Reviewer's report

Title: Labour force participation and the cost of lost productivity due to cancer in Australia

Version: 0 Date: 06 Nov 2017

Reviewer: Brendan Walsh

Reviewer's report:

This paper examines the labour force participation of people aged 25-64 with cancer in Australia and compares participation to those with no health condition and those with "any other" health condition. The study uses a large nationally representative dataset, the 2015 Survey of Disability, Ageing and Carers. The authors also attempt to determine the indirect costs due to lost productivity from, as a result of lower full time employment and examine the inequality in the distribution of labour force absence

Each of these questions is vitally important, especially since little up to date information exists on these topics in Australia. However, I think that the ability to examine many of these questions is significantly curtailed in this study due to the data available and techniques used.

Overall, future iterations of this study should focus on understanding of labour force participation amongst cancer survivors, subject to concerns I raise below.

The attempt to estimate the cost of labour force absence in this study is severely limited. Very different statistical analysis would be required to estimate the cost of labour force absence, and likely a different dataset also.

Main Points:

1) The chief concern I have with the manuscript relates to the lack of information and detail on the how the data in used and the techniques used to analyse the data. Specifically, there is a clear lack of detail on how participants in the Survey of Disability, Ageing and Carers were assigned the health categories chosen: no health condition, cancer, any other health condition.

- It would be useful to include how the question, which the authors use to categorise participants into the three health condition groups, was asked in the Survey. This would help clear up the questions related to the categories I include below:

- No health condition. Does this mean that these includes had no condition based upon proving a negative response to the question?
- Cancer. Does this mean the participants currently have cancer, or have had cancer in the past but are no longer undergoing treatment? I note that there is a question on treatment within the survey, however only breast and colorectal cancer are included as options in said question.

- Any other health condition. What are the most prominent conditions within this category? This would allow the reader to understand the severity of conditions in this group.

2) Prevalence of cancer in the survey.

- In addition to the point above on how cancer is recorded in the survey, what are the main neoplasms recorded. In particular, what proportions of cancer survivors have non-malignant skin cancer?

- There are 355 individuals with the cancer health condition. Compared to absolute numbers in other surveys internationally, this is not that small, but they still only constitute 1% of the overall sample. This needs to be discussed.

3) Display of results and estimate population representation.

- The inclusion of weighted numbers in Tables 1, 2, and 4 gives a false impression to the reader regarding the numbers included in the analysis. Surely it is more correct to applying weights when estimating both the descriptive statistics and regression coefficients, rather than implying that the 355 individuals with cancer represent 108,900 people with cancer in the general population.

- In Table 2, with a prevalence of 0%, the authors estimate 300 individuals with cancer are unemployed, which is simply spurious accuracy.

- Similarly, in the text, the inclusion and discussion of the estimate population numbers should be removed. The authors have over 34,000 observations in the 25-64 age-group which is very large sample in any internationally for a survey such as this.

- A descriptive table on the breakdown of the health condition categories should be provided. Are cancer patients older, poorer, live in urban areas?
- In the text beneath table 2, the authors discuss results that are not included in any table or figure. I think these are potentially important findings, and warrant illustration and explanation for the reader. For example, the authors provide an odds ratio of 3.00 (95% CI 1.96-4.57), and this is not included in any table. This is a noteworthy issue to examine and discuss.

- If the confidence interval range for this estimate is 1.96 - 4.57, would this not correspond to an odds ratio of 3.265?

- The authors also state: "those with any other long-term health condition were 2.15 times more likely to be employed full-time (95% CI 1.41-3.28; p=0.0004) than those with cancer".

- If the confidence interval range is 1.41-3.28, would the odds ratio would be not be 2.345?

4) Regression analyses.

- In the regression analyses only age, sex, and rurality are included as control variables. In a dataset that is as rich as the Survey of Disability, Ageing and Carers, a range of other variables are available. Why were these not included?

- Why are the results for age, sex, and rurality not provided?

- Was age included as a continuous variable or as dummy variables?

- The title of Table 3 is incorrect, "being not in the labour force":

- "Multiple models estimating the odds of being not in the labour force for those with cancer, compared to those with no health condition and any other long-term health condition"

5) Concentration indices used to examine the proportionate distribution of labour force absence amongst people with cancer.

- A concentration index plots the cumulative proportion of a variable (health) against the cumulative proportion of the population ranked by a socioeconomic variable, for example income. This socioeconomic ranking variable, Ri, used to estimate a concentration index is
ordinarily continuous or includes a number of different categories. This allows for the population to be ranked from poorest (R=0) to richest (R=1) in a cumulative manner. The smaller the number of groups used to calculate the ranking variable, the less accurate the concentration index, as within group inequality cannot be accounted for due to the inability to distinguish those with the same socioeconomic value. There are a large number of papers which discuss this issue in detail (see Appendix A.3 here: http://www.sciencedirect.com/science/article/pii/S0167629611001603). In this study, the ranking variable is binary - with individuals having a tertiary education or not, which leads to all individuals having one of 2 values, 0 or 1.

It is incorrect therefore to estimate a concentration index using a binary variable in this instance. It would be more appropriate to regress education on labour force absence using a multivariate analysis, which has the added benefit of controlling for other factors such as age and sex which will impact the relationship between workforce absence and education.

6) I have considerable worries regarding how the cost of labour force absence associated with cancer is measured.

- I am unclear as to how the cost of labour force absence is measured in this study. The authors' state: "An estimated 50,100 Australian adults of working age (25-64 years) with cancer were not in the labour force, thereby reducing Australia's GDP by approximately $1.7 billion (Table 4)." How was this estimated? There is little detail on this. As this figure is the main finding of this present study, the reader has no information on how it was calculated.

- As stated in previous points, are the authors stating that between the ages of 25 and 64, that there are 50,100 cancer survivors not in the labour force? Are there any official estimates from a cancer registry in Australia that this figure is correct? Would it not be more appropriate to estimate the average cost per cancer patient and subsequently impute this estimate on the actual number of cancer survivors in Australia aged 25-64?

- For what year was GDP taken.

- Is the same average cost used for males and females, tertiary and non-tertiary education, and major cities and other areas. Productivity, employment, and hours worked will all differ across these categories.

- Are individuals aged 15+ included in the estimate cost?
The technique discussed hastily in Aim 2 has been used previously for the cost of arthritis in Australia (Med J Aust. 2008 Oct 20;189(8):447-50), but that study is not referenced here. As this study is using a similar/identical technique, but applying it to cancer, why was this not included as a reference, particularly when discussing estimated costs?

7) There are a number of studies of international health systems that focus on the productivity cost of cancer. However, little space is given to these studies within this analysis.

**Are the methods appropriate and well described?**
If not, please specify what is required in your comments to the authors.

No

**Does the work include the necessary controls?**
If not, please specify which controls are required in your comments to the authors.

No

**Are the conclusions drawn adequately supported by the data shown?**
If not, please explain in your comments to the authors.

No

**Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?**
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I recommend additional statistical review

**Quality of written English**
Please indicate the quality of language in the manuscript:

Acceptable
Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal