JOURNAL
OF THE CANADIAN ACADEMY OF CHILD AND ADOLESCENT PSYCHIATRY / DE L'ACADÉMIE CANADIENNE DE PSYCHIATRIE DE L'ENFANT ET DE L'ADOLESCENT

RESEARCH ARTICLE

Application of Transactional (Cross-lagged panel) Models in Mental Health Research: An Introduction and Review of Methodological Considerations

Danielle A. Baribeau MD,PhD[1,8]; Simone Vigod MD[1,2,3]; Heather Brittain MA, MSc[4]; Tracy Vaillancourt PhD[4]; Peter Szatmari MD[1,5]; Eleanor Pullenayegum PhD[6,7]

Abstract

Transactional models employing cross-lagged panels have been used for over 40 years to examine the longitudinal relations and directional associations between variables of interest to child and adolescent mental health. Through a narrative synthesis of the literature, we provide an accessible overview of cross-lagged panels with attention to developing a research question, study design and assumptions, dynamic effects (including the random-intercept cross-lagged panel model), and reporting and interpretation of results. Implications and critical appraisal guidelines for readers are discussed throughout. Overall, several key points are highlighted, with particular emphasis on the intended level of inference, model and measure selection, and timing of assessments. Despite limitations in establishing causation, cross-lagged panel models are fundamental to non-experimental epidemiologic research in child mental health and development.

Key Words: *longitudinal studies, statistics, structural equation modelling*

Résumé

Les modèles transactionnels qui emploient des panels à décalage croisé sont en usage depuis plus de 40 ans dans le but d'examiner les relations longitudinales et les associations directionnelles entre les variables d'intérêt pour la santé mentale des enfants et des adolescents. Grâce à une synthèse narrative de la littérature, nous offrons une vue d'ensemble accessible de panels à décalage croisé en portant attention à l'élaboration d'une question de recherche, à la conception de l'étude et aux hypothèses, aux effets dynamiques (y compris le modèle du panel à décalage croisé à interception aléatoire), et le rapport et l'interprétation des résultats. Les implications et les guides d'évaluation critique pour les lecteurs sont discutés tout au long. En général, plusieurs points principaux sont soulignés et l'accent est mis sur le niveau voulu

[1]Department of Psychiatry, University of Toronto, Toronto, Ontario
[2]Department of Psychiatry, Women's College Hospital and Women's College Research Institute, Toronto, Ontario
[3]Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ontario
[4]Counselling Psychology, Faculty of Education, University of Ottawa, Ottawa, Ontario
[5]Centre for Addiction and Mental Health and The Hospital for Sick Children, Toronto, Ontario
[6]Child Health Evaluative Sciences, The Hospital for Sick Children Research Institute, Toronto, Ontario
[7]Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario
[8]Holland Bloorview Kids Rehabilitation Hospital, Toronto, Ontario

Corresponding E-Mail: Danielle.baribeau@mail.utoronto.ca

Submitted: October 5, 2021; Accepted: April 8, 2022

d'inférence, la sélection et la mesure du modèle, et la chronologie des évaluations. Malgré les limites de l'établissement d'une cause, les modèles des panels à décalage croisé sont fondamentaux pour la recherche épidémiologique non expérimentale en santé mentale et développement de l'enfant.

Mots clés: *études longitudinales, statistiques, modélisation d'équation structurelle*

Introduction

In the fields of child development and mental health research, there is longstanding interest in understanding how biological, psychological, and social factors influence each other over time, to identify both predictors of future outcomes and potential targets for intervention. Longitudinal studies have been central to consider how a child's temperament, family environment, and relationships may interact to influence their future proclivity to substance use, criminal justice involvement, mental disorders, and/or academic achievement, for example [e.g., (1-3)]. Many researchers considering these etiopathological questions employ cross-lagged panel models (CLPMs) to illustrate potential developmental transactions or cascading effects between variables over time (3, 4).

CLPMs require data from two or more variables collected at two or more time points, effectively creating a data 'panel' (Figure 1). An example would involve surveying the same cohort of adolescents each year about depressive symptoms and bullying. CLPMs can then be used to examine co-developmental processes unfolding between variables, by estimating the effects the variables have on themselves (autoregressive or stability paths) and each other both within-time and longitudinally (cross-lagged paths) (Figure 1) (5). To some extent, CLPMs (also referred to as transactional models, autoregressive cross-lagged models, cross-lagged path models, and cross-lagged regression models) can facilitate hypothesis formation with respect to potential causal associations from observational data, where randomized controlled trials are not possible or justifiable. Recently, the Random Intercepts Cross-Lagged Panel Model (RI-CLPM) has become popular, in part because it addresses some of the limitations of other models with making individual level inferences. Many statistical programs used for CLPMs are based on structural equation modelling (SEM), which is a sophisticated and versatile mathematical approach to examine data networks, clusters, and associations (6).

There are numerous recent examples of publications applying these methods in top journals [e.g., (7-10)]. While several excellent review papers and books provide an in depth look at the theory and application of SEM with respect to modelling of developmental processes and transactions (6, 11-16), our aim is to provide an introductory overview of these methods as they apply to CLPMs, for a non-statistical audience including clinicians. The goal is to facilitate the accessible interpretation of results and critical appraisal of these methods more broadly.

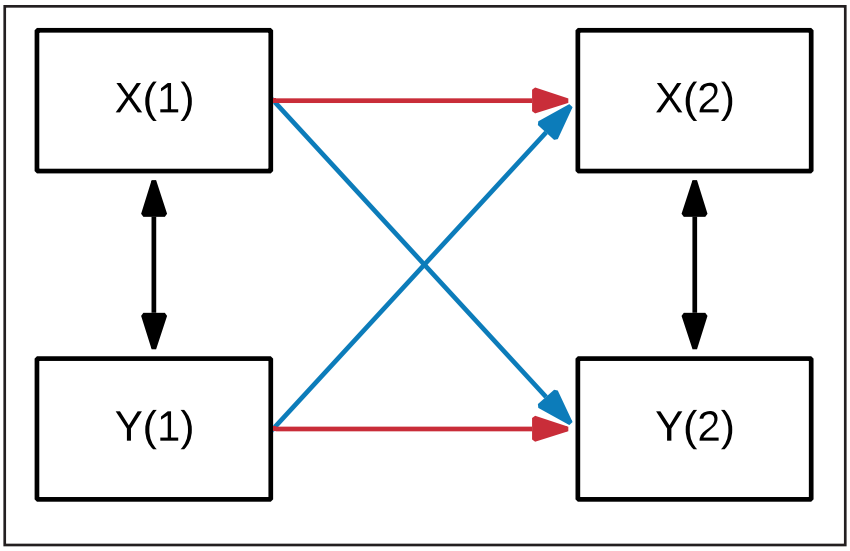
1. Research Questions and Hypotheses in CLPMs

CLPMs can help answer several 'chicken or egg' research questions simultaneously; for example, how an outcome changes over time, the extent to which it is dependent on prior observations, how different outcomes are associated at a single time-point, and whether they are 'cross-dependent.' Statistically speaking, variables with cross-dependent transactions have significant cross-lagged associations (Figure 1, blue arrows); effectively, the magnitude of a variable at one time point is associated with the magnitude of a different variable at a later time point, while accounting for, and/or separating out, the within- and across-time associations (black and red arrows, respectively). Although each question could be addressed in separate models, CLPMs can estimate all these relations simultaneously.

CLPMs are most useful in situations where researchers have some a priori ideas about what they would like to examine. For example, when the existing literature has already identified relations between the variables or constructs of interest, yet questions remain about the nature of their longitudinal associations. It has been suggested that established cross-sectional relations ought to be at least moderate in size (i.e., $r > 0.30$) in order to justify CLPMs (17). Specific examples of mental health research questions from studies include: (i) whether excessive screen time is associated with adverse physical and cognitive outcomes in children (9, 10), (ii) how bullying by peers, disordered eating behavior, and depressive symptoms are associated over adolescence (8), and (iii) how social anxiety and social deficits are directionally associated during childhood (18).

When approaching a paper employing CLPMs, readers should first identify the overall study objective and whether specific theoretical models/hypotheses were put forward based on a priori knowledge. The study objective and hypotheses are also important to guide the choice of statistical

Figure 1: Simplified cross-lagged panel model



Note: This figure presents a simplified 2-wave (i.e., two time-point), two variable, cross-lagged panel model (CLPM). X and Y refer to two different variables assessed at times 1 and 2. Here, we show within-time associations between variables using a double headed black arrow. Red arrows display the associations between the same variable across two time points (sometimes called auto-regressive, or stability paths), while blue arrows show the cross-lagged paths. All of these associations are estimated in the same statistical model, therefore cross-lagged effects in blue can be conceptualized as occurring in addition to, and while accounting for, the within- and across-time associations in black and red, respectively.

model (discussed further below). Readers can consider whether the phenomenon under investigation is most relevant at a variable level (comparing between individuals in the study) or individual level (examining effects within an individual in the study). As an example, a variable level (between-person) approach was used to examine broadly whether social deficits tend to precede social anxiety symptoms (or vice versa) in school-aged children (18). A RI-CLPM was used to examine individual level (within-person) effects, showing that when a toddler's screen time use increases relative to their typical pattern of use, they make fewer developmental gains in the future (9).

2. What is structural equation modelling (SEM)?

CLPMs are often built using structural equation modelling (SEM), which emerged in the 1960s and 1970s as complex computing became increasingly accessible (19-21). SEM can be conceptualized as an extension of traditional regression analysis. The term 'structural' in SEM refers to the fact that a specific theory about the patterns of association (covariance structure) of the variables using algebraic equations is being modelled and then the degree to which the 'real' data fit this hypothesized 'structure' is tested. Broadly, the two most common applications of SEM are measurement models and path analyses. Measurement models involve statistically studying how variables cluster together, for example through confirmatory factor analysis. This process is commonly used to derive and/or validate subscales on standardized surveys (e.g., the internalizing subscale on a parent report of child behaviour). Path analyses involve studying how variables are associated with each other; CLPMs are a type of longitudinal path analysis. While path analyses can be conducted without using SEM, an important feature of SEM is the modelling of unobserved (latent) constructs, factors, or variables (discussed further below). Most statistical programs are now capable of conducting SEM (6, 14). One barrier to the wider application (or even interpretation) of CLPMs, however, is their reliance on unique syntax, programming language, and path diagrams. Luckily, several authors have made their code available in supplementary materials (12, 22).

3. What are latent variables?

In SEM, variables are categorized into those that are 'observed' and those that are 'latent.' Observed (also called 'manifest' or 'measured') variables are variables that are measured; in mental health research they are commonly scores on surveys/standardized instruments. Latent variables are those that are theorized to exist, and may be derived mathematically, but are not actually measured.

In measurement models, latent variables are often overarching constructs of interest (e.g., anxiety, stress), derived in some capacity from the observed variables. A classic example of a latent construct is intelligence. A selection of observed variables from standardized cognitive subtests (e.g., memory, processing speed), and/or academic achievement, occupation prestige, or school grades might be combined in some capacity to provide an overall proxy for 'intelligence'. Theoretically, the latent construct of intelligence is thought to cause the variability on the observed measures. Alternatively, as discussed above, measurement models can be used to identify and define 'latent' subscales on a survey.

When appraising an article using CLPM, measurement models are often included as important preliminary analyses. Readers should check whether the authors have chosen previously validated measurement instruments or subscales that accurately reflect their constructs of interest and whether they have included some preliminary item level analyses (e.g., factor analysis, and/or analyses showing longitudinal measurement invariance) to show that the constructs are valid, distinct, and stable across time (23, 24). For example, prior to examining the directional association between social anxiety and social deficits, the authors first tested the degree to which items on both measures separated into distinct constructs at each time period under study (18).

Classic CLPMs can be expanded through SEM where the observed variables are broken down into various latent components of interest. Examples include separating out the 'true' estimate of the variable from measurement error, and in the case of RI-CLPMs, separating out stable (trait-like, between individual) components from the dynamic (state-like, within individual) components over time. Both measurement error and the RI-CLPM are discussed in more detail below.

4. Other factors to consider in CLPMs

4.1 *Data distribution:* Most SEMs use a default statistical approach that assumes the data are continuous and follow a normal distribution, and historically, most CLPMs have used continuous outcome measures. More recent advances in SEM allow researchers to employ a variety of statistical estimators or other extensions in the models to account for non-normal distributions or categorical data (25, 26). Practically, surveys with at least 4 or 5 response levels per item, when summed or averaged, often approximate a normal distribution (27). Readers can also consider whether the chosen instruments may be expected to have ceiling or floor effects when applied to their population of interest. A survey yielding results that are normally distributed in the general population may show different distributions (e.g., 'hitting a ceiling') when applied in a clinical cohort, for example.

4.2 *Measurement error:* Measurement error refers to both the systematic and random imprecisions of any measured variable. The possibility of correlated measurement error merits special attention in CLPMs, as failure to account for it can bias results (28). For example, consider a study where at a single time point, the same informant (e.g., a parent reporting on their child) completes two separate surveys consecutively while at a study visit. If responses from one survey somehow influence responses on the second survey, this could drive a false association from measurement error correlation (also called 'a halo effect,' 'shared-method variance' or 'informant bias'). Across two time points, error terms could also be correlated should the participant recall and be influenced by their previous responses (sometimes called 'conditioning') (28). There are statistical approaches in SEM that can account for correlated measurement error to some extent if needed. Several aspects of the study design can also help minimize error correlation: specifically, appropriate spacing of assessment time points, seeking out measures with high reliability, validity, and responsiveness (29, 30), and carefully considering how different informants (e.g., self vs. parent vs. clinician) may affect results.

4.3 *Linear associations:* Third, CLPMs generally assume that relations between variables of interest are linear (unless otherwise modeled). This means that for a unit increase in an independent variable, there is a measurable and predictable increase in the dependent variable that is of the same magnitude across the entire range of the independent variable. In mental health research, we know this is not often a true assumption; there may be thresholds above or below which screen time, bullying, or social difficulties may be benign or increasingly harmful. If the association between variables is expected or observed to be highly non-linear, sensitivity analyses may be suggested using models with non-linear associations (e.g., quadratic, cubic) (31), or with categorical outcomes (32).

4.4 *Assessment time and spacing*: Fourth, there are several time-related assumptions for CLPMs. First, all models assume synchronicity; in essence, all assessments at Time

1 occur at the same time and Time 2 at the same time etc. (17). This requires additional consideration if there are staggered data collection procedures as is common in larger studies (e.g., some interviews followed by mail or online surveys later).

The spacing of assessment time-points is another important factor (33-35). Essentially, we need to consider the extent to which variables may be influencing each other on much shorter (or longer) time spans than that which we are measuring (4, 14, 35). Investigators may come to different conclusions about the magnitude and even direction of associations by simply varying the time intervals between assessments (13). Readers can consider whether assessment spacing is logical with respect to the phenomenon under study (e.g., monthly tests of language acquisition in toddlers and preschoolers, but yearly in school-aged children) (15, 36). Equality of time point spacing is another factor to consider, since estimates are time dependent. If the spacing between assessments is unequal, this should factor into the interpretation of results, or may need to be accounted for with novel methods such as continuous time modelling (CTM) (13).

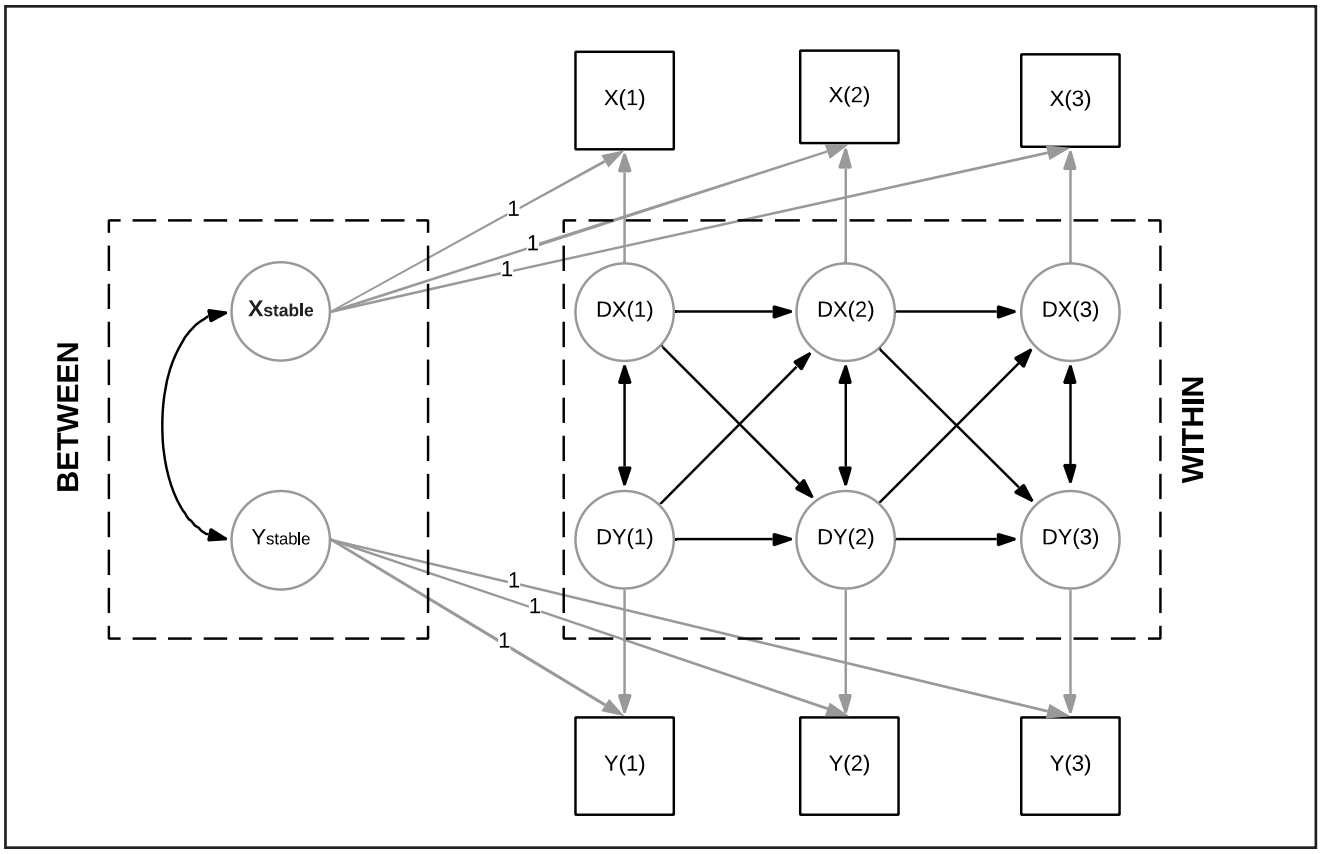
4.5 *Other covariates:* Fifth, as with all studies, we assume that other important variables that might be influencing the association under study (e.g., socioeconomic status, rurality, comorbid conditions, sex and gender, etc.) are either negligible, or are accounted for in the model as covariates. Readers are also directed to the *British Medical Journal* series on critical appraisal of cohort studies, for further discussion on ways to address confounding more broadly (37-39). There may also be interest in examining whether cross-lagged estimates are affected or "moderated by" by specific covariates; for example, whether the association between disordered eating and depression varies between boys and girls (8). In this setting, a multi-group analysis can be added to examine categorical moderators. Specifically, a model is built where the effects of sex in the model is initially allowed to vary freely (i.e., one estimate for boys and one for girls) and then another model is built where the effects of sex are constrained to be equal between boys and girls. The fit of the models is then compared using a statistical test (often the chi-squared difference test) to determine which model best fits the data and whether sex is a statistically significant moderator. It is inappropriate to simply split the sample into boys and girls, and then examine the significance of the resulting parameters, as estimates in the two models may be different for other reasons (chance, repeat testing, sample size differences, other confounders).

4.6 *Population heterogeneity:* Standard CLPMs (sometimes called an autoregressive cross-lagged models) assume everyone in the cohort tends to follow a similar trajectory of change over time, such that results derived from the whole cohort can be generalized to the individual level (5, 12). For example, most children are expected to have an increasing number of friends, social skills, and an increasing vocabulary over childhood. However, in situations where individual trajectories are suspected to vary in direction over time (e.g., depressive symptoms, bullying, and screen time use may increase or decrease depending on the person) more complex models that account for this may be needed, depending on the research question. Curran and Hancock review various statistical approaches that can account for some degree of population heterogeneity, for example by incorporating statistical terms to model differences in starting point (intercept), trajectory slope (e.g., the latent curve model with structured residuals or the autoregressive latent trajectory with structured residuals), or average over time (as in RI-CLPM).

4.7 Sample sizes: A sufficient sample size is needed for results of a CLPM to be reliable. Without a big enough sample, the model might not run, or might not have enough power to detect significant associations, especially with more complex models. Smaller 'true' cross-lagged associations, larger stability/autoregressive paths, larger within-time associations, smaller variance in data distributions, and less reliable instruments necessitate larger sample sizes to detect cross-lagged effects (17). Sample sizes of at least several hundred, but often several thousand, are ideal. A general rule of n=200 has sometimes been proposed for CLPMs with two waves (11, 40). That said, when the measures are reliable and valid, the underlying cross-lag associations are strong, and the models are not overly complex, smaller sample sizes may be sufficient (40). Simulation studies (e.g., the Monte Carlo Method) can also be used to estimate sample sizes and power a priori (41). In practice, power is only one important factor to consider when estimating sample sizes in CLPMs (along with bias, the expected magnitude of the weakest path, and the anticipated amount of missing data) (42).

4.8 *Missing data:* Although earlier researchers suggested CLPMs ought to be based on participants with complete data only (17), more recent simulation studies and advances in statistical procedures permit analyses using all available information. Most popular is the Full Information Maximum Likelihood (FIML) estimating procedure, which does not replace or impute missing data, but instead estimates the population parameters from known associations with non-missing variables. Simulation studies have suggested FIML is less biased for dealing with missing data than other missing data management techniques (43-45). To apply the FIML procedure, missing data patterns should

Figure 2: Modelling static vs. dynamic effects



Note: The Random Intercepts Cross-lagged Panel Models (RI-CLPM) is one example of how including latent variables (here shown in circles) can be useful to help separate out the static, trait-like aspects of a construct or symptom (X_{stable} and Y_{stable}), and the changing, dynamic aspects (DX and DY) from the measured variables (X and Y). Specific estimates and error terms have been omitted from this diagram for clarity.

be 'ignorable;' that is, the likelihood of a data point being missing should be independent of the value of that data point (or can be predicted from other non-missing data). In other words, missing data in a longitudinal anxiety survey cannot be ignored if some people suddenly, and unpredictably, become more anxious and then drop out of the study. Readers should examine the degree to which missing data are reported and described and how the missing data have been handled in the model. It is also important to consider that data could be missing over time due to attrition or within time because a survey was not fully completed (as one example) or both.

5. Random Intercepts Cross-lagged Panel Model (RI-CLPM)

A major criticism of traditional CLPMs is that they do not adequately separate out between- and within-person effects over time, resulting in estimates that can be difficult to interpret. Between-person effects are those that vary from one individual to another, that is, everyone's baseline tendency toward anxiety, exercise, or academic achievement, for example. These between-person effects are often conceptualized as static over time and 'trait-like' in nature. Within-person effects are the dynamic, time-dependent changes or 'states' that vary from measurement to measurement within the same individual. Thus, everyone in a study has both a baseline 'static' predisposition to anxious traits; and, within each individual, they may then experience episodic 'state-like' fluctuations in anxiety. Failure to account for these stable 'trait-like' between-person differences can bias the resulting cross-lagged estimates, effectively allowing stable traits to sneak in and confound (or even reverse) the directional effects (11, 12).

To address this concern, the Random Intercepts Cross-lagged Panel Model (RI-CLPM) (11, 12) uses SEM to separate out the measured variables into two latent components: a stable baseline between-person 'random intercept' component, and a dynamic, within-person component (Figure 2). Using the example of screen time effects on child development, cross-lagged paths in a RI-CLPM model would

Table 1. Sample fit statistics for cross-lagged models

Fit statistic	Target range for a 'Good fit'	Notes
Chi-square (χ^2)	P-value >0.05	χ^2 can be used as a global 'reject-or-support' statistic for the entire model. Small sample sizes tend to be underpowered for this test. It is often used to compare nested models (non-significant difference indicates models are not different).
Root mean square error of approximation (RMSEA)	< 0.10; ideally < 0.06	Favors parsimony
Comparative fit index (CFI)	> 0.90; ideally >0.95	One of the measures least affected by sample size
Standardize root mean square residual (SRMR)	< 0.08; ideally < 0.06	Standardizes the scale of RMR
Goodness of fit index (GFI)	> 0.90	Sensitive to sample size
Adjusted goodness of fit index (AGFI)	> 0.90	Accounts for sample size to some extent by incorporating degrees of freedom into the calculation - Tends to favor more parsimonious models (i.e., with fewer parameters)
Normalized chi-square (χ^2)	> 2.0	Accounts for sample size by incorporating degrees of freedom into the calculation (effectively a ratio of chi-squared value to degrees of freedom)
Root mean square residual (RMR)	Smaller is better	Unit is scale dependent
Akaike information criterion (AIC)	Smaller is better	Unit is scale dependent; often used to compare non-nested models
Bayes information criterion (BIC)	Smaller is better	Used to compare non-nested models

Note: For a detailed discussion on fit indices and model building, see Kline (2015), chapter 12 (6) and Hu et al., 1999(49).
Kline (2015) recommends including the following minimum set of fit statistics: χ^2, RMSEA, CFI, and SMR.

show how a change in screen time use affects a change in future development (and vice versa), relative to a child's own unique baseline. While the RI-CLPM has emerged as quite popular recently, there are several other statistical options available for trying to disentangle static vs. dynamic effects in a CLPM, which are thoroughly reviewed elsewhere (5, 11).

Readers should consider whether the study objectives and the interpretation of results are consistent with the model selected. If the goal is to understand and describe the directionality of previous cross-sectional associations, standard CLPMs may be sufficient. When results may inform recommendations applied to individuals (e.g., restricting preschool screen time exposure, suggesting a target for treatment), or when the study population is expected to be highly heterogeneous in baseline symptom severity or trajectories overtime (intercepts and slopes), it is worth considering a model that can disentangle trait, state, and individual effects.

6. Model selection and fit

There are numerous fit statistics or indices which help an investigator decide the degree to which the mathematical structure they have proposed is accurately reflected in their data (Table 1) (6, 14, 48, 49). Most authors will present a variety of fit indices to support their final chosen model, or to compare various model structures of increasing complexity. Commonly, models including only stability paths are compared to those with the addition of within-time, and then also cross-lagged associations (Figure 1). Models are said to be *nested* if the more restricted/parsimonious model is encompassed in the more complex models (like in the multi-group moderator analysis). Conversely, non-nested models may have different paths or variables that were not included in prior, less complex iterations. When comparing models, the choice of fit index varies depending on whether models are nested.

Although guidelines for acceptable ranges of fit indices exist (Table 1), most statisticians agree that there are no absolute cut-offs; interpreting model fit is a subjective process,

that depends in part on the research question, context, implications and setting for results (48). Many fit indices are also sensitive to sample size and may differentially favor model parsimony (i.e., less complex models with fewer estimated parameters) (6, 49). When critically appraising a study, readers should ensure that several fit indices are presented for each model, with the criteria for defining a "good enough" fit specified.

7. How do I interpret the results?

When interpreting the results of transactional models employing CLPMs, most investigators will present a final figure showing the cross-lagged panel and resulting estimates from the best/selected model. At times, investigators will show all paths (18), irrespective of significance or may exclude non-significant associations for figure clarity (8). Conceptually, all path estimates in transactional models are measures of association. They can be interpreted as correlation coefficients (within time) and regression coefficients (across time). Across time path coefficients are actually 'partial' coefficients because they incorporate and control for the multiple partial effects of several predictors.

Some investigators will present only standardized estimates in the final model, others will present standardized and unstandardized estimates together. Unstandardized estimates show the association with respect to the original scales or units of the variables (i.e., the expected change in Y for each unit change in X). Since measurement instruments usually differ in their scales, standardized estimates are used to anchor the parameter estimates to units of standard deviation (i.e., the expected change in standard deviation units for Y per change in one standard deviation unit of X). Effectively this puts the association parameters all on a scale of -1.0 to 1.0. Standardized estimates are useful to compare different paths (stability, within-time and cross-lags) within the same model, while unstandardized estimates are better when comparing across different sample populations or different models because the variance might be different.

Unlike traditional effect size estimates (e.g., Cohen's *d* of 0.9 = 'large effect'), cross-lagged paths have been adjusted for stability paths and within-time associations, and therefore absolute values that are much smaller can still be meaningful. In general, effect sizes of cross-lagged paths should be interpreted relative to the magnitude of the bivariate within-time correlations and stability paths (50). By inspecting the cross-lagged diagrams, we get an impression about whether the chosen variables influence each other over time, as well as the magnitude, and direction of the associations. The magnitude of the estimates of non-significant results can also be informative at times. In the study examples discussed previously (8, 9, 18), standardized estimates for stability paths ranged from 0.30 to 0.75 across time points, while significant cross-lagged estimates between variables had standardized estimates between 0.06 and 0.25.

There has been debate about whether CLPMs depict 'causal associations' (11, 17, 21). Recent studies have softened their interpretation of findings away from causality, considering the growing recognition of the limitations and sensitivities of SEM to a variety of factors. More common contemporary terminology includes 'reciprocal effects,' 'directional effects,' 'cross-domain effects,' 'antecedent symptoms' or 'influence' (11). Readers can judge the degree to which conclusions match the results, are not overstated and are anchored to the study time intervals.

8. Summary and guidelines for critical appraisal

Overall, through this introductory review of methodological and statistical procedures related to transactional models, several key points with respect to study design, analysis, and reporting have emerged, which may be helpful for non-statistical consumers of mental health research literature to consider. These are summarized in Table 2, where we provide a modified 12-item checklist based on the CASP checklist for cohort studies (51). This accessible tool can help with study interpretation and appraisal and can be useful for educational purposes.

Readers should begin by identifying the primary research objectives, intended level of inference, and any hypotheses formulated based on the existing literature. Measures and their constructs should be carefully described, and details of their measurement properties, validity (52), and longitudinal invariance in the study population discussed. Preliminary analyses may be needed to justify their use. Consideration of statistical procedures which can disentangle static 'trait-like' effects from dynamic 'state-like' effects, or account for data heterogeneity may be needed depending on context. Some missing data can be accommodated with newer statistical techniques. Missing data at each time point ought to be clearly reported. In general, sample sizes in the hundreds are needed, although this can vary highly depending on several factors. Multi-method, multi-informant instruments can help minimize measurement bias. Methods and results should present how the selected models were chosen and whether they meet the prespecified criteria for a good fit. Final chosen models and their diagrams should be clearly labeled and interpretable to non-expert audiences. Caution is advised to minimize causal language. Limitations should

Table 2. CASP checklist for interpretation and appraisal of cohort studies (51) adapted with further consideration for cross-lagged models

Question	CASP question	Further hints and cascade model considerations
Section A: Are the results of the study valid?		
1.	Did the study address a clearly focused issue?	- Are the research questions clearly articulated with respect to the variables, risk factors, outcomes and populations of interest? - Is there sufficient existing research and reason to support longitudinal cross-lagged modeling?
2.	Was the cohort recruited in an acceptable way?	- Was the cohort representative of a defined population? - Was everyone included who should have been included?
3. and 4.	Were the exposures and outcomes accurately measured to minimise bias?	- What were they trying to measure? - Do the chosen measures/ instruments capture what they wanted them to? Who completed the measures? - Were the chosen measures previously validated, in the target population? - Are the measures and constructs reliable and stable over the time intervals used, and are they likely to capture real change?
Section B: What are the results?		
5.	Have the authors identified all important confounding factors? Have they taken account of the confounding factors in the design and/or analysis?	- Are there variables that might have been missed that could explain any associations found or not found? - Are relevant confounders, mediators or moderators adjusted for or examined in the cross-lag model?
6.	Was the follow up of the subjects complete enough and long enough?	- What proportion of people were lost to follow up, and did they differ from those who were not in some way? - How were missing data accounted for? - Were the time intervals between assessment reasonable to address the research question?
7.	What are the results of the study?	- How strong are the associations between variables of interest? - Do they report standardized or unstandardized associations? Are they significant? - How good are the fit indices?
8.	How precise are the results?	
9.	Do you believe the results?	- Do the analyses separate static and dynamic effects? - How big are the effects? - Do the results make sense? (consider Bradford Hills criteria: time sequence, dose-response gradient, biological plausibility, consistency) - Does the magnitude and direction of effects make sense with respect to the hypotheses and phenomenon under study?
Section C: Will the results help locally?		
10.	Can the results be applied to the local populations?	- Do the participants in the study reflect your local setting?
11.	Do the results of the study fit with other available evidence?	- Are they consistent with previous longitudinal studies? Why or why not?
12.	What are the implications of this study for practice?	- One observational study rarely provides sufficiently robust evidence to recommend changes to clinical practice or within health policy decision making. - For certain questions, observational studies provide the only evidence. - Recommendations from observational studies are always stronger when supported by other evidence. - What are the next steps in terms of testing and validating an intervention based on the results?

be clearly described with respect to statistical assumptions not met or accounted for.

For clinicians, CLPMs are helpful to understand bio-psycho-social processes where randomized trials would be unfeasible or unethical, for example, regarding the impacts of spanking, discrimination, substance use, or specific early behavioural traits on future mental health outcomes (53-55). CLPMs can build on prior observed cross-sectional associations and begin to tease apart potential sequential transactions and co-developmental cascades which may be amenable to future studies of specific targeted interventions. CLPMs can be placed high on the hierarchy of evidence that can be derived from observational research. There are disadvantages to CLPMs as well; they represent a simplification of complex developmental processes and are contingent upon the quality of the underlying hypotheses, as well as the data from which they are derived. Specific trajectories of change and growth are not easily evident from panels; continuous processes are somewhat artificially broken down into discreet time points; and population heterogeneity not easily appreciated.

That said, developmental transactions and cascades are common, naturally occurring phenomena in mental health and development and statistical approaches using SEM and CLPMs are a powerful way to uncover them. Despite multiple limitations to these methods, they are fundamental to non-experimental and epidemiological research in mental health.

Conflicts of Interest:

DB, HB, TV, and EP have no conflicts of interest to declare. SV reports royalties from UpToDate Inc for authorship of materials related to antidepressants and pregnancy. PS receives royalties from Guilford Press and from Simon and Schuster.

Acknowledgements:

This study was supported by the Canadian Institutes of Health Research (CIHR) (P.S., grant numbers HDF-70333 and FDN 93621). Dr. Baribeau was supported by a CIHR doctoral award and the O'Brien Scholars Program.

References

1. Dussault F, Brendgen M, Vitaro F, Wanner B, Tremblay RE. Longitudinal links between impulsivity, gambling problems and depressive symptoms: a transactional model from adolescence to early adulthood. Journal of Child Psychology and Psychiatry. 2011;52(2):130-138.
2. Eiden RD, Lessard J, Colder CR, Livingston J, Casey M, Leonard KE. Developmental cascade model for adolescent substance use from infancy to late adolescence. Developmental psychology. 2016;52(10):1619-1633.
3. Burt KB, Obradovic J, Long JD, Masten AS. The Interplay of Social Competence and Psychopathology Over 20 Years: Testing Transactional and Cascade Models. Child development. 2008;79(2):359-374.
4. Masten AS, Cicchetti D. Developmental cascades. Development and psychopathology. 2010;22(3):491-495.
5. Curran PJ, Hancock GR. The Challenge of Modeling Co-Developmental Processes over Time. Child development perspectives. 2021;15(2):67-75.
6. Kline RB. Principles and practice of structural equation modeling: Guilford publications; 2015.
7. Yang J, Xu M, Sullivan L, Taylor HG, Yeates KO. Bidirectional Association Between Daily Physical Activity and Postconcussion Symptoms Among Youth. JAMA Network Open. 2020;3(11):e2027486-e.
8. Lee KS, Vaillancourt T. Longitudinal Associations Among Bullying by Peers, Disordered Eating Behavior, and Symptoms of Depression During Adolescence. JAMA psychiatry. 2018;75(6):605-612.
9. Madigan S, Browne D, Racine N, Mori C, Tough S. Association Between Screen Time and Children's Performance on a Developmental Screening Test. JAMA pediatrics. 2019;173(3):244-250.
10. McArthur BA, Browne D, McDonald S, Tough S, Madigan S. Longitudinal Associations Between Screen Use and Reading in Preschool-Aged Children. Pediatrics. 2021;147(6).
11. Hamaker EL, Kuiper RM, Grasman RP. A critique of the cross-lagged panel model. Psychological Methods. 2015;20(1):102-116.
12. Berry D, Willoughby MT. On the Practical Interpretability of Cross-Lagged Panel Models: Rethinking a Developmental Workhorse. Child Development. 2017;88(4):1186-1206.
13. Kuiper RM, Ryan O. Drawing Conclusions from Cross-Lagged Relationships: Re-Considering the Role of the Time-Interval. Structural Equation Modeling: A Multidisciplinary Journal. 2018;25(5):809-823.
14. McArdle JJ, Nesselroade JR. Longitudinal data analysis using structural equation models: American Psychological Association; 2014.
15. Mulder JD, Hamaker EL. Three extensions of the random intercept cross-lagged panel model. Structural Equation Modeling: A Multidisciplinary Journal. 2021:1-11.
16. Usami S. On the differences between general cross-lagged panel model and random-intercept cross-lagged panel model: Interpretation of cross-lagged parameters and model choice. Structural Equation Modeling: A Multidisciplinary Journal. 2021;28(3):331-344.
17. Kenny DA, Harackiewicz JM. Cross-lagged panel correlation: Practice and promise. Journal of Applied Psychology. 1979;64(4):372.
18. Pickard H, Rijsdijk F, Happe F, Mandy W. Are Social and Communication Difficulties a Risk Factor for the Development of Social Anxiety? Journal of the American Academy of Child and Adolescent Psychiatry. 2017;56(4):344-351 e3.
19. Bohrnstedt GW. Observations on the measurement of change. Sociological Methodology. 1969;1:113-133.
20. Duncan OD. Introduction to structural equation models: Elsevier; 2014.
21. Heise DR. Causal inference from panel data. Sociological

Methodology. 1970;2:3-27.

22. Lee KS, Vaillancourt T. The role of childhood generalized anxiety in the internalizing cluster. Psychological medicine. 2019:1-11.
23. Vandenberg RJ, Lance CE. A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. Organizational research methods. 2000;3(1):4-70.
24. Xu J, Zhang Q, Yang Y. Impact of violations of measurement invariance in cross-lagged panel mediation models. Behavior Research Methods. 2020;52(6):2623-2645.
25. Lai K. Estimating Standardized SEM Parameters Given Nonnormal Data and Incorrect Model: Methods and Comparison. Structural Equation Modeling: A Multidisciplinary Journal. 2018;25(4):600-620.
26. SAS S. STAT 9.3 User's guide. Cary, NC: SAS Institute Inc. 2011.
27. Norman G. Likert scales, levels of measurement and the "laws" of statistics. Adv Health Sci Educ Theory Pract. 2010;15(5):625-632.
28. Reddy SK. Effects of ignoring correlated measurement error in structural equation models. Educational and Psychological Measurement. 1992;52(3):549-570.
29. Kirshner B, Guyatt G. A methodological framework for assessing health indices. Journal of Chronic Diseases. 1985;38(1):27-36.
30. Beaton DE, Bombardier C, Katz JN, Wright JG. A taxonomy for responsiveness. Journal of Clinical Epidemiology. 2001;54(12):1204-1217.
31. Grimm KJ, Ram N, Hamagami F. Nonlinear growth curves in developmental research. Child Development. 2011;82(5):1357-1371.
32. He F, Teixeira-Pinto A, Harezlak J. Autoregressive and cross-lagged model for bivariate non-commensurate outcomes. Statistics in Medicine. 2017;36(19):3110-3120.
33. Duncan OD. Some linear models for two-wave, two-variable panel analysis. Psychological Bulletin. 1969;72(3):177.
34. Greenberg DF, Kessler RC. Equilibrium and identification in linear panel models. Sociological Methods & Research. 1982;10(4):435-451.
35. Gollob HF, Reichardt CS. Taking account of time lags in causal models. Child Development. 1987:80-92.
36. Vaillancourt T, Brittain H, Krygsman A, Davis A, Farrell A, Desmarais R, et al. Assessing the quality of research examining change in children's mental health in the context of COVID-19. University of Ottawa Journal of Medicine. 2021;11(1).
37. Rochon PA, Gurwitz JH, Sykora K, Mamdani M, Streiner DL, Garfinkel S, et al. Reader's guide to critical appraisal of cohort studies: 1. Role and design. British Medical Journal. 2005;330(7496):895-897.
38. Normand SL, Sykora K, Li P, Mamdani M, Rochon PA, Anderson GM. Readers guide to critical appraisal of cohort studies: 3. Analytical strategies to reduce confounding. British Medical Journal. 2005;330(7498):1021-1023.
39. Mamdani M, Sykora K, Li P, Normand SL, Streiner DL, Austin PC, et al. Reader's guide to critical appraisal of cohort studies: 2. Assessing potential for confounding. British Medical Journal. 2005;330(7497):960-962.
40. Iacobucci D. Structural equations modeling: Fit indices, sample size, and advanced topics. Journal of Consumer Psychology. 2010;20(1):90-98.
41. Muthén BO, Curran PJ. General longitudinal modeling of individual differences in experimental designs: a latent variable framework for analysis and power estimation. Psychological Methods. 1997;2(4):371.
42. Wolf EJ, Harrington KM, Clark SL, Miller MW. Sample Size Requirements for Structural Equation Models: An Evaluation of Power, Bias, and Solution Propriety. Educational and Psychological Measurement. 2013;76(6):913-934.
43. Enders CK, Bandalos DL. The relative performance of full information maximum likelihood estimation for missing data in structural equation models. Structural Equation Modeling. 2001;8(3):430-457.
44. Newman DA. Longitudinal modeling with randomly and systematically missing data: A simulation of ad hoc, maximum likelihood, and multiple imputation techniques. Organizational Research Methods. 2003;6(3):328-362.
45. Larsen R. Missing data imputation versus full information maximum likelihood with second-level dependencies. Structural Equation Modeling: A Multidisciplinary Journal. 2011;18(4):649-662.
46. Haukoos JS, Newgard CD. Advanced statistics: missing data in clinical research--part 1: an introduction and conceptual framework. Acad Emerg Med. 2007;14(7):662-668.
47. Rubin DB. Inference and Missing Data. Biometrika. 1976;63(3):581-592.
48. Hooper D, Coughlan J, Mullen MR. Structural equation modelling: Guidelines for determining model fit. Electronic Journal of Business Research Methods. 2008;6(1):53-60.
49. Hu Lt, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural equation modeling: a multidisciplinary journal. 1999;6(1):1-55.
50. Adachi P, Willoughby T. Interpreting effect sizes when controlling for stability effects in longitudinal autoregressive models: Implications for psychological science. European Journal of Developmental Psychology. 2015;12(1):116-128.
51. Critical Appraisal Skills Programme. CASP Cohort Study Checklist 2018 [Available from: https://casp-uk.net/wp-content/uploads/2018/01/CASP-Cohort-Study-Checklist_2018.pdf.
52. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Quality of Life Research. 2010;19(4):539-549.
53. Gibbons FX, Gerrard M, Cleveland MJ, Wills TA, Brody G. Perceived discrimination and substance use in African American parents and their children: a panel study. Journal of personality and social psychology. 2004;86(4):517.
54. Gershoff ET, Lansford JE, Sexton HR, Davis-Kean P, Sameroff AJ. Longitudinal Links Between Spanking and Children's Externalizing Behaviors in a National Sample of White, Black, Hispanic, and Asian American Families. Child Development. 2012;83(3):838-843.
55. Fergusson DM, Boden JM, Horwood LJ. Tests of causal links between alcohol abuse or dependence and major depression. Archives of general psychiatry. 2009;66(3):260-266.