

## PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below. Some articles will have been accepted based in part or entirely on reviews undertaken for other BMJ Group journals. These will be reproduced where possible.

### ARTICLE DETAILS

<b>TITLE (PROVISIONAL)</b>	Lung cancer in South Africa: A forecast to 2025 based on smoking prevalence data
<b>AUTHORS</b>	Winkler, Volker; Mangolo, Nosimanana; Becher, Heiko

### VERSION 1 - REVIEW

<b>REVIEWER</b>	Pamela Groenewald Medical Research Council, South Africa
<b>REVIEW RETURNED</b>	09-Dec-2014

<b>GENERAL COMMENTS</b>	<p>This study attempts to forecast lung cancer mortality by ethnic group in SA, based upon changes in smoking prevalence. However, it is difficult to understand the authors method of establishing smoking prevalence. From the methods section, I understood that they used the SADHS 2003 to establish prevalence of smoking in 2003 as the baseline, and then used Ng et al yearly prevalence estimates to project smoking prevalence between 1995 and 2010. However it is not clear from the article what smoking prevalence they used for 2003. Table 2, line 48 shows the prevalence from Ng et al but this differs from the SADHS 2003 (Men 35.1% and Women 10.2%). The smoking prevalences by ethnic group as published in Table 3 are quite different from those reported in the 2003 SADHS. The smoking prevalence rates are much higher according to the 2003 SADHS for all population groups. The authors should clarify which smoking prevalences were used for the baseline in 2003 and how these were estimated, as this will obviously influence the estimated lung cancer mortality rates.</p> <p>A recent article published in the Lancet looks at the differences among SA population groups in smoking attributed mortality using South African vital registration data. This study produced relative risks of death for lung cancer by population group and discusses the differences. It also confirms the 1998 SADHS smoking prevalence rates. I am surprised that this study was not referenced in this paper. Sitas F et al. Lancet 2013; 382:685-93.</p> <p>I suggest that the authors review their smoking prevalence estimates and either re-estimate LC mortality with baseline smoking prevalence rates that are closer to those found in the 2003 SADHS or justify why they have used such low smoking prevalences for their forecasts.</p> <p>The standard of written English could be improved with editing.</p>
-------------------------	--

<b>REVIEWER</b>	Matteo Malvezzi IRCCS-Istituto di Ricerche Farmacologiche 'Mario Negri', Milan, Italy
-----------------	--

<b>REVIEW RETURNED</b>	14-Dec-2014
------------------------	-------------

<b>GENERAL COMMENTS</b>	<p>This is an interesting and valuable paper on lung cancer mortality in South Africa. The authors use modelling techniques to estimate and predict mortality for lung cancer using detailed smoking prevalence data, and then compare results to data from GLOBOCAN and the South African National Burden of Disease study. This work is valuable in that quality data for low and middle income countries is scant and any approach that gives reasonable estimates is welcome. There are, however, a few issues that need to be addressed.</p> <p>The first is on the modelling front, the authors use a model tested and validated (validation on high income country data may not be adequate for use on low or middle income countries, but data on these is hard to find) in previous publications, but comparison to other similar models is somewhat lacking, a reference to a methodological paper such as this (Holford TR, Levy D. Comparing the adequacy of carcinogenesis models in estimating us population rates for lung cancer mortality. Risk Anal. 2012 July ; 32(Suppl 1): S179–S189. doi:10.1111/j.1539-6924.2011.01734.x.) with a brief commentary is needed.</p> <p>The second is on comparison with known data, authors use data from GLOBOCAN and GBD, but do not compare with data from the WHO mortality database. From these data my calculation give a world standardised mortality rate of 23.3 /100,000 men and 7.3/100,000 women in 2009, coverage for this dataset is only about three quarters of the population, but should, none the less, be a pretty good estimate. Comparing to the authors work towards estimates for 2003 we see that for women results are very close while for men there seems to be a strong underestimation. These data need to be presented and commented on.</p> <p>Other issues are minor, wording in presenting results should be made even more cautious since predictions from estimation models are only as good as the assumptions they're made on (even if in this case they are reasonable and justified in the literature). Finally even though the paper is generally well written and comprehensible, there are a few sentences with grammatical or syntactical issues that need correcting by a language editor.</p>
-------------------------	--

<b>REVIEWER</b>	Gill Nelson University of the Witwatersrand, Johannesburg South Africa
<b>REVIEW RETURNED</b>	21-Dec-2014

<b>GENERAL COMMENTS</b>	<p>This is an very interesting and important piece of work. I have only a few minor suggestions for improvement.</p> <p>One of the major study limitations is the small sample sizes for some of the ethnic groups. Although mentioned in the discussion, this is not listed under the Strengths and limitations section. The authors should also mention, as a 'strength' the availability of smoking prevalence data in SA compared to other developing countries.</p> <p>There are several grammatical errors throughout the manuscript. While these are minor, they do detract from the quality of the paper.</p> <p>Finally, I assume that that the smoking prevalence and other data</p>
-------------------------	--

### VERSION 1 – AUTHOR RESPONSE

Reviewer Name Pamela Groenewald

Institution and Country Medical Research Council, South Africa

Please state any competing interests or state None declared: None declared

Please leave your comments for the authors below

This study attempts to forecast lung cancer mortality by ethnic group in SA, based upon changes in smoking prevalence. However, it is difficult to understand the authors method of establishing smoking prevalence. From the methods section, I understood that they used the SADHS 2003 to establish prevalence of smoking in 2003 as the baseline, and then used Ng et al yearly prevalence estimates to project smoking prevalence between 1995 and 2010. However it is not clear from the article what smoking prevalence they used for 2003. Table 2, line 48 shows the prevalence from Ng et al but this differs from the SADHS 2003 (Men 35.1% and Women 10.2%).

The smoking prevalences by ethnic group as published in Table 3 are quite different from those reported in the 2003 SADHS. The smoking prevalence rates are much higher according to the 2003 SADHS for all population groups. The authors should clarify which smoking prevalences were used for the baseline in 2003 and how these were estimated, as this will obviously influence the estimated lung cancer mortality rates.

---- To estimate baseline smoking prevalence of 2003 we used the SADHS 2003. The differences described by the reviewer are explained by the following:

- Age-specific differences in prevalence between our table 3 and table 13.1 and 13.2 in the SADHS 2003 report (<http://dhsprogram.com/pubs/pdf/FR206/FR206.pdf>) are explained by fact that the category of current smokers and ex-smokers are defined differently. Our categories follow as explained in the methods section our modelling procedure. Therefore, the category current smokers also includes individuals who stopped smoking less than five years ago. In consequence, the category ex-smokers are restricted to individuals who stop smoking for at least 5 years. The SADHS 2003 classifies individuals according to the interview data only. We added a footnote to table 3 to clarify our categorization.

- Differences for total prevalence are due to the fact that the SADHS 2003 reports crude prevalence and we report age-adjusted prevalence. We also indicated that more clearly in table 3.

A recent article published in the Lancet looks at the differences among SA population groups in smoking attributed mortality using South African vital registration data. This study produced relative risks of death for lung cancer by population group and discusses the differences. It also confirms the 1998 SADHS smoking prevalence rates. I am surprised that this study was not referenced in this paper. Sitas F et al. Lancet 2013; 382:685-93.

---- We thank the reviewer for this hint and added the manuscript to the discussion.

I suggest that the authors review their smoking prevalence estimates and either re-estimate LC mortality with baseline smoking prevalence rates that are closer to those found in the 2003 SADHS or justify why they have used such low smoking prevalences for their forecasts.

---- We tried to clarify this – see above.

The standard of written English could be improved with editing.

---- We improved the language quality.

Reviewer Name Matteo Malvezzi

Institution and Country IRCCS-Istituto di Ricerche Farmacologiche  Mario Negri , Milan, Italy  
Please state any competing interests or state  None declared : None declared

Please leave your comments for the authors below

This is an interesting and valuable paper on lung cancer mortality in South Africa. The authors use modelling techniques to estimate and predict mortality for lung cancer using detailed smoking prevalence data, and then compare results to data from GLOBOCAN and the South African National Burden of Disease study. This work is valuable in that quality data for low and middle income countries is scant and any approach that gives reasonable estimates is welcome. There are, however, a few issues that need to be addressed.

The first is on the modelling front, the authors use a model tested and validated (validation on high income country data may not be adequate for use on low or middle income countries, but data on these is hard to find) in previous publications, but comparison to other similar models is somewhat lacking, a reference to a methodological paper such as this (Holford TR, Levy D. Comparing the adequacy of carcinogenesis models in estimating us population rates for lung cancer mortality. Risk Anal. 2012 July ; 32(Suppl 1): S179S189. doi:10.1111/j.1539-6924.2011.01734.x.) with a brief commentary is needed.

---- We thank the reviewer and added the reference with a commentary to the discussion part as follows:

A recent study compared the adequacy of different models in estimating LC mortality rates.[ Holford TR, Levy 2012] This study also considered a model [Knoke 2004, Knoke 2011]using non-smoker LC mortality rates which are close to our baseline rates.[Winkler 2011] The relative risks for the smoking categories of current smokers in our model were directly obtained from the considered model.[ Knoke 2004] In summary the considered model results in very good agreement for period and cohort trends, but in weaker agreement with age for younger individuals.[ Holford TR, Levy 2012]

The second is on comparison with known data, authors use data from GLOBOCAN and GBD, but do not compare with data from the WHO mortality database. From these data my calculation give a world standardised mortality rate of 23.3 /100,000 men and 7.3/100,000 women in 2009, coverage for this dataset is only about three quarters of the population, but should, none the less, be a pretty good estimate. Comparing to the authors work towards estimates for 2003 we see that for women results are very close while for men there seems to be a strong underestimation. These data need to be presented and commented on.

---- We added this interesting comparison to the results part and included the observed number of lung cancer death as well as age adjusted lung cancer mortality for 2010 from the most recent WHO mortality database to Figure 1 and discussed the result.

Other issues are minor, wording in presenting results should be made even more cautious since predictions from estimation models are only as good as the assumptions theyre made on (even if in this case they are reasonable and justified in the literature). Finally even though the paper is generally well written and comprehensible, there are a few sentences with grammatical or syntactical issues that need correcting by a language editor.

---- We improved the language quality and changed the wording of some sentences in the discussion part.

Reviewer Name Gill Nelson

Institution and Country University of the Witwatersrand, Johannesburg  
South Africa

Please state any competing interests or state  None declared : None declared

Please leave your comments for the authors below

This is an very interesting and important piece of work. I have only a few minor suggestions for improvement.

One of the major study limitations is the small sample sizes for some of the ethnic groups. Although mentioned in the discussion, this is not listed under the Strengths and limitations section. The authors should also mention, as a 'strength' the availability of smoking prevalence data in SA compared to other developing countries.

---- We expanded the strengths and limitations according to the reviewer's suggestions.

There are several grammatical errors throughout the manuscript. While these are minor, they do detract from the quality of the paper.

---- We improved the language quality.

Finally, I assume that that the smoking prevalence and other data are in the public domain. This should be clarified, nonetheless.

---- SADHS 2003 is not in the public domain. See <http://www.dhsprogram.com/data/available-datasets.cfm>

#### VERSION 2 – REVIEW

<b>REVIEWER</b>	Pam Groenewald Medical Research Council, South Africa
<b>REVIEW RETURNED</b>	26-Jan-2015

<b>GENERAL COMMENTS</b>	<p>The authors have addressed the issues that I raised.</p> <p>I have picked up two minor revisions required in Table 3</p> <ol style="list-style-type: none"> <li>1. the current smoker prevalence differs from the text for Coloured men 38.4% vs 34.4%</li> <li>2. the superscript * after total in row 7 of Table 3 should be replaced by a 3.</li> </ol> <p>I also think the article would benefit from copy editing.</p> <p>I am not qualified to comment on the statistical methods.</p>
-------------------------	---