The validity of self-reported cancer in an Australian population study

Loh, Venurs; Harding, Jessica; Koshkina, Vira; Barr, Elizabeth; Shaw, Jonathan; Magliano, Dianna

Published in:
Australian and New Zealand Journal of Public Health

DOI:
10.1111/1753-6405.12164

Published: 01/01/2014

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
The validity of self-reported cancer in an Australian population study

Venurs Loh,1 Jessica Harding,1,2 Vira Koshkina,1 Elizabeth Barr,1,3 Jonathan Shaw,1 Dianna Magliano1,2

Accurate and current estimates of cancer incidence and prevalence are required to quantify the health burden of cancer and to direct current and future allocation of health care funds.1 In recent decades, cancer registries have become the most accurate source of cancer data.2 However, the acquisition of cancer registry data is not always feasible for use in epidemiological studies, as data linkage to the registry can be costly and time consuming. In particular, there may be ethical restrictions on the use of cancer data or data may only be provided in an aggregated fashion. As an alternative, collection of cancer data by self-report may be a reasonable surrogate to cancer registry data. However, before self-reported cancer data can be used in research, it is important to know the accuracy of such data compared to that recorded on registries, as well as the patient characteristics that predict accurate self-reports.

The sensitivity of self-reported cancer data compared to cancer data obtained from a registry has been assessed in several studies in other developed countries, and ranges from 61% to 90%, varying considerably by cancer site. High sensitivities have been observed for self-reported breast (85.4-90%), prostate (78.9%), and lung cancers (74.1%), while low sensitivity has been observed for self-reported bowel cancer (16%-60.4%) and melanoma of the skin (36.9%).3,4,5 However, only a few large population-based studies have evaluated the characteristics of people who self-report wrongly.6,7 Furthermore, the findings from these studies may not be generalisable to the Australian context, as they only include women and were conducted in other countries. Given the differences in health care systems between countries, the validity of self-reported cancer may be different in Australia.

In Australia, cancer is a notifiable disease by law and detailed cancer data are routinely collected by the state and territory cancer registries. The cancer data are then collated by the Australian Cancer Database (ACD), a national registry established in 1983. The objective of this study is to determine the validity of self-reported all-cause and site-specific cancer in a national, population-based study of Australian men and women, compared to the ACD. If self-reported cancer data are shown to be sufficiently reliable, their use may be more efficient and cost-effective for the monitoring of cancer incidence in a range of research projects.

Methods

Study sample

The Australian Diabetes, Obesity and Lifestyle (AusDiab) study is a national population-based study of 11,247 Australian men and women aged 25 years or older. Methods

Abstract

Objective: The aim of this study is to determine the validity of self-reported cancer data by comparing it to the Australian Cancer Database (ACD).

Methods: Self-reported data were obtained from the Australian Diabetes, Obesity and Lifestyle (AusDiab) study, which were then linked to the ACD up until 31 December 2010. Positive predictive value, negative predictive value, sensitivity and specificity were calculated. Cohen’s kappa coefficient (κ) was also calculated to assess the agreement between self-reported cancer and the ACD. Logistic regression was used to examine the determinants associated with false negative and false positive reporting.

Results: The overall sensitivity of self-report cancer was 71.1%, and sensitivities showed great variation by cancer site. Higher sensitivities were observed for breast (90.7%), bowel (77.8%) and prostate (77.1%) cancers, whereas the lowest sensitivity was observed for melanoma of the skin (36.9%). Similarly, the kappa coefficient analysis showed substantial agreement for self-reported breast cancer (κ= 0.79) and moderate agreement for melanoma (κ= 0.45) against the ACD. Years since cancer diagnosis and older age were associated with false negative reporting and older age was associated with false positive reporting.

Conclusions and implications: The use of self-reported cancer to collect cancer outcomes has varying reliability, depending on cancer type and population. The findings presented here may assist medical researchers in making informed decisions when conducting research using self-reported cancer data in Australia where the acquisition of registry data is not feasible.

Key words: sensitivity, specificity, cancer, self-report
and response rates have been described previously.6 In 1999/2000, census collector districts (clusters) were randomly selected from each of the six Australian states and the Northern Territory (total =42). Participants were interviewed at the household and then were invited to attend a health examination in 2005 and all participants were invited to complete a follow-up postal health survey conducted in May 2010. Of the 11,247 participants at baseline, 915 (8.1%) were ineligible (214 were excluded due to chronic illnesses and 701 were deceased), and 1,442 (12.8%) were lost to follow-up (593 refused to be contacted, 759 had no valid contact details, and 90 were living overseas). The remaining 8,890 participants were asked by mail to complete a questionnaire pertaining to their history of previous cancer diagnosis (Have you ever been told by a doctor or nurse that you have cancer?) and were also asked for the type of cancer and the date of diagnosis for each cancer if more than one was reported.

**Australian Cancer Database (ACD)**

Each Australian state and territory maintains a cancer registry that provides information on incidence of and survival from cancer. The registry data are maintained by the Australian Institute of Health and Welfare (AIHW). It is a statutory requirement in all Australian states and territories for public and private hospitals, departments of radiation oncology, nursing homes, pathology laboratories, outpatient departments and day procedure centres to notify the registry of any cases of malignant neoplasms. Non-melanoma skin cancer is not a notifiable disease in the ACD and was therefore excluded from the current analysis. The ACD data are supplemented with information from hospital inpatient notifications and, in some states, on an ad hoc basis via links to other cancer registries as well as the National Death Index (NDI), if it is known or suspected that the person with cancer has moved to another state. All recorded cancers were coded from the International Classification of Diseases, Ninth revision (ICD-9) to Tenth revision (ICD-10) as appropriate.

To ascertain the validity of self-reported cancer outcomes, all 11,247 AusDiab participants were linked to the ACD from 1983 up until 31 December 2010. The AIHW uses an internally developed software system known as Record Matcher (REMA) for linkages to the NDI and ACD. REMA applies the general framework of Fellegi and Sunter,10,11 and includes a number of standard extensions to this framework, such as frequency-based agreement weights for names and approximate string comparators. REMA uses a number of linkage passes to give matches the best chance of being linked. Matching was performed using computer software (SAS Enterprise Guide 5.1 software) that utilises an ‘in house’ linkage algorithm of last name, first name, second name, third name, gender and date of birth to locate ‘linkage pairs’.

Potential ‘linkage pairs’ were given a weight with higher values reflecting a greater likelihood of a correct match. Based on clerical review of a sample of these matches, it is expected that matches with a weighting of 22.19 (low), 26.19 (medium) and 31.21 (high) would correspond to a link accuracy (positive predictive value, PPV) of 87.57%, 97.92% and 100%, respectively.12 For this study, we chose a medium cut-off point with a lower bound of 26.19 and predictive value of 97.92%. Possible matches from the ACD with a linkage weight below 26.19 were assumed to be non-matches.

**Statistical analysis**

Positive predictive value (PPV), negative predictive value (NPV), sensitivity and specificity were calculated for all cancers using the ACD as the gold standard. In addition, separate sensitivity estimates for breast, bowel, prostate and melanoma skin cancers were calculated, as these sub-types represent the most common cancers in Australia. Participants (3.1%) who reported ‘skin cancer’ were excluded as non-melanoma skin cancer is not a notifiable disease. Sensitivity, specificity, PPV and NPV for breast and prostate cancer were calculated among women and men, respectively. In addition, Cohen’s kappa coefficient (k) was calculated to provide a measure of agreement between self-reported data and ACD while accounting for the agreement expected by chance. False positive reporting occurred when a cancer was self-reported by the participant in the absence of a cancer notification by the ACD and false negative reporting occurred in the absence of a self-reported cancer with a cancer notification in the ACD. Logistic regression was used to examine the determinants associated with false negative and false positive self-reporting, which included: age, gender, educational level (≤6 or >6 years) and years since cancer diagnosis (only for false negative reporting). False negatives were compared to true positives, false positives and true negatives; and false positives were compared to true positives, false negatives and true negatives. All statistical analysis was performed using STATA version 12.1.

**Results**

Among the 8,890 participants who were invited to complete the survey, 7,352 (82.7%) returned their survey and responded to the cancer questions. A total of 636 (8.6%) reported at least one confirmed cancer diagnosis and, among those, 606 (98.3%) indicated cancer type and 349 (54.8%) reported a date of cancer.

The overall sensitivity of self-reported cancer was 71.1% (Table 1). Estimates of sensitivity varied considerably by cancer site and were highest for breast (90.7%), followed by bowel (77.8%) and prostate cancer (77.1%) with the lowest sensitivity for melanoma of the skin (36.9%). The specificity of self-reported cancer was 97.0%, 99.0%, 99.6% and 98.9% for breast, bowel, prostate and melanoma, respectively. The percentage agreement (Kappa coefficient) for all cancers was 94.7% (k = 0.65). The percentage agreements for site-specific cancers were 99.3% (k = 0.79), 99.4% (k = 0.73), 99.2% (k = 0.73) and 98.3% (k = 0.45) for breast, bowel, prostate cancer and melanoma, respectively. Among those with a reported date, 84.2% were within one year of the registry date.

There were 170 (2.3%) participants who gave false negative reports and a further 218 (2.9%) who gave false positive reports. The predictors of false negative and false positive reporting of cancers are presented in Table 2. Older age was a significant predictor for both false negative and false positive reporting, even after adjusting for years since cancer diagnosis. Additionally, there was a 4% increased risk of false negative reporting for each year since cancer diagnosis. Gender and educational level were not related to false positive or false negative reporting.

**Discussion**

Using the ACD as the gold standard, we show that the sensitivity of self-reported cancer is modest (71.1%), and varies considerably by cancer type with the highest sensitivity being for breast cancer (90.7%) and the lowest for melanoma (36.9%). Our results also show a uniformly high specificity (≥97.0%) overall and for each site-specific cancer, as
Cancer

that breast cancer is more frequently
prostate cancer, which may be confusing to
do not give a clear-cut diagnosis of
an elevated prostate surface antigen (PSA)
procedures are ambiguous.14,15 For example,
reported than cancers whose diagnostic
breast cancer, are more likely to be self-
with clear-cut diagnostic criteria, such as
study. It has been suggested that cancers
There are several reasons that may explain
the reporting patterns observed in our study.
It has been suggested that cancers
with clear-cut diagnostic criteria, such as
breast cancer, are more likely to be self-
reported than cancers whose diagnostic
procedures are ambiguous.14,15 For example,
an elevated prostate surface antigen (PSA)
does not give a clear-cut diagnosis of
prostate cancer, which may be confusing to
patients. Other studies16,17 have reported
that breast cancer is more frequently
disclosed to patients compared to other
cancer types due to greater public awareness
and destigmatisation of breast cancer in
the community. Similarly, this study shows
a moderately high sensitivity for prostate
cancer, possibly reflecting an increased
public awareness of prostate cancer within
the last decade.18 Our study demonstrated
much higher sensitivity for bowel cancer
as compared to previous studies in Spain
(17.4%)19 and Japan (14%).20 This may be
due to the nation-wide implementation of
bowel cancer awareness by health authorities
as it is the second most common cancer in
Australia.13
Unlike breast, bowel and prostate cancer,
melanoma is usually treated in an outpatient
setting and therefore this cancer is less
likely to be reported to the ACD than other
cancers.1 In addition, the low sensitivity of
self-reported skin cancer in our study could
be due to participants’ inability to accurately
report the type of skin cancer. According to
Aitken et al.21 the word ‘melanoma’ is
still poorly understood by the public. Our
data showed that several participants had
a melanoma registered on the ACD but
were not included in this analysis because
their questionnaire response noted only
’skin cancer’ and it was therefore not known
if this was a melanoma or non-melanoma
skin cancer. This may have resulted in an
underestimation of melanoma and thus a
decline in sensitivity.
Further, advancing age at the time of
questionnaire completion was positively
associated with false negative and false
positive reporting in our study. Previous
studies14,15 reported that the main reason for
false negative and false positive reporting
among older persons was declining cognitive
functioning. For some, cancer may also
be perceived as a ‘taboo’ subject that is
associated with fear and stigma, thus making
difficult for doctors to acknowledge and
discuss their diagnosis.22 Alternatively, it
could also be due to poor understanding of
their cancer diagnosis.23 Another explanation
could be that people tend to forget,
regardless of age or declining cognitive
function. For example, participants who have
been diagnosed with an in situ cancer may
not know whether their cancer was in fact
a frank cancer. This kind of confusion may
influence individual’s ability to accurately
report their cancer.
A longer period of time between cancer
diagnosis and the completion of the
questionnaire is associated with less accurate
self-reporting of cancer. The implication
of this is that accuracy can be improved if
self-reported questionnaire was performed
prospectively, the time between having been
diagnosed with the cancer and being asked
about the cancer would be shorter and thus
participants would be more likely to recall
their event and the validity of self-report
would most likely be much better. However,
our result contradicts two other similar
studies,7,19 which reported that recall was
more accurate in those with a longer time
between cancer diagnosis and self-report
of questionnaire. The reasons for this are
not known; it has been suggested that the
inconsistency of findings may reflect the
different methods employed to ascertain self-
reported cancer diagnosis.
This validation study has offered a unique
opportunity to overcome some of the
limitations encountered in previous studies.
First, the study uses a large population-
based Australian sample, and includes both
men and women. Other validation studies

Table 1: PPV, NPV, sensitivity and specificity of self-reported cancer.

<table>
<thead>
<tr>
<th>Cancer</th>
<th>True positive reports (N)</th>
<th>Sensitivity (%)</th>
<th>Positive predictive value (%)</th>
<th>True negative reports (N)</th>
<th>Specificity (%)</th>
<th>Negative predictive value (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All sites</td>
<td>418</td>
<td>71.1</td>
<td>65.7</td>
<td>6,546</td>
<td>97.0</td>
<td>97.5</td>
</tr>
<tr>
<td>Breast a</td>
<td>98</td>
<td>90.7</td>
<td>72.1</td>
<td>3,931</td>
<td>99.0</td>
<td>99.7</td>
</tr>
<tr>
<td>Bowel</td>
<td>63</td>
<td>77.8</td>
<td>70.0</td>
<td>7,244</td>
<td>99.6</td>
<td>99.7</td>
</tr>
<tr>
<td>Prostate b</td>
<td>84</td>
<td>77.1</td>
<td>70.0</td>
<td>3,130</td>
<td>98.9</td>
<td>99.2</td>
</tr>
<tr>
<td>Melanoma</td>
<td>52</td>
<td>36.9</td>
<td>60.5</td>
<td>7,177</td>
<td>99.5</td>
<td>98.8</td>
</tr>
</tbody>
</table>

a. Women only
b. Men only

well as positive predictive values ranging
from 60-72%, the highest being for breast
cancer. This result is supported by the Kappa
agreement analysis, with self-reported breast
cancer showing substantial agreement of 0.79
with the ACD and self-reported melanoma
showing moderate agreement of 0.45.
Advancing age at the time of questionnaire
completion is a significant predictor of both
false negative and false positive self-reporting
in our study. In addition, years since cancer
diagnosis is also positively associated with
false negative self-reporting in this study.
There are several reasons that may explain
the reporting patterns observed in our study.
It has been suggested that cancers
with clear-cut diagnostic criteria, such as
breast cancer, are more likely to be self-
reported than cancers whose diagnostic
procedures are ambiguous.14,15 For example,
an elevated prostate surface antigen (PSA)
does not give a clear-cut diagnosis of
prostate cancer, which may be confusing to
patients. Other studies16,17 have reported
that breast cancer is more frequently
disclosed to patients compared to other
cancer types due to greater public awareness
and destigmatisation of breast cancer in
the community. Similarly, this study shows
a moderately high sensitivity for prostate
cancer, possibly reflecting an increased
public awareness of prostate cancer within
the last decade.18 Our study demonstrated
much higher sensitivity for bowel cancer
as compared to previous studies in Spain
(17.4%)19 and Japan (14%).20 This may be
due to the nation-wide implementation of
bowel cancer awareness by health authorities
as it is the second most common cancer in
Australia.13
Unlike breast, bowel and prostate cancer,
melanoma is usually treated in an outpatient
setting and therefore this cancer is less
likely to be reported to the ACD than other
cancers.1 In addition, the low sensitivity of
self-reported skin cancer in our study could
be due to participants’ inability to accurately
report the type of skin cancer. According to
Aitken et al.21 the word ‘melanoma’ is
still poorly understood by the public. Our

Table 2: Predictors of false negative and false positive reporting of cancer.

<table>
<thead>
<tr>
<th>Variable</th>
<th>False negative</th>
<th>p-value</th>
<th>False positives</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>0.69 (0.48-1.01)</td>
<td>0.052</td>
<td>0.91 (0.69-1.19)</td>
<td>0.490</td>
</tr>
<tr>
<td>Age (per year)</td>
<td>1.02 (1.01-1.03)</td>
<td>0.019</td>
<td>1.03 (1.02-1.04)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Years since diagnosis c</td>
<td>1.04 (1.02-1.07)</td>
<td>&lt;0.001</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Educational level:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None or ≤5 years</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>&gt;6 years</td>
<td>0.77 (0.53-1.11)</td>
<td>0.171</td>
<td>1.04 (0.78-1.38)</td>
<td>0.779</td>
</tr>
</tbody>
</table>

Data were based on 7352 participants who responded to the survey on past history of cancer diagnosis. False negatives (n=170) were compared to true positives, false positives and true negatives (n=7134).
a. Adjusted for age (years), gender, years since cancer diagnosis and education (≤6 or >6 years).
b. OR, odds ratio
c. Defined as the interval between date of cancer diagnosis from ACD and date of questionnaire. Years since diagnosis was also analysed as a categorical variable (<4, 4-<7, 7-<11, 11-<17, >17 years) and was also significantly associated with false negative (OR=1.26 (1.11-1.44), p for trend =<0.0001). Data only available for false negative reporting as no dates were recorded in the ACD for the false positive reporting group.
Conduct conducted in Australia reporting accuracy of self-reported cancer in population-based registries has primarily focused on either a cohort of elderly women in a specific state or specific cancer types in specific states. Second, the ACD has nation-wide coverage of all malignant cancers since 1983 and previous reports show that this data is of high quality. Despite these strengths, this study is subject to potential limitations. First, probabilistic data linkage is limited by errors in personal identifiers such as names and dates of birth. However, the linkage software largely accommodates minor spelling differences and errors in dates, thus minimising errors where possible. Second, the self-reporting of date of cancer was missing in 45% of cases; however, when it was present, it was relatively accurate.

Conclusions and implications
In summary, the use of self-reported cancer in research studies may be acceptable under certain circumstances as a surrogate measure in the absence of available registry data. Self-reported cancer has high sensitivities for breast, prostate and bowel cancers; however, self-report melanoma is less likely to be accurate and is recommended to be supplemented with other sources of information. Self-report cancer data are more accurate among younger individuals and those with a more recent cancer diagnosis. Self-report cancer data may also be sufficiently reliable to be used to exclude those with cancer, or as a covariate to adjust for confounding. Also, for particular cancer types and analyses where the focus is on risk factors for cancer, it is likely that the lack of sensitivity of self-report cancer would not materially affect the relationship between the risk factor and the outcome. Self-reported cancer data may therefore be very useful as an outcome. The findings presented here may assist medical researchers to make informed decisions when conducting future research using self-reported cancer data in Australia where the acquisition of registry data is not feasible.

Acknowledgements
The AusDiab study co-ordinated by the Baker IDI Heart and Diabetes Institute, gratefully acknowledges the generous support given by: National Health and Medical Research Council (NHMRC grant 1002663), Australian Cancer Database, Australian Government Department of Health and Ageing, Abbott Australasia Pty Ltd, Alphapharm Pty Ltd, AstraZeneca, Bristol-Myers Squibb, City Health Centre-Diabetes Service-Canberra, Department of Health and Community Services – NT, Department of Health and Human Services – Tas, Department of Health – NSW, Department of Health – WA, Department of Health – SA, Department of Human Services – Vic, Diabetes Australia, Diabetes Australia NT, Eli Lilly Australia, Estate of the Late Edward Wilson, GlaxoSmithKline, Jack Brockhoff Foundation, Janssen-Cilag, Kidney Health Australia, Marian & FH Flack Trust, Menzies Research Institute, Merck Sharp & Dohme, Novartis Pharmaceuticals, Novo Nordisk Pharmaceuticals, Pfizer Pty Ltd, Pratt Foundation, Queensland Health, Roche Diagnostics Australia, Royal Prince Alfred Hospital, Sydney, Sanofi Aventis, sanofi-synthelabo, and the Victorian Government’s OIS Program. Also, for their invaluable contribution to the set-up and field activities of AusDiab, we are enormously grateful to A. Allman, B. Atkins, S. Bennett, A. Bonney, S. Chadban, M. de Courten, M. Dalton, D. Dunstan, T. Dwyer, H. Jahangir, D. Jolley, D. McCarty, A. Meehan, N. Meinig, S. Murray, K. O’Dea, K. Polkinghorne, P. Phillips, C. Reid, A. Stewart, R. Tapp, H. Taylor, T. Whalen and F. Wilson. The co-operation of other Australian Cancer Registries regarding procedures, coding, duplication and demarcation issues, and of staff of the Australian Cancer Database at AIHW, Canberra, is acknowledged as playing a vital part in ensuring data quality and comparability.

Funding sources
VHYL is supported by the National Health and Medical Research Council (1022923). JLH is supported by the Monash University APA scholarship and a Baker IDI top-up scholarship. JES is supported by the National Health and Medical Research Council Fellowship (586623). DJM is supported by the National Health and Medical Research Council (1002663).

References