Commentary: Is structural equation modelling a step forward for epidemiologists?

Yu-Kang Tu

One of the major challenges for epidemiologists is to understand causal relationships between risk factors and health outcomes by analysing data from observational studies. Epidemiologists know too well that correlations, such as those in regression analysis, rarely mean causation, and it would be very desirable if there is a methodology for observational studies, analogous to randomization in experimental studies, which can discover causes and effects amongst variables or at least confirm or refute the proposed causal relationships. Randomized controlled trials (RCTs) are certainly the ‘gold standard’ of establishing causes and effects, but quite often it is either unethical or unfeasible to conduct RCTs to test causal relations in epidemiological research. Epidemiologists need a methodology, which is sort of a combination of the directed acyclic graphs (DAGs) for conceptual construction of causal models and regression analysis for testing those models. It is therefore surprising that structural equation modelling (SEM) has not been so frequently used in epidemiology as in the social sciences, given that both epidemiologists and social scientists want to delineate causes and effects from observational data. The difference between DAGs and path diagrams in SEM is almost trivial to epidemiologists: the path between two variables can only have one direction in DAGs. An individual path in SEM is tested in the same way the regression coefficient is in regression analysis, and model fit indices provided by SEM software packages help the analysts to assess the adequacy of the proposed causal model compared with the observed associations in the sample data.

Why then is SEM still under-utilized in epidemiology? This is a question posed by a commentary in this journal a few years ago. The answers cited included unfamiliar terminology (SEM theory is formulated in Greek), restriction in the assumptions of variables (outcome variables need to be continuous), difficulties in testing interaction and non-linear relationship (it can be quite tedious to set up SEM models to do these), and equivalent models (two different causal models imply the same correlation structure and consequently, it is impossible to tell which is better). Recent advances in SEM theory and software development has resolved some of these issues: new estimation methods do not require the strict assumption of multivariate normality, and the outcome variables can be binary, ordinal or counts.

The basic rationale behind SEM is rather simple: multiple linear equations are used to specify causal relationship between variables some of which are manifest variables (i.e. observed and collected by the researchers), whilst others are latent variables (i.e. derived from the observed variables by specifying their relations using equations), such as those in factor analysis. The multiple equations in each causal model entail a certain structure of correlational relationships between observed variables which is usually given as a correlation (or covariance) matrix Σ. The estimation procedure is to minimize the difference between Σ and the observed correlation/covariance matrix S formulated by a likelihood function. The χ²-test is then used to evaluate the difference between these two matrices by taking into account the number of the estimated parameters in the proposed model. When the χ²-value is large (i.e. the difference between the two matrices is large) relative to the model’s degree of freedom, the proposed model is rejected, i.e. something in the causal relationships specified by the proposed model is not quite right and require a second thought. When the χ²-value is small, we fail to reject the model or tentatively accept the model as adequate. Because of the possibility of equivalent models, our model may still be wrong in terms of the causal relations amongst variables but happen to entail the same correlation structure (the same S) of the ‘true’ model. So SEM seems to be an endeavour to search for the truth by approximation, and in this aspect, doing SEM is quite similar to the process of ‘Conjectures and Refutations’ as described by the great philosopher of science, Sir Karl Popper.

All the traditional linear regression models can be specified within the framework of SEM, and there is
no doubt that SEM has many advantages over traditional linear regression. However, as SEM can be seen as an extension of linear regression, it also suffers some caveats that are commonly found in linear regression. The study by Dahly et al.\(^1\) found a significant, inverse relationship between latent foetal environment (LFE) and systolic blood pressure (SBP) at age 20–22 in a Filipino cohort. The standardized estimates (equivalent to standardized regression coefficients) are around \(-0.3\) for both females and males. This is a relatively strong association compared with those between birth sizes and SBP in the literature.\(^1\) However, as current weight is on the causal pathway from LFE to SBP in their model, this introduced the same problem of reversal paradox in the interpretation of the direct path from LFE to SBP (path \(a\) in their figures) as that in the interpretation of the association between birth sizes and SBP when current body sizes are adjusted for.\(^1\) This seems to suggest that the association between LFE and SBP would have been obscured if CW was not adjusted for. If we ignore birth weight and the measurement model part of the SEM model in their study, the estimate for path \(a\) can be interpreted as the regression coefficient for LFE when CW is adjusted for in multiple regression, i.e. path \(a\) is the relationship between LFE and SBP conditional on CW. Although path \(a\) is usually interpreted as the ‘direct effect’ in SEM literature, it would be a mistake to interpret the estimate for path \(a\) as the change in SBP caused by the change in LFE by one unit, because we need to take the indirect effect/path (i.e. path \(c\) and \(d\)) into account. The change in SBP caused by the change in LFE by one unit is the sum of the direct and indirect effects, i.e. 
\[
-0.18 \times a + 0.28 \times 0.68 \times c \times d = 0.01 \quad \text{for females} \\
-0.26 \times 0.57 \times 0.62 = 0.09 \quad \text{for men}.
\]
As a result, the overall effect of LFE on SBP turns out to be positive.

Whether a variable on the causal path can be adjusted for and how the results from such an adjustment should be best interpreted is a contentious issue, as the answer may be context-specific.\(^1\)-\(^8\) In their defence of including CW in the model, Dahly et al. argued that ‘another advantage is that we can include CW in the SEM without invoking a growth interpretation, because a one unit change in LFE, holding CW constant, does not imply that the individual grew any more or less’. This is debatable. LFE is a latent variable for foetal nutrition, and one of its contributing variables is mother’s height. Furthermore, LFE is assumed to affect BW which is also correlated with CW in the model. Though LFE is not a direct measurement of birth sizes, LFE is a proxy variable for the foetal environment which has been hypothesized by the foetal origins hypothesis to be vital to foetal growth. We can simply name it ‘Foetal Growth’ without specifically referring to length or weight. Thus, LFE may not free us from growth interpretation.

The base SEM model in their study had a very small \(\chi^2\)-value, indicating an excellent model fit, and hence Dahly et al. concluded that their ‘theoretical model adequately explained the observed relationships among the model’s observed variables’. However, a warning needs to be issued for researchers who after reading the paper wish to try their hands in SEM that it is more than often that the carefully formulated model does not have a good fit and require modifications. All SEM software provides model modifications index which automatically search for ways to improve the model fit, but the caveats of this approach are well documented.\(^4\)-\(^6\) On the other hand, the statistical power of the \(\chi^2\)-test to reject a model increases with the sample size.\(^4\) If the sample size had been 26 530 rather than 26 539, the base SEM model in Dahly et al. might well be rejected. So should we give up our carefully formulated model and modify our theory, or should we blame the sample size and try to identify other factors, such as random sampling errors, for the poor fit? It is not always straightforward to practise Popperian conjectures and refutations in SEM.

Despite of all these discussions, we tend to agree with Dahly et al. that SEM is a step in the right direction. As another philosopher of science, Nancy Cartwright, argued ‘no causes in, no causes out’,\(^1\) epidemiologists will never prove or refute a causal relationship from observational studies, if causes and effects are not part of their statistical models. Causality has to be incorporated into statistical analysis, if we want to make causal inference from the results. SEM is a useful tool for establishing causes and effects, but like any tool, things can go terribly wrong if we misuse it. Epidemiologists need to learn how to use it and when to use it. The study by Dahly et al. serves as an excellent example of how to carefully formulate causal models using SEM and how to test them.

**Conflict of interest:** None declared.

**References**


12 Tu YK, Gilthorpe MS, Ellison GTH. What is the effect of adjusting for more than one measure of current body size on the relation between birthweight and blood pressure? *J Human Hypertens* 2006;20:646–57.


16 Tu YK, Gunnell DJ, Gilthorpe MS. Simpson’s paradox, Lord’s paradox, and suppression effects are the same phenomenon—the reversal paradox. *Emerg Themes Epidemiol* 2008;5:2.
