Author's response to reviews

Title: Independent external validation of cardiovascular disease mortality in women utilising Framingham and SCORE risk models: a mortality follow-up study

Authors:

Louise GH Goh (g.goh1@postgrad.curtin.edu.au)
Timothy A Welborn (officetimwelborn@iinet.net.au)
Satvinder S Dhaliwal (s.dhaliwal@curtin.edu.au)

Version: 2  Date: 13 August 2014

Author's response to reviews: see over
13 August 2014

Professor Peter O’Donovan
Executive Editor
BMC Women’s Health

Dear Sir,

Re: Independent external validation of cardiovascular disease mortality in women utilising Framingham and SCORE risk models: a mortality follow-up study

We are submitting a revised version of the manuscript addressing the reviewers’ comments. We have attached a tracked and clean copy of the revised manuscript.

We conducted an independent external validation of three cardiovascular risk score models (Framingham risk score model and SCORE risk charts developed for low-risk regions and high-risk regions in Europe) on a prospective cohort of 4487 Australian women with no baseline history of heart disease, diabetes or stroke.

Participants were selected from the population-based National Heart Foundation third Risk Factor Prevalence Study and the baseline data were linked with the National Death Index to determine the causes of death during the 10 years follow-up. The 10-year risk of cardiovascular disease mortality was calculated using the Framingham risk score and SCORE models and the predictive accuracy of the three risk score models were assessed using both discrimination and calibration.

The discriminative ability of the Framingham and SCORE models were good (area under the curve > 0.85). Although all models overestimated the number of cardiovascular deaths by greater than 15%, the Hosmer-Lemeshow test indicated that the Framingham risk score and SCORE-Low models were calibrated and hence suitable for predicting the 10-year cardiovascular mortality risk in this Australian population. An assessment of the treatment thresholds for each of the three models in identifying participants recommended for treatment were found to be inadequate, with low sensitivity and high specificity resulting from the high recommended thresholds. Lower treatment thresholds of 8.7% for the Framingham risk score model, 0.8% for the SCORE-Low model and 1.3% for the SCORE-High model were identified for each model using the Youden index, at greater than 78% sensitivity and 80% specificity.

Framingham risk score model and SCORE risk chart for low-risk regions are recommended for use in the Australian women population for predicting the 10-year cardiovascular mortality risk. These models demonstrate good discrimination and calibration performance. Lower treatment thresholds are proposed for better identification of individuals for treatment.

We declare that no competing interests exist. Ethical approval for the study data was obtained in advance from the Australian Institute of Health Interim Ethics Committee, after consultation with the Commonwealth Privacy Commissioner. This study was approved by the Human Research Ethics Committee at Curtin University, and complies with the Declaration of Helsinki.

We will be grateful for your careful consideration of this revision.

Yours faithfully,

Professor Satvinder S. Dhaliwal
Epidemiology & Biostatistics
School of Public Health, Faculty of Health Sciences
Curtin University, GPO Box U1987, Perth, Western Australia 6845

Tel +61 8 9266 2949
Fax +61 8 9266 2958
Mobile +61 (0)409 882 259
Email s.dhaliwal@curtin.edu.au

Point-by-point response to reviewers:

Reviewer's report 1
Title: Independent external validation of cardiovascular disease mortality in women utilising Framingham and SCORE risk models: a mortality follow-up study
Version: 1 Date: 15 July 2014
Reviewer: Olga Khavjou

Reviewer's report:

This is an important paper that assesses the performance of three CVD risk score calculators in a cohort of Australian women. The authors find that all three calculators perform relatively well in the low risk population but that they overestimate the risk in the higher-risk group. There is a large number of various CVD risk calculators available for various populations thus it is important to identify the ones most appropriate for specific populations and I believe this study addresses this issue for the population of Australian women.

I think this is a very well organized and written paper. I only have a few minor discretionary revisions to suggest.

1. line 49: if that the number of deaths of 45622 in Australia? please specify.
Response: Yes, there were 45622 deaths in Australia recorded in 2011. We have revised the manuscript to reflect this.

2. line 51: I'd change "or" at the end of the line to "and"
Response: We have made the revision.

3. line 59: same as above, I'd change "or" at the beginning of the line to "and"
Response: We have made the revision.

4. line 59: I'd change guidelines to plural
Response: We have made the revision.

5. line 60: also add "race" to include "race/ethnicity"
Response: We have made the revision.

6. line 61: change to: "high blood pressure, high cholesterol,..."
Response: We have made the revision.

7. paragraph starting on page 55: I'd like to see a brief discussion of how the various risk calculators are different. Do they include different risk factors? Are they derived for different populations?
Response: We have included a brief discussion to address the reviewer’s comments.

“These risk score models were developed in the USA (Framingham, Reynolds and general CVD) and Europe1 (NSW) (SCORE, ASSIGN and QRISK). Age, sex, systolic blood pressure (SBP), total cholesterol (TC) level, high-density
lipoprotein cholesterol (HDL-C) level and smoking status were included in all six models. Diabetes status was included in all the models except the SCORE model. The Reynolds risk score model was the only model initially developed from a female population and it contains biomarkers in its calculation of the 10-year CVD risk [1]. The QRISK score model includes more risk variables i.e. body mass index (BMI) (which is also found in a simpler version of the general CVD risk score model), family history (also found in the Reynolds and ASSIGN model), Townsend deprivation score (a measure of social deprivation is also found in the ASSIGN model), use of antihypertensive medication (also found in the general CVD model), self-assigned ethnicity, rheumatoid arthritis, chronic renal disease and atrial fibrillation, compared with other models [2]."

We have also included this statement in the manuscript to address the first comment from reviewer 2 and reviewer 3. "The Framingham and SCORE models were selected and validated as they have similar endpoints. Both models predict 10-year CVD death risk. In addition, the risk variables used to calculate the 10-year CVD risk for the Framingham and SCORE models were collected in the National Heart Foundation (NHF) Risk Factor Prevalence Study. Some risk score models did not have recommended treatment thresholds for identifying women at increased risk of CVD and their performance could not be assessed."

8. line 75: "are" should be replaced with "is"
Response: We have made the revision.

9: line 199: I'd like to see the number of deaths reported as the number (152) and as percent of total population
Response: These 152 deaths represented approximately 3.4% of the sample and 0.3% of the total deaths in the Australian female population in 1989 [3], at the start of the study, and 0.2% in 1999 [4], at the end of the 10-year follow-up.

10: lines 248-249: please also report the number of deaths by which they were estimated (in addition to percent)
Response: This information is contained in Table 2 and we have also further clarified the number of deaths in the Results section. We have included this statement “The observed number of deaths is 28 (Table 2).”

Level of interest: An article of importance in its field
Quality of written English: Acceptable
Statistical review: Yes, and I have assessed the statistics in my report.
Declaration of competing interests:
I declare that I have no competing interests
Reviewer's report 2
Title: Independent external validation of cardiovascular disease mortality in women utilising Framingham and SCORE risk models: a mortality follow-up study
Version: 1 Date: 31 July 2014
Reviewer: Ivanny Marchant
Reviewer's report:
Summary
In the article “Independent external validation of cardiovascular disease mortality in women utilising Framingham and SCORE risk models: a mortality follow-up study” by Goh et al., the authors report the findings of an external validation study designed to evaluate the applicability of three different risk estimation systems: Framingham, SCORE for low risk regions and SCORE for high risk regions in a cohort of 4487 Australian adult women (age range 20 – 69 years) by comparing the risks predicted and the risks observed over ten years of follow up.

The manuscript is well written and adheres to relevant standards. It contains, however, some shortcomings that merit to be revised.

Major Compulsory Revisions
(which the author must respond to before a decision on publication can be reached)

- Methods:
Why did the authors choose the Framingham and SCORE models? As the authors have stated, there is a number of published risk predictors that could be used in cardiovascular primary prevention. They should introduce why they chose the Framingham and SCORE risk equations to evaluate their applicability in the study population.

Response: We thank the reviewer for the comment. We have added this statement into the manuscript. “The Framingham and SCORE models were selected and validated as they have similar endpoints. Both models predict 10-year CVD death risk. In addition, the risk variables used to calculate the 10-year CVD risk for the Framingham and SCORE models were collected in the National Heart Foundation (NHF) Risk Factor Prevalence Study. Some risk score models did not have recommended treatment thresholds for identifying women at increased risk of CVD and their performance could not be assessed.”

- Discussion
1. It appears difficult to know whether this paper represent a useful contribution to the field given the shortage of arguments in the discussion. Focused on calibration aspects, the present results are compared with the results obtained from an important number of validation studies in many different populations. In contrast with the big number of references, the discussion seems rather superficial; with no explanations on the possible causes of calibration defaults; or the eventual similarities/ differences between the cohort study population and the populations they refer to.

Response: The authors have included the following sentences in the limitations section. “It is possible that calibration defaults are present due to differences in the populations used to develop the Framingham and SCORE models. Though differences exist, these models have continued to exhibit calibration when externally validated against other populations and have been used extensively in the assessment of CVD risk in previous studies [5-8].

2. Regarding discrimination thresholds, their implications on the size of treatment target populations and expected benefit from treatment should be discussed as well as its economic consequences.

Response: We have included this statement in the manuscript. “The higher treatment thresholds currently used for risk score models, under-treat individuals in higher risk groups. Lower treatment thresholds, as proposed, would improve the identification of individuals who require treatment, at the expense of increasing the numbers to treat and increasing associated costs. Determining thresholds using diagnostic measures to maximise sensitivity and specificity is still the preferred approach.”

3. The authors made the assumption that no one presented with left ventricular hypertrophy in their cohort because the data was not available. Such an assumption may have introduced some bias in the risk estimates leading to underestimation of CVD risk by the Framingham risk predictor. Ideally, the magnitude of this potential gap could have...
been estimated or at least be considered as one of the limitations of the study. It is likely that the prediction by Framingham will be improved by taking into account the presence of left ventricular hypertrophy in Australian populations.

Response: We thank the reviewer for pointing this out. We have included this as a study limitation. “Participants were assumed not to have LVH as ECG was not undertaken in the NHF third Risk Factor Prevalence Study and this could affect risk prediction of the Framingham risk score model. The absence of ECG-LVH was handled similarly in a number of studies [7, 9]. It has been reported that while baseline ECG data to determine LVH were not available, this is unlikely to have a significant impact on risk prediction as ECG-diagnosed LVH is very rare in the general population without CVD [9].”

4. It would be desirable/useful to have a comparison between the performance -and implications of use - of the two risk scores retained (Framingham and SCORE-low). This would add value to the article since more interesting research questions would be addressed.

Response: We thank the reviewer for this comment but this is beyond the scope of our manuscript, which is on the external validation of these models. Thank you.

- Conclusion

1. The authors indicate that Framingham and SCORE-low models are recommended for use in the Australian population. Since the study population consisted of Australian women aged 20 to 69 years, conclusions could only be extended to the adult women Australian population.

Response: Yes, the findings of this study would be applicable to the Australian women population. We have revised the manuscript to reflect this.

Minor Essential Revisions
(such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

Discretionary Revisions
(which are recommendations for improvement but which the author can choose to ignore)

Figures:
- Error bars for the risk estimates should be included in order to judge for the accuracy of the estimates within each category.

Response: Calibration is assessed over all quintiles and not at individual quintiles. Hence, the Hosmer-Lemeshow $X^2$ test is more appropriate, as presented in Table 2.

- Why did the authors use lines to join the predicted risks points? As these are point estimates, there is no need to connect them.

Response: Joining the points gives the reader an indication of the gradient of risk, which is continuous.

Level of interest: An article whose findings are important to those with closely related research interests
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
Reviewer's report 3

Title: Independent external validation of cardiovascular disease mortality in women utilising Framingham and SCORE risk models: a mortality follow-up study

Version: 1 Date: 4 August 2014

Reviewer: Ruth Coleman

Reviewer's report:

Major Compulsory Revisions

1. In Background section of the manuscript the authors list various risk scores but then proceed to only validate Framingham and SCORE without explaining why they have not taken this opportunity to validate any of the other models mentioned.

Response: We thank the reviewer for the comment. We have added this statement into the manuscript. “The Framingham and SCORE models were selected and validated as they have similar endpoints. Both models predict 10-year CVD death risk. In addition, the risk variables used to calculate the 10-year CVD risk for the Framingham and SCORE models were collected in the National Heart Foundation (NHF) Risk Factor Prevalence Study. Some risk score models did not have recommended treatment thresholds for identifying women at increased risk of CVD and their performance could not be assessed.”

2. The Framingham risk scores are based on a population aged 30-74 (SCORE = 19-80. Were the participants in this analysis aged 20-69 years as these were the only ages in this cohort or were limits imposed and if so why?

Response: Yes, these were the only ages in this cohort. Information on the cohort was from the NHF third Risk Factor Prevalence Study.

3. The mean age of this cohort is much younger than that used to develop the risk models, are women aged 20-35 with no history of CVD or diabetes often given 10yr CV risk profiles? If not, how do we interpret the results of this manuscript? Would the modified treatment levels for fatal CVD be useful in those who are regularly given CVD risk profiles?

Response: The American College of Cardiology Foundation and American Heart Association recommends all asymptomatic women to undergo a global CVD risk assessment [10]. In the 2010 American College of Cardiology Foundation and American Heart Association Guidelines for Assessment of Cardiovascular Risk in Asymptomatic Adults, published in Circulation in 2010, this recommendation went even further to calculate the life-time risk for young adults and consider “a global risk score possibly worthwhile even in persons as young as age 20” [10]. The Framingham equations have also been used in this age group, < 30 years [11-15]. The SCORE models are applicable to women who are aged 19-80 years. The modified treatment levels for fatal CVD would improve the identification of women who require treatment (higher sensitivity).

Discretionary Revisions

1. The discussion makes no comment on the differences in the populations potentially explaining differences in estimation. One of the most interesting points in this paper is figure one & yet the only reference to it is the brief paragraph in the results section.

Response: We have included this statement in the manuscript. “Possible causes of differences in estimation include differences in the populations used to develop the risk score models and different risk variables used to estimate risk. Generally, age, sex, SBP, TC level, HDL-C level and smoking status are included into risk score models.”

2. Could the authors discuss whether fatal CVD calculation is more or less useful than CVD risk estimation in general?

Response: The focus on 10-year fatal CVD events ensures the comparability of the risk score models. These events are definite; these events are defined uniformly using the International Classification of Diseases codes; the method used to ascertain these events is consistent thus ensuring reproducibility [16] and these events are well documented in the national mortality databases, thus information is complete and available [17]. Non-fatal CVD events, however, are often self-reported and not formally adjudicated.
Level of interest: An article of importance in its field
Quality of written English: Acceptable
Statistical review: Yes, and I have assessed the statistics in my report.
Declaration of competing interests:
I declare that I have no competing interests