Author's response to reviews

Title: Estimation of Progression of Multi-state Chronic Disease Using the Markov Model and Prevalence Pool Concept

Authors:

Tao-Hsin Tung (ch2876@chgh.org.tw)
Pesus Chou (ht tung@ms66.hinet.net)
Hui-Chuan Shi h (arden.shih@msa.hinet.net)

Version: 2 Date: 28 June 2006

Author's response to reviews: see over
Re: Tung et al “Estimation of Progression of Multi-state Chronic Disease Using the Markov Model and Prevalence Pool Concept” (MS: 8018443429409976)

Dear Editor,

Thank your for your letter and referee comments from June 5, 2006. We thank the reviewers and the editor for the useful comments and suggestions. We have revised the paper as requested. The important issues raised by the reviewers have been clarified, corrected, and elaborated. We hope the correction of the revised manuscript is satisfactory and meets the requirements of highly-reputed journal. Please find the revised manuscript and a detailed reply to the referee. We are happy to make further changes if required.

Yours Sincerely

Tao-Hsin Tung (Corresponding author)
Reply to Referee 1

I basically find this article to be a fairly well written and interesting methodologic paper. The authors are using a combination of Markov modeling techniques along with an assumption of approximate stable prevalence to assist in estimation of chronic disease progression.

However, given the audience of this journal, I don't think it is appropriate. It is basically a methodologic paper with an application of the method shown for illustration purposes. It seems to me that articles in this journal should have an application as the main focus, and articles with a primarily statistical/epidemiologic methodologic focus belong elsewhere.

Agree. Thanks for the reviewer’s useful comments. We made apology submit the manuscript to the inadequate journal. The submitted manuscript has now been transferred to BMC Medical Informatics and Decision Making.
Reply to Referee 2

A common method to assess the impact of healthcare interventions over time involves using multistate markov models to track the changes in incidence, prevalence and costs of disease states and survival. The biggest challenge in using these models is to obtain reasonable estimates for transition probabilities between states from existing data sources that often represent partially observed cohorts at 2 or 3 cross sectional time points.

This is an interesting paper that presents a fairly technical discussion of a method for calculating transition rates based on both complete and partially observed data at several follow-up times. The authors argue that their approach using an E-M algorithm based on prevalence pools and Markov process models is simpler than previous parametric methods that they cite.

General points:

1. I would be interested in them expanding a bit on the conceptual differences between their approach and the others cited in the introduction. Also in some reading I did to try to better understand what they were doing, I came across another cite that seems like it would be worth commenting on. (Welton N, 2005, Medical Decision Making 24:633-645).

While I have in fact published some markov models using much cruder approaches to derive transition probabilities from cohort data, this is not my primary area of interest or expertise and it is difficult for me to evaluate the mathematical details of their approach. I trust that they are correct. However, I would suggest that despite my lack of expertise specific to their question, I am still probably among the more numerate members of any audience that the authors are likely to find and I nevertheless found this article hard to follow.
Ans. Thanks for the reviewer’s useful comments. In this manuscript, in fact, we proposed an illness-and-death Markov model together with prevalence pool concept to estimate the progression rates of multi-state disease. Missing data on interval cases have been taken into account. The proposed method not only stabilizes estimation with respect to parameters of multi-state transitions but is also useful for data on screening for chronic disease without interval cases. An illustration using type 2 diabetes was given in this study. As described above, we think the manuscript is worth to consider for publication.

2. It is not that it is badly written, and in fact both the introduction and conclusion are very well set up in terms of trying to set out the rationale and motivation for the approach and listing and discussing some of the advantages over the existing approaches. It is just that for someone who is not a population epidemiologist or specialist in Markov modeling, the mathematical descriptions of the likelihoods are too long and the conceptual explanations are too terse and ask for too much effort from even an interested and motivated audience from other fields.

For example, the introductory paragraph "Firstly, time to pre-clinical screen-detectable phase for prevalent screen-detected cases (identified in the first screen) is more uncertain than that for incident screen cases (identified in later screens) because prevalent screen-detected cases are left-censored whereas incident screen-detected cases are interval-censored. The latter usually provides more information on occurrence rates than the former." It needs another sentence or two to explain how the terms apply to the cohort data used (it does become clearer later) and to expand on the statement that "The latter usually provides more information on occurrence rates than the former".

Or this statement in the discussion, "In addition, simultaneous estimation of
lambda 1 and lambda 2 may encounter a collinearity problem due to a high correlation between two parameters. This phenomenon may be observed when there is no data on interval cases, which are sometimes unavailable for unregistered conditions such as Type 2 diabetes.” is just slightly too opaque. I think it is simply saying that if you don't have much information on the intermediate states then it is hard to disaggregate the overall rate into distinct rates for each individual state transition. Another sentence to that effect would make it more accessible.

**Ans.** Thanks for the reviewer’s useful comments. We made apology use the confused expressions. The expressions have been corrected. Please see page 3, line 2-4 and page 17, line 15-17.

3. I think some of the technical details might be moved to an appendix and even expanded so that it would be easier to replicate their methods (in some of the Welton cites they actually include BUGs code for their examples) and more time then spent on conceptual explanations.

   Obviously, if the authors picked a journal whose audience was composed primarily of quantitative population epidemiologists this might not be necessary, and that might be one approach to follow as I am not really sure that this article belongs in BMC Family Practice. It does not fit with the type of articles that have been published there as I look through the last few issues and is way too technical for most clinical audiences. The fact that a diabetes example is used to illustrate the technique is almost completely incidental the way the article is written at present. This is a technical methods piece.

   If the authors want to publish in an open source journal, Why wouldn't they publish in Population Health Metrics? There are some other related articles that are in that journal (e.g. Patten SB, Lee RC. Describing the longitudinal course of major
depression using Markov models: data integration across three national surveys. Popul Health Metr. 2005 Nov 15;3:11) and I think they would have an audience that would be much better suited to the material. BMC HSR is another possibility although not as good a fit I would think. Obviously places like Medical Decision-making would be another possibility.

But for a clinical or even a general health services journal they are going to have to go through the piece and add a more accessible conceptual explanation of what they are doing and why.

**Ans.** Thanks for the reviewer’s useful comments. We made apology submit the manuscript to the inadequate journal. The submitted manuscript has now been transferred to BMC Medical Informatics and Decision Making. In addition, in order to simplify the manuscript, we had moved some explanations to appendix in the original manuscript and we think the related statistical experts could read easily. However, if the reviewer still prefers the technical details might be moved to another appendix, we are happy to make a change.

4. Now there were a few aspects of the models they were examining that seemed to raise some questions for me. First, for example, given the perceived increase in population rates of type two diabetes throughout the world, is a prevalence pool approach justified? Second, in the mortality model why would the only path to death be through the symptomatic state of diabetes? Type two diabetes occurs in older populations and death is a significant competing risk and I would think you would want a model where both persons without disease and with asymptomatic disease might die.

**Agree.** 1. The reviewer is absolutely right. However, the concept of the prevalence pool was used by Rothman and Greenland (1998). Brookmeyer (1995) also applied
this concept to estimate progression rates associated with HIV and AIDS. It states that, in a steady population, the number of people entering the prevalence pool is balanced by the number exiting from it. We think the approach could also apply to chronic diseases such as type 2 diabetes.

2. The reviewer is absolutely right. Type 2 diabetes occurs in older populations and death is a significant competing risk and we should model both persons without disease and with asymptomatic disease to death. However, we did not have large enough sample sizes to estimate the transition rates from normal or asymptomatic phase to death exactly in the present study. Further long-term studies should be conducted to explore the different parameters of natural history for this chronic disease.

**Specific points**

p15 top- Is the OGTT sample random or selected based on history or other lab tests.

**Ans.** Based on the population-based screening for type 2 diabetes in the present study, all the overnight fasting and 2h serum and plasma samples (preserved with EDTA and NaF) were collected and kept frozen (-20°C) until analysis. Fasting plasma glucose concentrations were determined using the hexokmase-glucose-6-phosphate dehydrogenase method with a glucose (HK) reagent ldt (Gilford, Cberlin, OH). The descriptions have been corrected. Please see page 14, line 6-10.

**Minor Essential Revisions**

1. p16 top "lambda 1 was estimated as 0.0107594" is this a typo, in the estimate for lambda 2, lambda 1 is referred to as 0.107594

**Agree.** We made apology use the wrong number. The estimation has been corrected.
Please see page 15, line 7.

2. p19 top fragment "These could be expected to have 5-year cumulative death rate from the, of"

**Ans.** Thanks for the reviewer’s useful comments. We made apology use the confused expressions. The expressions have been corrected. Please see page 18, line 4.