Author’s response to reviews

Title: Is There a Role for Expectation Maximization Imputation in Addressing Missing Data in Research Using WOMAC Questionnaire? Comparison to the Standard Mean Approach and a Tutorial

Authors:

Hassan MK Ghomrawi (hag2008@med.cornell.edu)
Lisa A Mandl (Mandll@hss.edu)
John Rutledge (jrr2002@med.cornell.edu)
Michael M Alexiades (AlexiadesM@HSS.EDU)
Madhu Mazumdar (mam2073@med.cornell.edu)

Version: 3 Date: 1 March 2011

Author’s response to reviews: see over
We thank the reviewers for their comments. Additions based on their comments have greatly strengthened the discussion section. All additions to the text have been underlined for ease of identification by the editors.

This document includes the authors’ response to the reviewers’ comments. The authors’ response is in *italics* below each of the reviewers comments.

**Reviewer 1:** Gabriel Baron

**Reviewer’s report:**
The WOMAC questionnaire is known to produce missing data when computing global scores. The objective of this report was to compare different imputation methods. Authors compared 3 methods: mean imputation, expectation-maximization (EM) method and method recommended by the authors in the original paper. The EM methods provided better results in term of bias.

The authors have performed a well-done work, applying appropriate methods. However, I have some comments.

Minor Essential Revisions

1- The sample is biased because the study concern only patients with complete data (1339 of 2062 patients). Authors should compare baseline characteristics of the 1339 with complete WOMAC vs. 723 with incomplete WOMAC to ensure that the 2 populations are the same on controlled factors.

*We agree that this kind of comparative analysis will be absolutely necessary when we present analysis of the cohort of 2062 patients and answer a clinical question. However, in this tutorial paper, we are not presenting an analysis of the original data. We are presenting analysis of only simulated data with the complete data cohort of 1339 patients and borrowing the missingness percentage observed in the 2062 patient cohort as our guide. Therefore we don’t think this analysis is necessary here.*

2- I wonder why the authors did not use the multiple imputation method?

*The authors are suggesting EM as only one of the available advanced methods that can be used instead of the mean-substitution method and not the only method. Our choice of this method is based on its relative ease of use in already available packages and the fact that EM relies on fewer assumptions than the multiple imputation method. Multiple Imputation methods take longer to run and somewhat more complex to explain to practitioners. However, we have added a sentence to the discussion about their existence (page 11, line 9).*
3- Absolute mean bias and/or relative mean bias should be computed to quantify bias introduced by missing data. These estimates in association with boxplot will give complete information on bias introduced by missing data.

Absolute mean bias and relative mean bias being just transformations of the ‘true score’ and ‘estimated mean score’ show similar results. We prefer to keep the metric of score as it is more appealing to general practitioners who are more familiar with the score as presented.

Discretionary Revisions

Page 10 discussion: Using one of the short validated version of the WOMAC function is also one way to reduce the amount of missing data.

Although short forms will have less chance of having missing data, if the missing data was observed, the methods discussed here will still be of value. We have added a comment in the introduction about this issue (page 4, line 6).

Reviewer 2: Tim Spelman

Reviewers report:
Sample attrition and missing data are common characteristics of many real-life studies. Although the first response to expected attrition should rightly involve maximizing response rates within the logistic and financial constraints of an investigation, there is a clear need for researchers to be informed on available and suitable statistical contingencies for dealing with missing data.

This paper is a well-written, concisely presented addition to the ongoing development and refinement of probabilistic modeling, particularly in the field of quality of life orthopaedics where there is a paucity of formal testing of imputation techniques. Providing an alternative simulation at a higher rate of attrition expands the generalisability of the paper and the provision of a tutorial is an excellent idea.

Discretionary revisions

1) Whilst I have no argument with the mechanics of what the authors have undertaken, I have reservations regarding the assumptions made. The results and conclusions here are only as good as these assumptions. Expectation maximization works on the assumption that the pattern of attrition is missing at random (i.e. the probability that subscale data is missing is not related to the outcome of interest). With percent missing ranging from 2.9% to 14.5% for the various WOMAC items I would be surprised if the pattern of attrition here is genuinely missing at random, as has been assumed in order to demonstrate the benefits of EM. To be fair the authors have cited this as a potential limitation, however if the context here is the use of EM in quality of life research such as WOMAC, with its reliance upon Likert scales, I would suspect data attrition would rarely be random. As such even though the authors have clearly demonstrated the advantages of EM I'm wondering just how broadly applicable EM could be in this type of research. Given a random pattern of attrition is one of the fundamental assumptions underlying EM it would be very useful to include a brief guide to assessing the pattern of missing data. This would greatly assist researchers in the process of deciding whether EM is a suitable
contingency for dealing with missing data and avoid erroneous or misguided applications of EM.

Expectation maximization is applicable whenever the data are ‘missing completely at random (MCAR)’ or ‘missing at random (MAR)’ but unsuitable when the data are ‘not missing at random (NMAR)’


Several procedures can be undertaken to establish whether the data are ‘missing completely at random’, ‘missing at random’, and ‘not missing at random’. First, for each variable, researchers can assess whether the data differs between individuals who responded to some variable and individuals who did not respond to some variable.

For example, a series of t-tests or a logistic regression analysis can be undertaken to assess whether individuals who generated responses on the WOMAC scale and individuals who did not generate responses on the WOMAC scale differ on age, or other relevant covariates. Non-significant findings indicate that, perhaps, missing data on this variable is random; otherwise, at least one variable should differ between individuals who responded to this variable and individuals who did not respond to this variable.

Similarly researchers could calculate Little's MCAR test [Little, R.J.A. & Rubin, D.B. (1987) Statistical analysis with missing data. New York, Wiley]. A non-significant finding is consistent with the assumption that data are ‘missing completely at random’ and hence expectation maximization is applicable. This test will appear by default when expectation maximization is undertaken in SPSS.

Establishing missing at random

If the data are not ‘missing completely at random’, they might nevertheless be ‘missing at random’. To establish this possibility, expectation maximization needs to be invoked in SPSS as recommended in this tutorial. User needs to proceed to the table labeled Separate Variance t Tests. If all the p values exceed .05 or alpha, the data are missing at random. Expectation maximization is thus warranted. Part of this writing and relevant references are now added to the tutorial to guide researchers in assessing the pattern of missing data (page 13, lines 27).

2) There is a similar issue here with the assumption that the study data follows a multivariate normal distribution, which the authors point out is rarely the case with Likert scale items. Did the authors consider testing non-normal EM algorithms? Although computationally more difficult, these would arguably be more realistic in the context of quality of life score data.
We thank the reviewer for this comment. Yes, the authors considered testing other forms of EM algorithms. Specifically, they attempted employing the EM algorithm for imputation with multinomial distribution which is the accepted underlying form for the Likert scale data. However, as the reviewer pointed out, the algorithm, which is computationally more difficult, took more than a day to run on a regular computer and did not result in much different outcome. While it may be a more realistic in the context of quality of life measures, it will be hard to use it routinely in practice. The authors, aiming to provide a user-friendly solution to the practitioner using the WOMAC instrument, believe that the benefit in estimating an EM algorithm with multivariate normal distribution is advantageous to the mean imputation method and outweighs the impediments of estimating a more computationally demanding model that may provide only a marginal benefit. We have added this language to the discussion section (page 10, line 22)