Author's response to reviews

Title: Description and validation of a Markov model of survival for individuals free of cardiovascular disease that uses Framingham risk factors.

Authors:

Dr Chris J Martin (chris.martin@laindonhc.org.uk)
Mark Vanderpump (vanderpump.sec@rfh.nthames.nhs.uk)
Joyce French (hamsterley@hotmail.com)

Version: 2 Date: 29 Mar 2004

PDF covering letter
Responses to review of the paper originally entitled:

Description and validation of a Markov model of survival

that uses Framingham risk factors and smoking status.

Reviewer: Alan Montgomery

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

Overall, I felt that the Methods and Results sections needed more detail in order for someone else with reasonable competency in decision modelling to be able to reproduce the analyses, and to understand the results presented. Specific points that require further clarification:

I would agree far more detail on the structure of the model was needed and the methods and results have been substantially enhanced to do so. This may resolve some of the other issues in the process, as it is clear that the structure of the model was not adequately communicated.

(1) The description of the model in the text and Figure 1 talk about ‘calculating’ or ‘estimating’ mortality due to different causes. In a Markov model these probabilities are defined by the modeler at the outset, using data from observational studies, routine statistics, randomised trials etc. The probabilities are entered into the model, and individuals progress through the model each time period (in this case one year). Probabilities are not ‘re-calculated’ during the model, though may be re-defined if they change as someone gets older. This may just require some re-phrasing, or there may indeed be some flaw in the model, but either way this should be clarified.

This model uses the Framingham risks which have to be calculated at each age for each individual and so must be calculated at each cycle. Calculating risks and probabilities is not unusual in Markov models.[1-3]

(2) As a further point, probabilities should strictly be sampled from a distribution during each time cycle, so that the model is truly stochastic rather than deterministic. A point estimate and characteristics of the distribution are defined (such as normal distribution with 95% CI), the probability of an event is sampled from this distribution on each cycle, and the uncertainty inherent in all estimates of event probabilities is built into the model. If the authors do not want to increase the sophistication of the model to this level, they should at the very least conduct some sensitivity analyses that demonstrate the impact (or perhaps lack of impact) of changing the probabilities used in the model. A table presenting the probabilities used in the base case scenario and any sensitivity analyses would be helpful. These should be annual probabilities, rather than relative, or any other representation of risk.
I agree a sensitivity analysis is needed here for the blood pressure and total cholesterol variation and this is now included. It is not possible to include tables of the annual Framingham risk as they are calculated with each cycle uniquely for the individual being modelled. In total, there are 1,000 different Framingham risks calculated for each separate individual modelled and so will be different each time the model is used. There are also 2,000 smoking related, non-cardiovascular disease, annual, absolute risks and 200 other annual risks used in the model that are fixed and known in advance. It would be possible, but rather impractical to present these in a paper and so I have clarified the description of their derivation so that readers can easily reproduce them with a spreadsheet and access to the source data described.

(3) It wasn’t clear from either the description of the model or Figure 1 whether someone is allowed to have more than one MI or stroke, and if so, which one is assumed to be fatal.

This Markov model is evaluated using cohort simulation rather than monte-carlo methods, so there are no individuals as such in the model, just proportions (0.0-1.0) in each state. I have clarified the description with a state transition diagram (Figure 1) and enhanced the text to make the structure clearer.

(4) Validation. Almost 1400 patients from the cohort are excluded from the analysis. Some of these will be >65 years, but excluding many other cases because of missing data may introduce bias into the analysis. This should be discussed in the ‘limitations’ section.

I agree that this needed clarifying. Many cases were excluded as they were under 35 years old also. Only 8 had missing data. The rest had existing cardiovascular disease.

(5) As I understand it, Table 2 presents the proportions of individuals still alive after 20 years as estimated by the model and observed in the cohort, the difference between these proportions, a 95% confidence interval, and a p-value for the test of the null hypothesis that the difference equals zero. The final column shows the power of the given sample size to detect a 20% difference between the two proportions.

(a) By conducting repeated hypothesis tests, the authors increase the likelihood of finding a ‘significant’ difference by chance alone.

(b) I can’t find any reasoning behind conducting post-hoc power calculations based on a 20% difference between the estimated (model) and observed (cohort) 20 year survival. For agreement, if one is going to do such calculations at all, it makes sense to specify a difference beyond which one would not accept that the model makes a good estimate of survival, and I would have thought that this difference would be much smaller than 20%. For a smaller difference, say 5%, this is obviously going to drastically reduce the power of the study. However once the analyses have been done, there is really no need to do such post-hoc power calculations. Simply present the proportions in the model and cohort, difference and 95% CI, and possibly the p-values, and allow the reader to interpret these. The
I have removed this from the analysis as it causes more confusion than illumination. The use of the post-hoc power calculation was as a statistical test. For explanation – the 20% was a percentage of the mortality, not the total survival. So for the whole sample this amounted to 20% of the 25% who died or 5% of the total. By doing a post-hoc power calculation with a power of 95%, we were able to say that we were confident with a 95% certainty that if the model was not inaccurate by as much as 5% of the total, we would have detected the inaccuracy. A valid analysis, but as clear as mud, so best left out as it contributes little.

(6) Apart from a couple of categories with small sample sizes, the agreement between the model and observed survival strikes me as remarkably close. This re-iterates the need to make the description of the model and the sources of all the probabilities much clearer.

I find it quite awesome and uncomfortably accurate. I have tried to make the derivation of the probabilities used much clearer and have added a table with derivations not declared in the original draft. (Table 2.)

(7) Results paragraph 2 should be in the methods section. This also wasn’t completely clear to me. “… cases were sorted into ascending order of estimated survival at 20 years.” I thought that individuals were classified as either ‘alive’ or ‘dead’ after 20 years, so on what basis were they sorted and arranged into 40 groups? Figure 2 should make clear what each data point represents (i.e., what characteristics). Given that maximum survival can only be 1.0, the figure axis should be changed.

As individuals can only be alive or dead rather than a fraction dead, the cohort had to be put into groups large enough for a reasonably reliable mean average survival to be calculated. I have added other correlation analysis in response to the other reviewer’s comments, but in the original paper they were grouped according to their model estimated risk: the 40 at highest risk, then the next 40 and so on.

(8) Where did the p-values quoted in the Discussion section paragraph 1 come from? Also see point above about whether a 20% discrepancy between the model and cohort really represents acceptable agreement.

See above. Now removed.

(9) Discussion page 11 and Figure 3. Is this a comparison of two 35 year old men, one with existing CVD and one without and stays CVD-free throughout his life? Or is it an estimate of treating a 35 year old man’s high blood pressure, high cholesterol and getting him to stop smoking? This basis of comparison should be clearer, should be described in the methods section, and results presented in the results section.

This was a result of modeling the survival of a single 35-year-old man with the risk factors described and giving the improvement in survival in the imaginary
situation that CVD was eliminated entirely as a cause of death. This has been extensively revised to make it clearer and extended in response to the other reviewer's comments.

(10) Although this model makes assumptions about smoking behaviour, it is certainly possible to model quit/relapse rates and associated changes in risk in a more sophisticated model.

Yes. This is an area of further work and is a lot more complicated than it seems.

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

(11) The title of the paper specifically mentions smoking status. However the analysis is not particularly focussed on smoking behaviour.

The title was intended to reflect the fact that the model went far beyond a simple implementation of the Framingham risk factors and that more extensive use was made of the smoking status. The model was originally designed to aid decisions about intervention, but in practice it is mainly used to demonstrate the effects of smoking to smokers. However, the point is taken and the mention of smoking removed from the title. Taking the implied point that the title should be more tightly related to the context of the study, a statement about the application of the model to those free of CVD at the outset is now included.

(12) Although one could probably guess from the Background section, the aims of the study are not clearly stated.

Now clarified in the abstract.

(13) The key to the formula to calculate absolute risk in smokers (p6) contains definitions for NS and RN but these terms are not actually included in the formula.

Typo – sorry. Corrected.

(14) Is the proportion 0.803 (p7 para 4) a typing error? A 20% reduction from 0.754 would be 0.603. However consider removing this section (see above).

Not a typo. See above. It is a 20% reduction in the mortality, ie

20% of (1 - 0.754) = 0.0492

0.754 + 0.0492 = 0.8032
Reviewer: Steven A Grover

Reviewer's report: General

This is an interesting and original manuscript that provides useful insights into the shortcomings of current risk assessment models and suggests a useful alternative approach backed up by a serious attempt at model validation. It should be published.

---------------------------------------------------------------

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

The authors have made a serious attempt to validate their model on a data set that is independent of the data used to develop the model. This is a major improvement over most other papers in this area. However, I do note some difficulties with the validation given the limited time horizon of 20 years where it appears that approximately 50% of patients enjoy a survival rate of 80% or more.

A good point, but not one that is alterable without access to data sets extending beyond 20 years. This will be sought in future.

Accordingly, the Whickham cohort data tend to cluster in the right hand corner of figure 2 limiting the usefulness of the validation exercise where a high correlation is inevitable given the restricted time horizon. It would be interesting to see how ranking individuals by only age would correlate with the observed results.

Further analysis has been carried out to rank the subjects according to age, blood pressure, risk and total cholesterol to give a broaden the scope of the analysis.

It also follows that the estimates of error presented on page 8 are probably overly optimistic given the forced data clustering in figure 2. The authors should consider this in their evaluation of model validation and at very least present the results discussed in the opening paragraph of the discussion section in a transparent fashion... a table or figure would be useful providing more detail than simple p values.

The confusing statistical testing using power calculations has been removed.

If they can identify alternative validation exercises using the Whickham data this would greatly improve their paper.

The validation exercise has been extended to an analysis of the correlations between the model estimates of risk and the actual survival when the sample is ranked and grouped according to systolic blood pressure, age and total cholesterol.
The authors should also compare their estimates of survival benefit in figure 3 with those published by others... it appears that eliminating CVD results in approximately an additional 3 years of life expectancy which is small compared to other published estimates.

This part of the analysis has been extended. The original potential gain in life expectancy (PGLE) was for a single arbitrary individual and so differs for any other estimate of PGLE I have been able to find as the published estimates tend to relate to a population. Consequently, I have included averages for all individuals aged 35 and 55 years old to facilitate more direct comparison with the published literature.

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

All results presented in the abstract should be presented in the body of the manuscript.

Done.

Greater detail in presenting the results would be helpful to the reader.

The results section has been greatly elaborated and is hopefully clearer now.

A reference for the Framingham model that was used in page 5 is essential.

The reference to the Framingham equations have been added at this point.

A summary table describing the risk factors of the Whickham cohort would be useful.

Added.

-----------------------------------------------------------------------------------------------
Discretionary Revisions (which the author can choose to ignore)

I would include the supplemental tables in the manuscript.

Done.

I would also clarify in the first sentence of page 4 whether the model they refer to is that cited in reference 10 or refers to the model that is presented in the present paper.

Clarified in the text.
Reference List

