Author's response to reviews

Title: Cancer awareness and socio-economic position. Results from a population-based study in Denmark

Authors:

Line Hvidberg (LINE.HVIDBERG@ALM.AU.DK)
Anette F. Pedersen (AFP@ALM.AU.DK)
Christian N. Wulff (CHRIWULF@RM.DK)
Peter Vedsted (P.VEDSTED@ALM.AU.DK)

Version: 4 Date: 23 July 2014

Author's response to reviews: see over
The Editor
BMC Cancer

Dear Editor,

RE: ‘Cancer awareness and socio-economic position. Results from a population-based study in Denmark’, Manuscript ID number 1697943968122463.

Thank you for the opportunity to revise this manuscript. We are very pleased to learn that you found the article of relevance for the Journal. We truly appreciate the reviewers’ comments and have done our best to respond and adapt the text. Please find below a detailed description of our response to each of the points made by the reviewers.

We hope that you will consider the revised manuscript and that it will be suitable for publication. Do not hesitate to contact us if there are any further issues to be handled.

Yours sincerely,

Line Hvidberg, Anette Fischer Pedersen, Christian Nielsen Wulff and Peter Vedsted
Point-by-point response to reviewer Lindsay Forbes’ comments:

**Minor essential revisions**

1) *The meaning of the term ‘survivability’ in the abstract was not at first glance clear to me. Suggest find a better term.*

   This has now been changed into ‘perceived 5-year survival from bowel, breast, ovarian, and lung cancer’. Thus, the sentence in the abstract and it the background section now reads: “The aim of this study is to assess awareness of cancer symptoms, risk factors and perceived 5-year survival from bowel, breast, ovarian, and lung cancer in a Danish population sample and to analyse the association between these factors and socio-economic position indicators”.

2) *Background: second sentence a bit unclear – do the variations disproportionately affect people with lower SEP or in fact do they have worse cancer survival? I realise it is the latter but not very clear in this sentence.*

   We agree that this sentence was unclear. The sentence has been altered into: “Even within countries, survival rates vary much between patient groups with the same type of cancer, and for most cancers people with lower socio-economic position (SEP) have poorer outcomes than their socioeconomically more affluent counterparts”.

3) *Second paragraph, final sentence, not very clear. Second to last sentence, I think ‘ultimately’, would be a better adverb than ‘lastly’ – ‘ultimately’ implies importance, ‘lastly’ implies insignificance.*

   We have re-written this paragraph, which now runs as follows: “Cancer awareness accordingly seems to be a potentially modifiable contributor to the variations seen in healthcare seeking and, ultimately, survival. It is therefore important to assess cancer awareness among the general population and to investigate possible associations within different subgroups”.

4) *Third para – ‘only’ redundant.*

   Thank you for making us aware about this, we have now deleted ‘only’ in the sentence.

5) *Methods – second para – I think you mean ‘death’, not ‘decease’.*

   Before the change, the sentence was: Lastly, an additional 55 (0.1%) persons were excluded just before the data collection began, either because of a newly established research protection status, emigration from Denmark, or decease.

   Instead of changing ‘deceased’ into ‘death’ we have changed this into ‘passed away’: “(...) either because of a newly established research protection status, emigration from Denmark, or because the person had passed away”. Similarly, this has also been altered in Figure 1.
6) **Final sentence – not entirely sure what you meant by ‘available for participation’.**

We have altered the sentence: “In total, 47,066 persons (78.4% of the study base) were available for participation in the telephone interview (Figure 1)” to “In total, 47,066 persons (78.4% of the study base) were thus eligible for being contacted to answer the ABC measure (Figure 1)”.

7) **Page 8: second line a bit clumsy.**

The sentence: “This was estimated as the proportion of successfully completed interviews in relation to the total number of persons who were made contact to” has been altered to: “This was estimated as the number of completed interviews divided by the number of eligible persons made contact to.” Further, to clarify how the response rate was calculated, we have made several changes in the Results. This is noted in the response number 1 to reviewer Emily Power.

8) **I suggest comment on data completeness about SEP and other demographic variables – it is a key strength of this study.**

We agree that this is a key strength of the study and have now provided more detailed information about the SEP indicators in the Method section. Hence, the following paragraph has been included: “For these register-based SEP indicators, missing data ranged from 0% for age, gender and information on cancer diagnosis to 3.9% for information about educational level for the study base of 60,000 persons.”

Furthermore, in the discussion on strengths and limitations, we have also elaborated on the issue of data completeness: “A key strength of the present study was the use of the Danish CRS. All Danish residents are registered in the CRS which contains complete information on any Danish resident’s date of birth, gender, migration, etc. Owing to our use of the CRS, we were able to define a study base of 60,000 persons, a representative sample of the entire Danish population aged 30 years and older. Furthermore, the use of the CRS and the data linkage to a range of register-based SEP indicators provided us with precise and valid insight into variables that may be related to cancer awareness.”

9) **Tables 2, 3 and 4 – I suggest improve the titles of the tables, putting the subject e.g. cancer symptom awareness, awareness of risk factors and awareness that cancer risk increases with age earlier in the title. It took me a while to work out which was which.**

We have now put the theme of the table in front. Due to a new table on response, the table numbers are now 3-6. The titles of the tables are:

- **Table 3.** Recognition of less than nine symptoms of cancer. Prevalence ratios (PR) with 95% confidence intervals (CI) for association with socio-economic position (SEP) indicators.
- **Table 4.** Recognition of less than nine risk factors for cancer. Prevalence ratios (PR) with 95% confidence intervals (CI) for associations with socio-economic position (SEP) indicators.
**Table 5.** Lack of awareness of growing risk of cancer with age. Prevalence ratios (PR) with 95% confidence intervals (CI) for associations with socio-economic position (SEP) indicators.

**Table 6.** Underestimation/overestimation of 5-year survival from bowel, breast, ovarian and lung cancer. Prevalence ratios (PR) with 95% confidence intervals (CI) for associations with socio-economic position (SEP) indicators.

10) I was interested in how educational level, occupation and household income vary together – I suspect they are highly correlated, and I see that the analysis has acknowledged this by not adjusting them for each other. They are all indicators of the latent variable SEP, but of course perhaps have their own influence, especially education. However, some more comment on this in the results and discussion sections would be useful.

We agree with the reviewer that this is an important issue and have therefore addressed this issue more thoroughly in the discussion in the paragraph on strengths and limitations. After the paragraph given in response number 8, we have inserted the following: “Naturally, the SEP indicators capture correlated aspects. Still, since the correlation is not a hundred percent, each indicator contributes with unique information about the association with cancer awareness.”

11) Awareness of 5 year survival – I think ‘depicts’ is the wrong word to describe a table. More substantively, I think that Figure 2 is very interesting and deserves more comment. However I am not convinced that Table 5 is illuminating – partly because those who under and over-estimated were grouped together, so where there are statistically significant results, they are difficult to interpret. Some more comment on this in the discussion would be useful.

‘Depicts’ has now been changed with ‘shows’. We agree that Figure 2 is very interesting, and it is a very simple figure. We have therefore added the following sentence to the paragraph on awareness of 5-year survival from the four different types of cancer: “For ovarian and lung cancer, a large majority (86 and 78%, respectively) of the respondents overestimated the 5-year survival, whereas almost half of the respondents underestimated survival from breast cancer.” We have also elaborated on this in the discussion. Thus, at the end of the discussion, the following paragraph has been included on page 12 of the manuscript: “Awareness of the 5-year survival from bowel and breast cancer was fairly high; however, only a small percentage of the respondents correctly identified the 5-year survival from ovarian and lung cancer. This may be due to inadequate communication about the chances of survival for these cancer types. However, the results for lung cancer may also be partly explained by end-aversion bias, i.e. the tendency to avoid the extremes of a scale.”

In respect to Table 6, the aim of the paper is to assess awareness of the 5-year survival from four cancers; we therefore find it appropriate to combine under- and overestimation. Moreover, both over- and underestimation have been found to be associated with cancer-related behaviour.
12) *I think the results of the sensitivity analyses could be minimised, or even reduced to a single sentence.*

As we argue in the discussion section, using the median split procedure to categorise cancer awareness can come across as very arbitrary. Therefore, we think it is essential to show that our results are robust when using different splits. This is a key strength of this study compared with others in this field. Meanwhile, we have shortened the reporting of results.

13) *In the discussion, paras 2 and 3 are a bit confused. I think the authors should set out a bit more clearly the hypothesis that SEP influences awareness influences behaviour influences late stage influences survival (preferably in a less clumsy way than I have just done!), and address the extent to which this study contributes to that more systematically.*

We agree that we have probably not been sufficiently clear on this point. We have therefore clarified the discussion about the hypothesis, as requested. The changed text is in page 10 of the resubmitted manuscript.

14) *Minor point – laypeople, not laymen.*

Thank you. This has now been changed accordingly.
Point-by-point response to reviewer Emily Powers comments:

Minor essential revisions

1) Results - Response - it’s not 100% clear how the response rate was calculated from the text provided in this section and I don’t think it should be up to the reader to work it out. I would recommend following a standard formula such as the one used in the Forbes et al ICBP paper which will enable readers to compare across samples more easily. The estimated response rate represents the response rate after adjusting the size of the denominator for the likely proportion of households that were eligible. It is calculated by assuming that the proportion eligible among households of unknown eligibility is the same as the proportion of those tested for eligibility who were eligible (equivalent to American Association for Public Opinion Research response rate formula 3). It may also be a good idea to include a ‘minimum response rate’ as they did in that paper.

We are sorry to see that the way we have presented the information about the response rate is unclear. In the ICBP, two sampling methods were employed across the various countries. Thus, in some countries, telephone numbers were randomly selected using the random-digit dial (RDD) sampling method, whereas in others, e.g. Denmark, a random sample was selected from a list of all eligible adults listed on the national population registers and then matched for telephone numbers. The way of calculating the response rates in the ICBP paper reflects that the RDD methods was used, and the American Association for Public Opinion Research (AAPOR) report thus does not actually deal with the sampling methods used in Denmark. We therefore prefer not to use the response rate formulas given in this report.

However, we have, of course, made changes in the results that should clarify how the response rate was calculated. We believe that a table illustrating the process could help the clarification and have therefore included the following Table:

Table 1. Response rate

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of persons approached</td>
<td>11,297</td>
</tr>
<tr>
<td>Number of ineligible persons</td>
<td>1,697^a</td>
</tr>
<tr>
<td>Number of persons who could not be contacted after seven attempts</td>
<td>1,431</td>
</tr>
<tr>
<td>Number of persons eligible and made contact to</td>
<td>8,169</td>
</tr>
<tr>
<td>Number of persons who refused or did not complete the interview</td>
<td>5,169^b</td>
</tr>
<tr>
<td>Completed interviews</td>
<td>3,000</td>
</tr>
</tbody>
</table>

^a Incomplete/unobtainable number (n=1,328); wrong number (n=326); business/fax number (n=8); number barred (n=2); and unable to speak or understand Danish (n=33).

^b Refused to take part (before or after it was known whether or not it was the person eligible for study participation) (n=4,736); stopped the interview (n=154); the person eligible for study participation asked to be called back at a later date, but could not be contacted again (n=141); the persons answering the phone did not want to speak to the interviewer (n=92); another stated that the person eligible for the study was not available during data collection period (n=31); and the person stated that he/she was not in the age group anyway (n=15).

Furthermore, the text on the response rate has been clarified and now reads: To obtain inclusion of 1,000 respondents aged 30-49 years and 2,000 respondents aged 50 years or older, we approached a random sample of 11,297 persons. A response rate of 36.7% was achieved (Table 1); this was estimated as the number of completed interviews divided by the number of eligible persons made contact to”.
2) When describing how the study sample compares to the 60,000 drawn from the database it’s important to know which of the differences mentioned are statistically significant - I assume that all the ones mentioned are but it’s not clear. I think it’s also important to indicate these statistically significant differences in Table 1.

Indeed, all the mentioned differences are statistically significant at the 1% level. But to clarify, we have now inserted the results of the chi-square test in this table. As can be seen, there were differences between respondents and the study base on all SEP indicators except from being registered with a cancer diagnoses within the past 10 years. As noted in the table, the difference was tested between the 3,000 respondents and the study base minus the respondents (i.e. 57,000 persons independent of the respondents) in order to meet the assumptions of the chi-square test.

Discretionary revisions

3) Abstract - Results - recommend changing this sentence to 'No clear associations were found between social economic position and awareness of 5-year survival for bowel, breast....'

We can understand that the sentence reads more easily by changing ‘lack of awareness’ into just ‘awareness’. However, since the outcome is ‘lack of awareness’, we will like to keep this wording, but have changed “awareness on the 5-year survival’ into ‘awareness of 5-year survival’ as has been suggested.

4) Background - third para - you mention that previous studies have only included a small number of SEP indicators. I don’t think this is strictly true. For example, Robb et al included gender, age, marital status, ethnicity, and occupation. The only additional variables this study includes are education and income.

We acknowledge this and have therefore changed this paragraph accordingly. Thus, the paragraph have been changed from: “However, these studies are limited by including only a small number of SEP indicators, all relying on self-reporting” to “In these studies, all SEP indicators are based on self-reporting”.

5) Methods - it would be useful to know (for non-Danish readers) how representative the CRS is of the Danish population

In the revised manuscript, we have now elaborated on the Danish CRS (cf. response number 8 to reviewer Lindsay Forbes).

6) Statistical analysis - It would be useful to know why the analysis used unweighted data

It is true that we did not use weighted analyses. This choice was made because we prefer to report the possible bias that our non-response in the survey may have induced rather than adjusting for
non-response by weighting, and we assume that this would reduce the bias caused by non-response. Thus, a general assumption for applying non-response weight adjustment is that if it were possible to obtain the responses of the non-respondents within each cell (e.g. in term of educational level), their response would be exactly the same as the responses of the respondents within the same cell. We don’t know whether this is true or not. The analyses actually indicate that the respondents are not representative of the non-respondents, and we could introduce additional bias by using weighted analyses. Therefore, in the last paragraph of the discussion, we have instead chosen to state the limitation of the non-response and the possible effect of this. In the statistical analysis, we have now stated that “Statistical analyses were performed with unweighted data, because weighting may introduce additional bias”.

7) Results - response - It would be useful to include the response information in Figure 1 - e.g. the fact that 11,297 people were randomly selected from the 47,066 who you had telephone numbers for and that 1,697 of these were ineligible etc.

Cf. response number 1 above. We have made some significant changes to the Results to clarify how the response rate was calculated. We believe that the table we have included helps to illustrate this.

8) Results - awareness of symptoms - since the symptoms aren’t listed in the measures section it would be useful to have a full list of the symptoms that were asked about somewhere in the paper so that readers can put the awareness findings into context and don’t have to refer back to the ABC paper to do so. The same comment applies to the awareness of risk factor results.

From reading papers within this field, we know how frustrating it can be not to be given the exact details. Thus, in the description of the dependent variables, we have now listed the 11 symptoms and the 13 risk factors.

9) Similarly, so that readers don’t have to refer back to the methods section of the paper it would be useful to add ‘(out of 11)’ after describing the median number of symptoms and risk factors recognised.

We have added this to the Result section.

10) Respondents aged 50-69 and those with a close relative with cancer were also significantly higher awareness than younger respondents and those with no close relative with cancer - this should be referred to in the text as well as in the table otherwise I think it’s misleading.

In the description of Table 3 in the revised manuscript (i.e. awareness of cancer symptom), we have added that people with no close relatives with cancer were more likely to recognise less than nine symptoms of cancer than people with close relatives with cancer. Furthermore, in Table 4 (i.e. awareness of risk factors), we have inserted that being older was associated with recognising fewer than nine risk factors for cancer.
11) **Results** - awareness of risk factors - the sentence 'it should be noted that older age was associated with a higher propensity of recognising less than nine risk factors' makes it sound like this difference wasn’t statistically significant which could be misleading plus it’s awkwardly worded with a double negative, so I recommend making this more clear.

We have deleted this sentence since we have now included the information as stated above (cf. response 10).

12) **Awareness of 5-year survival** - second para - this sentence needs re-wording because it doesn’t make sense ‘After adjustments there were found significant associations between...’. I think it would also be useful to state who had lower/higher awareness e.g. those not working were more likely to over or underestimate the 5 year survival for breast cancer and men were more likely to over or underestimate the 5 year survival for lung cancer.

This second paragraph in the ‘awareness of 5-year survival’ has been altered into: “Table 6 shows the associations between the SEP indicators and underestimation/overestimation of the 5-year cancer survival. People outside the labour force were more likely than people within the labour force to wrongly estimate the 5-year survival for breast cancer, and men were more likely to wrongly estimate the 5-year survival for lung and ovarian cancer than women. Furthermore, people with a low and middle income and people with no close relatives with cancer were less aware of the 5-year survival from bowel cancer than people with a high income and with close relatives with cancer.”

13) **Discussion** - first para - recommend rewording this sentence to ‘the sensitivity analyses showed that the associations with low SEP increased with lower awareness, validating our analysis approach/methods/the findings’

This has been changed into: “The sensitivity analyses showed that the associations between SEP and the respondents’ awareness of symptoms and risk factors were independent of the median cut-off; thus, the findings appear to be robust”.

14) **You mention that females had lower awareness of the increasing risk of cancer with age here but it’s not been cited in the results which I think it should be, especially since you mention this in the abstract too. I also think you should mention the associations you found with age in this para e.g. that older people had lower awareness of risk factors**

The finding that females were more likely than males to lack awareness of the relationship between age and cancer risk has now been included in the Results: “Being a woman, having a low level of education, and a low income were associated with non-recognition of growing risk of cancer with age (Table 5)”. We have chosen not to mention the association between age and awareness of risk factors in the abstract due to the word limit.
15) **Discussion - third para - Recommend re-wording this sentence 'These factors include barriers to health-care seeking and beliefs...'**

The paragraph: “Recent research has also emphasized the role of other factors in the link between cancer awareness and cancer-related behaviour. These factors include among others barriers in relation to health-care seeking and beliefs about cancer” has been altered to: “Among others, it has been suggested that anticipated barriers to healthcare seeking and beliefs about cancer may mediate this link”.

16) **Discussion - fourth para - I think it would be useful to cite literature that highlights the lack of awareness of vague, non-specific symptoms and not realising the significant of these symptoms has been shown to be the biggest barrier to seeking help (Macleod et al 2009 review)**

At the end of fourth paragraph, we have included the following sentence: “Likewise, in a comprehensive review by Macleod et al, vague, ambiguous and more common symptoms were associated with a longer patient interval.”
Other revisions:

- The revised paper has been proofread by a native English-speaking person. We believe that this has improved the quality of the written English.

- Since the submission of the paper in February, we have had very valuable discussions about the SEP indicators. Regarding the indicator ‘marital status’, we believe that the important aspect is whether a person is living with someone or not, and not whether the person is married, cohabiting, or single. As a consequence, marital status now has two categories instead of three. This change has been incorporated in all the Tables. Regarding the indicator ‘household income’: To account for annual variation in household income, we have chosen to calculate the average household income for three years. Thus, measures at one point in time may fail to capture important information about income fluctuations and are sensitive to changes in life circumstances; thus, the advantage of using an average of three year income. It should be noted that the change in categorisation of marital status and the new measure of household income has not affected the results to a significant degree. However, it should be noted that in Table 4 (recognition of less than nine risk factors for cancer), retired has now become significant, and for Table 6 (underestimation/overestimation of 5-year survival from bowel, breast, ovarian, and lung cancers) more associations have now become statistically significant.