Reviewer's report

Title: Risk factors for non-communicable disease and healthcare expenditure in employees with private health insurance presenting for health risk appraisal: A cross-sectional study

Version: 2 Date: 16 October 2013

Reviewer: Adrianna Murphy

Reviewer's report:

This paper addresses a very important topic (expenditures associated with NCDs in low-and middle-income countries) is under-researched and I wish I could recommend it for publication. However, I think there are major weaknesses in the methods used and major revisions would be required to make the findings credible and valuable for policy-makers.

Major Compulsory Revisions

Introduction

1. The intro would benefit from at least some discussion of the morbidity/mortality attributed to NCD risk factors in South Africa (the most recent global burden of disease study seems an obvious reference for this) in order to highlight their importance as a public health problem in the country.

2. The authors state that describing the clustering of risk factors among those who volunteer for the HRA would help inform strategies to recruit new participants. How, exactly, would this do so? This should be made clear to the reader.

3. The intro lacks any discussion of the insurance context of SA. The stated aim is to look at clustering and expenditures among privately insured employees - are they they majority? What does this insurance cover? How do they compare with publicly insured individuals? Without this information it is unclear as to how the findings will inform policy.

Methods

1. There are potential biases introduced by this being a wellness day that is sponsored by either the insurer or the employer (it’s not clear from the manuscript who is actually paying/hosting this event and whether their role is made explicit to the employees). These need to be addressed.

2. Data on physical activity were collected over a 12 month period. Should any seasonal effects be considered?

3. There are many potential issues with the participants that need to be addressed. They were wellness participants from 68 companies and their
expenditure was compared to non-participants. What were these companies? Were the non-participants from the same company? If not, is there something different about the nature of their occupation that might affect their physical fitness and their expenditure? If there is anything specific to the companies that might affect any of the outcomes, clustering might need to be accounted for (e.g. with a multilevel model).

4. The authors state that demographic information was collected. What demographic information? There are so many potential factors that can affect health expenditure (gender, age, occupation, education, socio-economic status, marital status...). Data on these would need to be collected and controlled for but no details no this are given in the manuscript.

5. In reference to measurement, the authors state that self-reported physical activity was used. The reader would need to be convinced (i.e. with reference to literature) that this is a reliable measure (e.g. do people tend to over-report their physical activity?)

6. A sentence could be added about who the 'trained staff' are who are collecting the clinical measurements.

7. The statistical analysis needs major revision. Why was only a one-way ANOVA conducted? Why weren't any very relevant confounders (age, etc.) adjusted for (especially as it seems data were collected on these)? Had the authors considered using a matching approach? Also with the test of differences in risk factors between those who met physical activity guidelines and those who did not - there are so many confounders to consider and this is not addressed.

Results

1. The authors address logical constraints that caused non-participation such as being 'off-site' but what about other causes, such as ill health, that might actually affect the results?

2. The authors introduce the use of a regression model in the results section- this should be in the methods section. Moreover, the covariates controlled for in the model (if any??) should be identified. Lastly, it is used to show that physical activity results in increased odds of overweight or obesity, which is something that is known. Why not use it for the other (more interesting?) question on expenditure?

Minor Compulsory Revisions

1. In para 3 of the introduction, the word 'who' is missing (further insight into the characteristics of individuals WHO choose...)

2. The authors sometimes refer to the insurer as a 'private insurer' and sometimes as the 'national health insurer'. This would be confusing to those who are unfamiliar with the South African insurance context and who might assume 'national' means 'public'.
**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests.