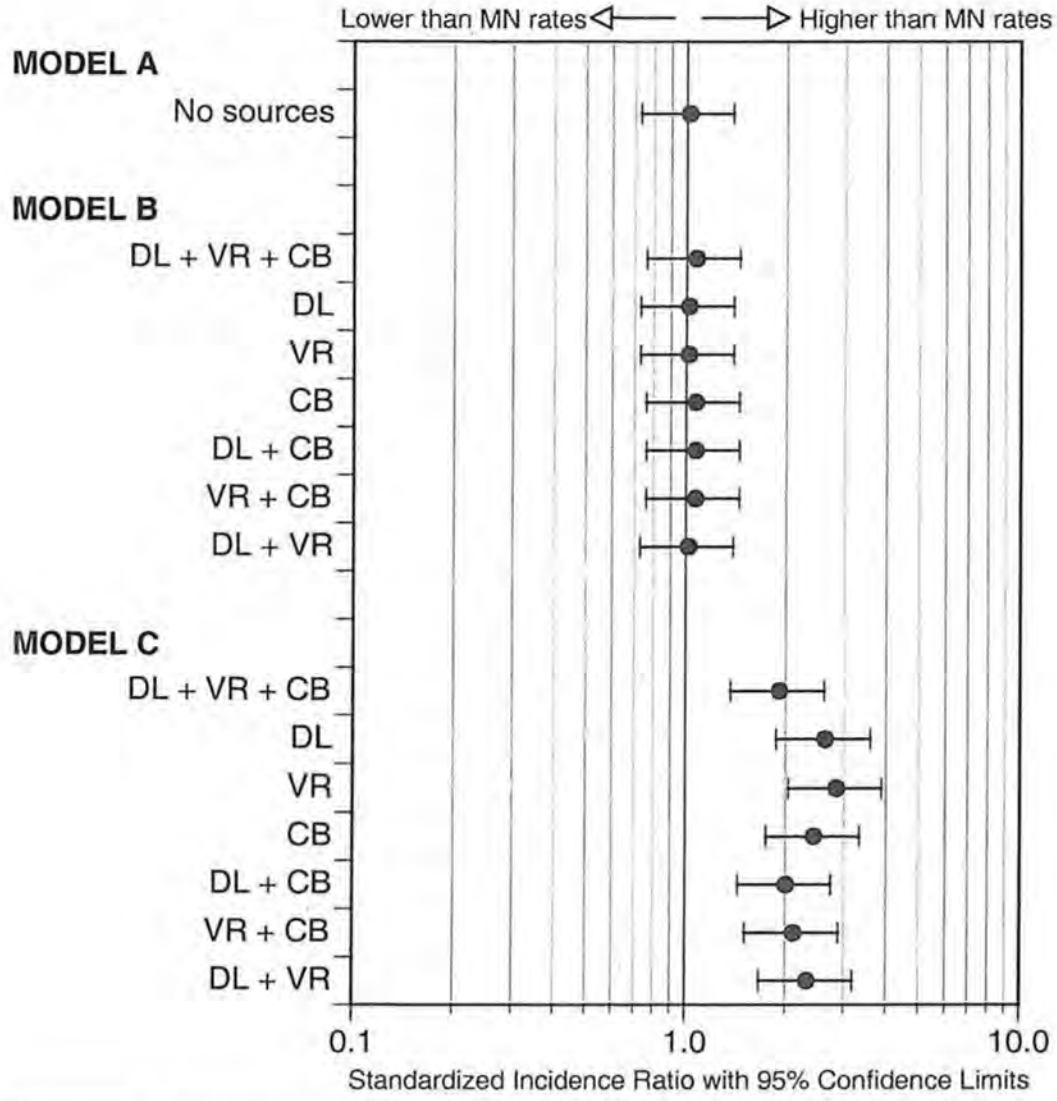


Information Sources:
DL=Drivers license records; VR=Voter registration records; CB=Credit bureau records

Residency Models:

- Model A assumes all were MN residents until the end of follow-up. No other sources were used.
- Model B assumes all were MN residents unless there was information to indicate otherwise.
- Model C assumes a 3% in- or out-of-state migration rate per year. (See text for explanation.)
- Models B and C also include residency information from cancer registry records and death records.

Figure 20. Mineral Board Worker Cohort Standardized Incidence Ratio (SIR) by Residency Model: Males, Lung and Bronchus

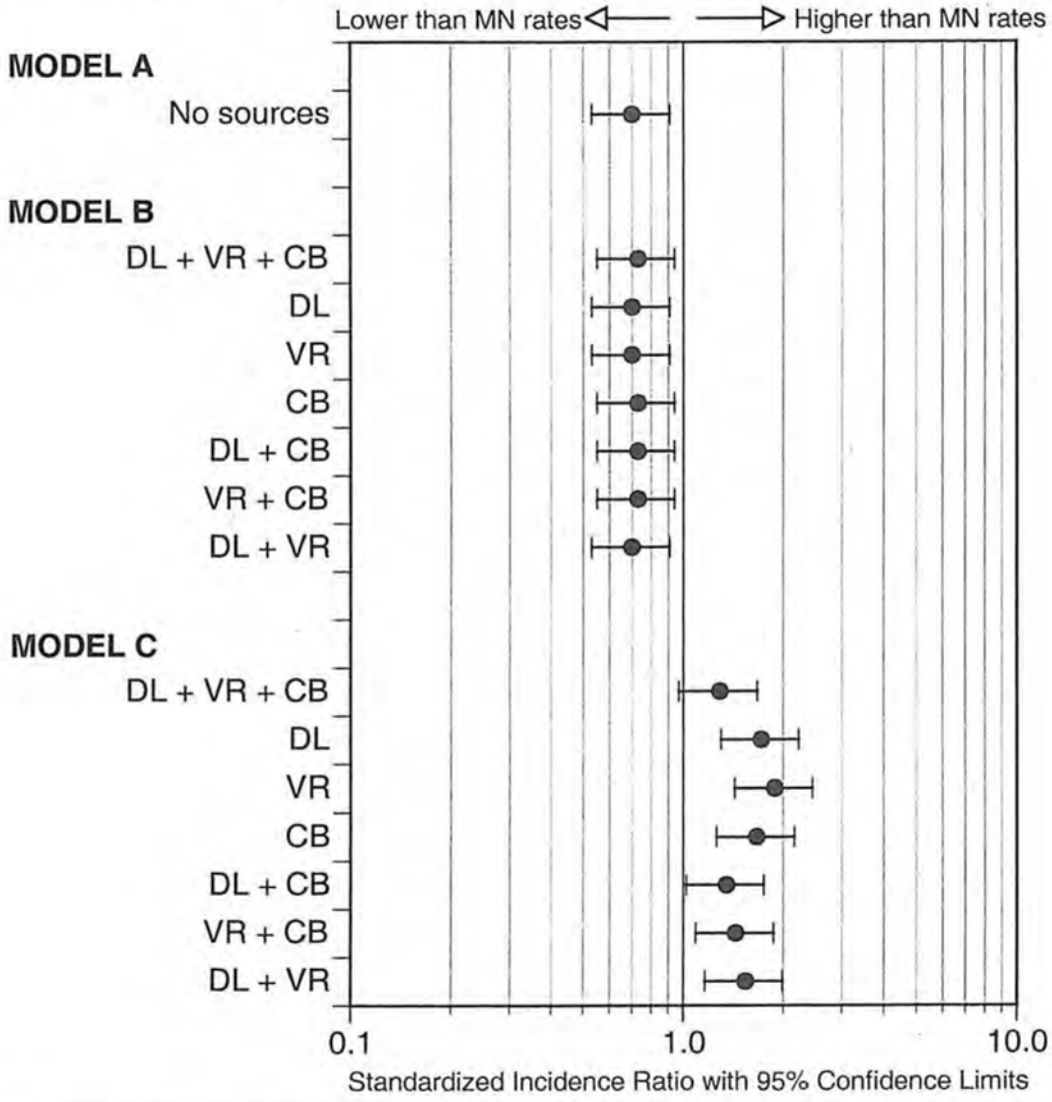


Information Sources:
DL=Drivers license records; VR=Voter registration records; CB=Credit bureau records

Residency Models:

- Model A assumes all were MN residents until the end of follow-up. No other sources were used.
- Model B assumes all were MN residents unless there was information to indicate otherwise.
- Model C assumes a 3% in- or out-of-state migration rate per year. (See text for explanation.)
- Models B and C also include residency information from cancer registry records and death records.

Figure 21. Mineral Board Worker Cohort Standardized Incidence Ratio (SIR) by Residency Model: Males, Prostate

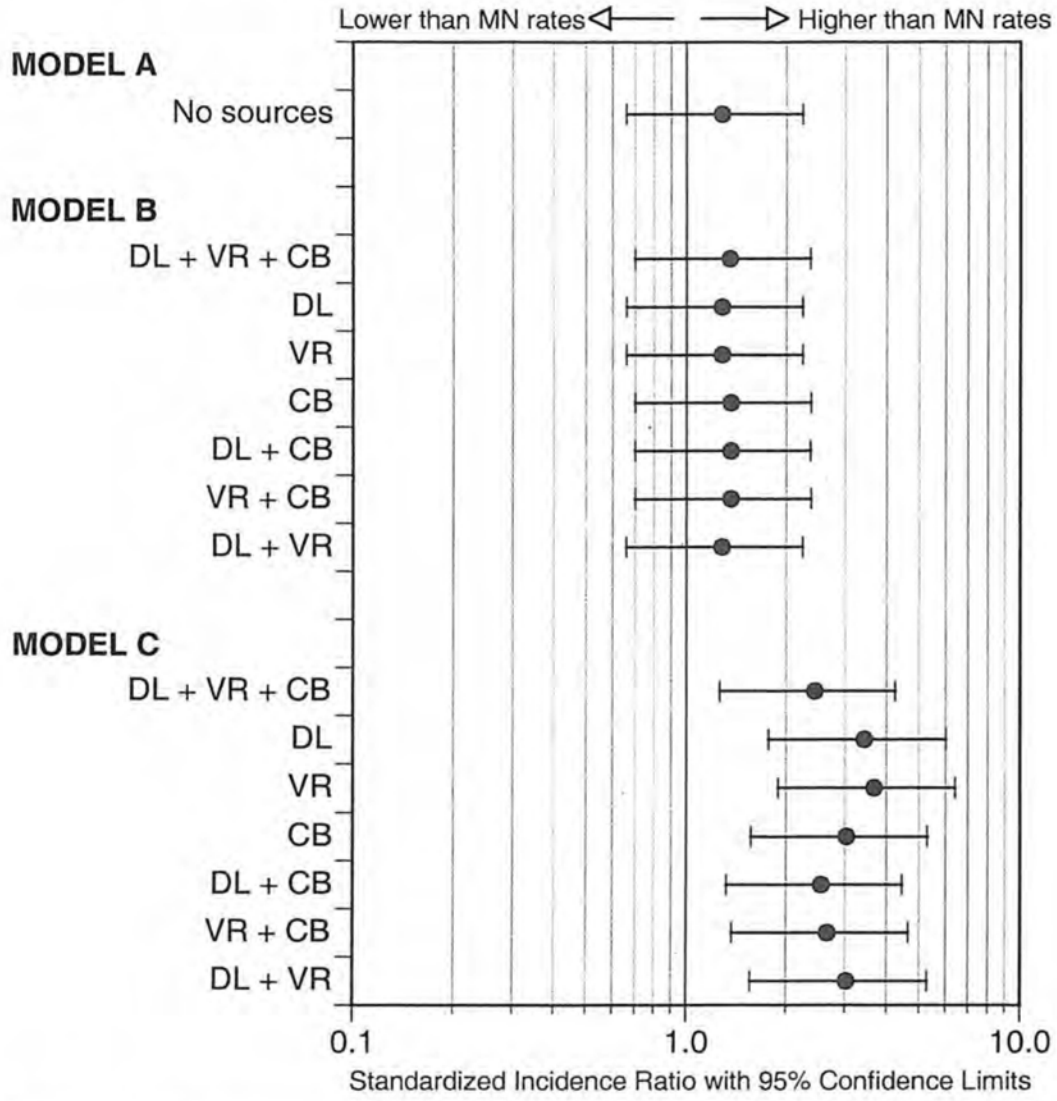


Information Sources:
DL=Drivers license records; VR=Voter registration records; CB=Credit bureau records

Residency Models:

- Model A assumes all were MN residents until the end of follow-up. No other sources were used.
- Model B assumes all were MN residents unless there was information to indicate otherwise.
- Model C assumes a 3% in- or out-of-state migration rate per year. (See text for explanation.)
- Models B and C also include residency information from cancer registry records and death records.

Figure 22. Mineral Board Worker Cohort Standardized Incidence Ratio (SIR) by Residency Model: Males, Kidney and Renal Pelvis

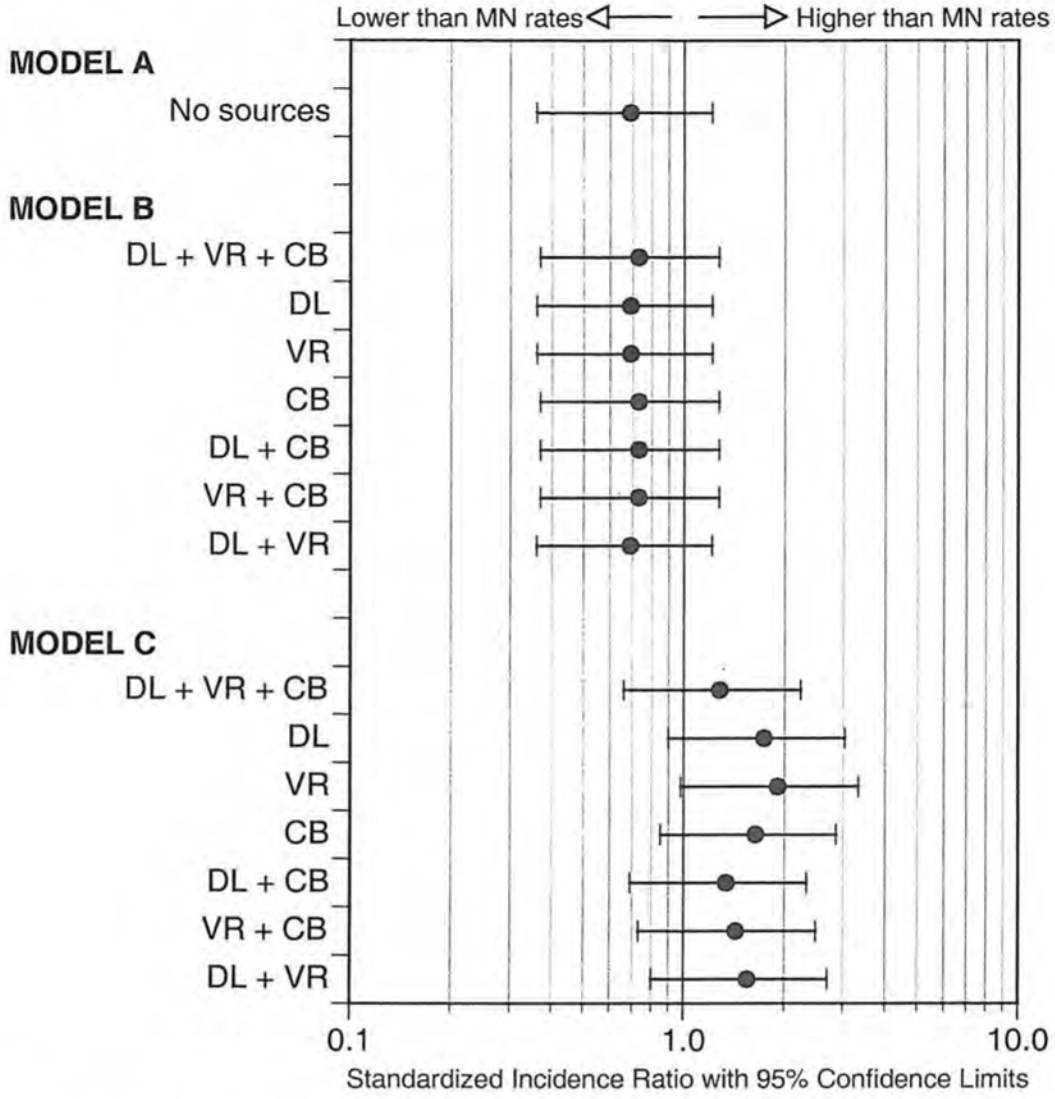


Information Sources:
DL=Drivers license records; VR=Voter registration records; CB=Credit bureau records

Residency Models:

- Model A assumes all were MN residents until the end of follow-up. No other sources were used.
- Model B assumes all were MN residents unless there was information to indicate otherwise.
- Model C assumes a 3% in- or out-of-state migration rate per year. (See text for explanation.)
- Models B and C also include residency information from cancer registry records and death records.

Figure 23. Mineral Board Worker Cohort Standardized Incidence Ratio (SIR) by Residency Model: Males, Urinary Bladder

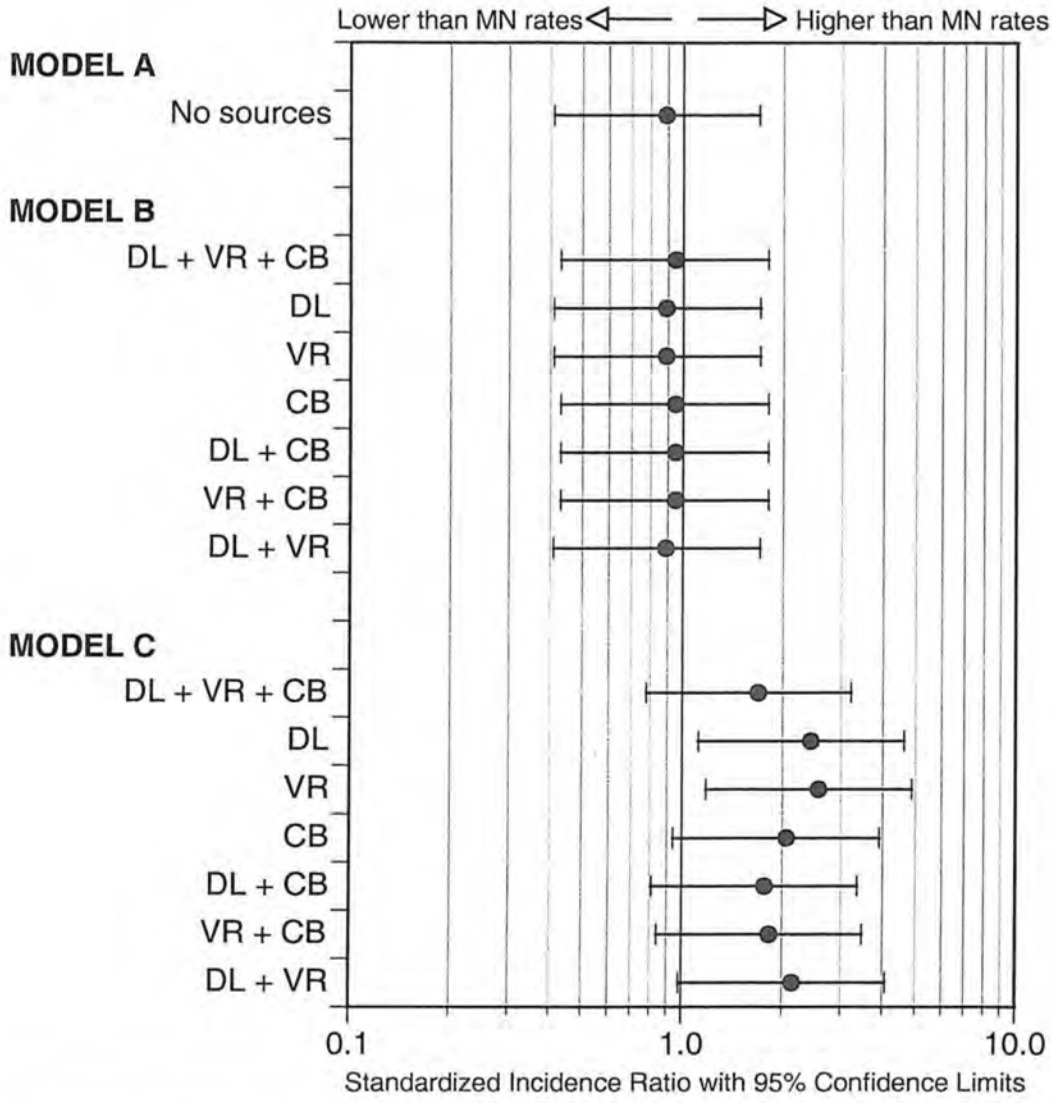


Information Sources:
DL=Drivers license records; VR=Voter registration records; CB=Credit bureau records

Residency Models:

- Model A assumes all were MN residents until the end of follow-up. No other sources were used.
- Model B assumes all were MN residents unless there was information to indicate otherwise.
- Model C assumes a 3% in- or out-of-state migration rate per year. (See text for explanation.)
- Models B and C also include residency information from cancer registry records and death records.

Figure 24. Mineral Board Worker Cohort Standardized Incidence Ratio (SIR) by Residency Model: Males, Melanomas of the Skin

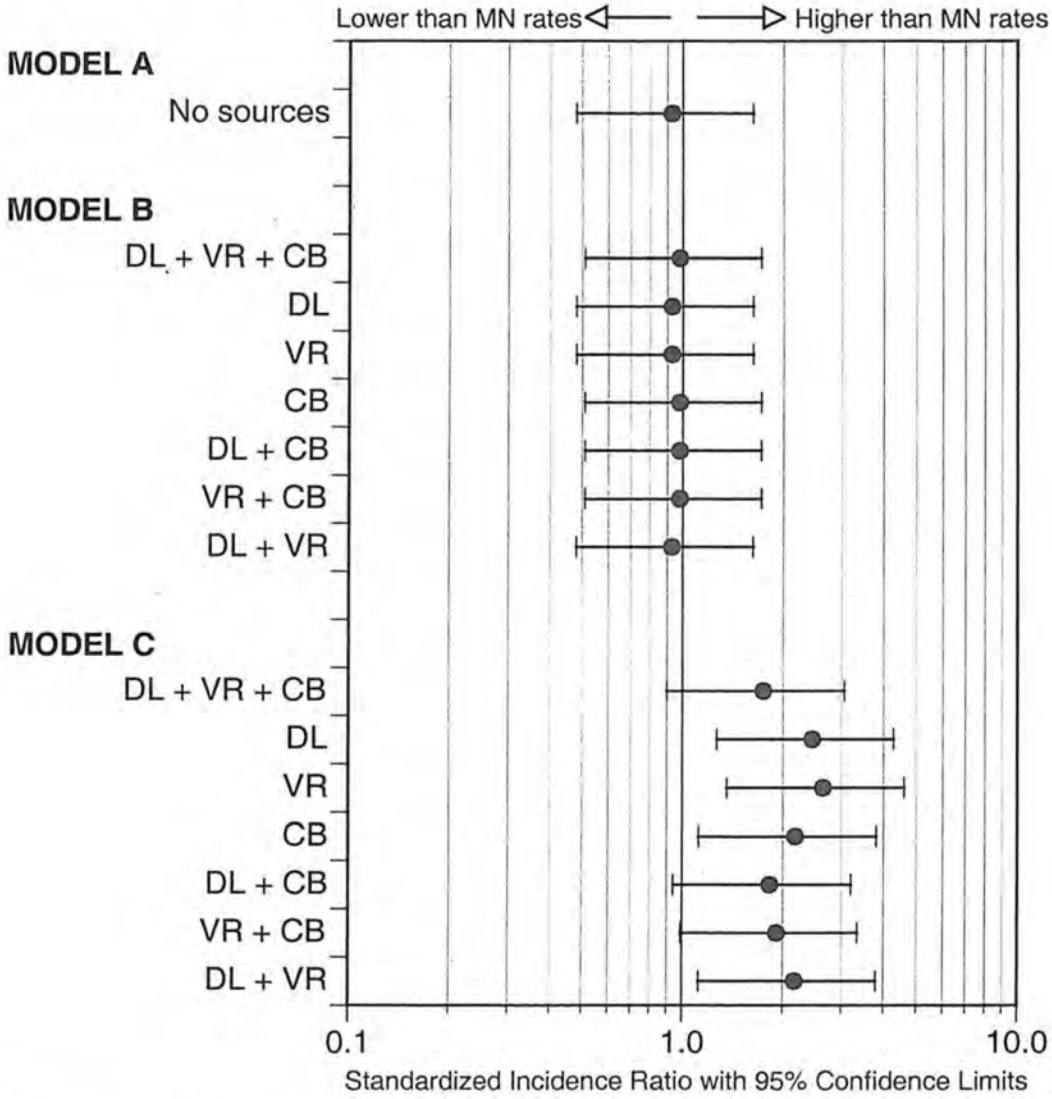


Information Sources:
DL=Drivers license records; VR=Voter registration records; CB=Credit bureau records

Residency Models:

- Model A assumes all were MN residents until the end of follow-up. No other sources were used.
- Model B assumes all were MN residents unless there was information to indicate otherwise.
- Model C assumes a 3% in- or out-of-state migration rate per year. (See text for explanation.)
- Models B and C also include residency information from cancer registry records and death records.

Figure 25. Mineral Board Worker Cohort Standardized Incidence Ratio (SIR) by Residency Model: Males, Non-Hodgkin's Lymphomas

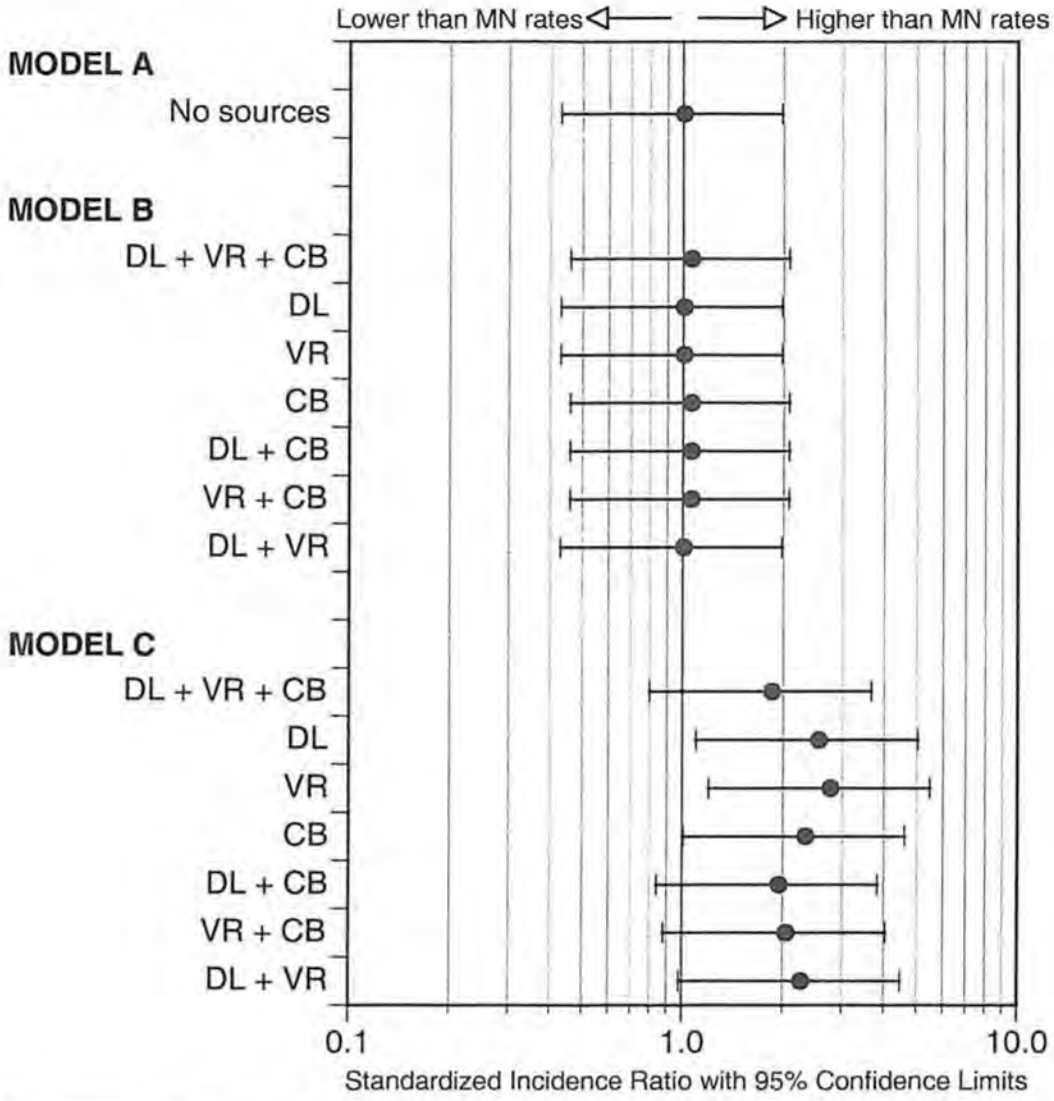


Information Sources:
DL=Drivers license records; VR=Voter registration records; CB=Credit bureau records

Residency Models:

- Model A assumes all were MN residents until the end of follow-up. No other sources were used.
- Model B assumes all were MN residents unless there was information to indicate otherwise.
- Model C assumes a 3% in- or out-of-state migration rate per year. (See text for explanation.)
- Models B and C also include residency information from cancer registry records and death records.

Figure 26. Mineral Board Worker Cohort Standardized Incidence Ratio (SIR) by Residency Model: Males, Leukemias



Information Sources:
DL=Drivers license records; VR=Voter registration records; CB=Credit bureau records

Residency Models:

- Model A assumes all were MN residents until the end of follow-up. No other sources were used.
- Model B assumes all were MN residents unless there was information to indicate otherwise.
- Model C assumes a 3% in- or out-of-state migration rate per year. (See text for explanation.)
- Models B and C also include residency information from cancer registry records and death records.

Table 19. Mineral Board Worker Cohort (Males) and Highway Worker Cohort (Males): Summary of Standardized Incidence Ratio Models A, B, and C

Cancer Model Type	Mineral Board Worker Cohort (Males)			Highway Worker Cohort (Males)		
	Observed	Expected	Ratio of Observed/ Expected	Observed	Expected	Ratio of Observed /Expected
Stomach						
Model A SIR	4	5.24	0.76	6	6.84	0.88
Model B SIR	4	4.99	0.80	6	6.71	0.89
Model C SIR	4	2.83	1.41	6	4.30	1.40
Colon						
Model A SIR	17	22.01	0.77	37	30.02	1.23
Model B SIR	17	21.00	0.81	37	29.45	1.26
Model C SIR	17	11.90	1.43	37	18.86	1.96*
Rectum & Rectosigmoid						
Model A SIR	6	10.45	0.57	14	12.69	1.10
Model B SIR	6	9.92	0.60	14	12.45	1.12
Model C SIR	6	5.54	1.08	14	8.13	1.72
Colorectal						
Model A SIR	23	32.46	0.71	51	42.71	1.19
Model B SIR	23	30.92	0.74	51	41.90	1.22
Model C SIR	23	17.44	1.32	51	26.99	1.89*
Pancreas						
Model A SIR	2	5.26	0.38	4	6.23	0.64
Model B SIR	2	4.99	0.40	4	6.11	0.65
Model C SIR	2	2.77	0.72	4	4.01	1.00
Lung & Bronchus						
Model A SIR	40	39.35	1.02	42	49.74	0.84
Model B SIR	40	37.49	1.07	42	48.79	0.86
Model C SIR	40	20.82	1.92*	42	31.87	1.32
Prostate						
Model A SIR	57	81.57	0.70*	95	119.74	0.79*
Model B SIR	57	78.28	0.73*	95	117.45	0.81*
Model C SIR	57	44.25	1.29	95	75.22	1.26*
Kidney & Renal Pelvis						
Model A SIR	12	9.36	1.28	9	10.19	0.88
Model B SIR	12	8.86	1.35	9	9.99	0.90
Model C SIR	12	4.92	2.44*	9	6.63	1.36

Table 19 (Continued). Mineral Board Worker Cohort (Males) and Highway Worker Cohort (Males): Summary of Standardized Incidence Ratio Models A, B, and C

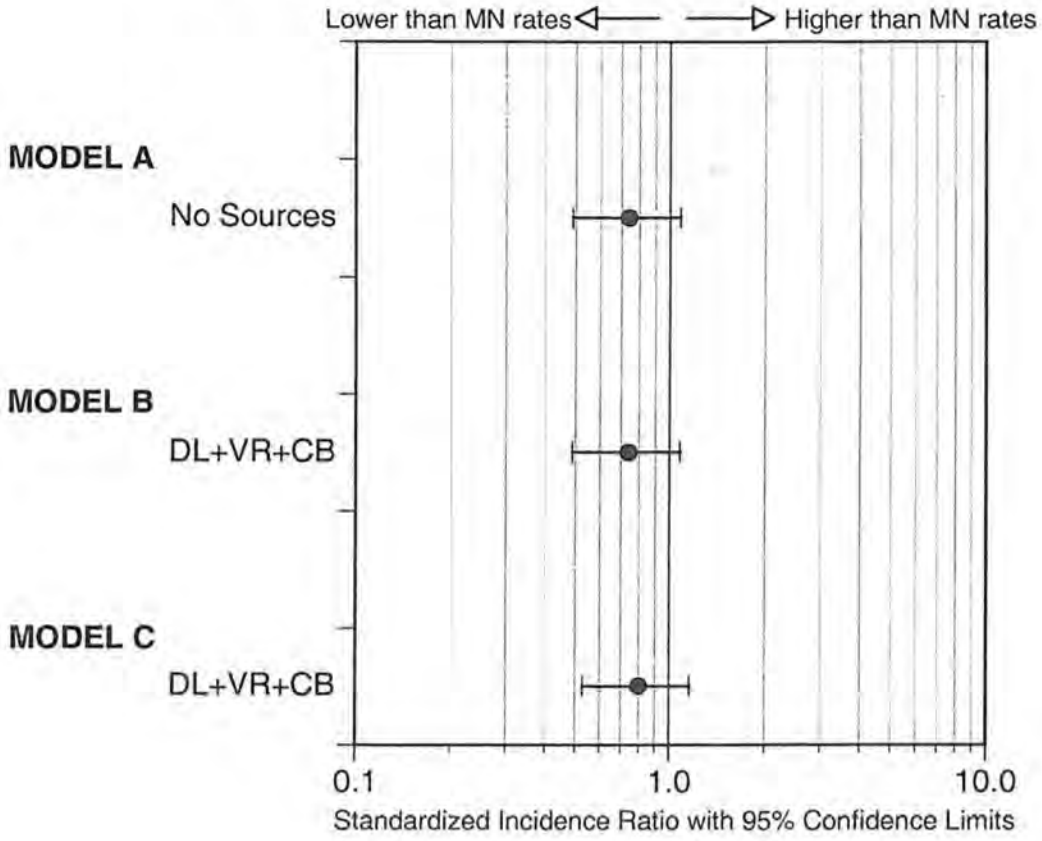
Cancer Model Type	Mineral Board Worker Cohort (Males)			Highway Worker Cohort (Males)		
	Observed	Expected	Ratio of Observed/ Expected	Observed	Expected	Ratio of Observed /Expected
Urinary Bladder						
Model A SIR	12	17.32	0.69	15	23.57	0.64
Model B SIR	12	16.53	0.73	15	23.13	0.65
Model C SIR	12	9.36	1.28	15	14.81	1.01
Melanomas						
Model A SIR	9	10.09	0.89	12	9.52	1.26
Model B SIR	9	9.50	0.95	12	9.34	1.29
Model C SIR	9	5.31	1.69	12	6.28	1.91
Non-Hodgkin's Lymphomas						
Model A SIR	12	12.88	0.93	15	14.53	1.03
Model B SIR	12	12.20	0.98	15	14.25	1.05
Model C SIR	12	6.85	1.75	15	9.35	1.60
Leukemias						
Model A SIR	8	7.96	1.01	7	9.94	0.70
Model B SIR	8	7.57	1.06	7	9.75	0.72
Model C SIR	8	4.30	1.86	7	6.31	1.11
All Cancers						
Model A SIR	209	274.19	0.76*	297	351.12	0.85*
Model B SIR	209	261.24	0.80*	297	344.44	0.86*
Model C SIR	209	146.92	1.42*	297	223.42	1.33*

*=$p < 0.05$

Residency Models:

- Model A assumes all cohort members were Minnesota (MN) residents until the end of follow-up. No other sources were used.
- Model B assumes all were MN residents unless there was information to indicate otherwise.
- Model C assumes a 3% in- or out-of-state migration rate per year.
- Models B and C also include residency information from drivers license records, voter registration records, credit bureau records, cancer registry records and death records.

Figure 27. Mineral Board Worker Cohort Standardized Incidence Ratio (SIR) by Residency Model: Females, All Cancers

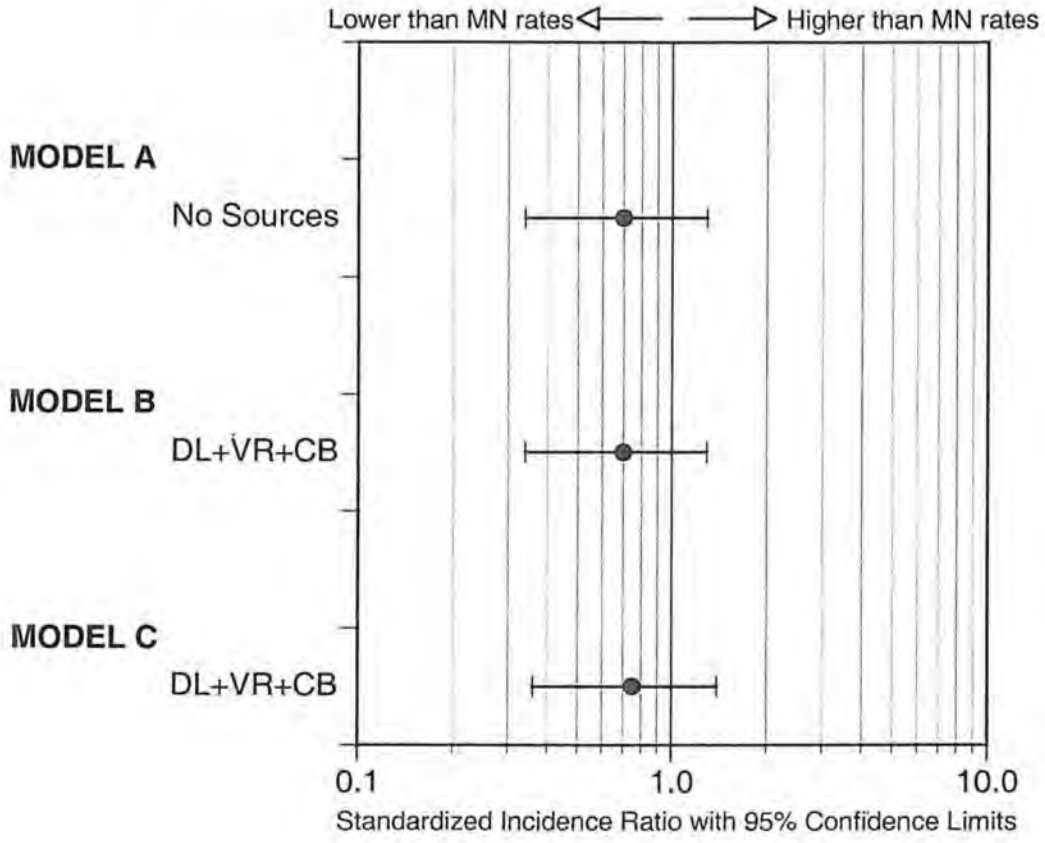


Information Sources:
DL=Drivers license records; VR=Voter registration records; CB=Credit bureau records

Residency Models:

- Model A assumes all were MN residents until the end of follow-up. No other sources were used.
- Model B assumes all were MN residents unless there was information to indicate otherwise.
- Model C assumes a 3% in- or out-of-state migration rate per year. (See text for explanation.)
- Models B and C also include residency information from cancer registry records and death records.

Figure 28. Mineral Board Worker Cohort Standardized Incidence Ratio (SIR) by Residency Model: Females, Breast

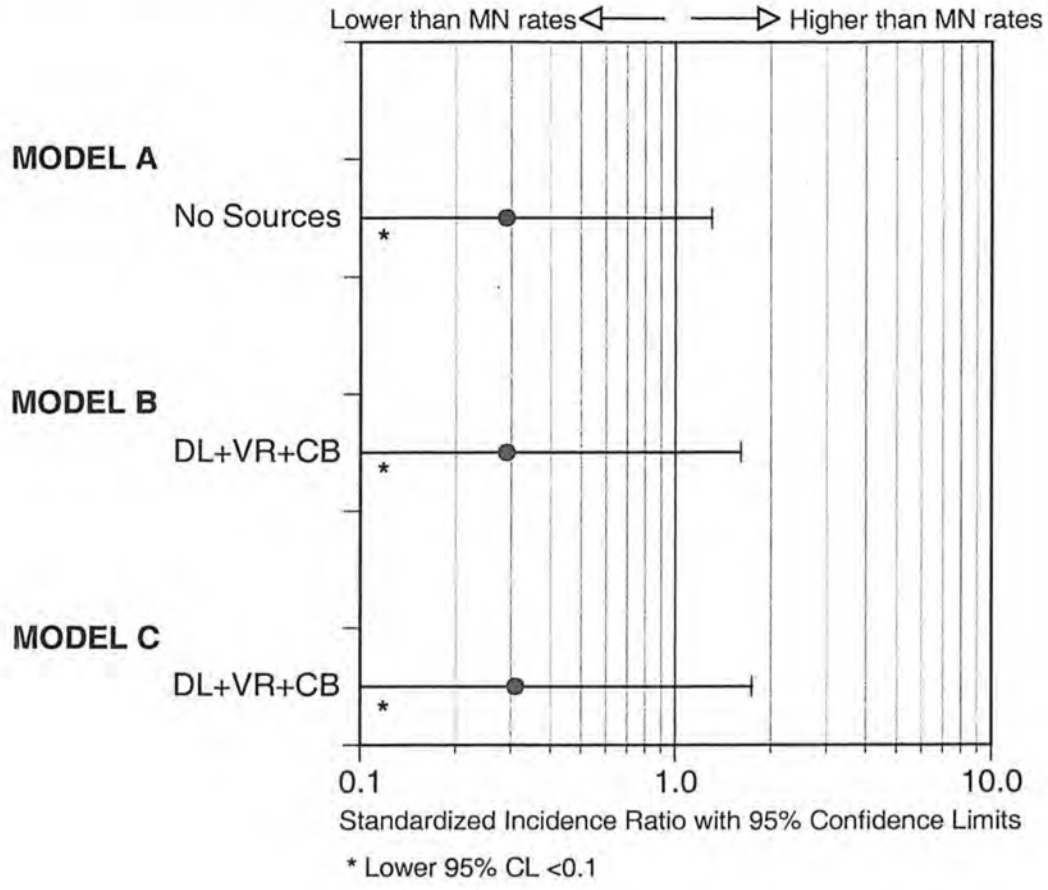


Information Sources:
DL=Drivers license records; VR=Voter registration records; CB=Credit bureau records

Residency Models:

- Model A assumes all were MN residents until the end of follow-up. No other sources were used.
- Model B assumes all were MN residents unless there was information to indicate otherwise.
- Model C assumes a 3% in- or out-of-state migration rate per year. (See text for explanation.)
- Models B and C also include residency information from cancer registry records and death records.

Figure 29. Mineral Board Worker Cohort Standardized Incidence Ratio (SIR) by Residency Model: Females, Lung



Information Sources:
DL=Drivers license records; VR=Voter registration records; CB=Credit bureau records

Residency Models:

- Model A assumes all were MN residents until the end of follow-up. No other sources were used.
- Model B assumes all were MN residents unless there was information to indicate otherwise.
- Model C assumes a 3% in- or out-of-state migration rate per year. (See text for explanation.)
- Models B and C also include residency information from cancer registry records and death records.

Table 20. Mineral Board Worker Cohort (Females): Summary of Standardized Mortality Ratio (SMR) and Standardized Incidence Ratio (SIR) Models

Cancer Model Type	Mineral Board Worker Cohort (Females) Observed	Expected	Ratio of Observed/Expected
Breast			
Model A SIR	10	14.25	0.70
Model B SIR	10	14.22	0.70
Model C SIR	10	13.27	0.75
Lung and Bronchus			
Model A SIR	1	3.46	0.29
Model B SIR	1	3.45	0.29
Model C SIR	1	3.20	0.31
All Cancers			
Model A SIR	27	36.38	0.74
Model B SIR	27	36.29	0.74
Model C SIR	27	33.78	0.80

Residency Models:

- Model A assumes all cohort members were Minnesota (MN) residents until the end of follow-up. No other sources were used.
- Model B assumes all were MN residents unless there was information to indicate otherwise.
- Model C assumes a 3% in- or out-of-state migration rate per year.
- Models B and C also include residency information from drivers license records, voter registration records, credit bureau records, cancer registry records and death records.

Minnesota Revenue Residency Information

Three different sets of assumptions (Models A, B, and C) about Minnesota residency status were tested to determine their effect on the total number of years as a Minnesota resident and the SIR. Although we believed that Model C was the closest to "truth" of the three models, we were unable to measure how it compared to any available "gold standard" without additional information.

The best source of residency status for occupational cohorts in Minnesota was thought to be Minnesota Revenue records. The Minnesota Department of Revenue had information about residency information on individuals who lived and/or filed a Minnesota tax return. Following a departmental review and subsequent approval of the study objectives and protocols, they did a count of the total number of workers who lived and/or filed a tax return for each of the years from 1988-1996. This was probably a slight undercount because (1) some individuals may not have filed a return; and (2) all counts were based on an exact social security number match. However, despite the slight undercount, it was probably quite close to the total number of cohort members living in Minnesota. Previous tracing efforts demonstrated that the Department of Revenue was the most complete source of tracing Minnesota residents and could find residents not found by other sources.

Tables 21 and 22 both show that Model C residency status totals closely reflect Minnesota Revenue estimates. Approximately eighty and sixty percent of the highway and mineral board worker cohorts, respectively, were determined to be Minnesota residents. Revenue estimates were higher in 1988 and gradually declined through 1996; the difference in the percentage of Minnesota residents between Model C and the Minnesota Revenue was somewhat higher in later years than in earlier years.

Although the percentage of highway workers with Minnesota residency declined from 1988-1996, a different pattern emerged for mineral board workers. This occurred in part because we were able to supplement the known residency status for highway workers with employment information in 1988 and 1989. This, in turn, affected the probabilities assigned for unknown residency years in Model C. Equivalent employment information was unavailable for the mineral board workers. Therefore, a comparatively lower percentage of mineral board workers were known to be Minnesota residents in the early years.

Table 21. Highway Worker Cohort (Males): Residency Status by Information Source and Year of Follow-Up

Model Type / Information Source	1988	1989	1990	1991	1992	1993	1994	1995	1996
Model A									
No Sources	3497	3425	3360	3297	3241	3179	3112	3054	3004
	(100)	(98)	(96)	(94)	(93)	(91)	(89)	(87)	(86)
Model B									
DL + VR + CB	3465	3374	3299	3223	3163	3101	3031	2976	2928
	(99)#	(96)	(94)	(92)	(90)	(89)	(87)	(85)	(84)
Model C									
DL + VR + CB	3007	2999	2919	2847	2822	2843	2888	2885	2837
	(86)	(86)	(83)	(81)	(81)	(81)	(83)	(82)	(81)
Minnesota Revenue Records									
SSN Match	2781	2735	2672	2624	2544	2509	2470	2417	2387
	(80)	(78)	(76)	(75)	(73)	(72)	(71)	(69)	(68)

Percent = Number of People Found / Number of Cohort members (N=3497) rounded to the nearest percent

Information Sources:
DL=Drivers license records; VR=Voter registration records; CB=Credit bureau records

Residency Models:
- Model A assumes all cohort members were Minnesota (MN) residents until the end of follow-up. No other sources were used.
- Model B assumes all were MN residents unless there was information to indicate otherwise.
- Model C assumes a 3% in- or out-of-state migration rate per year. (See text for explanation.)
- Models B and C also include residency information from cancer registry records and death records.

Table 22. Mineral Board Worker Cohort (Males & Females): Residency Status by Information Source and Year of Follow-Up

Model Type / Information Source	1988	1989	1990	1991	1992	1993	1994	1995	1996
Model A									
No Sources	5086	5039	5004	4961	4913	4868	4816	4774	4724
	(100)	(99)	(98)	(98)	(97)	(96)	(95)	(94)	(93)
Model B									
DL + VR + CB	4941	4748	4665	4611	4514	4463	4398	4354	4299
	(97)#	(93)	(92)	(91)	(89)	(88)	(86)	(86)	(85)
Model C									
DL + VR + CB	3012	3107	3133	3103	3135	3226	3323	3344	3281
	(59)	(61)	(62)	(61)	(62)	(63)	(65)	(66)	(65)
Minnesota Revenue Records									
SSN Match	3083	3093	3043	3022	2961	2911	2891	2848	2808
	(61)	(61)	(60)	(59)	(58)	(57)	(57)	(56)	(55)

Percent = Number of People Found / Number of Cohort members (N=5086) rounded to the nearest percent

Information Sources:
DL=Drivers license records; VR=Voter registration records; CB=Credit bureau records

Residency Models:
- Model A assumes all cohort members were Minnesota (MN) residents until the end of follow-up. No other sources were used.
- Model B assumes all were MN residents unless there was information to indicate otherwise.
- Model C assumes a 3% in- or out-of-state migration rate per year. (See text for explanation.)
- Models B and C also include residency information from cancer registry records and death records.

13. Comparison of Standardized Mortality Ratios (SMRs) and Standardized Morbidity Ratios (SIRs)

For all cancers combined, the ratio of the number of male workers diagnosed with a new pathology-based cancer compared to the number of cancer deaths was consistent with known population-based cancer statistics. Approximately twice as many cancers were identified from 1988-1996 through the cancer registry as were found on death certificates (i.e., National Death Index records; Minnesota death certificates). For male mineral board workers, there were 209 people diagnosed with cancer and 104 cancer deaths. For male highway workers, there were 297 people diagnosed with cancer and 153 cancer deaths. For female mineral board workers there were 27 cancer diagnoses and 11 cancer deaths, which is a ratio also proportionate to state and national data. Information depicting the SIR and SMR estimates for highway worker and mineral board cohorts are shown below in Figures 30-32 and Tables 23-24. Due to the low number of females in the mineral board worker cohort (N=736), only two cancers, lung and breast, as well as "all" cancers were evaluated.

Because of the increased number of cancer diagnoses, when compared with cancer deaths for the same time period, the confidence intervals for the SIR estimates were consistently narrower and thus more precise than the SMR estimates.

It was assumed that the occurrences, confidence intervals, and SMR, SIR estimates would be similar for cancers with very short relative survival rates. Among the ten cancers we evaluated, lung, pancreatic and stomach cancer had among the lowest survival rates of 13.3, 3.9, and 19.1 percent, respectively (5-year relative survival rate for white males, 1992-1997). Lung cancer estimates were nearly equivalent for highway workers and female mineral board workers, but SIRs were clearly higher than SMRs for male mineral board workers. Pancreatic cancer was relatively rare, so the confidence intervals for both SMR and SIR were quite broad. Although the actual point estimates were different, the confidence intervals overlapped. Stomach cancer estimates and confidence intervals were very similar in both cohorts, but the number of stomach cancer deaths among male mineral board workers, however, was very low (N=2) and resulted in a very wide confidence interval.

Lung and bronchus cancer SMR and SIR estimates were nearly equivalent for highway workers and female mineral board workers, but SIR estimates were clearly higher than SMR estimates for male mineral board workers. We examined the crossover between death certificate diagnosis of lung cancer and cancer registry diagnosis of lung cancer to see why these SIR and SMR estimates were different for male mineral board workers. There were 55 men identified with lung cancer from either source. Eighteen men were identified with lung cancer by both sources. Twelve men did not have any cancer diagnosis but were identified by death certificates as dying from lung cancer; 4 of these 12 men died out-of-state and of the remaining 8, the date of death was 1988 (2), 1989 (1), 1991 (2), 1992 (2), and 1995 (1). Fifteen men had a diagnosis of lung cancer by the cancer registry and had no date of death; 8 of these 15 men were known to have died after 1996 (from a partial Minnesota death certificate search) and the others were diagnosed in 1988 (1), 1989 (1), 1993 (1), 1995 (1) and 1996 (3). Seven men had a diagnosis of lung cancer by the cancer registry and the death certificate had a different cause of death (6 died from other endocrine and 1 died from non-cancer causes). Three men were diagnosed with prostate, penile,

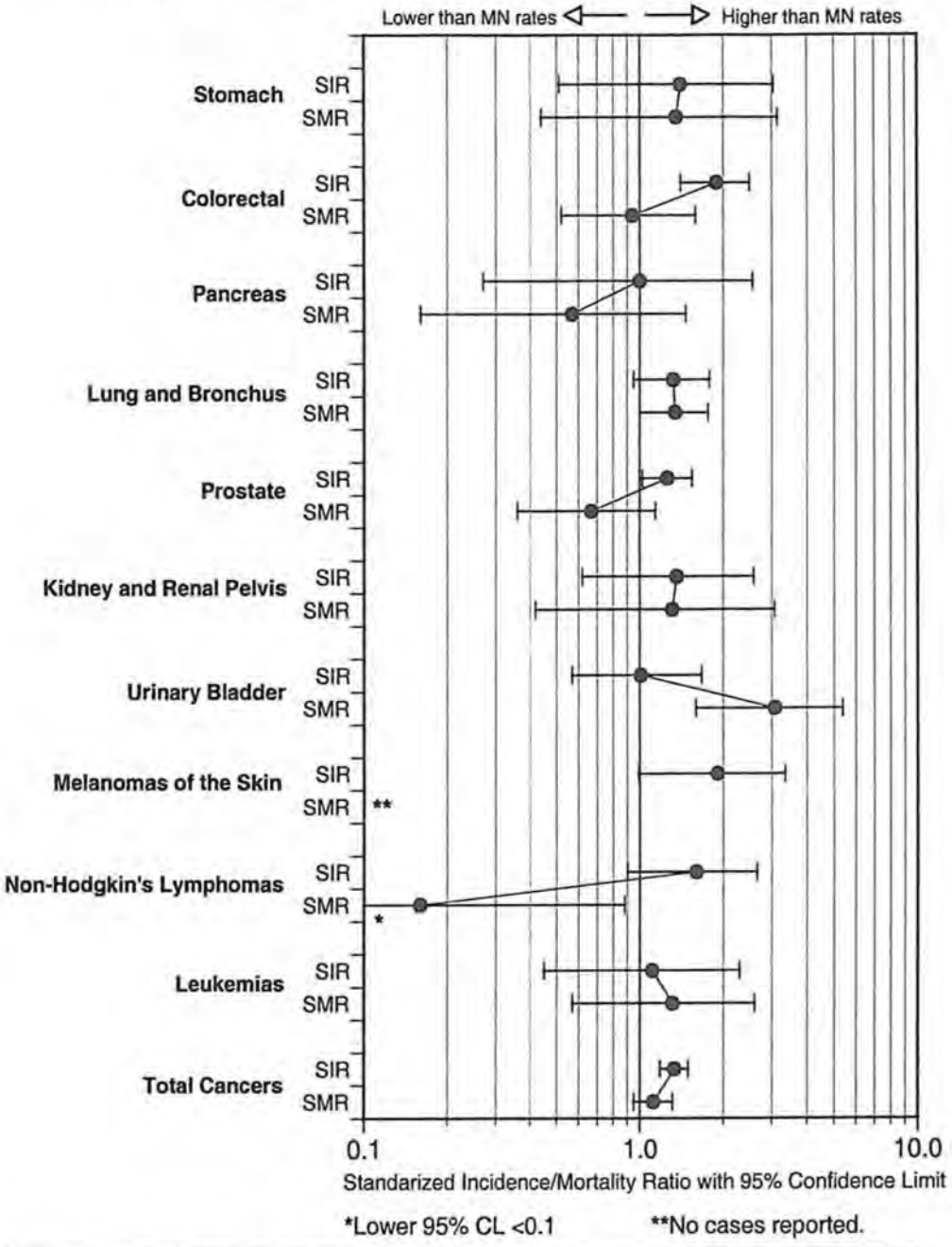
and other endocrine cancers by the cancer registry but the death certificate listed lung cancer as the cause of death.

It was assumed that occurrences, confidence intervals, and SMR/SIR estimates would be quite different for cancers with good, long-term survival such as prostate cancer, melanoma of the skin, urinary bladder cancer, and female breast cancer. Their respective 5-year relative survival rates for white males, 1992-1997 are 97, 86.6, and 83.9 percent. The 5-year relative survival rate for breast cancer in white females is 86.8 percent. As expected, the SIR for prostate cancer was higher than the SMR. Likewise, the SIR for melanomas of the skin was higher, although there were no deaths reported among male highway workers and only one death reported among male mineral board workers. The SIR for bladder cancer was higher than SMR for male mineral board workers, but the confidence interval fell within the confidence interval for the SMR. The opposite situation occurred among male highway workers with the SIR being much lower than the SMR. In addition, there was only a minimal degree of overlap between the SIR and SMR confidence intervals.

For female breast cancer, the SMR was higher than the SIR, but the confidence intervals were wide and largely overlapped.

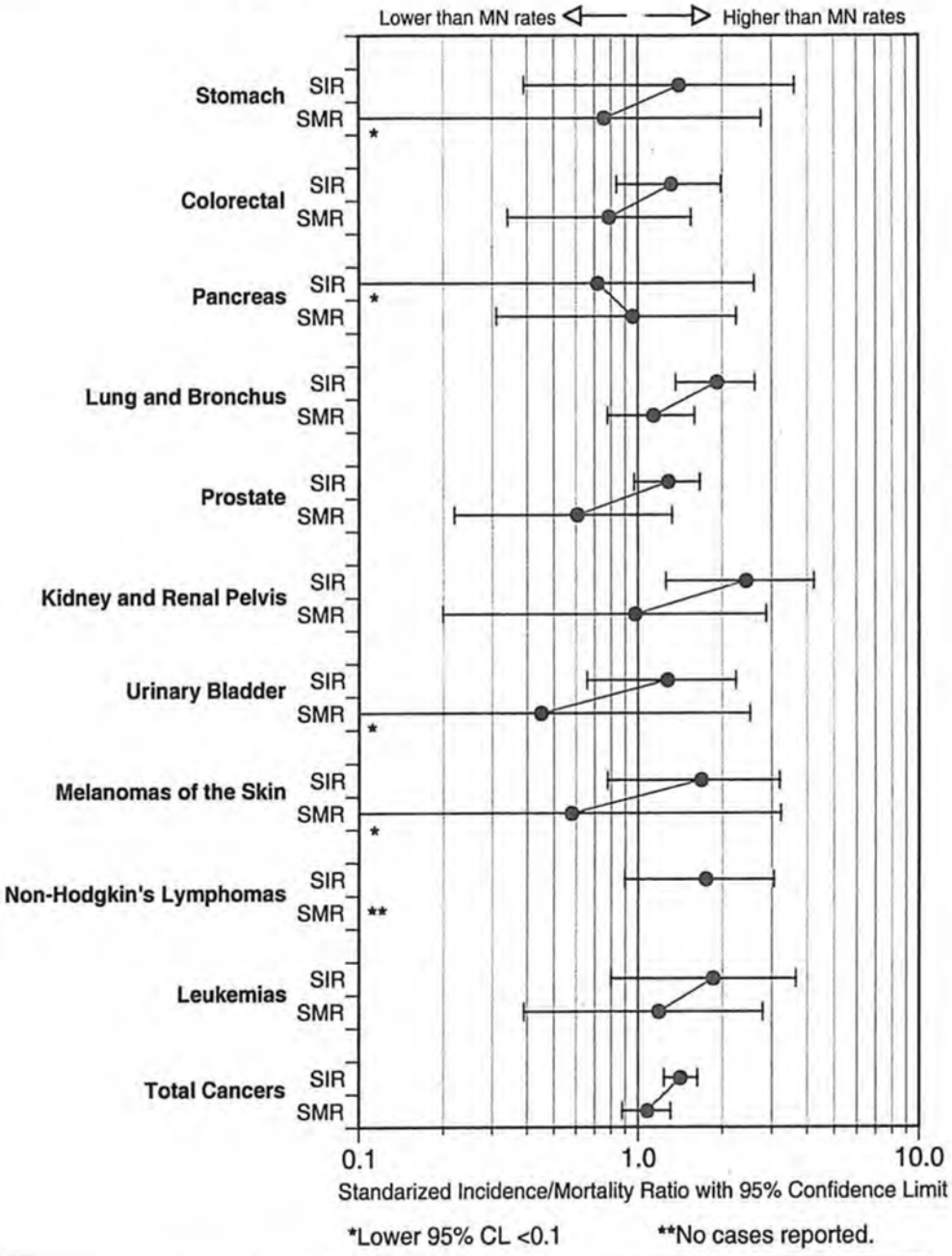
Differences were noted for which cancers the standardized mortality ratio (SMR) and standardized incidence ratio (SIR) reached the usual criteria of statistical significance ($p < 0.05$). Among male highway workers, colorectal, prostate, and "all" cancer SIRs were significantly elevated, while their corresponding SMRs were not significantly elevated. The opposite situation occurred for lung cancer, urinary bladder cancer, and non-hodgkin's lymphoma. Among male mineral board workers, lung cancer, kidney and renal pelvis cancer, and "all" cancer SIRs were significantly elevated, while their corresponding SMRs were not significantly elevated. There were no instances where the SMR was elevated significantly. Among female mineral board workers, none of the cancers were significantly different than their expectation.

Figure 30. Highway Worker Cohort Standardized Mortality Ratio (SMR) versus Standardized Incidence Ratio (SIR): Males, Selected Cancers.



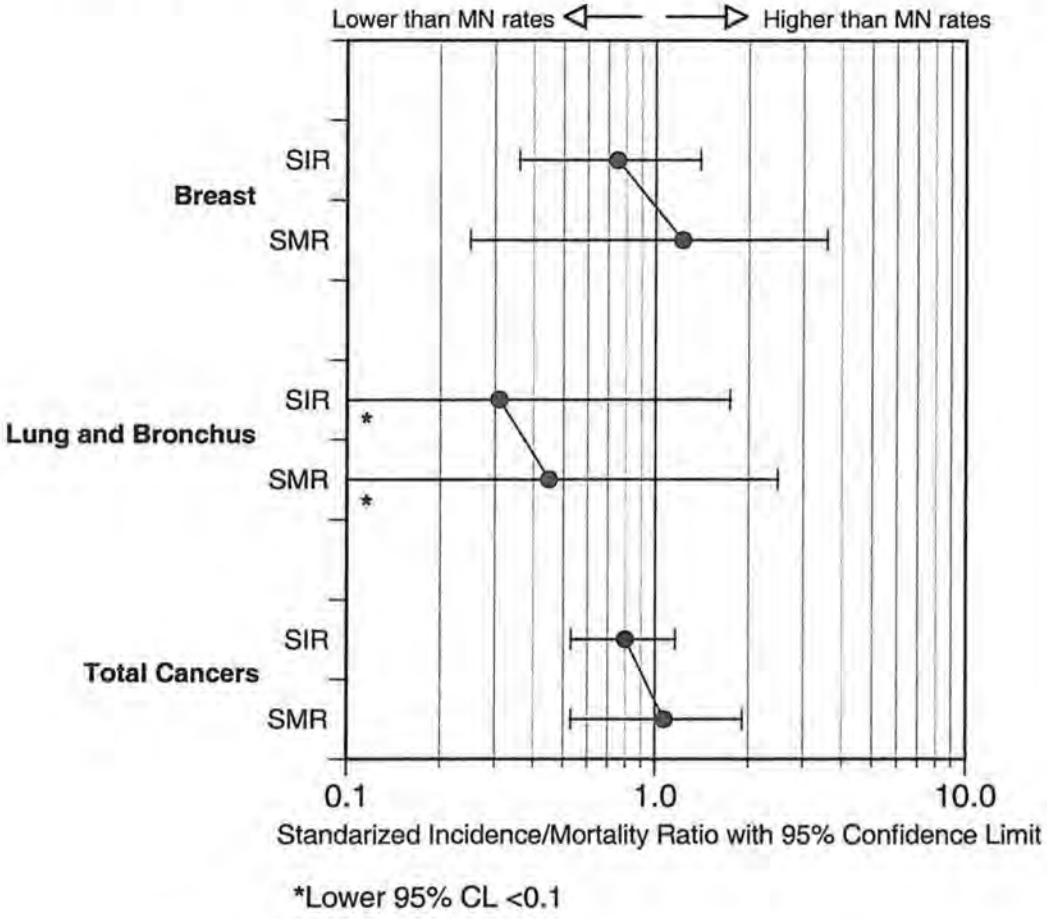
SIRs calculated using all information sources (Drivers license, Voter registration and Credit bureau records) and Residency Model C, which assumes a 3% in- or out-of-state migration rate per year. Model C also includes residency information from cancer registry records and death records.

Figure 31. Mineral Board Worker Cohort Standardized Mortality Ratio (SMR) versus Standardized Incidence Ratio (SIR): Males, Selected Cancers



SIRs calculated using all information sources (Drivers license, Voter registration and Credit bureau records) and Residency Model C, which assumes a 3% in- or out-of-state migration rate per year. Model C also includes residency information from cancer registry records and death records.

Figure 32. Mineral Board Worker Cohort Standardized Mortality Ratio (SMR) versus Standardized Incidence Ratio (SIR): Females, Selected Cancers



SIRs calculated using all information sources (Drivers license, Voter registration and Credit bureau records) and Residency Model C, which assumes a 3% in- or out-of-state migration rate per year. Model C also includes residency information from cancer registry records and death records.

Table 23. Mineral Board Worker Cohort (Males) and Highway Worker Cohort (Males): Summary of Standardized Mortality Ratio and Standardized Incidence Ratio Model C

Cancer Analysis Type	Mineral Board Worker Cohort			Highway Worker Cohort		
	Observed	Expected	Ratio of Observed/ Expected	Observed	Expected	Ratio of Observed /Expected
Stomach						
SMR	2	2.62	0.76	5	3.72	1.35
Model C SIR	4	2.83	1.41	6	4.30	1.40
Colorectal						
SMR	8	10.16	0.79	14	14.83	0.94
Model C SIR	23	17.44	1.32	51	26.99	1.89*
Pancreas						
SMR	5	5.19	0.96	4	6.99	0.57
Model C SIR	2	2.77	0.72	4	4.01	1.00
Lung & Bronchus						
SMR	33	29.07	1.14	53	39.48	1.34*
Model C SIR	40	20.82	1.92*	42	31.87	1.32
Prostate						
SMR	6	9.79	0.61	13	19.44	0.67
Model C SIR	57	44.25	1.29	95	75.22	1.26*
Kidney & Renal Pelvis						
SMR	3	3.05	0.98	5	3.82	1.31
Model C SIR	12	4.92	2.44*	9	6.63	1.36

Table 23 (Continued). Mineral Board Worker Cohort (Males) and Highway Worker Cohort (Males): Summary of Standardized Mortality Ratio and Standardized Incidence Ratio Model C

Cancer Analysis Type	Mineral Board Worker Cohort Observed	Expected	Ratio of Observed/ Expected	Highway Worker Cohort Observed	Expected	Ratio of Observed /Expected
Urinary Bladder						
SMR	1	2.21	0.45	12	3.90	3.08*
Model C SIR	12	9.36	1.28	15	14.81	1.01
Melanomas						
SMR	1	1.72	0.58	0	1.80	--
Model C SIR	9	5.31	1.69	12	6.28	1.91
Non-Hodgkin's Lymphomas						
SMR	0	4.62	--	1	6.31	0.16*
Model C SIR	12	6.85	1.75	15	9.35	1.60
Leukemias						
SMR	5	4.19	1.19	8	6.09	1.31
Model C SIR	8	4.30	1.86	7	6.31	1.11
All Cancers						
SMR	104	96.11	1.08	153	136.80	1.12
Model C SIR	209	146.92	1.42*	297	223.42	1.33*

*=$p < 0.05$

Model C assumes a 3% in- or out-of-state migration rate per year and uses all information sources (i.e., drivers license records, voter registration records, credit bureau records, cancer registry records and death records).

Table 24. Mineral Board Worker Cohort (Females): Summary of Standardized Mortality Ratio (SMR) and Standardized Incidence Ratio (SIR) Model C

Cancer Analysis Type	Mineral Board Worker Cohort (Females) Observed	Expected	Ratio of Observed/Expected
Breast			
SMR	3	2.45	1.22
Model C SIR	10	13.27	0.75
Lung and Bronchus			
SMR	1	2.23	0.45
Model C SIR	1	3.20	0.31
All Cancers			
SMR	11	10.31	1.07
Model C SIR	27	33.78	0.80

Model C assumes a 3% in- or out-of-state migration rate per year and uses all information sources (i.e., drivers license records, voter registration records, credit bureau records, cancer registry records and death records).

14. Stratified Analysis of Selected Cancers, Standardized Mortality Ratio and Standardized Incidence Ratios: Packyears of Smoking, 30-Year Employment Latency, Completeness of Residency Status

Although the total number of male workers diagnosed with a new, incident cancer compared to cancer deaths seemed consistent with known population-based cancer statistics, we wanted to check what the effect of smoking would be on lung cancer, prostate cancer and all cancers. Because exposure information was available for a subcohort of male mineral board workers that had participated in a voluntary health screening (N=874) a stratified analysis examining the impact of smoking was possible. Figure 33 shows that workers with greater than 20 packyears of smoking had elevated SMR and SIR lung cancer estimates. There was a much smaller difference noted for prostate cancer and a slightly greater difference noted for all cancers.

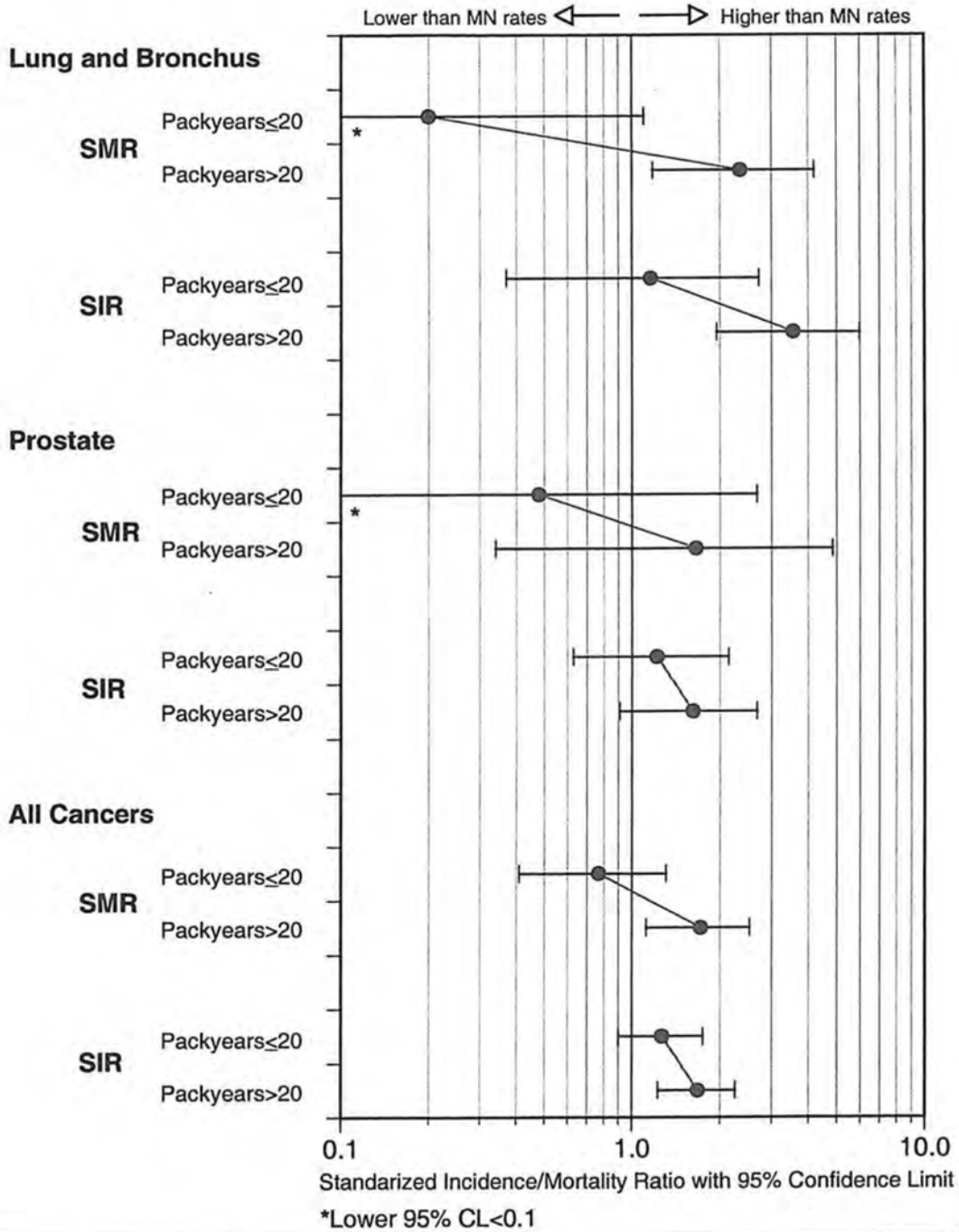
It was unclear how the subcohort of screened male mineral board workers compared with the full cohort of workers. A stratified analysis was conducted for lung cancer, prostate cancer and all cancers to examine potential differences between these two groups of men for workers with at least 30 years of latency (i.e., date of first employment to first date of follow-up (1988)). Figures 34 and 35 show that the SMR and SIR estimates for lung cancer went in different directions for the two groups; the workers with at least 30 years of latency had higher SMR and SIR estimates within the screened subcohort, but had lower estimates for the entire cohort. The confidence intervals for the two stratum (i.e., start year $\leq$ 1958 and start year > 1958) however, were quite wide for the workers in the latter group and basically encompassed those for the group with longer latency. The SMR for prostate cancer was elevated for the screened subcohort when compared to all males, but the SIR estimates were not qualitatively different due to the wide confidence intervals. The SMR and SIR stratum were fairly similar for all cancers.

Because of the problems associated with making assumptions about residency status, it might be tempting to conduct SIR studies using just the subset of individuals who have a complete residency history (i.e., all the years, 1988-1996, were determined to be Minnesota or non-Minnesota residency status through record-linkages with information sources). Also, it might be possible to conduct an SIR analysis using just the subset of individuals who have some residency information (i.e., some of the years, 1988-1996, were determined to be Minnesota or non-Minnesota residency status through record-linkages with information sources). A stratified analysis was completed to determine how large an effect these subset analyses would have on the SMR or SIR for lung cancer, prostate cancer, and all cancers for male highway workers.

There were several reasons why the highway workers found in the subset with (1) complete versus incomplete residency or (2) any residency versus no residency information might be quite different. First, workers who were employed in 1988 and 1989 were different than those who had retired or resigned at an earlier date. The more recent employees would have fallen into either the "complete" or "any" residency information categories. Second, workers who died would have been in either the "complete" or "any" residency information categories. If a worker died in 1988, he would have automatically had "complete" residency information. Third, the years of residency in Minnesota are clearly affected by the use of different modeling procedures to complete the unknown residency status. This would likely be worse for the subset of men with no residency information as compared to those with incomplete residency information. Lastly,

older age and older date of last employment corresponded with fewer matches among the information sources. Younger men who tended to work longer were more likely to be among the "complete" or "any" residency information categories. Figures 36 to 38 compare the SMR and SIR estimates for this stratified analysis. The differences among these subsets definitely affected the resulting estimates; the SIR and SMR estimates for the two strata are, for the most part, quite different.

Figure 33. Mineral Board Worker Cohort Stratified Analysis by Packyears of Smoking: Screened Subcohort, Males, Selected Cancers



SIRs calculated using all information sources (Drivers license, Voter registration and Credit bureau records) and Residency Model C, which assumes a 3% in- or out-of-state migration rate per year. Model C also includes residency information from cancer registry records and death records.

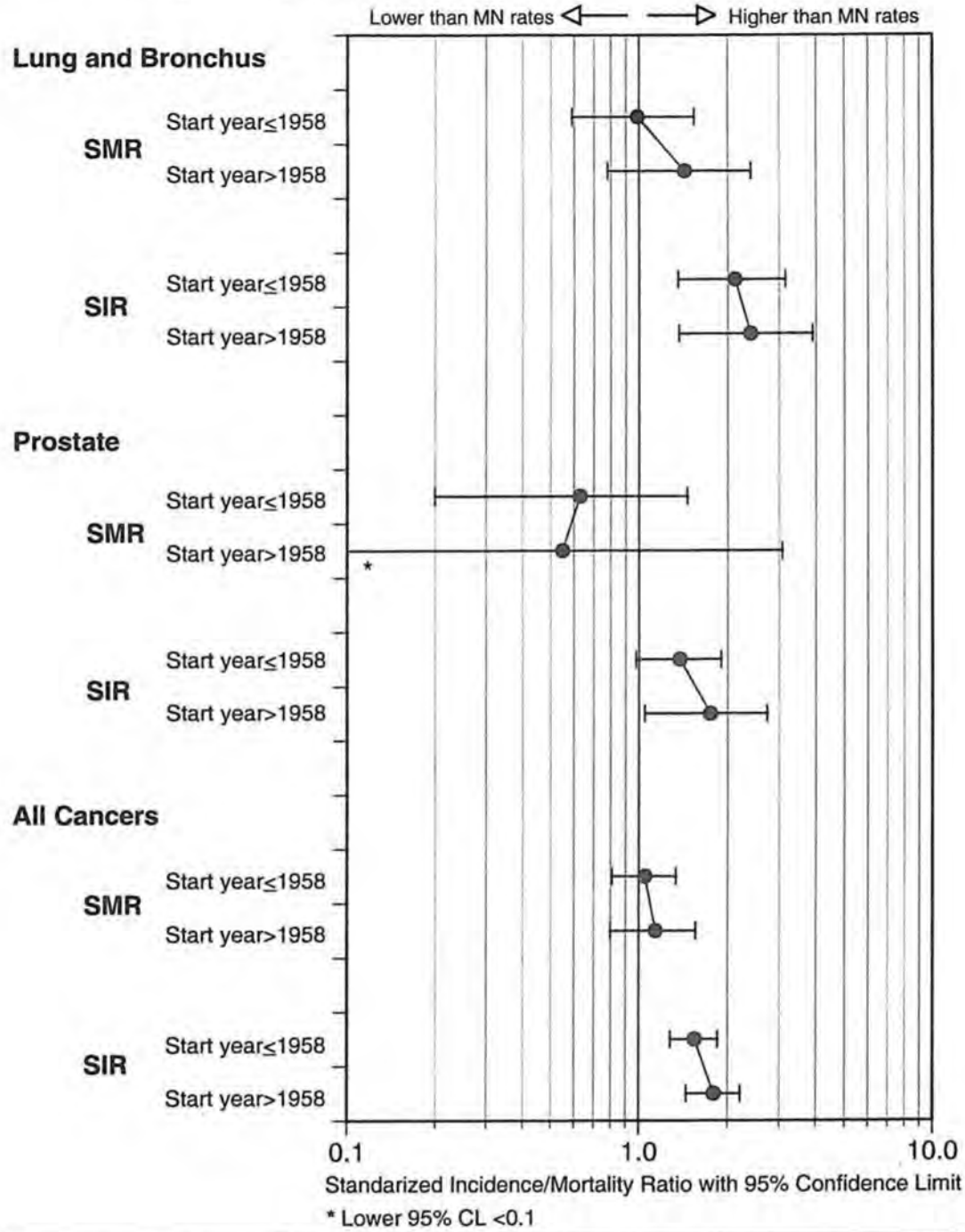
Table 25. Mineral Board Worker Cohort (Males) Stratified Analysis by Packyears of Smoking: Screened Subcohort, Males, Selected Cancers

Cancer Analysis Type Exposure Type	Mineral Board Worker Screened Subcohort (Males) Observed	Expected	Ratio of Observed/Expected
Lung and Bronchus			
SMR			
Packyears ≤ 20	1	5.07	0.2
Packyears > 20	11	4.67	2.35*
SIR			
Packyears ≤ 20	5	4.32	1.16
Packyears > 20	14	3.92	3.57*
Prostate			
SMR			
Packyears ≤ 20	1	2.08	0.48
Packyears > 20	3	1.81	1.66
SIR			
Packyears ≤ 20	12	9.80	1.22
Packyears > 20	15	9.25	1.62
All Cancers			
SMR			
Packyears ≤ 20	13	16.97	0.77
Packyears > 20	26	15.11	1.72*
SIR			
Packyears ≤ 20	38	29.89	1.27
Packyears > 20	45	26.77	1.68*

*$p < 0.05$

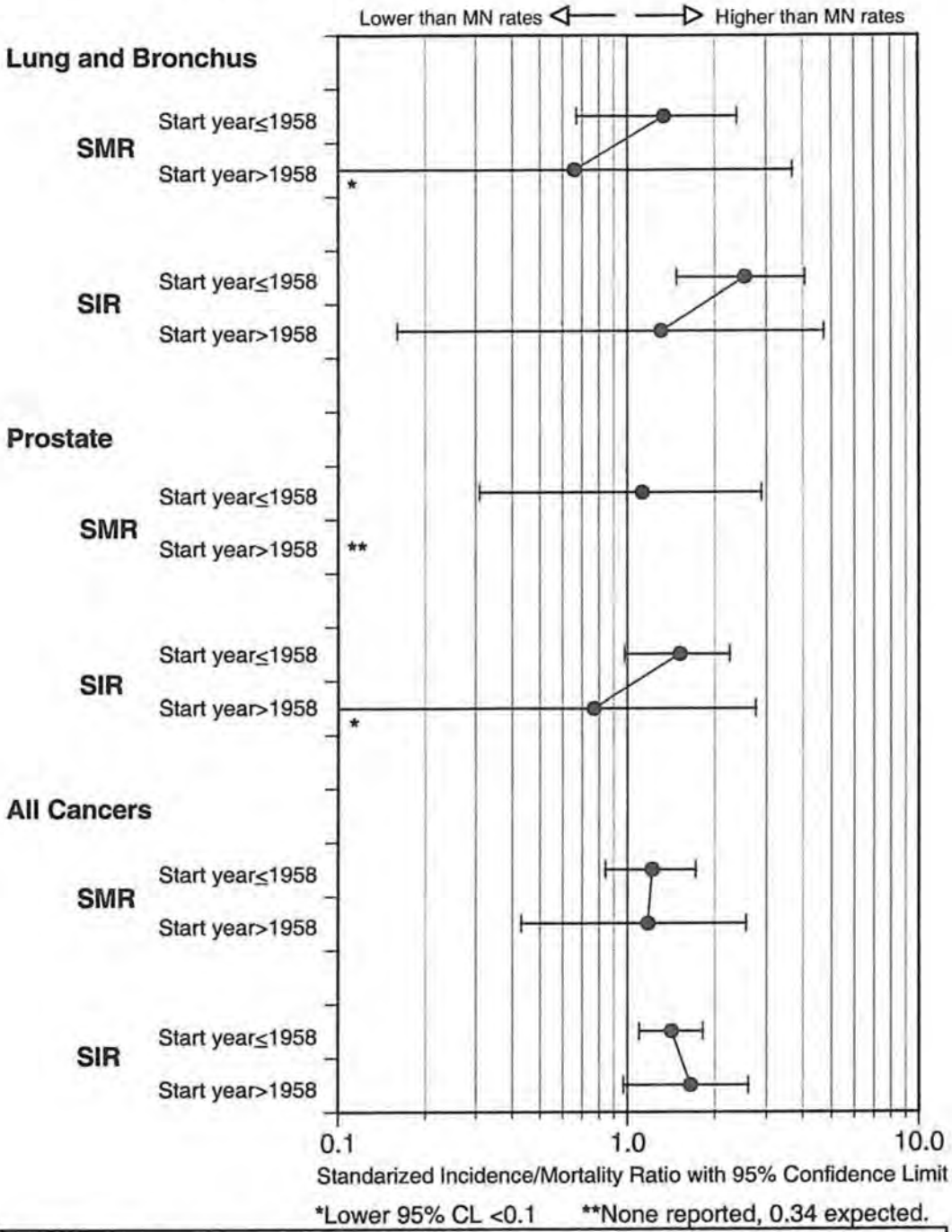
Model C SIR assumes a 3% in- or out-of-state migration rate per year and uses all information sources (i.e., drivers license records, voter registration records, credit bureau records, cancer registry records, and death records).

Figure 34. Mineral Board Worker Cohort Stratified Analysis by 30-Year Employment Latency: Males, Selected Cancers



SIRs calculated using all information sources (Drivers license, Voter registration and Credit bureau records) and Residency Model C, which assumes a 3% in- or out-of-state migration rate per year. Model C also includes residency information from cancer registry records and death records.

Figure 35. Mineral Board Worker Cohort Stratified Analysis by 30-Year Employment Latency: Screened Subcohort, Males, Selected Cancers



SIRs calculated using all information sources (Drivers license, Voter registration and Credit bureau records) and Residency Model C, which assumes a 3% in- or out-of-state migration rate per year. Model C also includes residency information from cancer registry records and death records.

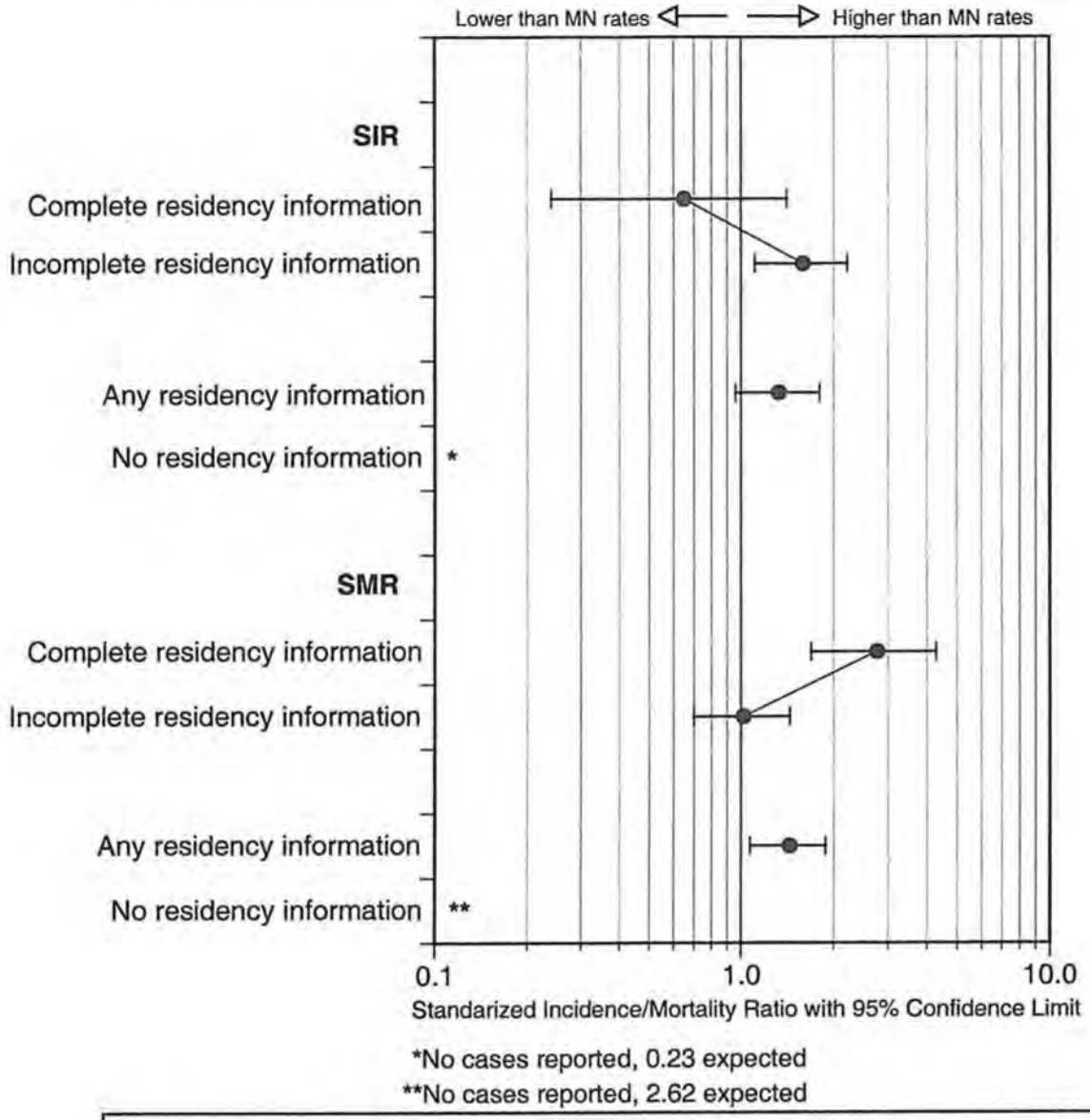
Table 26. Mineral Board Worker Cohort and Subcohort Stratified Analysis by 30-year Employment Latency: Males, Selected Cancers

Cancer Analysis Type Exposure Type	Complete Mineral Board Worker Cohort (Males)			Mineral Board Worker Screened Subcohort (Males)		
	Observed	Expected	Ratio of Observed/ Expected	Observed	Expected	Ratio of Observed /Expected
Lung and Bronchus						
SMR						
Start year ≤ 1958	19	19.26	0.99	11	8.22	1.34
Start year > 1958	14	9.79	1.43	1	1.51	0.66
SIR						
Start year ≤ 1958	24	11.33	2.12*	17	6.70	2.54*
Start year > 1958	16	6.66	2.40*	2	1.53	1.31
Prostate						
SMR						
Start year ≤ 1958	5	7.98	0.63	4	3.54	1.13
Start year > 1958	1	1.81	0.55	0	0.34	-
SIR						
Start year ≤ 1958	38	27.52	1.38	25	16.41	1.52
Start year > 1958	19	10.88	1.75*	2	2.61	0.77
All Cancers						
SMR						
Start year ≤ 1958	66	62.70	1.05	33	26.94	1.22
Start year > 1958	38	33.37	1.14	6	5.1	1.18
SIR						
Start year ≤ 1958	120	77.61	1.55	65	45.63	1.42*
Start year > 1958	89	49.55	1.80	18	10.93	1.65

*=$p < 0.05$

Model C SIR assumes a 3% in- or out-of-state migration rate per year and uses all information sources (i.e., drivers license records, voter registration records, credit bureau records, cancer registry records, and death records).

Figure 36. Mineral Board Worker Cohort Standardized Incidence Ratio (SIR) and Standardized Mortality Ratio (SMR) by Completeness of Residency Data: Males, Lung and Bronchus

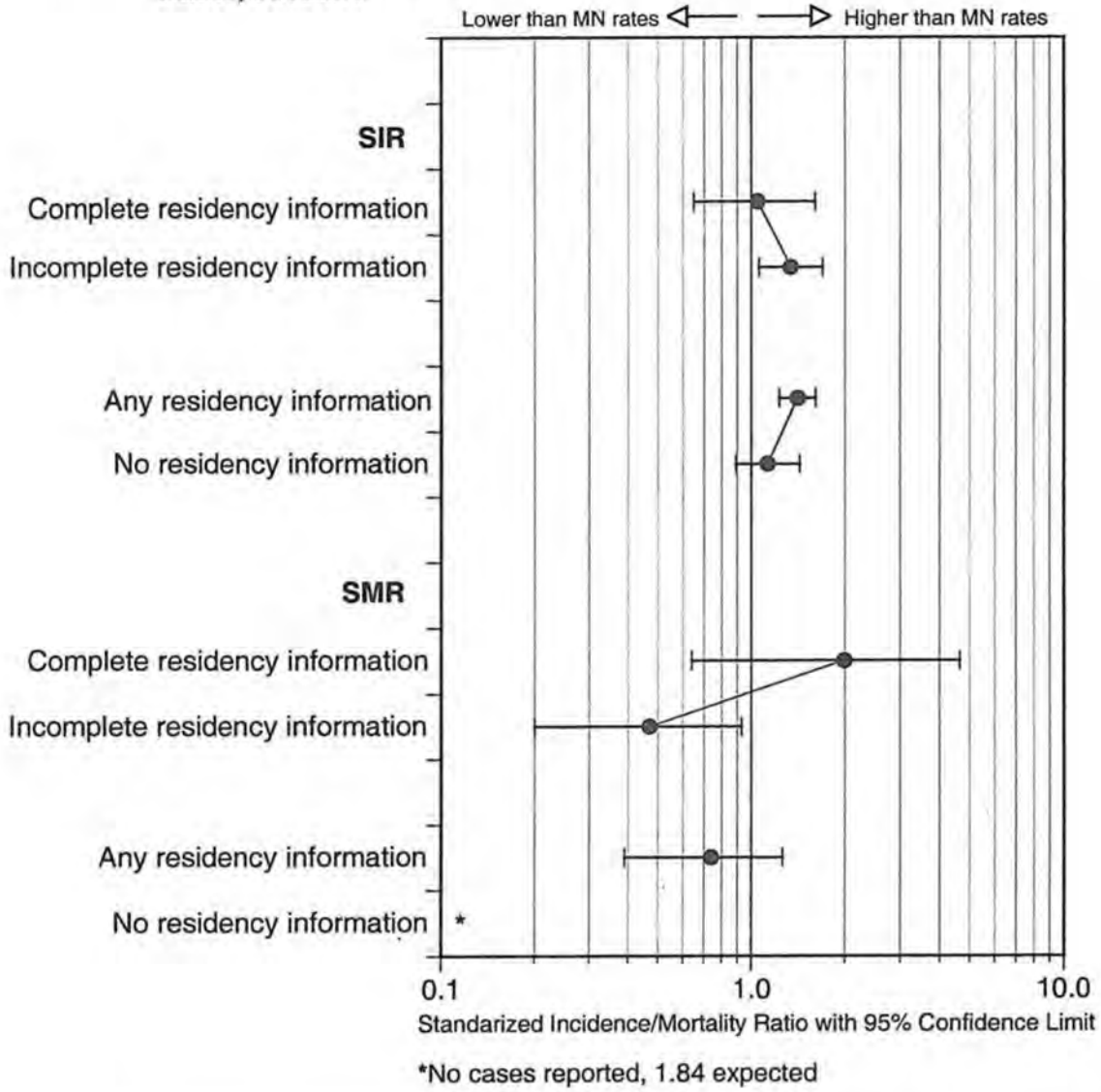


*No cases reported, 0.23 expected

**No cases reported, 2.62 expected

SIRs calculated using all information sources (Drivers license, Voter registration and Credit bureau records) and Residency Model C, which assumes a 3% in- or out-of-state migration rate per year. Model C also includes residency information from cancer registry records and death records.

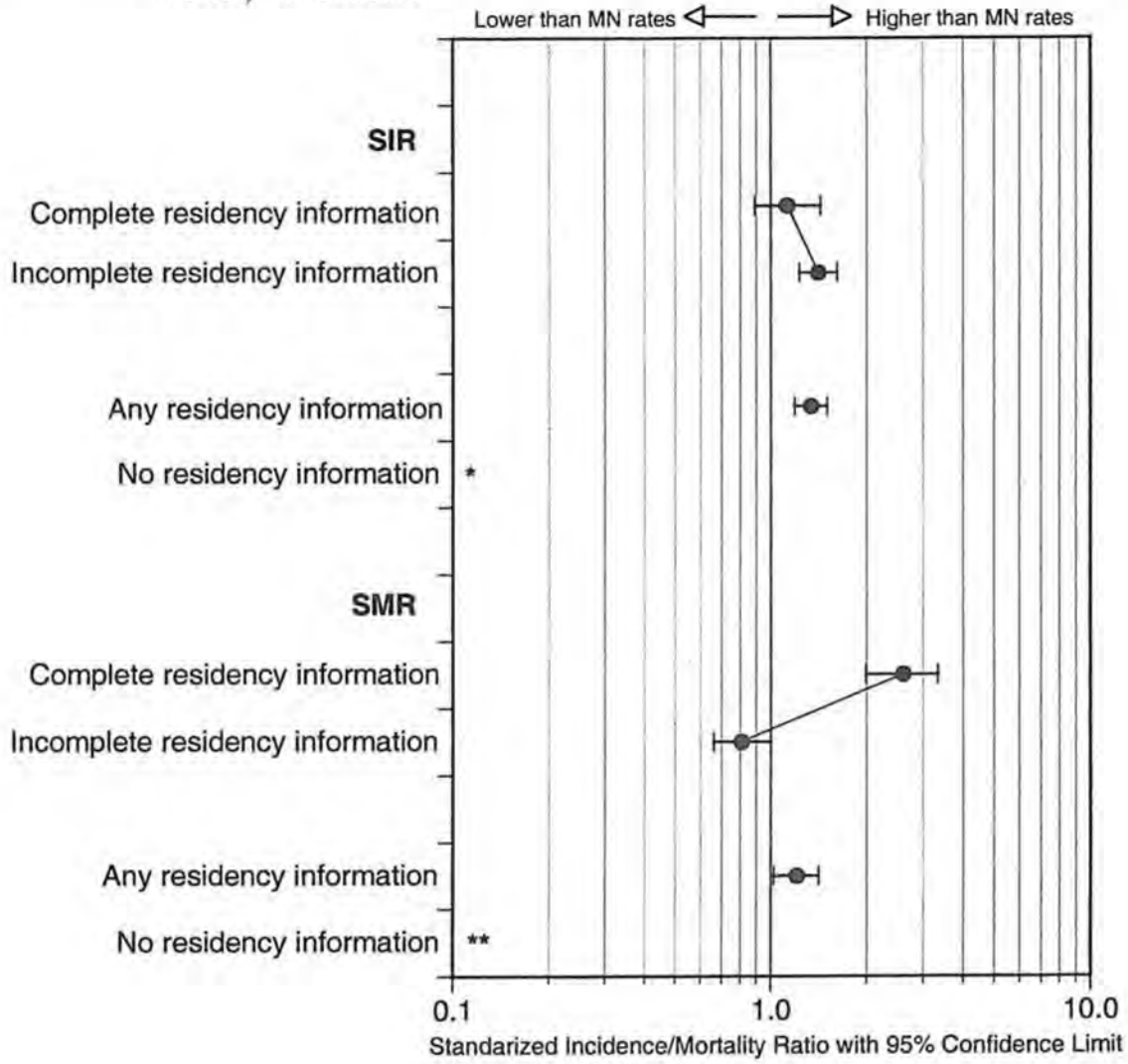
Figure 37. Mineral Board Worker Cohort Standardized Incidence Ratio (SIR) and Standardized Mortality Ratio (SMR) by Completeness of Residency Data: Males, Prostate



*No cases reported, 1.84 expected

SIRs calculated using all information sources (Drivers license, Voter registration and Credit bureau records) and Residency Model C, which assumes a 3% in- or out-of-state migration rate per year. Model C also includes residency information from cancer registry records and death records.

Figure 38. Mineral Board Worker Cohort Standardized Incidence Ratio (SIR) and Standardized Mortality Ratio (SMR) by Completeness of Residency Data: Males, All Cancers



*No cases reported, 1.6 expected

**No cases reported, 10.05 expected

SIRs calculated using all information sources (Drivers license, Voter registration and Credit bureau records) and Residency Model C, which assumes a 3% in- or out-of-state migration rate per year. Model C also includes residency information from cancer registry records and death records.

Table 27. Highway Worker Cohort Stratified Analysis by Completeness of Residency Information: Males, Selected Cancers

Cancer Analysis Type Exposure Type	Observed	Expected	Ratio of Observed/Expected
Lung & Bronchus			
SMR			
-Complete residency information	20	7.21	2.77*
-Incomplete residency information	33	32.27	1.02
-Any residency information	53	36.87	1.44*
-No residency information	0	2.62	-
SIR			
-Complete residency information	6	9.26	0.65
-Incomplete residency information	36	22.60	1.59*
-Any residency information	42	31.65	1.33
-No residency information	0	0.23	-
Prostate			
SMR			
-Complete residency information	5	2.52	1.99
-Incomplete residency information	8	16.92	0.47
-Any residency information	13	17.61	0.74
-No residency information	0	1.84	-
SIR			
-Complete residency information	21	20.09	1.05
-Incomplete residency information	74	55.11	1.34*
-Any residency information	95	74.68	1.27*
-No residency information	0	0.56	-
All Cancers			
SMR			
-Complete residency information	61	23.47	2.60*
-Incomplete residency information	92	113.33	0.81
-Any residency information	153	126.78	1.21*
-No residency information	0	10.05	-
SIR			
-Complete residency information	72	63.62	1.13
-Incomplete residency information	225	159.75	1.41*
-Any residency information	297	221.87	1.34*
-No residency information	0	1.6	-

*=$p < 0.05$

Model C SIR assumes a 3% in- or out-of-state migration rate per year and uses all information sources (i.e., drivers license records, voter registration records, credit bureau records, cancer registry records, and death records).

ACKNOWLEDGEMENTS

Additional support for the Occupational Cancer Surveillance through Record Linkage project was provided by the Chronic Disease Prevention and Control Division, Chronic Disease and Environmental Epidemiology Section, Minnesota Department of Health, Minneapolis, Minnesota.

This report was supported by Grant Number U75-CCU510693 from the Centers for Disease Control and Prevention. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention.

The authors with to thank George Maldonado, Ph.D., Associate Professor, University of Minnesota, Minneapolis, Minnesota for his advice on methodological issues and Ronald Regal, Ph.D., Professor, University of Minnesota, Duluth, Minnesota for his assistance with the capture-recapture analysis.

The authors wish to thank Cynthia Elias for her contributions in completing the record-linkage evaluations and adjudications in this study.

REFERENCES

Bailar JC, Ederer F. Significance factor for the ratio of a Poisson variable to its expectation. *Biometrics.* 1964;20:639-643.

Bender AP, Parker DL, et al. Minnesota highway maintenance worker study: cancer mortality. *Am J Ind Med.* 1989;15:545-556.

Bender AP, Williams AN, et al. Experiences of a state-sponsored notification and screening program for asbestos workers. *Am J Ind Med.* 1993;23:161-169.

Breslow NE, Day NE. *Statistical Methods in Cancer Research. Volume II—The Design and Analysis of Cohort Studies.* Lyon: IARC, 1987.

Cancer Surveillance and Control Program. *Case Completeness and Data Quality Audit: Minnesota Cancer Surveillance System 1994-1995.* Addendum issued February 1997. (unpublished report available from the MCSS); 1997.

Chen VW, Howe HL, Wu XC, Hotes JL, Correa CN (eds). *Cancer in North America, 1993-1997. Volume One: Incidence.* Springfield, IL: North American Association of Central Cancer Registries, April 2000.

Fleiss, JL. *Statistical Methods for Rates and Proportions.* New York: J Wiley and Sons, 1981.

Fellegi I, Sunter A. A theory for record linkage. *J Am Statistical Assoc.* December 1969;c-25:172-178.

Hill ID. Computing man years at risk. *Br J Prev Soc Med.* 1972;26:132-134.

Minnesota Department of Health. *The Minnesota Highway Maintenance Worker Mortality Study 1945-1984.* 1987.

Minnesota Department of Health. *Medical Screening for Asbestos-Related Lung Disease Among Conwed Corporation (Cloquet) Workers and Their Spouses: Preliminary Report to the Minnesota Legislature.* 1989.

Minnesota Department of Health. *Health and Safety of Minnesota Highway Maintenance Workers: Follow-up Report.* 1990.

Minnesota Department of Health. *The Minnesota Highway Maintenance Worker Mortality Study 1945-1989.* 1993.

Nelson D. *SEER Data Management System-II Linkage System Statistical User's Guide.* 1982.

Parker DL, Bender AP, et al. Minnesota highway maintenance worker cohort mortality study: methods and noncancer mortality. *Am J Ind Med.* 1989;15:531-543.

REFERENCES (Continued)

Punyko J, Bushhouse S, et al. *MCSS Record Linkage System*. Minnesota Department of Health, Minnesota Cancer Surveillance System. 1995.

Ries LAG, Eisner MP, Kosary CL, Hankey BF, Miller BA, Clegg L, Edwards BK (eds). *SEER Cancer Statistics Review, 1973-1998*. Bethesda, MD: National Cancer Institute, 2001.

Waxweiler RJ, Beaumont JJ, et al. A modified life-table analysis system for cohort studies. *J Occ Med*. 1983;25:115-124.

Williams AN. *An Epidemiologic Study of Radiographic Abnormalities Among Asbestos Ceiling Tile Workers and Their Spouses*. University of Minnesota. 1994

PUBLICATIONS: PRESENT AND ANTICIPATED FUTURE

Anticipated Publications:

Follow-Up Bias: Impact on Standardized Incidence Ratios. Debora Boyle, Wendy Brunner, Allan Williams and George Maldonado.

Comparison of Standardized Mortality Ratios and Standardized Incidence Ratios for two Occupational Cohorts. Debora Boyle, Allan Williams, Wendy Brunner, Alan Bender and George Maldonado.

Feasibility of Conducting Standardized Incidence Ratio Analysis. Wendy Brunner, Debora Boyle, Allan Williams and George Maldonado.

Evaluation of Quality of Administrative Data for Compiling Residence Information. Cyndi Elias, George Maldonado, Debora Boyle, and Wendy Brunner.

Presentations:

Occupational Cancer Surveillance through Record Linkage. 1999 Cancer Conference: Meeting the Challenges of Comprehensive Cancer Control, Atlanta, Georgia. September 9, 1999.

Occupational Cancer Surveillance through Record Linkage. NAACCR Cancer Surveillance and Control Program: Occupation/Industry Data Collection and Coding Issues, Costa Mesa, California. March 3, 1999.