Deriving Valid Population–Based Cancer Survival Estimates in the Presence of Nonnegligible Proportions of Cancers Notified by Death Certificates Only

Hermann Brenner¹ and Bernd Holleczek²

Abstract

Background: Studies of cancer survival by population-based cancer registries are a key component in monitoring progress against cancer. Patients notified by death certificates only (DCO) are commonly excluded from such studies. The validity of this “exclude DCO” approach has been questioned and an alternative “correct for DCO” approach has been proposed.

Methods: We assess the validity of both the “exclude DCO” approach and the “correct for DCO” approach using model calculations. We illustrate implications for population-based cancer survival analyses by analyses of 5-year relative survival of cancer patients in Saarland, Germany.

Results: The “exclude DCO” approach provides (too) optimistic survival estimates and the “correct for DCO” approach provides (too) pessimistic survival estimates under plausible assumptions. For example, in case of true survival of 50%, underascertainment of 5% of surviving patients and of 15% of dying patients (yielding a proportion of DCO cases of 7.7%), the two approaches would provide survival rate estimates of 52.8% and 48.8%, respectively. The difference of survival estimates obtained with both approaches increases with incompleteness of registration and the proportion of DCO cases. Trace back of DCO cases shifts survival estimates from the former toward the latter estimate.

Conclusions: In case of nonnegligible DCO proportions, cancer survival studies should not be exclusively based on either the “exclude DCO” or the “correct for DCO” approach. A combination of estimates from both approaches may be useful to delineate a plausibility range for true survival.

Impact: Our results may help to enhance validity and comparability of population-based cancer survival estimates. Cancer Epidemiol Biomarkers Prev; 20(12); 2480–6. ©2011 AACR.
colleagues (4) estimated the expected change in survival estimates if all DCO cases were successfully traced. They found survival estimates to become substantially lower, with the relative reduction in survival estimates being close to the proportion of DCO cases in the registry. Following this observation and applying a simple proportional hazards model, Silcocks and Thompson (9) concluded that a simple correction procedure, multiplying observed survival by 1 minus the proportion of DCO cases, provides an acceptable correction of 5-year survival in most cases (albeit not necessarily for short-term survival, such as 1-year survival). However, a problem inherent in this “correct for DCO” approach is that it selectively corrects for potential bias by underreporting of dying patients who are identified through death certificates (6), whereas no correction is employed for potential bias by missed cancer survivors who do not come to the attention of the cancer registry by death certificate. 

In this article, we assess the expected validity of the 2 analytic strategies, the “exclude DCO,” and the “correct for DCO” strategy, in estimating cancer survival in the presence of nonnegligible proportions of DCO cases. Our analyses are based on model calculations with systematic variation of key parameters regarding true survival and proportions of cancer cases missed at lifetime. Empirical examples applying both strategies are provided using data from the population-based cancer registry of Saarland, Germany.

Methods

Model calculations

Our model calculations are based on the following key parameters:

- The proportion of patients dying from their cancer within the follow-up period of analysis (e.g., 5 years for 5-year survival estimates), denoted pd.
- The proportion of patients missed by the registry at lifetime among those dying from their cancer within the follow-up period of analysis, denoted md.
- The proportion of patients missed by the registry at lifetime among those not dying from their cancer within the follow-up period of analysis, denoted ms.

Note that our model calculations focus on cause-specific survival estimates, which are commonly reported in clinical studies and which may be an alternative to the relative survival estimates commonly reported in population-based survival analyses (10, 11).

To keep the model as simple as possible, and to focus on the impact of the aforementioned parameters, the following simplifying assumptions (to be discussed in detail below) are made:

- Incidence of the cancers, true survival of cancer patients, and completeness of cancer registration at lifetime are constant over time.
- The cancer registry has been running for some time so that DCO cases resulting from diagnoses made prior to the era of cancer registration are negligible.
- All deaths due to the cancer during the follow-up time come to the attention of the cancer registry through death certificates.
- Loss to follow-up is negligible or at least unrelated to survival probability of cancer patients.

Under these assumptions, the expected proportions of cancer patients registered at lifetime equals

\[ pd \times (1 - md) + (1 - pd) \times (1 - ms), \]

and the expected proportion of cancer cases first notified by death certificate equals

\[ pd \times md. \]

The expected survival estimate for the “exclude DCO” approach, denoted \( S_E \), is

\[ S_E = (1 - pd) \times (1 - ms)/(pd \times (1 - md) + (1 - pd) \times (1 - ms)), \]

and the expected survival estimate for the “correct for DCO” approach, denoted \( S_C \), is

\[ S_C = (1 - pd \times ms) \times S_E = (1 - pd) \times (1 - ms)/(pd \times (1 - md) + (1 - pd) \times (1 - ms)). \]

A closer look at these expected survival estimates, applying simple algebra, easily discloses the following properties:

1. If no survivors are missed, that is, if ms equals 0, then \( S_C = 1 - pd \), that is, the “correct for DCO” approach provides a valid estimate of survival.
2. If relative underreporting is the same for patients who do and who do not die within the follow-up period of interest, that is, if ms = md, then \( S_E = 1 - pd \), that is, the “exclude DCO” approach provides a valid estimate of survival.

In practice, however, none of these extreme assumptions is likely to hold. In the following, we therefore provide expected survival estimates for various combinations of values of the key model parameters and compare them with true survival, 1 - pd. In a base case scenario, pd is assumed to be 0.5, and md and ms are assumed to be 0.15 and 0.05, respectively. Hence, a higher miss rate is assumed for dying patients than for surviving patients for reasons outlined above, with the overall proportion of cases missed at lifetime being 0.1. Starting from this base case scenario, key parameters are varied as follows in additional scenarios:

- pd is varied between 0.1, 0.5, and 0.9 to address cancers with different prognosis.
- Pairs of md and ms are varied between md = 0.2, ms = 0 and md = 0.1, ms = 0.1 to address the impact of the
degree of differentiality in miss rates between dying and surviving patients, while keeping the overall miss rate constant at 0.1.

- The overall miss rate is varied between 0.02 and 0.30, while keeping the ratio of $md$ and $ms$ constant at 3:1, the ratio chosen for the base case scenario.

Table 1. Expected survival estimates under various scenarios of underascertainment of surviving and dying patients at lifetime

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Proportion missed at lifetime</th>
<th>Patients registered at lifetime</th>
<th>DCO patients</th>
<th>Expected survival estimate (%) according to type of analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Surviving</td>
<td>Dying</td>
<td>Surviving</td>
<td>Dying</td>
</tr>
<tr>
<td>&quot;Perfect&quot;</td>
<td>0%</td>
<td>0%</td>
<td>500</td>
<td>500</td>
</tr>
<tr>
<td>&quot;Base Case&quot;</td>
<td>5%</td>
<td>15%</td>
<td>475</td>
<td>425</td>
</tr>
<tr>
<td>&quot;Selective&quot;</td>
<td>0%</td>
<td>20%</td>
<td>500</td>
<td>400</td>
</tr>
<tr>
<td>&quot;Equal&quot;</td>
<td>10%</td>
<td>10%</td>
<td>450</td>
<td>450</td>
</tr>
</tbody>
</table>

NOTE: All scenarios are based on 1,000 patients with 50% true survival.

Empirical analyses

Five-year relative survival was estimated for patients diagnosed with common forms of cancer in Saarland, Germany, in 1999 to 2003. Data from the population-based Saarland Cancer Registry were used, which is operating since 1968 and covers the federal state of Saarland in southwestern Germany with about 1.03 million inhabitants. The registry regularly contributes to descriptive and analytic studies in national or international collaborations (1–3, 12–14).

In Saarland, attempts are made to trace back cancers first notified to the cancer registry through death certificates by contacting clinicians and pathology laboratories. If successful, a new date of diagnosis is assigned to those cases. In our empirical analyses, 5-year relative survival was estimated using the following approaches:

- "Exclude DCO, no trace back," which equals "exclude DCO" in the absence of trace back.
- "Exclude remaining DCO" after trace back, which is the approach taken by the Saarland Cancer Registry in routine practice.
- "Correct for DCO before trace back" (equals "correct for DCO" in the absence of trace back).
- "Correct for DCO after trace back."

The following malignancies with strongly varying prognosis were used for illustration: cancers of the stomach, colorectum, lung, breast, and prostate, as well as leukemias (all forms combined). Analyses were carried out separately for age groups 15 to 74 and 75+ years, as well as for all age groups combined, with age adjustment using age groups and their weights from the International Cancer Survival Standards (ICSS; ref. 15). In all analyses of relative survival, expected survival was calculated from population life tables by the Ederer II method (16, 17). Analyses were done by a publicly available R program package for relative survival (18, 19) which is based on previously reported macros for cohort and period analysis of relative survival (20).

Results

Table 1 illustrates expected results of estimating survival proportions in a population of 1,000 cancer patients with true survival proportion of 50% ($pd = 0.5$). Obviously, this proportion would be expected to be estimated correctly in the absence of underascertainment of cases at lifetime. In the base case scenario, assuming 5% and 15% underascertainment of surviving and dying patients, respectively, the "exclude DCO" approach would be expected to slightly overestimate true survival (52.8%), whereas the "correct for DCO" approach would be expected to slightly underestimate true survival (48.8%). In the more extreme scenario with selective, exclusive underascertainment of 20% of dying patients, the "exclude DCO" approach would be expected to more substantially overestimate true survival (55.6%), whereas the "correct for DCO" approach would provide an unbiased survival estimate. If underascertainment equally affected surviving and dying patients, the "exclude DCO" approach would be expected to more substantially overestimate true survival (55.6%), whereas the "correct for DCO" approach would provide an unbiased survival estimate. If underascertainment equaled survival and dying patients, the "exclude DCO" approach would be expected to provide an unbiased estimate of true survival, whereas the "correct for DCO" approach would be expected to underestimate true survival rate (47.5%).

In intermediate scenarios (Table 2) with various levels of selectivity of miss rates in dying patients and surviving patients (while keeping overall miss rates constant at 10%), the expected proportion of DCO cases would vary between 5% and 10%, and the true survival proportion of 50% would consistently lie in between the survival estimates expected with the "exclude DCO" approach and the "correct for DCO" approach. The former approach would be expected to come closer to the truth in case of less...
selectivity of underascertainment, whereas the validity of the latter approach is expected to increase with increasing selectivity of underascertainment.

Table 3 shows survival estimates expected with the 2 analytic approaches for various levels of underascertainment of cases while keeping the selectivity of underascertainment constant at the level of the base case scenario (i.e., assuming a 3 times higher miss rate for dying patients than for surviving patients). For each level of completeness, the “exclude DCO” approach is expected to overestimate true survival, whereas the “correct for DCO” approach is expected to underestimate true survival. Expected DCO proportions and the expected degree of over- or underestimation increase with the levels of miss rates.

The pattern of expected overestimation of true survival with the “exclude DCO” approach and of underestimation of true survival with the “correct for DCO” approach is consistently seen for a wide range of levels of true survival, as illustrated in Table 4. In absolute terms, expected over- or underestimation of true survival is most pronounced for levels of true survival rate close to 50%. Whereas the miss rates of surviving and dying patients are the same in all scenarios shown in Table 4 (5% and 15%, respectively), the expected DCO rate is very small in case of high true survival and substantially increases with decreasing survival proportions.

Figure 1 shows the results of the empirical analyses. Overall, between 2% and 10% of cases of the various cancers were known to the registry from a death certificate only. For about half of the cases, additional notifications could be obtained by trace back. As expected from theory, estimates of 5-year relative survival were consistently highest with the “exclude DCO, no trace back” approach, and consistently lowest with the “correct for DCO” approaches. The latter yielded very similar results regardless of whether the correction was made prior to or after trace back. Differences in survival estimates from the

Table 2. Expected survival estimates for various levels of selectivity of underascertainment of surviving and dying patients

<table>
<thead>
<tr>
<th>Underascertainment, %</th>
<th>DCO, %</th>
<th>Expected survival estimate (%) according to type of analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surviving patients</td>
<td>Dying patients</td>
<td>Exclude DCO</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>5.0</td>
</tr>
<tr>
<td>9</td>
<td>11</td>
<td>5.8</td>
</tr>
<tr>
<td>7</td>
<td>13</td>
<td>6.7</td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td>7.7</td>
</tr>
<tr>
<td>3</td>
<td>17</td>
<td>8.6</td>
</tr>
<tr>
<td>1</td>
<td>19</td>
<td>9.5</td>
</tr>
<tr>
<td>0</td>
<td>20</td>
<td>10.0</td>
</tr>
</tbody>
</table>

NOTE: True survival of 50% and overall underascertainment of 10% of cases are assumed in all scenarios.

Table 3. Expected survival estimates for various degrees of (selective) underascertainment of surviving and dying patients

<table>
<thead>
<tr>
<th>Underascertainment, %</th>
<th>DCO, %</th>
<th>Expected survival estimate (%) according to type of analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surviving patients</td>
<td>Dying patients</td>
<td>Exclude DCO</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>1.5</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>4.6</td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td>7.7</td>
</tr>
<tr>
<td>7</td>
<td>21</td>
<td>10.9</td>
</tr>
<tr>
<td>9</td>
<td>27</td>
<td>14.1</td>
</tr>
<tr>
<td>11</td>
<td>33</td>
<td>17.5</td>
</tr>
<tr>
<td>13</td>
<td>39</td>
<td>20.9</td>
</tr>
<tr>
<td>15</td>
<td>45</td>
<td>24.3</td>
</tr>
</tbody>
</table>

NOTE: True survival is assumed to be 50% in all scenarios.
various approaches were generally modest and below 3% units for all assessed solid cancers among patients below 75 years of age. Much larger differences, ranging up to 9% units for the solid cancers (breast cancer, prostate cancer) and up to 15% units for leukemias, were observed for patients aged 75 years and older. Differences for all age groups combined were slightly higher than those seen among patients below 75 years of age. In all cases, trace back with subsequent exclusion of remaining DCO cases led to survival estimates that are approximately halfway between the “exclude DCO, no trace back” and “correct for DCO” estimates.

Discussion

In this study, we provide a simple model for assessing the impact of various analytic strategies on cancer survival estimates in the presence of nonnegligible proportions of DCO cases. Assuming higher miss rates among dying patients than among surviving patients, our analyses suggest that the commonly used “exclude DCO” strategy is expected to lead to too optimistic survival

Table 4. Expected survival estimates for various levels of true survival

<table>
<thead>
<tr>
<th>True survival rate, %</th>
<th>DCO, %</th>
<th>Expected survival estimate (%) according to type of analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Exclude DCO</td>
<td>Correct for DCO</td>
</tr>
<tr>
<td>5</td>
<td>14.3</td>
<td>5.6</td>
</tr>
<tr>
<td>10</td>
<td>13.6</td>
<td>11.0</td>
</tr>
<tr>
<td>20</td>
<td>12.1</td>
<td>21.8</td>
</tr>
<tr>
<td>30</td>
<td>10.7</td>
<td>32.4</td>
</tr>
<tr>
<td>40</td>
<td>9.2</td>
<td>42.7</td>
</tr>
<tr>
<td>50</td>
<td>7.7</td>
<td>52.8</td>
</tr>
<tr>
<td>60</td>
<td>6.2</td>
<td>62.6</td>
</tr>
<tr>
<td>70</td>
<td>4.7</td>
<td>72.3</td>
</tr>
<tr>
<td>80</td>
<td>3.1</td>
<td>81.7</td>
</tr>
<tr>
<td>90</td>
<td>1.6</td>
<td>91.0</td>
</tr>
<tr>
<td>95</td>
<td>0.8</td>
<td>95.5</td>
</tr>
</tbody>
</table>

NOTE: Underascertainment of 5% of surviving cases and 15% of dying cases is assumed in all scenarios.
estimates, whereas the previously suggested "correct for DCO" strategy would be expected to yield too pessimistic survival estimates. The difference between the 2 estimates is expected to be small with underreporting of cancers at lifetime below 10% but may become more substantial for higher miss rates. As true survival always lied between the estimates obtained by the "exclude DCO" and the "correct for DCO" approach, our analyses suggest that these estimates might serve to delineate a "plausibility range" for true survival.

An assumption implicitly made in the commonly used "exclude DCO" strategy is that survival of the remaining cases, that is, those notified at lifetime, is representative for survival of all cancer cases. Under these conditions (addressed in the scenario assuming equal miss rates among dying and surviving patients shown in Tables 1 and 2 in our analysis), the "exclude DCO" strategy would be expected to yield an unbiased estimate of survival. In practice, however, patients with a poor prognosis typically have a lower chance of registration during lifetime. For example, underreporting is notoriously more severe in older patients who less often receive curative treatment, such as surgery, in which case the registry is less likely to receive notifications from both clinicians and pathology laboratories.

In contrast, the previously suggested "correct for DCO" strategy would provide correct survival estimates if ascertainment of surviving patients was complete. This is unlikely to be the case in cancer registries with problems of case underascertainment though, and the problem can also not be overcome by trace back of DCO cases which selectively reduces underascertainment of cases who died of their cancer only. Hence, neither the "exclude DCO" approach nor the "correct for DCO" approach is expected to yield an unbiased estimate of true survival in practice. These different approaches may nevertheless be useful, however, to delineate a plausibility range for the unknown true survival.

Some cancer registries make additional attempts to trace back patients known from death certificate only. This way, only patients who remain DCO cases after (unsuccessful) trace back are finally excluded. Such an approach is also taken in the Saarland Cancer Registry that provided the data for our empirical illustrations. As shown in our analyses, successful trace back of about half of the initial DCO cases with a subsequent "exclude remaining DCO" analysis limited the optimistic "exclude DCO" survival estimates to some extent. In fact, the "exclude remaining DCO" analysis yielded survival estimates that were approximately half way between the "exclude DCO, no trace back" and the "correct for DCO" estimates. These results suggest that successful trace back of all initial DCO cases might have yielded survival estimates close to the "correct for DCO" estimates (assuming similar survival times of DCO cases with and without successful trace back), a pattern that closely resembles observations made in the aforementioned trace-back study conducted by the Thames Cancer Registry (4). This suggestion is further supported by the finding that the "correct for DCO" estimates derived before and after trace back were almost identical.

Our empirical analysis showing that trace back with exclusion of remaining DCO cases makes survival estimates less (over)optimistic is consistent with observations from previous trace back studies (4–8). The intensity and success of trace back are expected to determine how far the survival estimates are shifted from the optimistic to the pessimistic end of the plausibility range. Variation of the intensity of trace back of DCO across cancer registries may therefore be an additional source of heterogeneity in comparative cancer survival studies. Although high completeness of trace back appears to be most appealing in multiple respects, the possibility has to be kept in mind that it could make survival estimates even (too) pessimistic due to selective completion of ascertainment of dying patients.

Our model calculations were based on a number of simplifying assumptions. First, incidence, survival, and completeness of registration were assumed to be constant over time. This assumption had to be made because some DCO cases coming to the attention of the registry through death certificate during the period of diagnosis for which survival analyses are made may actually have occurred prior to that period. Assuming constant incidence, survival, and completeness of registration, the shift in date of diagnosis assigned to those DCO cases would not matter, as it is balanced by an analogous shift of cancer diagnoses from the period of investigation to a later date. However, given that many DCO cases typically only have a very short true survival time (4–8), the shift in apparent date of diagnosis is typically short, and only very strong short-term changes in incidence, survival, and completeness of registration, which are uncommon in practice, would be of concern regarding the validity of the model.

Second, our model calculations pertain to a situation in which cancer registration has been established for a number of years. During the build-up phase of a new cancer registry, the proportion of DCO notifications may be high simply because they pertain to patients whose diagnosis and first-line treatment were made prior to establishment of the registry and who therefore remained unregistered.

A third assumption was completeness of registration of cancer deaths. This assumption may be violated, for example, by less than perfect coding of causes of deaths (21, 22). However, misreporting of cancer causes of death has been found to be low compared to other causes of death, and our model calculation would still hold even in the presence of misreporting, as long as underreporting and overreporting of the cancer of interest on death certificate are approximately balanced. However, in situations where DCO cases are mostly due to false-positive diagnoses on the death certificates (a situation that might be encountered, for example, for cancers at sites of frequent metastasis, such as the lung or the liver), the "exclude DCO" approach might actually be the most
appropriate estimate, whereas the "correct for DCO" approach might lead to unjustified reductions of survival estimates due to correcting for false-positive apparent DCO diagnoses. Finally, the assumption of loss-to-follow-up being unrelated to prognosis is an assumption commonly made in all cancer survival estimates from population-based cancer registries.

Despite uncertainties due to potential violations of the underlying assumptions and in the degree of selectivity of underreporting of cases, our model calculations provide a theoretical framework for and support suggestions that the common practice of simply excluding DCO cases from population-based survival analyses may often lead to somewhat too optimistic survival estimates (5), at least if no trace back is undertaken. On the other hand, however, our results also suggest that the previously suggested "correct for DCO" approach may often lead to somewhat too pessimistic survival estimates. Varying completeness of registration and DCO proportions across registries or over time may also threaten the validity of international or regional comparison of cancer survival or of time trends in survival (5). Obviously, the best prevention of potential bias would be to minimize DCO proportions by maximizing completeness of registration in the first place. To the extent that this cannot be achieved despite major efforts, we propose to complement reporting of the (optimistic) "exclude DCO" survival estimates by reporting of the (pessimistic) "correct for DCO" survival estimates. Neither of these estimates is likely to be perfect, but their joint report might be used to delineate a plausibility range of true survival. This range would be wider for cancer registries with high DCO proportions than for cancer registries with low DCO proportions and might ensure fairer comparison of survival estimates between such registries than either "the exclude DCO estimate" or the "correct for DCO" estimate alone. In particular, the width of the plausibility ranges compared with differences in survival between registries might be an important criterion in judging to what extent the latter might be affected by differences in completeness of cancer registration.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Grant Support

This study was supported in part by a grant from the German Cancer Aid (Deutsche Krebshilfe; no. 108237).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received July 21, 2011; revised September 19, 2011; accepted September 22, 2011; published OnlineFirst September 28, 2011.

References

Deriving Valid Population–Based Cancer Survival Estimates in the Presence of Nonnegligible Proportions of Cancers Notified by Death Certificates Only

Hermann Brenner and Bernd Holleczek


Updated versionAccess the most recent version of this article at:
doi:10.1158/1055-9965.EPI-11-0697

Cited articlesThis article cites 16 articles, 3 of which you can access for free at:
http://cebp.aacrjournals.org/content/20/12/2480.full.html#ref-list-1

E-mail alertsSign up to receive free email-alerts related to this article or journal.

Reprints and SubscriptionsTo order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

PermissionsTo request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.