Reviewer's report

Title: Socioeconomic Patient Characteristics Predict Delay in Cancer Diagnosis: A Danish Cohort Study

Version: 1 Date: 12 November 2007

Reviewer: Richard D Neal

Reviewer's report:

General

This is an interesting and well-written and designed study that adds to the literature in a positive way. It is also a large cohort study. However there are a number of mainly design issues that I would seek clarification on before acceptance for publication.

Aims & Originality

The aims are laudable. To a degree this replicates work done in other centres, but using more robust methods.

Importance

Potentially, of importance. But only if likely to inform interventions that may lead to earlier stage presentation and therefore potential improvements in morbidity and mortality. More needs to be made of this in the background. (minor, essential)

Background

I would suggest that the authors refine their statement that 'delay in cancer diagnosis and treatment is an important factor in prognosis'. There is systematic review evidence that this is the case for breast cancer, but either no or mixed evidence in other cancers. For colorectal cancer for example, there is no clear association. (minor essential)

I would also suggest that the different factors that may contribute to the ‘delay’ in each part of the cancer journey be dealt with separately (i.e. factors relating to patient delay are really quite different to those for system delay). I would also suggest, to get this in perspective, a brief overview of the contribution of the different component delays to overall time to diagnosis. (minor, essential)

Methods

Analysis of all cancers together

Not convinced by their approach of including all cancers together. Most clinicians
and researchers would agree that individual cancers are very different diseases, and this includes the way that they present. It therefore seems somewhat unusual to analyse them together. For example, the ways in which different cancers present and the ‘times to diagnosis’ of different cancers are very different. The likely outcome of this approach is that simply because of the nature of the growth of the cancer, the proportions of different cancers in each short and long delay group will be hugely unequal. This cannot make sense? Further justification of this approach, or analysis of cancer by cancer is needed (and their dataset appears to be big enough to justify this). It may also help them make more sense of findings within the context of individual cancers because it is likely that the socio-demographic differences that they might find will be cancer specific, for good reason. At a minimum, we need numbers of each cancer in the dataset presented. (major, essential)

Statistics
As an outcome of the results of the regression model I was looking for a model showing how much significant factors contributed to it and how much of the variance this explained, but couldn’t find this. Apologies for my amateur grasp of stats but I think that the unadjusted and adjusted models need more explaining. Adjusted for what? The interacting effect of the other variables, or the effect of the GP with multiple patients? (minor essential)

Missing data
Figure 2 is most welcome, as it clearly explains the process of recruitment into the study and explains where data are lost. Moderate response is a problem because it is likely to bias sample; for example there is likely to be a death (and near-death) bias. Is there the possibility of any additional presentation of data to show how non-responders may have been different in any way? Indeed, it would be useful to have sight of the ‘submitted paper’ so that the process of recruitment clearer. For example, how complete are the HDR data, and why exclude those outside the time period if it increased the completeness of data? (minor essential)

Definition of delay
The doctor delay reference for ‘widely-used’ is from 1974. This needs retracting or updating and justifying more. How was the date of first symptom defined by the GP from the records? This is notoriously difficult (and probably quite inaccurate); so at a minimum the advice given to GPs and how they interpreted it needs to be presented in this paper. For patients who were not investigated in general practice, but referred directly, how was their system delay calculated? Working from the Figure 1 it would have been impossible. Similarly how calculated for emergency admissions? More information on this is needed. (minor essential)

The use of quartiles is presumably to get a statistical grasp of the data, but it does mean that what are in effect continuous data are categorised into a rather artificial ‘long’ and ‘short’ categories. More justification or alternatives please! I am also confused that the 75th centile for doctor delay was only 2 days. (minor
Choice of ‘factors’ for the questionnaire
No rationale for the inclusion of those factors included or indeed the non-inclusion of non-included factors is given. What about, for example, co-morbidity, or ease of access to medical facilities for example? (minor essential)

Timing of patient questionnaire
How soon after diagnosis was the questionnaire actually sent? This is important to get a handle on recall bias. Presentation of these data would be helpful. (minor essential)

Results
Table 1. This surely shows the effect of prostate compared to breast and gynae cancers, and to be expected. More justification to analyse cancer by cancer.

Tables 2 and 3. The striking issues here is how different the women and men are. Why should this be? Is there an issue about multiple statistical testing and perhaps adjustment for this that needs considering.

Table 4. Personally I think this is the most interesting, and almost to the point of including these results as the only results (at the expense of Tables 2 and 3). There are many data presented in this paper and it may lead to a clearer message if the focus is only on this. (all discretionary)

Discussion and conclusion
Despite the level of work that has gone into the collection of data from this cohort, there is something of a ‘so what’ feel to the main findings. This is hard to avoid, but may lead to more critical discussion of the difficulties of researching in this field, and some of the limitations of their methods (all cancers together, missing data, bias on questionnaire etc). This should not preclude publication, as this certainly adds to the increasing world literature in this area. (major essential)

Co-morbidity may be confounding alcohol and tobacco use. (minor essential)

Language and presentation etc
The authors use the word ‘delay’ throughout the paper. This in keeping with the literature but is a misnomer, because it implies that the time to diagnosis is always longer than it should be. Some recognition of this would be welcome. (discretionary)

What next?: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions
**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.

**Declaration of competing interests:**

'I declare that I have no competing interests'