The Chicken or the Egg: One Statistical Approach for Determining the Direction of Influence

Christa L. Lilly, PhD

In a previous volume of *The Journal*, Berge et al attempted a "which comes first, the chicken or the egg?" argument when it comes to parental feeding practices and young child eating behaviors.[1] Their statistical method of choice was a cross-lagged path analysis (CLPA). This appropriate statistical technique is appropriate for this research question, because the technique allows one to examine a feedback loop across repeated measurements. I present the fundamentals of CLPA and assume an understanding of basic regression techniques. My own education was influenced by Loehlin's introductory textbook; I draw heavily on this source throughout the overview (any errors, of course, are my own).[2]

Overview of Path Analysis and Structural Equation Modeling

Simply put, path analysis is regression that has multiple dependent variables. Given the wide use across fields of study, dependent variables may be referred to as endogenous or downstream variables, and independent variables may be referred to as exogenous or upstream variables. An endogenous variable may be both an independent and a dependent variable, whereas exogenous variables are only independent. For example, Berge et al use models that typically have baseline parent and child measures as exogenous variables, and all other time-point variables as endogenous.[1] A sample path analysis model is given in the Figure where the x1 and y1 variables are exogenous, and all other variables (ie, x2, x3, x4, y2, y3, y4) endogenous.

For those unfamiliar with structural equation modeling (SEM), it is worth pointing out that path analysis is 1 piece of that particular puzzle. Sometimes we want to account for measurement error while answering a research question. SEM does this via a combination of confirmatory factor analysis and path analysis. Confirmatory factor analysis takes into account differences in measurements by combining multiple measures of a single construct. We picture the construct as something we cannot precisely or directly measure, but we are sure the true construct exists. This true construct is a latent variable, which are variables that are not directly measured, but also estimated with what we can measure: the manifest variables. Once we take into account measurement error, we try to answer our research question, which is what we are most concerned with in path analysis (eg, are the exposure and outcomes related?).

SEM and associated techniques were intended to answer causal questions; unfortunately, it is hung on regression techniques that make those implicit causal tenets difficult to pull out (given that correlation does not equal causation). The application of causality to SEM and related methods has been debated for many years.[3] With that caveat, we can apply causal theory to SEM in an attempt to answer causal questions, specifically using causal diagramming, when the design and data allows for causal inference.[4,5]

Thus, a core feature of SEM and associated techniques includes graphically displaying the desired model (such as in the Figure). Typical conventions include using boxes to indicate manifest variables (eg, variables x and y at each of 4 time points in the Figure) and circles for latent (not shown in path analysis). Lines are used to denote a relationship between variables. If the line has an arrow at only 1 end, then we are regressing a variable onto an endogenous variable (eg, x and y at times 2 through 4); if the line has arrows at both ends, typically curved, it denotes a correlation (such as between x1 and y1). Small arrows pointing to endogenous variables denote error (residual arrows; zetas in the Figure). Although constructing the graphics in many user-friendly software packages seems to be simple (just drawing lines to boxes!), given the link with causal inference it should also be readily apparent that it is an easy tool to misuse.[6] Careful construction of the graphic can assist the reader in determining the validity of the causal inference of the results.

CLPA

Berge et al use multiple measurements.[1] Like all regression techniques, repeated measurements warrant specialized modeling, because it violates the assumption of independence of observations. One way we can take this into account in path analysis is via cross-lagged paths, or CLPA. CLPA accounts for the autoregressive nature of the data by allowing each variable to regress on the next nearest time point, and additionally adds the cross-lags, or relationships between variables over time. For example, paths b1 to b6 in the Figure are the cross-lags. Taking a closer look at one cross-lagged path, b1, shows a relationship between x at time 1 with y at time 2, controlling for y at time 1. The example by Berge et al uses 4 time points for each child and parent variable. With a good fitting model, causal implications can be examined by comparing standardized coefficients of the cross-lagged paths (hence answering the

From the Department of Biostatistics, West Virginia University, Morgantown, WV.
The author declares no conflict of interest.

https://doi.org/10.1016/j.jpeds.2020.08.060

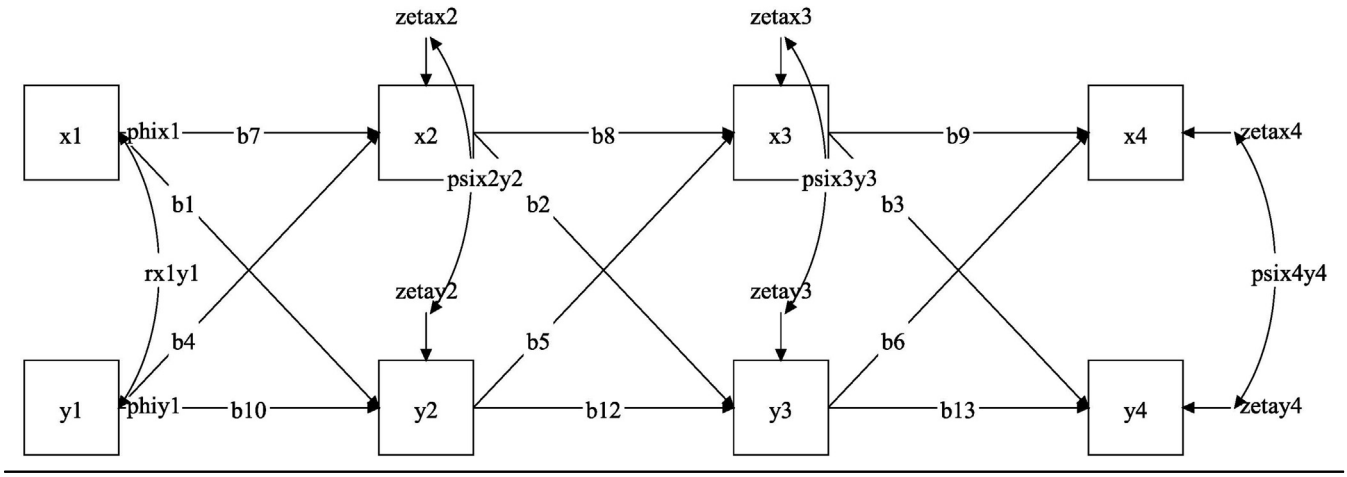

Figure. Example of a CLPA with 4 time points and 2 variables, x and y. Simplified, without assuming a mean structure or latent equivalents to the manifest variables, other parameters include exogenous variances (phi), correlations at the same time point across exogenous variables (r) or via the disturbances (psi), cross-lagged paths (b1 to b6), autoregressive paths (b7 to b12), and endogenous residuals (zeta).

chicken or egg question). Are the relationships consistently stronger between the cross-lags going from the parent (eg, the x variables in our simplified **Figure**) to the child (eg, the y variables), vice versa, or both? Nested models (ie, another model that includes or takes away paths) could be used to explore the best fit by dropping b1 to b3 (all x to y paths), for example, and comparing against a full model. If the model fit better without b1 to b3, this would indicate that the correct direction of the relationship would be y to x, rather than the reverse.

A key consideration is the time component. Were the measurements made at time points that allowed the relationship to be captured accurately? The traditional CLPA is a simple autoregressive model (ie, the relationship of each variable to itself over time), which is assumed stable and equidistant over time points. Time variability is assumed to only be related to (and fully mediated by) the next time-point in the series. Violations of this assumption will result in biased estimates, particularly if the constructs are trait like or time invariant.[7] Others critique this model for the inability to separate the between and within-person associations of the model.[8]

Model Fit and Interpretation

A powerful aspect of this particular tool (and all related SEM frameworks) is that it uses a wide range of model fit indices to help us determine the usefulness of our model.[2] Much like we interpret the omnibus ANOVA before determining where group effects occur, one should always look at the model fit before interpreting the parameter estimates included in the model. Model fit includes 4 different types: (1) an omnibus model fit measure, the χ^2, (2) incremental fit measures, akin to effect sizes, (3) absolute fit, akin to how "badly" fitting the model is, and (4) comparative fit, which allows us to test related models.

First, we look at the χ^2 for the model. It is an assessment of badness of fit, but it uses inferential testing via reverse hypothesis testing. A significant *P* value indicates significant differences between the correlations you obtained with your sample and the correlations implied by your model. Thus, a *P* of more than .05 indicates a good-fitting model. Unfortunately, given this is reverse hypothesis testing, very small differences in model fit are generally statistically significant if you have a well-powered model, meaning you may obtain a *P* value of less than .05 but your model may otherwise have good fit indices. The model χ^2 forms the basis for many other fit indices, and is useful when conducting a change in χ^2 test on nested models.

Incremental fit measures include a wide variety of "larger is better" measures, such as the comparative fit index used by Berge et al, where a poor fitting model has a standardized value of 0 and a perfect model has a value of 1.[1] It is called incremental (or sometimes relative) fit because it compares your model against a "null" model, or a model with all correlations set to 0, relative to the best possible model fit the data can have. It is akin to an R^2 value; it is a descriptive effect size measure. We generally look for values close to 1.00, and preferably larger than 0.95.[9] These are (to some extent unsatisfactory) rules of thumb.[10]

Absolute fit measures include the root mean square error of approximation used by Berge et al.[1] Models with small amount of error will have values closer to 0; it involves the χ^2 implied by the model, but takes into account the sample size and model parsimony, and is often a preferred fit index particularly for power analysis.[11] This fit measure is descriptive, but has an accompanying CI; we prefer values smaller than 0.06, and CIs that do not cross 0.10.[6]

Finally, comparative fit indices include the Akaike information criterion, where "smaller is better." The numbers by themselves are not interpretable, but rather should help us to select the best fitting model among nested models while taking into account parsimony.

A use of all these measures gives us a feel for how useful the model may prove to be (with the caveat that they should be interpreted with caution), and should be examined before looking at the parameter estimates included in the model.[10] They are most powerful when considering nested models. Once we know we have a useful model, parameter estimates may be interpreted as one would typically interpret regression estimates. R^2 values are typically given for each dependent variable in the model.

Conclusions

CLPA is an effective traditional technique for basic repeated measures data where a feedback loop should be examined, and where time effects are fairly stable, evenly spaced, and likely capturing the imagined feedback effect. Unfortunately, the models presented had generally poor fit indices, nested models not reported, and only three models had more than 1 significant cross-lagged coefficient, and no models showed consistent temporal feedback.[1] Despite the spotty model fit, the authors suggest the significant cross-lag paths give some clinical implications about parent and child feedback loops with pediatric eating behaviors. Given that a CLPA is not as flexible as more advanced models under the same SEM framework, the authors may well have also considered a parallel-growth latent growth curve (or trajectory) model, or a random intercept cross-lagged panel model.[7,12] There are many alternatives; some argue the underlying theory of time/change may best dictate statistical model selection.[13] A final cautionary reminder: this tutorial is a brief, summary introduction that gives a background of the statistical topic without any math proofs. With that being said, this tutorial intends to convey that CLPA provides an intuitive graphical approach and statistical tool clinicians may wish to use when thinking about feedback loops in their research. ■

Reprint requests: Christa L. Lilly, PhD, Department of Biostatistics, West Virginia University, School of Public Health, 64 Medical Center Dr, P.O. Box 9190, Morgantown, WV 26506-9190. E-mail: cllilly@hsc.wvu.edu

References

1. Berge JM, Miller J, Veblen-Mortenson S, Kunin-Batson A, Sherwood NE, French SA. Bidirectional analysis of feeding practices and eating behaviors with toddlers and parents from low-income and minority households. J Pediatr 2020;221:93-8.e2.
2. Loehlin JC. Latent Variable Models: An introduction to factor, path, and structural equation models. Mahwah (NJ): Lawrence Erlbaum Associates, Inc; 2004.
3. Biddle BJ, Marlin MM. Causality, confirmation, credulity, and structural equation modeling. Child Dev 1987;58:4-17.
4. Pearl J. The causal foundations of structural equation modeling. In: Handbook of structural equation modeling. New York: The Guilford Press; 2012. p. 68-91.
5. Selig JP, Little TD. Autoregressive and cross-lagged panel analysis for longitudinal data. In: Handbook of developmental research methods. New York: The Guilford Press; 2012. p. 265-78.
6. Steiger JH. Driving fast in reverse. J Am Stat Assoc 2001;96:331-8.
7. Hamaker EL, Kuiper RM, Grasman RP. A critique of the cross-lagged panel model. Psychol Methods 2015;20:102-16.
8. Berry D, Willoughby MT. On the practical interpretability of cross-lagged panel models: rethinking a developmental workhorse. Child Dev 2017;88:1186206.
9. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Modeling 1999;6:1-55.
10. Fan X, Sivo SA. Sensitivity of fit indices to model misspecification and model types. Multivariate Behav Res 2007;42:509-29.
11. Sivo SA, Fan X, Witta EL, Willse JT. The search for "optimal" cutoff properties: fit index criteria in structural equation modeling. J Exp Edu 2006;74:267-88.
12. Curran PJ, Howard AL, Bainter SA, Lane ST, McGinley JS. The separation of between-person and within-person components of individual change over time: a latent curve model with structured residuals. J Consult Clin Psychol 2014;82:879-94.
13. Ferrer E, McArdle J. Alternative structural models for multivariate longitudinal data analysis. Struct Equ Modeling 2003;10:493-524.