Practice of Epidemiology

Using Multiple Cause-of-Death Data to Investigate Associations and Causality between Conditions Listed on the Death Certificate

Matthew D. Redelings¹, Matthew Wise¹,², and Frank Sorvillo¹,²

¹ Data Collection and Analysis Unit, Los Angeles County Department of Public Health, Los Angeles, CA.
² Department of Epidemiology, School of Public Health, University of California, Los Angeles, Los Angeles, CA.

Received for publication September 25, 2006; accepted for publication December 8, 2006.

Death rarely results from only one cause, and it can be caused by a variety of factors. Multiple cause-of-death data files can list as many as 20 contributing causes of death in addition to the reported underlying cause of death. Analysis of multiple cause-of-death data can provide information on associations between causes of death, revealing common combinations of events or conditions which lead to death. Additionally, physicians report the causal train of events through which they believe that different conditions or events may have led to each other and ultimately caused death. In this paper, the authors discuss methods used in studying associations between reported causes of death and in investigating commonly reported causal pathways between events or conditions listed on the death certificate.

Abbreviations: EOR, exposure odds ratio; HIV, human immunodeficiency virus; MCOD, multiple cause-of-death; MOR, mortality odds ratio.

Death certificate data are generally analyzed by examining the underlying cause of death as reported on the death certificate by the medical examiner (1). However, deaths rarely have only one cause, and this method does not allow researchers to assess the role of conditions that were important contributors to death but were not reported as the underlying cause. Multiple cause-of-death (MCOD) data offer an alternative. MCOD data specify the underlying cause of death but also include data on other contributing causes of death (2). Thus, more complete information on factors and diseases causing death can be utilized.

MCOD data can be useful in a number of ways. In this paper, we focus specifically on the utility of MCOD data for the analysis of associations between causes of death. We describe methods that can be used to test associations between conditions in MCOD data and to suggest possible causal relations. In order to provide examples of the types of analyses we describe, we refer to our previously published analysis of the 114,380 deaths reported in the United States from 1990 to 2001 in which pressure ulcers were listed as a cause of death in MCOD coded data (3).

FREQUENCIES AND ODDS RATIOS

The simplest analysis of association between conditions in death certificate data involves examining the frequency with which two conditions are reported together. For example, an analysis of all 114,380 death certificates listing pressure ulcers as a cause of death reveals that septicemia is also reported on 45,374 (39.7 percent) of these death certificates. Since this percentage is far higher than we would expect to see among decedents without pressure ulcers, we may conclude that some association between pressure ulcers and septicemia is present in death certificate data. However, it may not always be clear whether the percentage of death certificates reporting comorbidity is different...
than we would expect. For example, 46,714 (40.8 percent) death certificates that report pressure ulcers as a cause of death also report heart disease as a cause of death. This figure is not substantially higher or lower than we would expect, so it is difficult to discern whether or not an association exists between reporting of pressure ulcers and reporting of heart disease in death certificate data.

In order to solve this problem, we can conduct a case-control study comparing deaths from the condition of interest with deaths from other conditions. The use of mortality odds ratios (MORs) and exposure odds ratios (EORs) in case-control studies with mortality data is preferable to the traditional use of the proportional mortality ratio (4). The MOR and EOR differ in that the MOR assumes the condition (pressure ulcers) used for case/control selection to be the outcome, while the EOR assumes it to be the exposure. In our data, we suppose that if a causal association exists between pressure ulcers and septicemia, pressure ulcers will act as the exposure and septicemia as the outcome. Thus, in order to calculate an MOR, we would have to use septicemia deaths as cases and choose nonsepticemia deaths as controls. In contrast, if we wanted to calculate the EOR, we would compare pressure-ulcer-associated deaths with a control sample of non-pressure-ulcer-associated deaths. The difference between the MOR and the EOR is largely a difference of terminology. Although the two figures are calculated using different sampling procedures, they are estimates of the same quantity. Thus, the MOR and the EOR will differ only to the extent that error is present in the selection of controls, and controls may be selected randomly so that the difference between MOR and EOR reflects only random error. Additionally, when testing possible causal associations between causes of death, it is not possible to prove which variable is the outcome and which is the exposure. Thus, the terms MOR and EOR reflect a researcher’s educated guess as to the direction of causation in the analysis being performed. To the extent that EOR and MOR differ, it is not possible to know which figure is preferable to use unless additional information is present. However, we mention the difference between MOR and EOR because correct identification of the exposure variable in case-control studies can be important in identifying potential confounders for stratification and in the selection of appropriate controls (4).

Stratification may be necessary to modify crude MOR and EOR figures because persons who die of pressure ulcers may be different in a number of ways from decedents without pressure ulcers. For example, they will probably be older and thus more likely to be reported as suffering from heart disease. The presence of factors, such as age, which exert an effect on the exposure (heart disease) and are related to the outcome (pressure ulcers) may introduce confounding. Various methods of stratification are available to control for confounding, including the use of the standardized mortality odds ratio (5). However, we focus here on the use of matched comparisons because of their utility even when data are sparse in some strata. In matched comparisons, a matched control is selected for each case from other deaths that are similar to the case with respect to age, sex, race/ethnicity, or any other factors for which we desire to control.

### Table 1. Pressure ulcer and comorbidity status of subjects in a matched case-control study

<table>
<thead>
<tr>
<th>Pressure ulcer-associated death</th>
<th>Matched non-pressure-ulcer-associated death</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comorbidity present</td>
<td>A</td>
</tr>
<tr>
<td>Comorbidity absent</td>
<td>B</td>
</tr>
<tr>
<td>Comorbidity present</td>
<td>C</td>
</tr>
<tr>
<td>Comorbidity absent</td>
<td>D</td>
</tr>
</tbody>
</table>

Matched mortality odds ratios (mMORs) or matched exposure odds ratios (mEORs) are then calculated in order to estimate the magnitude of the association between the conditions. While the selection of matched controls does not control for confounding on its own, the calculation of mMORs (or mEORs) does. mMORs (and mEORs) treat each individual matched pair as a single stratum and introduce stratification by the variables used in the matching process.

As with odds ratios in other types of matched pair studies (6, 7), matched statistics are calculated by examining discordant pairs, where either the case is exposed and the control is unexposed or the case is unexposed and the control is exposed (see table 1).

The following equations are used (6, 7):

\[ \text{mMOR} = \frac{\text{mEOR}}{\text{A}} \]

and

\[ \text{Var}[\ln(\text{mMOR})] = \text{Var}[\ln(\text{mEOR})] = \frac{1}{B} + \frac{1}{C}. \]

In our sample of pressure-ulcer-associated deaths (3), matched exposure odds ratios confirm our suspicion that a strong association exists between reporting of pressure ulcers and reporting of septicemia in death certificate data (mEOR = 11.3, 95 percent confidence interval: 11.0, 11.7). In contrast, a negative association was found between reporting of pressure ulcers and reporting of heart disease (mMOR = 0.5, 95 percent confidence interval: 0.5, 0.5).

### INTERPRETING STATISTICAL COMPARISONS

Generally, when we speak of two conditions as being “associated” with each other, we mean that the presence of one condition indicates a higher or lower probability that the other condition will also be present. However, physicians filling out death certificates only report conditions that they believe were instrumental in causing death. They do not report all prevalent conditions. Thus, death certificate data cannot indicate whether septicemia is more likely to be present in decedents with pressure ulcers than in decedents without them. They can only indicate whether septicemia is more likely to act as a cause of death when pressure ulcers are also a cause of death than when they are not. This information has real clinical significance, suggesting the importance of septicemia prevention and treatment in persons with pressure ulcers as a mechanism for preventing death. However, associations in death certificate data should not be
interpreted as reflecting associations in the general population. It is possible that persons in the general population with pressure ulcers are more likely to have or develop septicemia, but death certificate data do not measure the prevalence of either condition and can only provide indirect support for such a theory.

Additionally, selection bias can threaten the validity of inferences derived from a case-control study. Selection bias may arise when both outcome and exposure status influence the inclusion of subjects in a study. In case-control study designs, an association necessarily exists between the outcome and selection, since subjects are enrolled in the study on the basis of their disease status. Consequently, any association between exposure status and selection will lead to selection bias. In order to be included in a case-control study of mortality data, a person must have died. Thus, any exposure that increases the likelihood that someone will die also increases the probability that the person will be included in the study, generating a type of selection bias termed Berksonian bias (5). For example, suppose an investigator were interested in human immunodeficiency virus (HIV) as a potential risk factor for heart disease and sought to estimate the association between these conditions using death certificate data. Persons would initially be included in the study based on the presence of heart disease on the death certificate, creating an association between heart disease and selection. In addition, persons with the exposure, HIV, would also be more likely to be included in the study, since HIV infection increases the likelihood that a person will die. This would probably cause persons with HIV to be overrepresented in death data relative to the general population.

Since most medical conditions listed on death certificates ought to have contributed to death, it is reasonable to expect that some level of Berksonian bias will operate in nearly all associations estimated between causes of death listed on death certificates. Berksonian bias can be limited, however, by selecting controls who died from conditions thought to be unassociated with the exposure. A second form of selection bias known as detection bias (8) can occur if the exposure and outcome conditions are thought to be associated. In this case, persons with the exposure condition may be more likely to be tested for the outcome condition and thus will be more likely to be identified as cases in the study.

Death certificate data may also be affected by reporting bias. At the simplest level, reporting bias may result from incorrect diagnoses being recorded on the death certificate. However, a more systematic type of reporting bias is also possible. In our data, the number of death certificates that recorded both septicemia and pressure ulcers as causes of death may be an artifact, in part, of common medical opinions among physicians filling out death certificates. The biologic plausibility of a connection between pressure ulcers and septicemia is high. Thus, a physician filling out a death certificate for a person who died of septicemia and had a severe pressure ulcer might mistakenly assume that both conditions contributed to death. The opposite can also occur; most conditions (such as heart disease in the example above) are not thought of as being associated with the development or progression of pressure ulcers. Even if both heart disease and pressure ulcers were important contributors to death, physicians may be unlikely to list heart disease and pressure ulcers on the same death certificate, simply because of the low biologic plausibility of the two conditions’ acting in concert to cause death.

In the case of models using a matched design, a special type of selection bias known as overmatching bias may be present (4). While matching on true confounders will not cause this type of bias, matching on any factor which is affected by the exposure variable and is related to the outcome variable may introduce overmatching bias, and in some situations this bias cannot be eliminated analytically. Thus, it is important to select matching variables carefully. The common use of demographic variables such as age, race, and sex for matching should not introduce overmatching bias, since these variables are unlikely to be affected by the exposure.

**CAUSATION AND ENTITY AXIS CODES**

While matched odds ratios are unable to indicate causal relations between conditions, epidemiologic methods of causal inference are available (8, 9). Some types of evidence typically used to imply causation (such as a dose-response relation or temporality) cannot be examined using death certificate data, but others can. The strong biologic plausibility of a causal relation between pressure ulcers and septicemia provides some evidence that the association we observed may be causal. Additionally, the strength of the relation (mEOR = 11.3, 95 percent confidence interval: 11.0, 11.7) suggests that it could not have resulted from weak or moderate biases alone (however, strong biases may exist in death certificate data).

In addition to standard epidemiologic methods of causal inference, death certificate data contain information on the supposed causal train of events leading to death, as reported by the physician. The US Standard Certificate of Death has two sections in which cause-of-death data can be entered (figure 1). In part I, the physician is asked to describe the causal train of events which led to death. The condition reported on line 7 of part I is the immediate cause of death, the last condition in the causal train of events leading to death. It is caused by the condition or event (if any) reported on line 6, which in turn is caused by the condition or event (if any) reported on line 5.

This information is recorded in MCOD data in the entity axis code variables (10). Entity axis codes provide researchers with International Classification of Diseases codes for the conditions or events reported in part I and also note which line of part I the physician originally used to report each condition or event. Thus, in the above example, if a pressure ulcer is reported on line 7 and septicemia is reported on line 6, we may suppose that the physician believed the pressure ulcer to be a cause of septicemia, which in turn caused death. Entity axis data are not always clear. While only one condition should be reported per line in part I, this rule is not always followed. For example, a physician may report “septic pressure ulcer infection” on line 7 of part I. In this case, both septicemia and pressure ulcers will be coded in MCOD data as having been entered on line 7 and it will
not be possible, using the data alone, to determine the causal relation between the two. In our data, 1,072 (3.0 percent) deaths were identified for which pressure ulcers and septicemia were indicated on the same line of the death certificate, such that the direction of causation was unclear. However, in 33,102 (94.0 percent) of the 35,215 cases where pressure ulcers and septicemia were both listed in part I, septicemia was listed as being “due to or a consequence of” pressure ulcers. It is unlikely that such a strong trend would be observed by chance alone.

However, causal pathways reported in entity axis codes only reflect physicians’ best judgment about the causal train of events. It is not possible, using death certificate data, to determine which criteria the physician used to infer causation. The physician may have reasoned incorrectly or may have had insufficient information with which to adequately assess the causal train of events leading to death, and therefore caution must be exercised.

In certain cases, we may be able to recreate and validate the physician’s process of causal inference with a reasonable degree of certainty. For example, in 490 of the 498 (98.4 percent) cases for which multiple sclerosis and pressure ulcers were both reported in part I, pressure ulcers were listed as a result of multiple sclerosis. It is reasonable for a physician to assume that a person with an incapacitating case of multiple sclerosis would not have developed a severe pressure ulcer in the absence of multiple sclerosis. However, in some cases, the physician’s process of causal inference may be unclear, or it may even be apparent that the physician’s reporting cannot reflect true causation. Six death certificates were found in which multiple sclerosis was reported as having resulted from pressure ulcers. Current medical knowledge suggests that this is unlikely, and so few physicians reported this causal pathway that it is quite possible they did so unintentionally. To the extent that physicians’ reasoning is unclear or seems incorrect, causal pathways reported in entity axis codes should be given less credence.

While part I of the death certificate only allows for the reporting of a single train of events leading to death, causality may be extremely complicated in real situations. Deaths may be caused by multiple trains of events which may or may not converge. Thus, physicians are also provided with part II (figure 1), in which they can record other conditions or events that contributed to death but were not listed in part I. It is not possible to establish a causal relation between these conditions and those reported in part I, though such a relation may exist. Of the 45,374 death certificates that listed both pressure ulcers and septicemia as causes of death, either pressure ulcers or septicemia was listed only in part II for 10,159 (22.4 percent), such that no causal relation was indicated.

Additionally, data from entity axis codes are recoded to generate record axis code variables. While entity axis codes reflect data originally reported by the examiner, record axis codes reflect “cleaned” data, in which some conditions are recoded to better match the context of other conditions or events reported on the death certificate (11). For example, suppose that a medical examiner reports both alcohol dependence syndrome and cirrhosis of the liver without mention of alcohol as a cause of death. Both conditions will be shown in entity axis code data. However, in record axis code data, they will be replaced with a single condition: alcoholic cirrhosis of the liver. Unfortunately, record axis code data do not reflect the position in which conditions are recorded in the causal train of events leading to death.

**CONCLUSIONS**

Like all data, death certificate data are imperfect, and it is important to understand their limitations before attempting to interpret associations observed in the data. Multiple
sources of bias exist, and it is not possible to check the accuracy with which data were recorded or to easily access supplemental data, such as decedents’ medical histories, that would potentially shed light on information recorded in death certificate data.

However, while caution should be used in examining associations observed between conditions in MCOD data, the data are still of considerable value. MCOD data are free and provide an easily available source of population-based information on causes of death and associations between causes of death. While studies utilizing death certificate data cannot prove the existence of causal relations, entity axis code data provide an additional source of information on the causal pathways physicians believe to have resulted in death.

ACKNOWLEDGMENTS

This work was supported by the Los Angeles County Department of Health Services.

The authors thank Dr. Katherine J. Hoggatt for her comments on the manuscript.

Conflict of interest: none declared.

REFERENCES


