EDITORIAL

Quality of cancer registry data: completeness of TNM staging and potential implications

Mette Søgaard[1,2]
Morten Olsen[1]

[1]Department of Clinical Epidemiology, Aarhus University, Aarhus, Denmark; [2]Department of Clinical Microbiology, Aalborg Hospital, Aarhus University Hospital, Aalborg, Denmark

Abstract: Population-based cancer registries play a key role in cancer monitoring, and their utility relies heavily on the completeness and validity of the registered data. It is crucial to reduce incomplete stage information and to strengthen the role of TNM stage as key variables in the Danish Cancer Registry and other cancer registries. In this respect, distinction in cancer registries between evidence that staging was not performed clinically versus truly missing data would be an important next step. The present studies examined the completeness of the TNM staging data. In future studies it will be important to evaluate the accuracy as well.

Keywords: Danish Cancer Registry, TNM staging data, completeness, validity

Introduction

Population-based cancer registries play a key role in cancer monitoring. The primary objective of the registries is to provide continuously and systematically updated information on cancer incidence, prevalence, mortality, and survival of cancer in the underlying countries or regions.[1,2] However, the applications and use of cancer registry data have evolved beyond routine statistics on cancer occurrence and outcome. The modern cancer registries provide data to a number of other areas within cancer monitoring which, according to the World Health Organization,[3] includes:

- assessing the current magnitude of the cancer burden and its likely future evolution;
- providing a basis for research on cancer causes and prevention;
- providing information on prevalence and trends in risk factors; and
- monitoring the effects of early detection/screening, treatment, and palliative care.

Still, the utility of cancer registries and their ability to address these tasks rely heavily on the completeness and validity of the registered data.

Cancer registry data: completeness of TNM staging

The studies reported in the present supplement to *Clinical Epidemiology* examined the completeness and pattern of missing information on TNM stage for selected cancer sites recorded in the Danish Cancer Registry (DCR), which is the oldest nationwide cancer registry in the world. Cancer stage is a cornerstone in the management of cancer and has important uses in cancer monitoring. With a given TNM stage reflecting the severity and spread of cancer at the time of diagnosis, cancer stage is used to estimate prognosis, guide selection of primary or adjuvant treatment, and to evaluate

Correspondence: Mette Søgaard
Department of Clinical Epidemiology,
Aarhus University Hospital,
Sdr Skovvej 15, 9000 Aalborg,
Denmark
Tel +45 99 326 900
Fax +45 99 326 914
Email mette.soegaard@rn.dk

1

treatment results. In addition to this, secondary prevention through early detection and treatment is monitored, in part, by stage at time of diagnosis with, on average, earlier stage diagnoses over time indicating improved performance of the health care system.[4,5]

The DCR records incident cancer cases based on combined information from in- or outpatient hospital settings registered in the Danish National Patient Register, primary sector practitioners, the Danish Pathology Register, and death certificates.[6–9] Since 2004, notifications from hospitals and primary sector practitioners have been performed electronically, thereby integrating cancer registration into the patient administrative systems.[6] Information on cancer staging in the DCR has, since 2004, been recorded according to The American Joint Committee on Cancer's TNM classification.[6] The DCR is an important data resource for monitoring and research, but analyses of trends in stage-specific incidence and mortality are impaired if stage information is missing. Thus, it is important to examine the quality of the cancer registry data and to evaluate the characteristics of patients with unstaged cancer, including how they may differ from staged cases.[10] This will facilitate evaluation of the magnitude and direction of potential biases in studies using information on stage.

What can be learned from the reports in this supplement?

Overall, the reports demonstrated a substantial variation in the completeness of TNM stage registration in the DCR according to cancer site, and the proportions of missing stage information were generally not trivial. In particular, complete information on stage was not available for two-thirds of prostate cancer patients and more than half of bladder cancer patients. The studies consistently report the highest proportions of unstaged cancer cases in patients of old age and/or high level of comorbidity. For example, information on TNM stage was available for 75% of patients with prostate cancer aged 0–39 years compared to only 11% of those aged 80 years and older. This indicates a huge disparity of cancer care according to age and comorbidity. However, the causes of missing stage data are likely multifactorial, including other characteristics than age or comorbidity. Besides incomplete diagnostic assessment related to age and comorbidity, potential explanations include:[11–13]

- Apparent signs of distant metastases, obviating the need for staging in treatment decisions.
- The patient and/or the family have chosen not to pursue further investigations.

- Incomplete documentation of stage by the physician, notably failure to report the final TNM stage after completed clinical work-up.
- Errors in coding of the retrieved data.
- Other potential inconsistencies in the recording process.

Potential implications

The findings of these reports have important implications. If a large proportion of cancer cases are missing stage information, the aspects of cancer monitoring that rely on this information will be compromised and thereby limit the inferences that can be made, for example, about the effect of early detection strategies. The findings also have methodological implications. Previous studies have often excluded unstaged cancers, categorized them with late-stage cancers, or analyzed them in a separate category. However, exclusion of patients with missing data not only wastes valuable data, but can also lead to invalid results if the excluded group is a selected subsample from the source population as indicated in basically all reports in this supplement. In addition, it is clearly unfortunate to combine unstaged cancer cases into a single group, if those patients are in fact distributed across all stage categories. Recently, more efficient and less biased alternatives to complete case analysis, such as multiple imputation or inverse probability weighting, have been developed and included in statistical software packages, thus enabling standardized imputation of missing data on stage in cancer studies.[14,15]

What are the next steps?

An important question is whether the findings identify a problem with the quality of care being provided to older persons. It is difficult to determine this based solely on calculations of the proportions of patients with missing data on stage. According to the Surveillance, Epidemiology, and End Results registers in the United States, unstaged cancer is defined as "cancer for which there is not enough information to indicate a stage."[16] As noted, this may reflect either the absence of clinical work-up to yield this information, or the failure of surveillance to capture clinical information that does exist. Identifying the reasons for missing staging data is warranted. It is likely that the compromised health status of some elderly cancer patients results in fewer diagnostic procedures and treatment options, which may well be the most ethically sound practice. However, there is also growing recognition that increasing segments of the older population can tolerate and may benefit from cancer treatment.

It is crucial to reduce incomplete stage information and to strengthen the role of TNM stage as key variables in the DCR and other cancer registries. In this respect, distinction in cancer registries between evidence that staging was not performed clinically versus truly missing data would be an important next step. The present studies examined the completeness of the TNM staging data. In future studies it will be important to evaluate the accuracy as well.

Acknowledgments

Disclosure

The authors report no conflicts of interest in this work.

References

1. Jensen OM, Parkin DM, MacLennan R, Muir CS, Skeet RG, editors. *Cancer Registration: Principles and Methods*. Lyon: International Agency for Research on Cancer; 1991.
2. Curado MP, Edwards B, Shin HR, et al, editors. *Cancer Incidence in Five Continents, Volume IX*. Lyon: International Agency for Research on Cancer; 2007.
3. World Health Organization. *National Cancer Control Programmes. Policies and Managerial Guidelines*, 2nd ed. Geneva: World Health Organization; 2002.
4. de Camargo Cancela M, Chapuis F, Curado MP. Abstracting stage in population-based cancer registries: the example of oral cavity and oropharynx cancers. *Cancer Epidemiol*. 2010;34(4):501–506.
5. Klassen AC, Curriero F, Kulldorff M, Alberg AJ, Platz EA, Neloms ST. Missing stage and grade in Maryland prostate cancer surveillance data, 1992–1997. *Am J Prev Med*. 2006;30(Suppl 2):S77–S87.
6. Gjerstorff ML. The Danish Cancer Registry. *Scand J Public Health*. 2011;39(Suppl 7):42–45.
7. Lynge E, Sandegaard JL, Rebolj M. The Danish National Patient Register. *Scand J Public Health*. 2011;39(Suppl 7):30–33.
8. Bjerregaard B, Larsen OB. The Danish Pathology Register. *Scand J Public Health*. 2011;39(Suppl 7):72–74.
9. Helweg-Larsen K. The Danish Register of Causes of Death. *Scand J Public Health*. 2011;39(Suppl 7):26–29.
10. Vach W, Blettner M. Biased estimation of the odds ratio in case-control studies due to the use of ad hoc methods of correcting for missing values for confounding variables. *Am J Epidemiol*. 1991;134(8):895–907.
11. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol*. 2005;58(4):323–337.
12. Worthington JL, Koroukian SM, Cooper GS. Examining the characteristics of unstaged colon and rectal cancer cases. *Cancer Detect Prev*. 2008;32(3):251–258.
13. Koroukian SM, Xu F, Beaird H, Diaz M, Murray P, Rose JH. Complexity of care needs and unstaged cancer in elders: a population-based study. *Cancer Detect Prev*. 2007;31(3):199–206.
14. Eisemann N, Waldmann A, Katalinic A. Imputation of missing values of tumour stage in population-based cancer registration. *BMC Med Res Methodol*. 2011;11:129.
15. Marshall A, Altman DG, Royston P, Holder RL. Comparison of techniques for handling missing covariate data within prognostic modelling studies: a simulation study. *BMC Med Res Methodol*. 2010;10:7.
16. National Cancer Institue. SEER: glossary of statistical terms. Available from: http://seer.cancer.gov/cgi-bin/glossary/glossary.pl. Accessed May 10, 2012.