Cancer Epidemiology 59 (2019) 232–235

Contents lists available at ScienceDirect

Cancer Epidemiology

journal homepage: www.elsevier.com/locate/canep

Commentary

Improving the reporting of cancer-specific mortality and survival in research using cancer registry data

Michel D. Wissing*, Zoë R. Greenwald, Eduardo L. Franco

Department of Oncology, McGill University, Montréal, QC, Canada

ARTICLE INFO

Keywords:
Cancer-specific mortality
All-cause mortality
Cancer-specific survival
Overall survival
Cancer registries
US SEER database

ABSTRACT

Population-based registries are increasingly used in cancer research. In such studies, cancer-specific mortality or survival is frequently used as the primary outcome. To determine whether a putative cancer was part of the causal chain of events leading to death, cancer registries primarily rely on death certificates. Hence, they depend on the subjective interpretation of information available to medical examiners at the time of death. Misclassification may occur: studies report misclassification of cancer as a cause of death in 15%–35% of death certificates based on evaluation by expert panels and/or autopsy reports. Further misclassification may occur when coding death causes in the cancer registry. Researchers should be aware of potential misclassification bias when using cancer registry data. Differential misclassification may bias the results towards or away from the null hypothesis, depending on whether there is relative over- or under-reporting of cancer-related deaths in one group. Strategies to improve reporting of cancer-specific survival/mortality include (1) describing the procedure used to identify cancer-specific deaths; (2) considering the use of multiple definitions of cancer-related deaths (strict/liberal definitions of cancer-specific deaths, and/or addition of relative survival as an outcome); and (3) reporting cancer-specific survival/mortality together with the objectively measured parameters overall survival or all-cause mortality.

1. Introduction

The use of population-based registries for cancer research has grown exponentially in the past decades [1,2]. The most frequently used registry for cancer research is the Surveillance, Epidemiology, and End Results (SEER) program from the National Cancer Institute (NCI); its use in PubMed-indexed medical articles has increased by nearly tenfold since 2000, with almost 800 original research articles based on SEER data published in 2017 alone [3]. The SEER program has been collecting data from the United States (US) population since 1973, and its registry network currently encompasses data on approximately 28% of the US population. Every year, the American Cancer Society calculates the expected incidence and mortality rates for all cancers, primarily based on data from the SEER program [4]; these data are widely used for teaching, research, and policymaking.

In addition to being the mainstay of cancer surveillance, cancer registries have many advantages, as they provide academic researchers easy access to large datasets, enabling large-scale correlative studies and the investigation of rare diseases and outcomes [2]. It is a major information source for studies testing hypotheses that are too expensive or unethical for randomized clinical trials, and to evaluate practice outside the strictly guideline-regulated clinical trial setting. Cancer registries are constantly evolving to improve their scientific use; for example, EUROCOURSE aims to reduce fragmentation of cancer registries within Europe [5,6]. Hence, it can be expected that future scientific use of cancer registries will only expand.

It is imperative that researchers working with cancer-registry data are aware of the potential limitations of such data. In this commentary, we would like to underscore the limitations of cancer-specific mortality and survival when using cancer-registry data. We provide several suggestions for improving reporting of these outcome parameters. We often use the SEER program as an example, but the limitation applies to most, if not all cancer registries, and to other observational studies too.

2. Determining cancer-specific mortality and survival in cancer registry data

In cancer research, defining a valid outcome parameter as an indicator of clinical relevance has its challenges. Overall survival (or its related outcome measure, all-cause mortality) is generally considered

* Corresponding author at: Division of Cancer Epidemiology, Department of Oncology – McGill University, 5100 Boulevard de Maisonneuve Ouest, Office 720, Montreal, QC, H4A 3T2, Canada.

E-mail address: michel.wissing@mcgill.ca (M.D. Wissing).

https://doi.org/10.1016/j.canep.2019.02.004
Received 6 October 2018; Received in revised form 4 February 2019; Accepted 7 February 2019
Available online 02 March 2019

the gold standard for the demonstration of clinical benefit in cancer patients, as it is measured objectively, it is a direct clinical benefit, and it can be determined unequivocally [7]. However, when the studied disease has relatively high survival rates, and/or the studied population has high mortality rates from other causes (i.e., all deaths unrelated to the studied disease or intervention), differences in survival between groups due to the intervention may be overshadowed by discrepancies in deaths due to causes unrelated to the disease or intervention. This limitation is particularly evident in non-randomized cohort and case-control studies, in which imbalances in baseline characteristics are likely to occur. Therefore, cancer-specific mortality may be a more informative outcome in such populations, as it directly measures the effect of the intervention on disease-specific outcome [8,9]. Cancer-specific mortality and survival are frequently used as primary outcomes in studies using cancer registry data, often without concomitant reporting of all-cause mortality or overall survival.

Any outcome parameter other than overall survival has a level of subjectivity. Unlike overall survival, whose definition precludes the need to ponder on the cause, defining cancer-specific mortality and survival requires ascertaining the cancer in question or its consequential clinical outcomes as the cause of death. Such interpretation is dependent on the individuals evaluating the cause of death, as well the data available to the examiner, such as autopsy reports and a full medical history [10]. Furthermore, assessing cancer-specific deaths in settings where interventions are unblinded may lead to biases, as previously reported in cancer screening data. Black et al. reported that in various large randomized trials of cancer screening, use of cancer-specific mortality led to paradoxical interpretation of the effects of the screening interventions [11]. The authors ascribed the issue to misclassification of the cause of death, both as deaths from the disease had been attributed to other causes, and as deaths from other causes in screened patients had falsely been attributed to the disease detected upon screening. A well-known example of the latter is the transient increase in prostate cancer-related deaths in the US in the late 1980s and early 1990s after the introduction of screening for prostate-specific antigen.

Optimally, causes of deaths would be determined by in-depth assessment by a blinded panel of experts who have access to the full medical history of the patient, and if conducted, autopsy reports. However, such evaluations are generally limited to large clinical trials as this requires a massive investment in resources. The costs of conducting such analyses in cancer registries would outweigh the gain in accuracy of the reported cause of death. Furthermore, since cancer registry data is available to the public, it is crucial that confidentiality of the patients included in the registry is protected. Therefore, most cancer registries, including the SEER program, and most observational studies in general, collect mortality statistics from death certificates [12–14]. On death certificates, medical examiners report the direct cause of death. They may also report underlying causes of death, as well as other significant conditions that did not directly contribute to the chain of events leading to death, if known. If two underlying causes of death added to the direct cause of death, the least direct cause of death is reported as another significant condition. Such evaluations are increasingly complicated when patients have substantial comorbidity, especially when the medical examiner did not treat the patient during his/her lifetime and when no autopsies are conducted. When comparing causes of death as registered on death certificates with those in autopsy reports, major disagreements occurred in 15 to 35% of cancer cases [15,16]. Studies report similar discrepancy rates when comparing causes of death between death certificates and blinded evaluation by an expert panel: for instance, a recent Norwegian report found that over- and underreporting of prostate cancer deaths on death certificates occurred in 33% and 19% of all prostate cancer patients, respectively [17–19].

3. Improved reporting of cancer-specific mortality and survival using cancer registry data

Reporting on cancer-specific deaths can be improved by enhanced data collection, allowing more precise determination of the cause of death. For example, by expanding its coverage, SEER reduces missing information due to migration patterns within the US. Detailed information on administered therapies may provide useful information on the course of diseases. If registries such as SEER could collect information on cancer recurrence, which would be an interesting endpoint in itself, it would decrease inaccuracies in cause of death reporting: a patient without recurrence is less likely to have a cancer-related death relative to a patient who has been admitted to the hospital multiple times due to cancer recurrence.

When studying cancer registry data, the definition of cancer-specific mortality deserves consideration. The SEER program developed and validated an algorithm to optimize the use of death certificates to report cancer-specific deaths. Using this algorithm, cancer-specific deaths are not limited to patients whose death certificates report the cancer to be part of the causal chain leading to death, but to any cause of death that is likely related to the particular cancer, or a consequence of the cancer diagnosis. For instance, when a patient is diagnosed with endometrial cancer as a primary tumor, all deaths related to the uterus and ovaries are considered deaths related to endometrial cancer. This method is not perfect: the NCI acknowledges that mislabeling may occur, for example when a metastasis is not attributed to its primary tumor. Nevertheless, this algorithm will make ascertainment more conservative by improving the chances that cancer-related deaths will not be missed [14,20]. On the other hand, if researchers aim to prevent misclassification of any deaths unrelated to the studied cancer as cancer-related deaths, a stricter definition may be chosen. In that case, the value of other-cause mortality is limited, as the strict definition leads to misclassification of cancer-related deaths as other-cause mortality [9,21]. Thus, researchers should carefully consider misclassification due to their definition of cancer-related deaths, which may lead to under- and/or over-reporting of cancer-related deaths, depending on the definition chosen. The definition must be reported in the methods, as well as the rationale for any deviation from the SEER-definition of cancer-specific mortality.

Potential misclassification bias may occur when using cancer registry data. In comparative studies, non-differential misclassification will generally result in bias towards the null hypothesis. Importantly, differential misclassification may bias the results in either direction, depending on whether there is relative over- or under-reporting in one group. For this reason, standard reporting of cancer-specific mortality or survival should be done in parallel with all-cause mortality or overall survival, respectively. While differences in the magnitude of the effect in these two types of outcome measures are to be expected because of the dilution effect that comes from using all-cause mortality or overall survival [22], this approach provides the opportunity of verifying whether the directionality of the effects using these two outcomes is coherent with each other and with the underlying hypotheses.

A valid approach would be to report cancer-specific survival or mortality using different definitions of cancer-related deaths, akin to a sensitivity analysis, thus examining multiple scenarios for consistency of findings with underlying hypotheses. The most conservative definition would only include all deaths that have been directly attributed to the cancer being studied. A more liberal definition would be the SEER definition of cancer-specific mortality. Furthermore, opinions from clinicians and researchers specialized in the cancer of interest could be used to evaluate which codes in the International Statistical Classification of Diseases and Related Health Problems are definitely, plausibly, and/or possibly related to the studied cancer. Such definitions would be cancer- and case-specific: for example, bone metastases commonly occur in breast and prostate cancer, but are relatively uncommon in skin cancer patients; deaths related to bone disorders are therefore more likely to be related to the primary tumor in breast or prostate cancer patients than in skin cancer patients. Similarly, suicide is more likely cancer-related in a patient who just received a cancer diagnosis than in a patient who already used antidepressants prior to the cancer diagnosis and/or has been living with low-grade cancer for years without evidence of recurrence [23]. Finally, when possible, one should consider the reporting of relative survival in addition to cancer-specific mortality/survival. In recent years, use of this measure has decreased relatively to cancer-specific mortality/survival in studies using cancer registries such as SEER [24,25]. In relative survival, the observed survival rate is corrected on the basis of the expected survival rate in a comparable population without the disease. This measure has its own limitations, for example as non-matching life tables may introduce bias, particularly when comparing subgroups [26–28]. Nevertheless, its advantage is that it does not require the subjective interpretation of death causes. Therefore, concomitant reporting of cancer-specific survival and relative survival may complement each other.

To illustrate the benefit of this approach, in a recent observational study, we evaluated hormonal treatment among young women with early-stage, favorable-histology endometrial cancer as compared to surgery using SEER data [21]. When using the SEER definition of cancer-specific mortality, other-cause mortality was significantly higher in our population than the average mortality rate in US census data for an age-matched population, and other-cause mortality differed significantly between treatment groups, suggesting misclassification. When three physicians blindly scored which deaths were plausibly related to endometrial cancer, cancer-specific mortality increased substantially while other-cause mortality did not differ between groups and approximated that of the general population. Hence, by assessing a range of survival outcome definitions, we were able to obtain a more nuanced approach to evaluate therapeutic effects.

The aforementioned recommendations for cancer-registry studies can be generalized to all observational studies that use death certificates to determine the cause of death in patients. However, if medical files and other data sources are available in observational studies to ascertain likely cause of death, they should be used to obtain additional insights concerning cause of death, thus permitting assessment of any underlying associations using the full complement of outcome measures available, thereby improving the validity of interpretations.

4. Conclusions

Although at face value, cancer-specific mortality and survival may seem to provide the construct validity needed for epidemiologic surveillance or intervention studies of cancer prevention or therapy, there is misattribution of the cause of death to an underlying cancer, which may lead to serious biases. The potential for error may be larger in studies based on cancer registry data, in which attribution of cause is done via algorithms, thereby precluding the granularity that is possible when adjudication of the cause of death is completed by medical experts blinded to any intervention under study. We emphasize the need to explain in detail how cancer-specific deaths are determined in studies. Authors should consider conducting sensitivity analyses of cancer registry data using strict and liberal definitions for cancer-specific deaths. Concomitant reporting of relative survival and cancer-specific survival may strengthen researchers' findings. We also propose that cancer-specific mortality and survival should always be reported together with all-cause mortality and overall survival, respectively, as this would assist in identifying differential misclassification between prognostic factors or treatments being compared.

Financial support

MDW is funded by a Fellowship Award (201611MFE-382677-269048) from the Canadian Institutes of Health Research.

Conflicts of interests

The authors declare no potential conflicts of interest for this manuscript.

Author statement

All three authors were involved in the conception and design of the commentary. MDW drafted the manuscript, which was critically revised by the other authors before submission. All authors approve the final version that is submitted.

References

[1] M.S. Thong, F. Mols, K.D. Stein, T. Smith, J.W. Coebergh, L.V. van de Poll-Franse, Population-based cancer registries for quality-of-life research: a work-in-progress resource for survivorship studies? Cancer 119 (Suppl. 11) (2013) 2109–2123.
[2] H.E. Rice, B.R. Englum, B.C. Gulack, O.O. Adibe, E.T. Tracy, S.G. Kreissman, J.C. Routh, Use of patient registries and administrative datasets for the study of pediatric cancer, Pediatr. Blood Cancer 62 (9) (2015) 1495–1500.
[3] PubMed Search Engine, US National Library of Medicine, National Institutes of Health Web site, 2019 Accessed 2 February 2018 https://www.ncbi.nlm.nih.gov/pubmed/?term = SEER + cancer.
[4] R.L. Siegel, K.D. Miller, A. Jemal, Cancer statistics, 2018, CA Cancer J. Clin. 68 (1) (2018) 7–30.
[5] R. Zanetti, I. Schmidtmann, L. Sacchetto, F. Binder-Foucard, A. Bordoni, D. Coza, S. Ferretti, J. Galceran, A. Gavin, N. Larranaga, D. Robinson, L. Tryggvadottir, E. Van Eycken, V. Zadnik, J.W. Coebergh, S. Rosso, Completeness and timeliness: cancer registries could/should improve their performance, Eur. J. Cancer 51 (9) (2015) 1091–1098.
[6] S.F. Altekruse, G.E. Rosenfeld, D.M. Carrick, E.J. Pressman, S.D. Schully, L.E. Mechanic, K.A. Cronin, B.Y. Hernandez, C.F. Lynch, W. Cozen, M.J. Khoury, L.T. Penberthy, SEER cancer registry biospecimen research: yesterday and tomorrow, Cancer Epidemiol. Biomark. Prev. 23 (12) (2014) 2681–2687.
[7] R. Pazdur, Endpoints for assessing drug activity in clinical trials, Oncologist 13 (Suppl. 2) (2008) 19–21.
[8] W. van de Water, C. Markopoulos, C.J. van de Velde, C. Seynaeve, A. Hasenburg, D. Rea, H. Putter, J.W. Nortier, A.J. de Craen, E.T. Hille, E. Bastiaannet, P. Hadji, R.G. Westendorp, G.J. Liefers, S.E. Jones, Association between age at diagnosis and disease-specific mortality among postmenopausal women with hormone receptor-positive breast cancer, JAMA 307 (6) (2012) 590–597.
[9] P. Dell'Oglio, K. Boehm, V. Trudeau, Z. Tian, A. Larcher, S.R. Leyh-Bannurah, M. Moschini, U. Capitanio, S.F. Shariat, A. Briganti, F. Montorsi, F. Saad, P.I. Karakiewicz, Survival after conservative management versus external beam radiation therapy in elderly patients with localized prostate cancer, Int. J. Radiat. Oncol. Biol. Phys. 96 (5) (2016) 1037–1045.
[10] J.T. Santoso, C.M. Lee, J. Aronson, Discrepancy of death diagnosis in gynecology oncology, Gynecol. Oncol. 101 (2) (2006) 311–314.
[11] W.C. Black, D.A. Haggstrom, H.G. Welch, All-cause mortality in randomized trials of cancer screening, J. Natl. Cancer Inst. 94 (3) (2002) 167–173.
[12] Y. Song, J.E. Chavarro, Y. Cao, W. Qiu, L. Mucci, H.D. Sesso, M.J. Stampfer, E. Giovannucci, M. Pollak, S. Liu, J. Ma, Whole milk intake is associated with prostate cancer-specific mortality among U.S. male physicians, J. Nutr. 143 (2) (2013) 189–196.
[13] A.P. Polednak, Trend (1999–2009) in U.S. death rates from myelodysplastic syndromes: utility of multiple causes of death in surveillance, Cancer Epidemiol. 37 (5) (2013) 569–574.
[14] A.P. Polednak, Inaccuracies in oral cavity-pharynx cancer coded as the underlying cause of death on U.S. death certificates, and trends in mortality rates (1999–2010), Oral Oncol. 50 (8) (2014) 732–739.
[15] C. Percy, E. Stanek 3rd, L. Gloeckler, Accuracy of cancer death certificates and its effect on cancer mortality statistics, Am. J. Public Health 71 (3) (1981) 242–250.
[16] T. Kircher, J. Nelson, H. Burdo, The autopsy as a measure of accuracy of the death certificate, N. Engl. J. Med. 313 (20) (1985) 1263–1269.
[17] S. Loffeler, A. Halland, H. Weedon-Fekjaer, A. Nikitenko, C.L. Ellingsen, E.S. Haug, High Norwegian prostate cancer mortality: evidence of over-reporting, Scand. J. Urol. 52 (2) (2018) 122–128.
[18] E.L. Turner, C. Metcalfe, J.L. Donovan, S. Noble, J.A. Sterne, J.A. Lane, E. I Walsh, E.M. Hill, L. Down, Y. Ben-Shlomo, S.E. Oliver, S. Evans, P. Brindle, N.J. Williams, L.J. Hughes, C.F. Davies, S.Y. Ng, D.E. Neal, F.C. Hamdy, P. Albertsen, C.M. Reid, J. Oxley, J. McFarlane, M.C. Robinson, J. Adolfsson, A. Zietman, M. Baum, A. Koupparis, R.M. Martin, Contemporary accuracy of death certificates for coding prostate cancer as a cause of death: Is reliance on death certification good enough? A comparison with blinded review by an independent cause of death evaluation committee, Br. J. Cancer 115 (1) (2016) 90–94.
[19] N. Horeweg, R.J. van Klaveren, H.J. Groen, J.W. Lammers, C. Weenink, K. Nackaerts, W. Mali, M. Oudkerk, H.J. de Koning, Blinded and uniform cause of death verification in a lung cancer CT screening trial, Lung Cancer 77 (3) (2012) 522–525.
[20] N. Howlader, L.A. Ries, A.B. Mariotto, M.E. Reichman, J. Ruhl, K.A. Cronin, Improved estimates of cancer-specific survival rates from population-based data, J.

Natl. Cancer Inst. 102 (20) (2010) 1584–1598.
[21] Z.R. Greenwald, L.N. Huang, M.D. Wissing, E.L. Franco, W.H. Gotlieb, Does hormonal therapy for fertility preservation affect the survival of young women with early-stage endometrial cancer? Cancer 123 (9) (2017) 1545–1554.
[22] E.L. Franco, Perfect is the enemy of good: going to the war on cancer with less evidence than we could have, Epidemiol. Methods 4 (1) (2015) 23–28.
[23] S. Misono, N.S. Weiss, J.R. Fann, M. Redman, B. Yueh, Incidence of suicide in persons with cancer, J. Clin. Oncol. 26 (29) (2008) 4731–4738.
[24] PubMed Search Engine, US National Library of Medicine, National Institutes of Health Web site, 2019 Accessed 13 December 2018 https://www.ncbi.nlm.nih.gov/pubmed/?term=((cancer-specific+mortality)+OR+(cancer-specific+survival))+AND+SEER.
[25] C.M. Wheeler, S.K. Kjaer, K. Sigurdsson, O.E. Iversen, M. Hernandez-Avila, G. Perez, D.R. Brown, L.A. Koutsky, E.H. Tay, P. Garcia, K.A. Ault, S.M. Garland, S. Leodolter, S.E. Olsson, G.W. Tang, D.G. Ferris, J. Paavonen, M. Steben, F.X. Bosch, J. Dillner, E.A. Joura, R.J. Kurman, S. Majewski, N. Munoz, E.R. Myers, L.L. Villa, F.J. Taddeo, C. Roberts, A. Tadesse, J. Bryan, L.C. Lupinacci, K.E. Giacoletti, M. James, S. Vuocolo, T.M. Hesley, E. Barr, The impact of quadrivalent human papillomavirus (HPV; types 6, 11, 16, and 18) L1 virus-like particle vaccine on infection and disease due to oncogenic nonvaccine HPV types in sexually active women aged 16-26 years, J. Infect. Dis. 199 (7) (2009) 936–944.
[26] T. Blakely, M. Soeberg, K. Carter, R. Costilla, J. Atkinson, D. Sarfati, Bias in relative survival methods when using incorrect life-tables: lung and bladder cancer by smoking status and ethnicity in New Zealand, Int. J. Cancer 131 (6) (2012) E974–82.
[27] A. Okuyama, A. Shibata, H. Nishimoto, Critical points for interpreting patients' survival rate using cancer registries: a literature review, J. Epidemiol. 28 (2) (2018) 61–66.
[28] D. Sarfati, T. Blakely, N. Pearce, Measuring cancer survival in populations: relative survival vs cancer-specific survival, Int. J. Epidemiol. 39 (2) (2010) 598–610.