Author's response to reviews

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

Authors:

Hui-Chuan Shih (arden.shih@chgh.org.tw)
Pesus Chou (thtung@ms66.hinet.net)
Tao-Hsin Tung (ch2876@chgh.org.tw)

Version: 3 Date: 25 October 2006

Author's response to reviews: see over
Dear Editor,

Thank you for your letter and referee comments from September 14, 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**

The authors have changed the target audience by shifting to a informatics and medical decision-making journal which I think is a good decision.

Some of the responses to my suggestions still need some work.

1. I mentioned that another cite that seems like it would be worth commenting on or at least citing is Welton N, 2005, Medical Decision Making 24:633-645. I would include it or respond that you "looked at it and didn't find it to be useful or relevant for the following reasons..."

   **Agree.** The reference has been cited in the present study.

2. In the first review I wrote: "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 in estimating transition probabilities (it does become clearer later) and to expand on the statement that "The latter usually provides more information on occurrence rates than the former" Unfortunately the response was to add the following sentence and is not successful.

   page 3 para 1 "Due to the date of diagnosis is hard to identify for prevalent screen-detected cases, the incident screen-detected cases usually provides more information on occurrence rates." This new sentence is not grammatically correct or clear and I can't quite figure out what the authors are trying to say in order to correct it. The authors need to try again to explain this better, as I note below this is one of several
places where working with a good editor might help the clarity of the paper considerably. **Agree.** We made apology for the wrong English grammar. The descriptions have been corrected. Please see page 3.

3. In the first review I wrote: given the perceived increase in population rates of type two diabetes throughout the world, is a prevalence pool approach justified? The response to this was that "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." This response basically agrees that there might be a problem and then restates the point. I think the authors need to include a sentence in the paper as a limitation. It should say that this approach will only work when estimates can be made using a population where the rates of disease are assumed to be at a steady state and that this assumption may not necessarily apply to diabetes at this point in time given the rapid increase in the incidence of diabetes that has been observed. **Agree.** The limitation has been described. Please see page 19.

4. In the first review I wrote:"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. Again the response acknowledges this point but the authors also need to add a sentence in the paper, perhaps in section 3.1.1 noting that a more complete model would include competing mortality but that the data sources they were using as an example did not allow them to estimate
competing risks. I would also wonder whether it was just a data problem or whether adding competing risks might significantly complicate the method of estimation that they are proposing?

**Agree.** The limitation has been described. Please see page 19.

5. In the first review I wrote: "p15 top- Is the OGTT sample random or selected based on history or other lab tests." In the response, the authors misunderstood what I was asking for and provided technical details on the storage of specimens and the test method. Given that the analysis was restricted to those subjects who had complete OGTT data, I wanted to know if that subsample could be considered a random sample of the population or whether by design the OGTT testing was only done on a subset with specific characteristics. In the latter case, the use of the OGTT subset would produce parameter estimates that would not be representative of the population but only of the sub-population that had the characteristics that led to them having an OGTT.

**Agree.** We made apology for misunderstanding the reviewer’s comments. The limitation has also been described. Please see page 19.

6. There are a number of grammatical and language errors some of which are new (noted above in 2/) and a large number that I seemed to have missed the first time when I was concentrating on the conceptual issues. This is an interesting paper and I would really suggest that the authors get an editor to review the draft and help with the grammar and phrasing of some of the sentences so that they do not detract from the overall effect. I'm not completely confident that I caught all the language problems with the time I was able to give to this review.

page 3 para 4 "The use of the prevalence pool concept (Rothman and Greendland., 1998) that if prevalence is roughly constant." this is not a sentence.
"The numbers leaving the prevalence pool must be approximately equal to the numbers joining it may be amenable to the estimation when there is no data on interval cases." this is not a sentence.

"Suppose the natural history of a chronic disease to be defined" should be "...chronic disease is defined"

"preclinicable" should be preclinical

"Firstly, it is not computationally intensive as single stage estimation in the traditional Markov model." should be "Firstly, it is not as computationally intensive as a single stage estimation using the traditional Markov model."

Also the tense changes in the next sentence: "The reason for this was that the prevalence pool equation and the illness-and-death equation were integrated into the likelihood function to make estimation of parameters simpler." The tenses need to be fixed.

How about "Firstly, it is not as computationally intensive as a single stage estimation using the traditional Markov model. The parameter estimation is simplified by integrating the illness and death equation into the likelihood function." In addition, the rest of the paragraph could be improved considerably by working with a good editor.

**Ans.** We made apology for the bad English grammar. The manuscript has been revised and corrected by an English editor. The English problem should be improved.
Reply to Referee 2

General

This is an interesting paper which presents an innovative idea. Basically, the authors propose a Markov model with the added constraint that the number of individuals in the prevalence pool is constant over time.

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

Although this work is interesting. There were various issues which need clarification.

1. This idea of a prevalence pool concept needs more explanation. How does it differ from saying that a stochastic process is in equilibrium? Does one assume that it is a sensible assumption or can one examine whether it is a reasonable assumption based on empirical evidence (i.e., through hypothesis testing)?

   **Ans.** Thanks for the reviewer’s useful comments. The idea of a prevalence pool concept has been further discussed in the introduction section. Please see page 1-5.

2. I found the E-M algorithm development somewhat confusing. Equation (7) treats the parameters as random variables. Is the estimation procedure using a Bayesian approach? At the end of the process are estimators Maximum likelihood estimators?

   **Ans.** The reviewer is absolutely right. The estimation procedure in the present study is using a Bayesian approach and at the end of the process are estimators Maximum likelihood estimators.

3. The authors suggest that Markov estimation without the Prevalence Pool Concept is
difficult and unstable. It is not clear why? Approaches such as Kalbfleish and Lawless (1985) are fairly easy to implement. It would be very helpful to show an analysis of the Type 2 diabetes using a standard Markov modeling approach. A comparison which shows the advantages of the authors approach over the standard Markov approach would help readers to understand the merits of this new approach.

**Ans.** Thanks for the reviewer’s useful comments. We agree that approaches such as Kalbfleisch and Lawless (1985) are fairly easy to implement using a standard Markov modeling method. However, from the best our knowledge, it is not computationally intensive as single stage estimation in the traditional Markov model. The reason for this was that the prevalence pool equation and the illness-and-death equation were integrated into the likelihood function to make estimation of parameters simpler. The traditional three-state model usually estimated \( \lambda_1 \) and \( \lambda_2 \) simultaneously using a full likelihood function. Therefore, the likelihood function in the traditional method is more complicated than that in the present study.