Author’s response to reviews

Title: Simultaneous evaluation of abstinence and relapse using a Markov chain model in smokers enrolled in a two-year randomized trial

Authors:

Hung-Wen Yeh (hyeh@kumc.edu)
Edward F Ellerbeck (eellerbeck@kumc.edu)
Jonathan D Mahnken (jmahnken@kumc.edu)

Version: 2 Date: 30 March 2012

Author’s response to reviews: see over
March 29, 2012

RE: MS: 3153360496190477 - Simultaneous evaluation of abstinence and relapse using a Markov chain model in smokers enrolled in a two-year randomized trial

Dear Editor:

We have made revision on our manuscript based on reviewers’ comments. The detailed point-by-point response to reviewer’s comments is attached to the end of this letter. Our response and revision in the manuscript are marked in blue to ease review. The materials have not been copyrighted, published, or under consideration for publication by another journal.

Sincerely yours,

Hung-Wen Yeh

Assistant Professor
Reviewer's report #1

This manuscript presents an interesting analysis of data from a smoking cessation trial. The use of Markov modeling is fairly novel and its use here is the main strength of the paper. The presentation is generally clear and technically correct. The data are sufficient in size and quality to support the findings and the discussion is reasonable and does not go beyond the results. There are several weak points to the manuscript, however. It is not true that most smoking studies apply cross-sectional analysis to longitudinal data anymore. Where it is done, it is usually an appropriate application in that the research question involves a comparison of abstinence rates at a key assessment point. A second problem with the premise of this work is that in many studies the goal is to achieve abstinence. While relapse is common, it is often not the primary focus and once one knows the abstinence rates, obviously one knows the relapse rates. The authors should make it clear that this approach is not the same as (but may be complimentary to) a longitudinal mixed-effects model.

Response: We thank the reviewer’s two comments. We completely agree to the first comment about cross-sectional analysis in longitudinal studies. We have revised the first paragraph of Introduction using more careful wording to avoid the confusion. We also agree to the second comment and have revised the Introduction (second paragraph, p. 3 – 4) and Discussion (second paragraph, p. 12 – 13) to illustrate the differences between mixed-effects models and transition models.

The paragraphs discussing missing data are correct but that information is widely known now and I’m not certain it is worth the space. Also, it is unclear why so much print is given this topic when the authors deleted subjects. While the percent of data deleted is not large, it runs counter to good statistical practice.
Response: We agree with the reviewer about the paragraphs discussing missing data and have deleted them. The reviewer is absolutely right that subjects should not have been deleted if their responses could have been imputed. We have re-analyzed the data using all subjects, and identified one more predictor variable for the abstinence model. We have revised the manuscript correspondingly.

The analysis used “the backward elimination procedure.” What exactly is this? If it means removing variables based on an algorithm, such stepwise procedures are a poor choice. Frank Harell’s text on regression spells out the issues which include the F test statistics do not have the claimed distribution, the standard errors of the parameter estimates are too small and p-values are too low, due to multiple comparisons.

Response: We apologize for not making the point clear. As the reviewer pointed out, the conventional selection procedures based on F statistic are problematic. The “backward procedure” refers to the sequence of models that we have considered, but the comparison criterion was not the F statistic but the Bayesian Information Criterion (BIC). We have made this point clear in the text (p. 8, first paragraph).

The main problem I have with the analysis is that it is fairly constrained. Why did the authors not consider more complex sets of Markov model? An example of such an approach can be found in Delucchi and Weisner’s paper modeling the presence of problem drinking (Delucchi & Weisner, 2007). The authors also fail to cite other work using similar modeling in smoking (Hughes, Keely, Fagerstrom, & Callas, 2005; Martin, Velicer, & Fava, 1996; Velicer, Martin, & Collins, 1996).

Response: We thank the reviewer for the reference that we overlooked. To respond to the comment, we applied the Latent Transition Analysis (LTA) using a SAS procedure PROC LTA (Lanza and Collins, 2008) which produces equivalent results as Mplus (Muthén and Asparouhov, 2011). To investigate the performance of the procedure, we conducted a simulation with the same design as the KanQuit trial (750 individuals, all smoking at enrollment, with 4 follow-up time points) and, without losing generality, used parameters of:
(a) Homogeneous transition probabilities: 0.1 from smoking to non-smoking and 0.2 from non-smoking to smoking;

(b) The false self-reporting rates: 0.3 probability of self-reported non-smoking given actually smoking, and 0.1 probability of self-reported smoking given actually not smoking.

With 1,000 simulation runs, the averages, standard errors (in parentheses), and true parameter value [in bracket parentheses] of estimates are

<table>
<thead>
<tr>
<th></th>
<th>Month 6 proportion</th>
<th>To Month 12</th>
<th>To Month 18</th>
<th>To Month 24</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Smoking</td>
<td>Not smk.</td>
<td>Smoking</td>
<td>Not smk.</td>
</tr>
<tr>
<td>Smoking</td>
<td>0.82 (0.14) [0.9]</td>
<td>0.84 (0.13) [0.9]</td>
<td>0.86 (0.11) [0.9]</td>
<td>0.86 (0.11) [0.9]</td>
</tr>
<tr>
<td>Not smk.</td>
<td>0.18 (0.14) [0.1]</td>
<td>0.25 (0.23) [0.2]</td>
<td>0.20 (0.18) [0.2]</td>
<td>0.20 (0.16) [0.2]</td>
</tr>
</tbody>
</table>

False and true positive and negative probabilities

<table>
<thead>
<tr>
<th></th>
<th>Latent state</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-reported state</td>
<td>Smoking</td>
</tr>
<tr>
<td>Smoking</td>
<td>0.64 (0.08) [0.7]</td>
</tr>
<tr>
<td>Not smk.</td>
<td>0.36 (0.08) [0.3]</td>
</tr>
</tbody>
</table>

The simulation results suggest reasonably accurate and precise estimates. We then confidently applied PROC LTA to our data. The estimates of the LTA models (including different covariates do not alter these estimates) fitted to the KanQuit data are shown below:

<table>
<thead>
<tr>
<th></th>
<th>Month 6 proportion</th>
<th>To Month 12</th>
<th>To Month 18</th>
<th>To Month 24</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Smoking</td>
<td>Not smk.</td>
<td>Smoking</td>
<td>Not smk.</td>
</tr>
<tr>
<td>Smoking</td>
<td>0.843</td>
<td>0.925</td>
<td>0.940</td>
<td>0.906</td>
</tr>
<tr>
<td>Not smk.</td>
<td>0.157 &lt; 0.0001</td>
<td>&gt; 0.9999</td>
<td>0.129</td>
<td>&lt; 0.0001</td>
</tr>
</tbody>
</table>

False and true positive and negative probabilities

<table>
<thead>
<tr>
<th></th>
<th>Latent state</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-reported state</td>
<td>Smoking</td>
</tr>
<tr>
<td>Smoking</td>
<td>0.970</td>
</tr>
<tr>
<td>Not smk.</td>
<td>0.030</td>
</tr>
</tbody>
</table>
The results indicate that there is nearly no chance (< 0.0001) of relapse, which is very unlikely. Moreover, given the latent status was smoking, the respective chance that one reported her-/himself as smoking and non-smoking was estimated to be 0.97 and 0.03; namely, participants were 97% honest or provided correct self-reported smoking status. On the other hand, if an individual was not smoking, the chance that this person reported the status to be smoking was 0.13. It is also unlikely that the false positive probability would be higher than the false negative in smoking studies.

Furthermore, the BIC of the LTA model is not superior to the BIC of the conventional Markov model with no covariate (2117.0 vs. 2080.1). Therefore, we chose to apply the conventional Markov model rather than the latent model.

We have added a paragraph in Discussion (p. 14) and explained the rationale we did not apply the latent Markov models. We briefly discussed when these methods might be necessary and cited the articles that the reviewer suggests except Hughes et al. (2005). Please note that Hughes et al. mentioned but did not perform LTA. Therefore, we did not cite this work in our revised manuscript.

Finally, it is not clearly explained to the average reader what information this form of modeling provides over and above the more standard approach.

Response: We regret we did not make this clear in the first version. We have revised the second and third paragraphs in Discussion (p. 12 – 13) to elaborate the knowledge earned from this modeling.

Reference


Reviewer's report #2

Summary: A first-order Markov chain model is applied to analyze the transition of smoking status measured every 6 months in a 2-year randomized smoking cessation trial. A sensitivity analysis is conducted to evaluate the effect of missing values due to non-response.

Strength: This is a well-written article. Researcher interested in the longitudinal data analysis with dealing with missing values may find it useful.

1. The authors use a first-order Markov transition model to evaluate the intervention effects in a randomized smoking cessation trial. This model provides flexibility to identifying factors that might work differently in one direction (e.g., from smoking to abstinent) versus the other (from abstinent to smoking). This is the limitation for the GEE and GLMM which work only on one direction.

2. A comprehensive sensitivity analysis is performed to evaluate whether the missing values affect the results. This includes (a) available data only (as the baseline), (b) missingness depends on whether the smoking status is observed at the previous time point, the actual smoking status at previous time and current time point, and (c) additional dependency on time and treatment factors.

3. The longitudinal smoking status (Markov chain model) and the missing process are modeled and estimated by the EM algorithm with detailed mathematical justification.

4. Results demonstrate the strength of Markov chain model and show a differential treatment effect from one direction to the other direction. For example, HDM has a smaller p value in the abstinence model than in the relapse model. In contrast, MDM has a larger p value in the abstinence model than in the relapse model. Such findings provide researchers the opportunity to examine why both arms works differently.
5. Results also highlight the importance of sensitivity analysis for missing values. P value was higher in the analysis with missing value imputed than in the analysis with data only available. It allows us to better understand the impact by missing values.

Minor issues:

1. p. 12: (0.05<p<0.07). The p value should be reported (e.g., p=0.06).

   Response: We thank the reviewer’s comment. After we took another reviewer’s comment to analyze the complete data without excluding any participants, now all p-values are significant and we hence revised the results section and deleted this sentence.

2. There are many nice features of the jointed model (Markov chain model + missing process). It would be great if the SAS code is provided as an appendix to share with the research community for broad application.

   Response: We have added SAS codes in Appendix as suggested.
Reviewer's report #3

This research methodology is important and timely. The authors' first paragraph is right-on. As is their second. Critics will object to the memoryless quality of the Markov model (MM); but it is a more innocuous simplification than many others that are made by others. The focus on missingness, and how to handle it, is, I think, overdone; it is important how we impute status to drop-outs, of course. In the Ellerbeck study that provided the data for this analysis, significant differences vanish if it is assumed that those who are non-responders have reverted to smoking (a more reasonable imputation, imho). Only in the 'secondary analyses', with 'no imputation', were significant differences found. Thus, the nominal success or failure of a major study hinges on how one treats drop-outs.

Response: We thank the reviewer for the comment on the issue of missing values. We would like to emphasize that the way we handled non responses follows the recommendation of Hall et al. (2001). We have re-analyzed the data without excluding any subjects according to another reviewer's comment. Now the sensitivity analyses provide consistent conclusions as that of analysis using available data.

Minor essential revisions.

page 4, do a better job of clarifying the kinds of missingness. (Do not rely on citations 3 and 4 to do that job for you).

Response: We thank the reviewer for the comment. According to another reviewer’s comment that the content in paragraphs about missingness in Introduction is well known, we have deleted these paragraphs accordingly.

Make contact with the work of CK Enders (e.g., 2006) on modern techniques for missing data imputation.

Response: We have cited Enders (2006) in the Discussion about multiple imputation (p. 14, [19]).
Make contact with the hoary Kaplan-Meier "missing at random" bias.

**Response:** We thank the reviewer's comment and have mentioned this in the last paragraph of Introduction.

Eq 2, and the general treatment, is elegant and thorough.

Perhaps too much so; I would have preferred a model-comparison approach, with time-invariant transition probabilities (over all groups) as a "null hypothesis", and then evaluation of the information gained by recognizing the different groups. But I understand that is not the authors' tradition.

**Response:** We agree with reviewer's viewpoint regarding testing the transition probabilities across time points. However, we believe Eq. (2) is more versatile and allows evaluating the impact of missing values potentially not missing at random, which cannot be answered by the “model-comparison approach”.

I recognize that the authors are married to the KanQuit data set; but I think they do their modeling an injustice, by restricting it to these data. Because of the liberality of the intervention, the intervention effects are weak. (Even in the best of cases, of course, smoking interventions are relatively weak). Nonetheless, nothing new is learned from all of the analyses. The authors should do a better job of resonating all of the points that they do make with prior discoveries—that is make better contact with the traditional literature. They should do themselves a favor and apply this modeling to data-sets with stronger effects (e.g., perhaps the Piper et al studies of varenicline, etc).

**Response:** We regret that we did not explain clear enough about what is learned in the transition model. We have revised the Discussion, mainly the second and the third paragraphs (p. 12 – 13), to make connection to the prior discoveries and elaborate what’s new in the current work. In addition, we have re-analyzed the data without excluding any individuals and found significant intervention effects and confirmed with sensitivity analysis. Certainly, we will continue applying our method to our research when interventions have similar or even greater effects.
They should also make better contact with applications of MM to cessation data. For instance, they talk about the cost of interventions, but there are many good articles using MM to analyse the monetary benefits from successful cessation programs, harm reduction, etc. (e.g., Hurley et al; Orme et al; etc etc etc). They certainly must make contact with the BENESCO model.

**Response:** We thank the reviewer for the suggestion, and agree that analyses of monetary benefits and harm reduction are so important and deserve further research. In our understanding, the Benefits of Smoking Cessation on Outcomes (BENESCO) model requires the knowledge of morbidity status at multiple time points as where smoking status was evaluated. In this trial, however, morbidity was only collected at the baseline but not at the follow up time points. Hence, the design doesn’t allow modeling transition among disease status by the BENESCO model.

I have applied my MM using the average published parameters for individuals in cessation programs, and placebo controls, to the data from the minimal PM group and maximal HDM group (bottom rows of my Table 1). It does a surprisingly good job of predicting these data (with the exception of the 24 month PM group). I attach pictures. The authors, excellent statisticians, can take this farther than I, if they care to. (Of course, given the radical difference between validated cotinine levels and self reports (the former being half the latter) in the source data, much of this attention to detail may be mere sophistry).

**Response:** We thank the reviewer’s suggestion on creating the prediction curves for cessation rates, which provides visual presentation and definitely is strength for the original work of Ellerbeck et al. that focuses on cessation rates. However, this work focuses on transition of states. Hence, we added Figure 1 to present the transition probabilities instead. As for the issue of discrepancy between validated saliva levels and self-reported states, we have added in Discussion (p. 14) about using the latent modeling approach.