Author's response to reviews

Title: Latent variables and structural equation models for longitudinal relationships: an illustration in nutritional epidemiology

Authors:

Michel Chavance (michel.chavance@inserm.fr)
Sylvie Escolano (sylvie.escolano@inserm.fr)
Monique Romon (mromon@univ-lille2.fr)
Arnaud Basdevant (arnaud.basdevant@htd.aphp.fr)
Blandine de Lauzon-Guillain (blandine.delauzon@inserm.fr)
Marie-Aline Charles (marie-aline.charles@inserm.fr)

Version: 2 Date: 31 March 2010

Author's response to reviews: see over
Reviewer's report
Title: Latent variables and structural equation models to model longitudinal relationships: an illustration in nutritional epidemiology
Version: 1 Date: 9 December 2009
Reviewer: Fruhling V Rijsdijk
Reviewer's report:
- Major Compulsory Revisions
  Terminology and Model-fitting analyses:
The basics of latent variables and SEM model-fitting are well explained. However, it’s not clear to me whether the authors have fitted the proposed model on the correct data, or if they have just used ill-chosen terminology. What do they mean exactly with ‘change’ scores in Adp indicators? Do they mean the difference between baseline and time 2 scores or simply the ‘Time 2’ scores? Some sentences imply the former, others don’t, e.g.
on Pg 6: ‘Finally, we fitted a model for each sex group considering simultaneously the baseline measurements and their changes and assuming the same relationships between latent adiposity and its four indicators at baseline and two years later; this model constrained the four loadings, i.e. the regression coefficients, to be identical for baseline adiposity and adiposity change’.
This is a crucial point. There is some controversy regarding the use of difference scores in SEM, which is not mentioned nor justified. More over, the mix of base-line and difference scores can lead to even bigger problems. If fitting SEM models on difference scores, critics could argue that these change or difference scores must have some variability to function as good predictors (or outcomes), which they often do not, and that they frequently correlate with the initial level of the characteristic measured. If they have fitted the SEM model on a combination of baseline and difference scores, the variance covariance structure of the data most likely will not be multivariate normal, an important assumption in SEM. In addition, the variance/covariance structure of the ‘change scores’ themselves might be totally different from the baseline measures, and, thus, the model (same loadings of change on Adp time 2 as baseline on Adp at time 1, same errors with neg signs) would not make sense at all. As a consequence of these problems, I would be reluctant to use these scores. If the authors have a strong case for doing this, this unusual approach should be discussed (with relevant references) and not glossed over. This is important as the aim of the paper is to introduce SEM to this research field.
If I’m wrong, and the authors have modelled simply the time 1 and time 2 variables, I would strongly recommend to delete all references to ‘change scores’ in the text, and simply state that they are the Time 2 scores of the same variables.

On the other hand, if the authors are really interested in modelling ‘change’ as such, they will require fitting entirely different models, e.g. growth curve models, in which both the means an variance / covariance structure are modelled as a function of growth parameters such as intercept and slope, which is not possible with only two time points of data, by the way. However, as far as I know, modeling only two-time-point data (pre- and posttest change) in SEM is also achieved to some extent by analyzing the actual post scores but specifying them as a function of the pre-scores, e.g.

PRE1 = X + e1,
PRE2 = X + e2,
Our approach is justified by the fact that the distribution of the measurement at time 1 and time 2 can be expected to be similar when the delay between the two times is not too large, which is the case here. As stated in the appendix, identical loadings for the measurements at time 1 and time 2 imply identical loadings for the baseline measurements and the changes, even though the covariance structure is not the same: the model assumes different variances for the baseline measurements and their changes. If each distribution is normal, the difference should also be normal. The approach used by Steyer et al. makes sense when the correlation between the repeated error terms is low. Here we can expect that the repartition of the fat mass of a given subject between two measurements will not change very much: someone with much of her/his fat in the abdominal compartment at time 1 will not have it in her/his lower limbs at time 2, even if her or his adiposity changed. Thus, the assumption of independent errors will be far from reality. As mentioned in the article, the measurement changes are not independent from the baseline measures, and the errors also are correlated, but the correlations may be expected to be lower, in absolute value, when the measurement change is used instead of the measurement observed at time 2. Actually, when we used the model of Steyer et al., the global fits of the models were not satisfying (RMSEA=0.23, NFI=0.84 for females and RMSEA=0.20, NFI=0.87 for males).

Continuing on the topic of terminology: on page 7, it is suggested that ‘adiposity and CRS changes are modeled conditionally on their baseline values’ What do they mean by ‘conditional on’?

Expression "conditional on" was unclear; it has been replaced by "adjusted for" the effect of adiposity and CRS changes are adjusted for their baseline values (i.e., both the baseline value and its change appear in the same equation)

Also very puzzling and unusual are all the adjustments made e.g. ‘…the effect of baseline adiposity on CRS change was adjusted for baseline CRS and thus freed from the factors confounding the cross-sectional effect’ and ‘The direct effects of baseline CRS on adiposity and CRS changes were also adjusted for baseline adiposity and freed of the cross-sectional confounding effects’.

What do they mean by ‘adjusted for’?

See above.

Level of interest: An article whose findings are important to those with closely related research interests
Quality of written English: Acceptable
**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**
I declare that I have no competing interests. I
Reviewer's report

Title: Latent variables and structural equation models to model longitudinal relationships: an illustration in nutritional epidemiology
Version: 1 Date: 22 January 2010
Reviewer: Emil Kupek

Reviewer's report:
The manuscript illustrates the use of structural equations modelling (SEM) to determine direct and indirect effects of baseline adiposity on adiposity change 2 years after. Adiposity is defined as latent variable based on 4 observed anthropometric variables which measure different aspects (locations) of body fat and are highly correlated. The mediating variable is a test score reflecting “cognitive restraint” with respect to eating behaviour. The test was applied both for baseline and follow-up evaluations.

The theme is relevant for nutritional epidemiology but needs some clarifications and additional statistical analysis to justify its conclusions. Overall, this is an interesting methodological issue for many researchers who deal with indirectly measured outcomes without a gold standard and want to assess a mediating effect in a longitudinal perspective.

Major Compulsory Revisions

1. Explain the advantages of adiposity as a latent variable as opposed to four observed variables IN STATISTICAL TERMS. This is a central idea of the paper and deserves a better elaboration than just saying “One way to take advantage of all these measurements is to combine them into an adiposity latent variable within a structural model. This approach avoids the drawbacks of either arbitrarily choosing a single adiposity measurement or performing separate analyses on each fat-mass indicator, which leads to multiple testing problems and interpretation difficulties when analyses are not sufficiently consistent.”

   We now refer explicitly to familywise error rate and we added a reference to this concept. If one tests separately the effect of restrained eating on each measurement, the familywise error rate [10], i.e. the probability of making any error in this family of tests when restrained eating has no effect on adiposity, is higher than the size $\alpha$ of each test. By contrast, combining the four measurements into an adiposity latent variable within a structural model avoids the drawbacks of either arbitrarily choosing a single adiposity indicator.

First, why would separate analyses be such a “drawback”? It is perfectly justified to analyse how each of 4 different adiposity measures change in time and how cognitive restraint influences these changes, remembering that the measures reflect (by and large) different locations of fat in the body. A table presenting the results with this standard regression approach compared to the latent variable analysis would provide some grounds to compare the two.

These results are given in Table 6 (ex Table 5)

Second, it is not clear to me what was meant by “multiple testing problems”. The analysis presented is not a multi-group ANOVA comparison where such problems may arise (but there are ways of dealing with them). As long as your observed
outcomes are not exchangeable, I see no multiple testing problem in the way it is usually defined.

The sentence was modified and a new reference was added. Indeed, the issue we had in mind was the control of the familywise error rate.

Third, what was meant by the phrase “when analyses are not sufficiently consistent” and what type of “interpretation difficulties” would it imply?

This sentence does not appear anymore in the introduction.

2. Maximum likelihood estimation assumes multivariate normality; departure from it may have large impact on standard errors and all the inferences based on these parameters. Did you test this assumption?

Indeed, BMI; waist circumference and skinfold thickness had skewed distribution which became more symmetric after log transformation. The Kolmogorov-Smirnoff statistics improved in all cases, even if it was already non significant before transformation for some of the variables, as this test is known to lack power. This point is now mentioned in the text.

3. There are two basic components in your SEM: the factor analysis for latent adiposity (measurement model) and the relationship between the variables in the model (structural model). It is of great importance that the measurement model fits well before you can move to the inferences about the structural relationship. For example, what were the communalities, KMO and percentages of the variance explained by the latent factor for the measurement models? How did the transformation such as log affect these measures?

When using BMI, waist circumference and sum of skinfold without log transformation, we observed that RMSEA and NFI worsened lightly for females but improved for males. For this reason (added to answer to point 2 above), we decided to use the log transformation for these 3 adiposity measurements. We have now added a table (Table 2) with details on the measurement models, including percentage of variance explained ($R^2$) as suggested. Unfortunately, SAS CALIS procedure does not provide communalities and Kaiser-Meyer-Olkin adequacy measure.

4. Did you try a measurement model for adiposity after 2 years instead of change scores? As far as I can see, the latter should be more variable than the former due to the multiplicity of factors influencing the change. Were there problems with the model converging when original follow-up measurements were used?

In a previous submission, we had modelled 1999 and 2001 measurements according to 1999 and 2001 adiposity but used as explanatory variables baseline CRS and change CRS in order to avoid problems of interpretation and colinearity with highly correlated variables as CRS1999 and CRS2001. Indeed, 2001 measurements were better explained by 2001 adiposity than measurement changes are explained by adiposity changes. But a reviewer suggested an identical treatment of CRS and adiposity, and we followed that recommendation. We feel that the effect of 1999 CRS on adiposity change, adjusting for 1999 adiposity is easier to interpret than its effect on 2001 adiposity adjusting for 1999 adiposity. As mentioned in the answer to the other reviewer, we also estimated a model were 2001 measurements were explained by...
1999 adiposity change and by adiposity change. The global fit of the models was not satisfying (RMSEA=0.23, NFI=0.84 for females and RMSEA=0.20, NFI=0.87 for males).

5. To justify the equality constrains on p.9, you can test formally whether their imposition reduced the deviance beyond a pre-specified level; can you show the results?

The likelihood ratio tests to compare models with and without loadings equality constraints was significant for females (p=0.014) and non significant for males (p=0.16). We chose to use the constrained models for both groups for the following reasons
- The fit was reasonably good and actually better for females than for males.
- Our aim was not to build the “best model” for the baseline measurements and their changes, but to test a couple of hypotheses and the interpretation of the coefficients is simpler under the constrained model. In particular what is the interpretation of the latent adiposity change if different loadings are used for 1999 and 2001 measurements?

6. On p.9, RMSEA for males (0.16) is not very convincing. Can you show which part of the model (measurement or structural) did not fit well? Why don’t you show 95% confidence interval for RMSEA?

Table 2 gives the 95% CI for RMSEA. We discuss the decrease of fit of the structural model as compared to the measurement models in the first paragraph of the discussion section.

7. Did you try a sub-group analysis by baseline adiposity level, such as BMI<=20 versus BMI>20 or perhaps using BMI of 25 as a cut-off point? It seems reasonable to assume that cognitive restraint may influence adiposity differently for people who are overweight or obese compared to those who are not.

BMI <=20 selected only 9% and 2% of individuals, in female and male groups respectively, which was clearly insufficient to perform further analyses.
When using BMI=25 as cut point, females were split into 162 normal weight and 94 overweight, while men were split into 86 normal weight and 115 overweight. Analyses performed in these subgroups did not show any new results, except loss of significance for some relationships due to lack of power. That is why this analysis by subgroups was not mentioned in the manuscript.

Minor Essential Revisions

1. In conclusion, both in abstract and in discussion, the statement that “The latent variable modeling approach enabled presentation of synthetic results rather than four parallel analyses...” should state exactly what was meant by “four parallel analyses”. I suppose it meant four separate multivariate regressions for the change in each anthropometric measure of adiposity but am not sure how the interclass correlations would be treated (e.g. multi-level analysis), whether stratification by sex would be maintained, etc.

The sentence was changed as suggested (ie: "parallel" replaced by "separate"). Indeed, we had in mind four separate regressions for each sex group, but without considering any correlation
2. On p.9, the phrase “closely linked” can mean many things; please specify (e.g. “highly correlated”).

The sentence was changed as suggested.

3. On p.9, substitute “was” for “were” in the sentence “The model fit for the baseline measurements and their changes was only slightly modified...”.

The sentence was changed as suggested.

4. On p.10, 2nd paragraph, “and” seems more appropriate than “or” in “As expected, adiposity or CRS changes...”.

The sentence was changed as suggested.

5. On p. 13, use “the” instead of “common” in “The indirect effect through CRS change is at least partly due to common regression to the mean...”.

The sentence was changed as suggested.

6. What was the effect size for the hypothesis of interest? What does it mean in practical terms?

This point is now discussed (top of p 13).

7. In table 3, you can add t-values and flag statistically significant ones.

Table 3 was changed as suggested, t-values were written in bold when significant at the 5% level.

Discretionary Revisions

1. Title may sound better if the phrase “to model” is substituted by “for”.

Title was changed to

**Latent variables and structural equation models for longitudinal relationships: an illustration in nutritional epidemiology**

In order to avoid repetition of model/modeling. We are ready to accept “for modeling of longitudinal relationships” if it is preferred by the editors.

2. On p.3, the phrase “in which the common problem is how to handle questionnaires” is too vague and probably meant “how to deal with psychometric properties of the questionnaires”.

The sentence was changed as suggested.
3. On p.7, the phrase “the effect of baseline adiposity on CRS change was adjusted for baseline CRS and thus freed from the factors confounding the cross-sectional effect” seems too strong because you can never be sure of controlling all the confounders as the phrase implies. Even adjusting for the baseline values works only for the linear relationship between the baseline and follow-up. However, many test scores have a threshold indicating a clinically relevant change above that point, which makes the relationship a non-linear one. For example, it seems plausible that cognitive restraint exerts a clinically relevant effect on adiposity only above certain level of the test score. Was this relationship tested?

The sentence was modified and now states that "[the effect was] freed, at least partially, ...". Our male and female groups are quite homogeneous with respect to CRS (26.9 +/- 19.7 and 40.4 +/- 21.3, respectively, while the theoretical range of CRS is 0 - 100). The point is interesting but does not apply here because our dataset is a sample of a community-based study so we did not observe subjects with extreme CRS values (nor with extreme adiposity measurements).

4. On the bottom of p.7, the word “deterministic” may well be dropped out. The interpretation of the direct effect which follows may be better suited for discussion.

Modifications have been done as suggested.

5. On p.8, 2nd paragraph, you may drop “in practice”.

The sentence was changed as suggested.

**Level of interest:** An article of importance in its field  
**Quality of written English:** Needs some language corrections before being published  
**Statistical review:** Yes, and I have assessed the statistics in my report.  
**Declaration of competing interests:** I declare that I have no competing interests