SMGroup

SM Journal of Biometrics & Biostatistics

Research Article

Mediation Analysis of Co-occurring Conditions for Complex Longitudinal Clinical Data

Douglas D. Gunzler[1]*, Nathan Morris[2], Adam Perzynski[1], Deborah Miller[3], Steven Lewis[1] and Robert A. Bermel[3]

[1]*Case Western Reserve University, Center for Health Care Research & Policy, Metro Health Medical Center, Cleveland, Ohio, USA*

[2]*Case Western Reserve University, Department of Epidemiology and Biostatistics, Cleveland, Ohio, USA*

[3]*Mellen Center for Multiple Sclerosis Treatment and Research, Cleveland Clinic, Cleveland, Ohio, USA*

Article Information

Received date: Nov 29, 2016
Accepted date: Jan 18, 2017
Published date: Jan 25, 2017

***Corresponding author**

Douglas D. Gunzler, Case Western Reserve University, Center for Health Care Research & Policy, Metro Health Medical Center, 2500 Metro Health Drive, Cleveland, Ohio, 44109-1998, USA, Tel: 216-778-2764; Fax: 216-778-3945; Email: dgunzler@metrohealth.org

Keywords Structural equation modeling; Latent growth modeling; Mediation analysis; Differential item functioning; Multiple indicator multiple cause modeling; Multiple sclerosis

Abbreviations DIF - Differential Item Functioning; MS - Multiple Sclerosis; PRO - Patient Reported Outcome; PHQ-9 - Patient Health Questionnaire-9; EHR - Electronic Health Records; SEM - Structural Equation Modeling; LGM - Latent Growth Modeling; MIMIC - Multiple Indicator Multiple Cause; KP - Knowledge Program; PS - Performance Scales© ; MLR = (MPlus Option for) Maximum Likelihood with Robust Standard Errors; FIML - Full Information Maximum Likelihood; BIC - Bayesian Information Criterion; AIC - Akaike Information Criterion; RMSEA - Root Mean Error of Approximation; CFI - Comparative Fit Index; EFA - Exploratory Factor Analysis; WLSMV - (MPlus Option for) mean and variance adjusted weighted least squares estimator (WLSMV); ICC - Intraclass Correlation Coefficient; IE - Indirect Effect; CI - Confidence Interval

Abstract

Background: Symptoms or test results may be common to two or more co-occurring conditions. This problem of symptom overlap makes it challenging for clinicians to determine a focus for treatment in a patient given changes in the severity of either condition.

Methods: Structural equation modeling methods can be used to disentangle some of the complexities of disease symptom etiology, given co-occurring conditions, and support treatment decision making. These techniques provide the flexibility to deal with specific challenges present in data as extracted from Electronic Health Records (EHR) (i.e. individually varying follow up times, irregular follow up, missingness, systematic error in patient reported outcomes, lack of clear temporal precedence between measures). Specifically, a proposed latent growth modeling approach accounting for differential item functioning along with the Monte Carlo simulation method for assessment of mediation can be used to investigate how one condition leads to a co-occurring condition, adjusted for the overlapping symptoms of both conditions.

Results: This paper uses an example investigating how Multiple Sclerosis (MS) leads to depression in patients in which depressive symptoms overlap with other symptoms of MS, such as fatigue, cognitive impairment and physical impairment to illustrate the methods. It was demonstrated that not adjusting for this overlap can lead to different results.

Conclusions: Developing methods for mediation analysis of co-occurring conditions for more complex longitudinal clinical data as recorded at a typical patient visit can help clinicians make improved use of data bases such as EHR to support clinical decision making in real time.

Background

In mediation, we consider an intermediate variable, called the *mediator*, that helps explain how or why an independent variable influences an outcome [1-3]. Gaining an understanding of which particular disabilities act as mechanisms of change (or mediators) between co-occurring conditions would allow for more focused treatment for a patient given changes in severity of both conditions. In many cases of co-occurring conditions, however, temporal precedence for mediation may be unclear due to the unknown origin of symptoms in the conditions (i.e. cognitive impairment in multiple sclerosis and depression symptoms). Thus, in such cases, the mediator and the outcome measured across multiple time points may be viewed as separate parallel processes [4]. As a result, the *mediational process* can be defined as the independent variable influencing the growth of the mediator, which, in turn, affects the growth of the outcome [4].

Further, symptoms of the co-occurring conditions may overlap. This can lead to inappropriate clinical decisions (medication selection, escalation, etc.) and to incorrect inferences regarding treatment effectiveness. Scales and diagnoses for patients with co-occurring conditions may be especially problematic as they may suffer from *criterion contamination* due to the overlap. Criterion contamination occurs when the criterion measure is affected by "construct-irrelevant" [5] factors that are not part of the criterion construct. Methods have been previously proposed to adjust scales to better represent the underlying dimension of the criterion measure for patients with co-occurring conditions using cross-sectional data [6,7]. Here we propose methods which should improve our ability to perform mediation analysis and distinguish symptom changes over time in patients with co-occurring conditions.

Overlapping symptoms of co-occurring conditions leads to a type of systematic error in scales or diagnoses known as *Differential Item Functioning* (DIF). DIF can occur when people from different groups (e.g., levels of Multiple Sclerosis (MS)-related fatigue) with the same latent trait (level of depression) have a different probability of giving a certain response on a questionnaire or test (e.g., items for sleep problems and fatigue of the PHQ-9).

OPEN ACCESS
ISSN: 2573-5470

How to cite this article Gunzler DD, Morris N, Perzynski A, Miller D, Lewis S and Bermel RA. Mediation Analysis of Co-occurring Conditions for Complex Longitudinal Clinical Data. SM J Biometrics Biostat. 2017; 2(1): 1004.

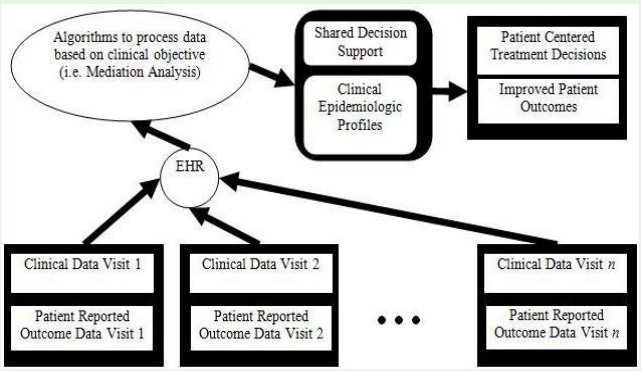

Figure 1: Integrating Patient Reported Outcomes and Clinical Data Using Electronic Health Records (EHR) to Improve Care and Outcomes.
Visit 1 up to Visit *n* are the individually time varying follow up visits for each individual from baseline up to *n* possible time points.

Incorporating the Patient Reported Outcomes (PROs), such as the PHQ-9, collected at a typical patient visit and stored in big databases such as Electronic Health Records (EHR), into patient care and research protocols is an attractive approach to patient centered medicine. However, PROs while informative in measuring the self-reported health state of the patient directly, lead to additional systematic error [8]. Each individual patient may have a different view of how to fill out a test questionnaire. Further, EHR data bases are also filled with complex clinical data characterized by individually time varying follow up times, irregular follow up, missingness, and systematic error. Methods developed and applied in this paper provide a look forward at how trajectory information about symptoms of co-occurring conditions based on PROs and collected in an EHR database can then be used to support clinical decision making in real time (see Figure 1).

Using Structural Equation Modeling (SEM) [9,10] we can perform mediation analysis to examine the complex relationships between different symptoms of the co-occurring conditions (i.e. patients' physical and mental health states) while accounting for (1) the overlapping symptoms of the co-occurring conditions and (2) features of complex longitudinal clinical data. Latent Growth Modeling (LGM) is a practical application of SEM for longitudinal data to estimate growth trajectories [11,12]. The LGM framework allows us to build a statistical model corresponding precisely with the present study's conceptual framework leading to clear hypothesis articulation and enhanced statistical power while examining the relationships between multiple time varying measures [12]. The Multiple Indicator Multiple Cause (MIMIC) model, a measurement model with covariates [10,13-15], permits detection and adjustment for DIF. Given unclear temporal precedence between co-occurring conditions, evaluation of the mediational process can be carried out using the parallel process LGM approach [3,4] while simultaneously using the MIMIC model at repeated measures to account for overlapping symptoms of our co-occurring conditions.

Our original contribution is thus two-fold: (1) extending mediation theory for application on co-occurring conditions (2) developing methods designed for use with complex longitudinal clinical data as found in large data bases such as EHR to support clinical decision making in real time. Clinicians may use information provided by PROs and other disability measures collected at a typical patient visit to help aid in appropriately tailoring care management.

We start, in Methods, with background information about a MS-depression study for our application of these SEM methods. This is followed, by an introduction to the SEM framework for mediation analysis of co-occurring conditions for complex longitudinal clinical data, as well as an introduction to the SEM techniques (MIMIC modeling, LGM, mediation analysis) used in these analyses. In Results and Discussion, we provide an extension and application of our methods on the MS-depression study. We make a comparison in this section of our model accounting for overlapping symptoms of co-occurring conditions to the same model without accounting for the overlapping symptoms. Conclusion summarizes the paper. We provide the MPlus code for our example in the Appendix.

Methods

Multiple sclerosis and depression example and study design

Multiple sclerosis (MS) is the most common progressive neurological disease of young adults and affects approximately 400,000 persons in the United States [6,16]. Depression is the most frequent psychiatric diagnosis in MS patients, with lifetime risk estimated at ~50% [16,17]. Patients with MS show increased severity of depressive symptoms compared to patients with other chronic neurological conditions [18]. Symptoms of depression in MS overlap with other common MS symptoms including fatigue, cognitive impairment and physical disability [18,19].

Cleveland Clinic's Knowledge Program (KP) [20] links patient-reported PHQ-9 data to its EPIC Electronic Health Record (EHR), yielding powerful opportunities to study and improve patient care and clinical research. The Mellen Center [21] for Multiple Sclerosis manages more than 20000 visits and 1000 new patients every year for MS treatment. The KP tracks illness severity and treatment efficacy over time across the Mellen Center population.

We used a retrospective cohort study design. The inclusion criteria for our sample included patients making at least one visit to the Mellen Center with measurements of PHQ-9 score and a timed 25-foot walk available. Data were available for 3507 MS patients from 2008-2011 that met inclusion criteria at baseline. If a follow-up visit was less than one month later, we either did not consider it in the longitudinal data set or merged any new recordings to fill in missing data for the prior visit. The reason we collapsed visits less than one month apart was that these might not have been new visits, but just additional information added in the EHR database about the patient. Further, these might have been partial visits for the purpose of clinical surveillance of a more acute problem. Patients were seen an average of 3.9 times (SD = 1.5) during the KP to date. Most (77%) of the patients returned for a second visit in the available data window, and just over four-fifths (81%) of those patients made a third visit. Similar drop-off patterns emerged through the first eight visits, and 402 patients had at least seven follow-up visits. Visits to the Mellen Center after the first occurred irregularly, with about half of the patients seen again within six months.

More severely disabled MS patients might be inclined to visit the Mellen Center more frequently, thus leading to the possibility

of nonignorable missing data patterns. However, our inclusion criteria of a recorded timed 25-foot walk eliminated anyone who was completely immobile from this dataset (also randomly eliminated some patients without a timed walk recorded in the EHR system). Further, we also examined if number of visits per patient was correlated with symptom severity (baseline total PHQ-9 score, MS-related fatigue, MS-related cognitive impairment, timed walk and peg test) and the Pearson correlation was less than 0.10 in all five cases.

Measures assessed

The PHQ-9 [22] screens for and monitors depression. A self-reported multiple item depression screening tool, the PHQ-9 is meant to be used in connection with expert clinical judgment and/or further rating tools [22] and not as an actual depression diagnosis. Patients specify frequency in the past two weeks (0 = not at all to 3 = every day) of nine symptoms, yielding a total score (range: 0-27). Scores of 5, 10, 15, and 20 are validated thresholds for mild, moderate, moderately severe and severe depression. Scores on this self-reported instrument are often used to guide treatment decisions [23]. In particular, a PHQ-9 ≥ 10 has been previously established as a screening cutoff for depressive disorder [19,23]. The PHQ-9 has been validated using multiple modes for administration, clinical populations, and diverse race/ethnicity groups [24].

The KP collects MS Performance Scales© (PS) [25,26] which are patient-reported disability measures. Single-item PS was originally developed for eight domains of function (mobility, hand function, vision, fatigue, cognition, bladder/bowel, sensory, and spasticity) [26,27]. To increase content validity, three more measures were added in 2001 to assess disability associated with pain, depression, and tremor/coordination [26,27]. Each PS has six ordinal responses except for the mobility PS which has seven. Reliability, criterion and construct validity have been established for these domains in previous studies of MS patients [25,26]. The fatigue and cognitive domains of the PS are used as our main covariates for self-reported MS-related fatigue and cognitive decline.

In addition to patient-reported measures, our measures for MS physical disability, the timed 25-foot walk and 9-hole peg test, are single item objective performance measures of lower (timed 25-foot walk) and upper (9-hole peg test) extremity function [28]. The timed 25-foot walk is a test of quantitative mobility and leg function performance, while the 9-hole peg test is a brief, standardized, quantitative test of arm and hand function. These two measures together present a measure of functional impairment [29-31].

If a MS patient can clearly distinguish MS symptoms from depressive symptoms, then the MS-disability scales would only capture MS-related symptoms, while the PHQ-9 would only capture depression-related symptoms. However, symptoms such as fatigue, cognitive impairment and physical impairment are greatly more complex, due to the multiple dimensions, such as physical, mental and emotional, that may be involved in describing MS-related symptoms [32]. Thus, the experience of fatigue, cognitive impairment and physical impairment in a MS patient includes both separate and intersecting properties of MS and depression. As a result of this complexity in categorizing these MS disabilities, we must account for the ensuing systematic error in the form of DIF when analyzing these scales.

MS patient-specific disease characteristics are assessed from two measures in our database. Firstly, baseline time since symptom onset is a measure of disease duration [33]. Secondly, MS type at baseline (relapsing or progressive) defines disease phenotype, where progressive forms are characterized by progressive neurologic decline between acute attacks without the definite periods of remission that occurs in relapsing forms.

Demographic characteristics of the MS population in the KP data base

We compile descriptive statistics to summarize demographic information and scores at baseline on our measures of depression and MS symptoms (performance scales fatigue, cognitive, timed 25-foot walk, 9-hole peg test, and PHQ-9). The sample mirrors the United States' MS population in that MS is typically diagnosed in patients in their early 30s, Caucasians are of highest risk and females are twice as likely as males to develop MS [6,16]. In our baseline sample, 73% were female, 83% were white, and the average age was 46 (SD = 12). These patients had their first MS diagnosis an average of 10 (SD = 9) years ago with 81% relapsing and 16% progressive with the remaining patients falling into other categories, or under evaluation for a potential MS diagnosis. We leave these patients who are not relapsing or progressive (N= 70) out of our analyses based on MS type, due to our uncertainty about their diagnosis (Table 1).

Nearly 30% (n=1005) of patients had PHQ-9 ≥ 10 at their entry to the KP. The distribution of PHQ-9 scores represents a wide range of depression severity levels. We summarize the characteristics of the

Table 1: Characteristics of the Mellen Center MS population.

	PHQ-9 < 10	PHQ-9 ≥ 10	
	n = 2502	n = 1005	p
PHQ-9	3.64 ± 2.75	15.26 ± 4.40	<0.001
MSPS fatigue	1.62 ± 1.25	3.35 ± 1.12	<0.001
MSPS cognitive	0.86 ± 0.96	2.23 ± 1.30	<0.001
25-foot timed walk	7.85 ± 10.56	8.83 ± 7.61	0.002
9-hole peg test	23.68 ± 10.66	26.82 ± 12.48	<0.001
age	46.12 ± 11.88	44.47 ± 11.20	<0.001
baseline time since symptom onset	11.80 ± 10.00	10.89 ± 9.37	0.016
gender, n (%)			0.879
female	1836 (74)	740 (74)	
male	666 (27)	265 (26)	
race, n (%)			0.070
caucasian	2112 (85)	821 (82)	
african-american	225 (9)	114 (11)	
other	144 (6)	65 (7)	
MS type, n (%)			0.067
relapsing	2045 (84)	787 (82)	
progressive	383 (16)	177 (18)	

Mean ± standard deviation for continuous measures and number of subjects in each category for discrete measures with p-values reported from t-tests and chi-square tests where appropriate. Adapted from Gunzler *et al.* (2014). Disentangling Multiple Sclerosis & Depression: An Adjusted Depression Screening Score for Patient-Centered Care. *Journal of Behavioral Medicine.*

Mellen Center MS population by the PHQ-9 binary threshold of 10 (screening cutoff for depressive disorder) in Table 1.

We define $\alpha = 0.05$ for our level of significance in all statistical tests throughout this paper. All statistical tests are two-tailed. Analyses were carried out using Mplus Version 7 [34], SAS Version 9.2 [35] and R [36].

Longitudinal mediation model path diagram

The mediational process used for investigating how one condition influences a co-occurring condition, adjusted for the overlapping symptoms of both conditions should have a basis in expert derived a priori theory. A path diagram could be used to visually depict the conceptual model for the hypothesized (1) overlap between the co-occurring conditions (2) growth of each condition and (3) meditational process investigating how growth of one condition influences the growth of a co-occurring condition. In Figure 2 we express one simplified example of this type of path diagram.

As an illustrative example, corresponding to a path diagram such as Figure 2, we might have at multiple time points measures of three items on a depression screening scale (i.e. mood, anhedonia, cognition) for an underlying latent factor of *depression* and an observed single item MS-related cognitive impairment scale measuring cognitive decline, a symptoms of MS. We hypothesize that growth of MS-related cognitive impairment mediates the relationship between patient-specific disease characteristics at baseline (i.e. time since symptom onset) and growth of *depression*. However, in order to perform the mediation analysis we would need to account for the potential overlap between *depression* (cognition item of the depression screening scale) and MS-related cognitive impairment.

The general hypothesized longitudinal mediation model depicted in Figure 2 involves a latent response and another observed measure (the mediator) describing a symptom from a co-occurring condition. We use causal paths to describe the overlap, which can be referred to as *DIF paths* or *DIF effects*. Without loss of generality, we assume all measures in our model are on an interval scale. Assuming items or measures are categorical can be done using the framework we describe along with general latent variable modeling theory [37,38].

To work up to the full model in our application to MS, we first discuss the simpler model from Figure 2 with three observed indicators of latent factor η_{it}, representing symptoms of Condition A, and an observed covariate z_{it} (the mediator), representing a symptom of Condition B at $t = 1, 2, \ldots, T$ repeated measures for subject i, $i = 1,\ldots, n$. Adding in a feature of the complex longitudinal clinical data from an EHR data base, the measurements are time-varying for each individual and the t^{th} measurement for the i^{th} individual was made at τ_{it}. The potential overlapping symptoms of A and B are modeled through the average DIF path (γ_y) from regressing $[\text{Item 1}]_{it}$ on z_{it}. To understand in practical terms the average DIF path, from our illustrative example, we are directly regressing the item for cognition from our depression screening scale on the MS-related cognitive impairment scale to account for the potential overlap between *depression* and MS-related cognitive impairment.

Note here that we have assumed that the measurement model for the latent factor along with the DIF relationship is invariant over time, although it need not be. While z_{it} is an observed covariate with respect to η_{it}, it is still an endogenous variable in the structural equation model corresponding to Figure 2, since it is regressed on x_i.

As mentioned above, for an example, η_{it} can represent three items in a depression screening scale for an underlying latent factor of depression and z_{it} can represent an observed single item MS-related cognitive impairment scale with potential overlap between the two scales. Figure 2 can be extended for more items, specified correlations between items, latent factors and cross-loaded items. Thus an even more general type of model, beyond the scope of discussion in this paper, with multiple latent factors is a multi-factor Confirmatory Analytic (CFA) model with cross loaded items (Brown, 2006). Extending to the more general model becomes straightforward using SEM theory but demands more sample size and computation power with more model complexity.

While simultaneously modeling this overlap from Condition A and Condition B from Figure 2, a parallel process LGM approach is used to model the growth in z_{it} and η_{it} for the T repeated measures. We hypothesize longitudinal mediation in that the growth in the mediator z_{it} affects the growth in η_{it}. In this case we conceptually hypothesize that the growth in z_{it} mediates the relationship between an observed independent variable x_i and the growth of the latent response η_{it}. x_i may be some measure of disease severity. In the case of MS, without a good measure of disease severity we might use a patient-specific disease characteristic such as time since symptom onset for x_i.

SEM techniques for Longitudinal Mediation Model

We provide a brief overview of the individual SEM techniques used in this paper, corresponding to Figure 2. MIMIC modeling with DIF paths is a technique for modeling the relationship between symptoms of co-occurring conditions, while accounting for the overlapping symptoms of both conditions. For example,

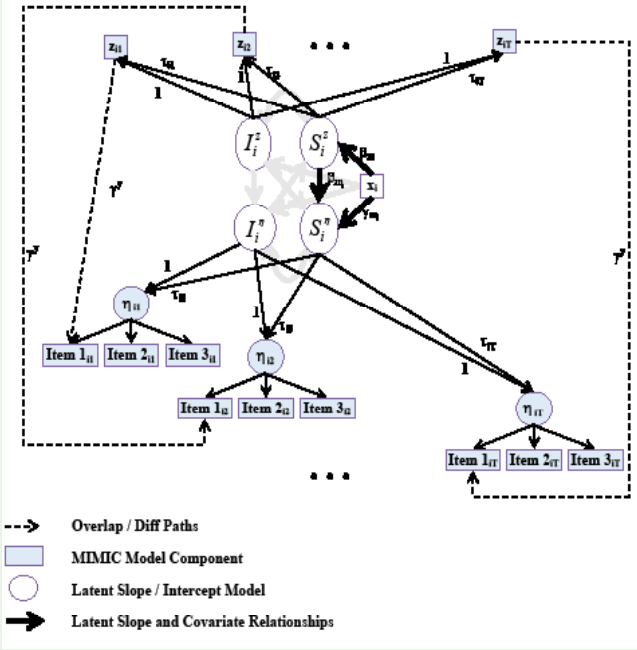

Figure 2: Simplified path diagram for longitudinal mediation model of overlapping symptoms of two conditions.

the approach can be used to model the relationship between MS symptoms and depressive symptoms while accounting for overlap via DIF path effects. Latent growth modeling is a technique used to model the change in the co-occurring conditions over time (i.e. growth of depressive symptoms and growth of MS symptoms). Finally mediation analysis, in the context of SEM, is a technique used to investigate how one condition leads to a co-occurring condition. For example, here we might be interested in understanding how MS patient-specific disease characteristics lead to depression.

MIMIC modeling with DIF paths

The MIMIC model is an important special case of SEM. It is a measurement model (i.e. factor model) with observed covariates [10,14,15]. The observed covariates explain the latent construct in the MIMIC Model. A MIMIC model can include DIF paths which can be used to model the overlap between the co-occurring conditions.

A general form of a MIMIC model with DIF paths (in the matrix $\boldsymbol{\Gamma}_y$):

$$\begin{aligned} \boldsymbol{y}_i &= \boldsymbol{\nu}_y + \boldsymbol{\Lambda}_y \boldsymbol{\eta}_i + \boldsymbol{\Gamma}_y \boldsymbol{X}_i + \boldsymbol{\varepsilon}_i, \\ \boldsymbol{\eta}_i &= \boldsymbol{\nu} + \boldsymbol{\Lambda} \boldsymbol{\eta}_i + \boldsymbol{\Gamma} \boldsymbol{X}_i + \boldsymbol{\varepsilon}_i^* \end{aligned} \tag{1}$$

Here we discuss a form that is applicable to cross-sectional data for subject i, $i = 1,\ldots, n$, for observed variables $\mathbf{y}_i$ and $\mathbf{x}_i$ and a latent unobserved variable $\boldsymbol{\eta}_i$ measured by $\mathbf{y}_i$. Our unknown parameters to be estimated are $\boldsymbol{\nu}_y, \boldsymbol{\Lambda}_y, \boldsymbol{\Gamma}_y, \boldsymbol{\nu}, \boldsymbol{\Lambda}$, and $\boldsymbol{\Gamma}$. x_i is a $q \times 1$ vector of independent variables. $\boldsymbol{\eta}_i$ is a $m \times 1$ vector of unobserved latent *endogenous* variables which are measured by the $p \times 1$ vector of observed variables $\mathbf{y}_i$. The equation for $\mathbf{y}_i$ includes a $p \times 1$ vector of intercepts $\boldsymbol{\nu}_y$, a $p \times m$ matrix of slopes $\boldsymbol{\Lambda}_y$, a $p \times q$ matrix of slopes $\boldsymbol{\Gamma}_y$, and a $p \times 1$ vector of corresponding random error terms $\boldsymbol{\varepsilon}_i$. $\boldsymbol{\Lambda}_y$ is often referred to as a loading matrix, while $\boldsymbol{\Gamma}_y$ is a matrix of slopes representing DIF paths. $\boldsymbol{\nu}$ is a $m \times 1$ vector of intercepts and $\boldsymbol{\Lambda}$ is a $m \times m$ matrix of slopes relating the endogenous latent variables to each other, $\boldsymbol{\Gamma}$ is $m \times q$ matrix of slopes for the observed independent variables and $\boldsymbol{\varepsilon}_i^*$ is a $m \times 1$ vector of random error terms for the unobserved endogenous latent variables.

As an example of the MIMIC model, the observed variables $\mathbf{y}_i$ could represent the items in a depression screening scale. Here the independent variables x_i could describe MS symptoms. Thus, the matrix of slopes representing DIF paths $\boldsymbol{\Gamma}_y$ would be parameters to estimate of the magnitude of overlap between symptoms of MS and depression.

Latent growth modeling

Latent growth modeling is a technique that can be used to model changes in a condition over time, such as the growth of depression. The general form of a latent growth model at repeated measures index t, $1 \le t \le T$, for a individually varying time point τ_{it} for subject i, $i = 1,\ldots, n$ is discussed in this section. If for an individual the number of observed repeated measures $t = T^*$ and if $T^* < T$, then the last $T - T^*$ repeated measures will be considered missing. This represents a typical case with real world hospital observational data (such as the EHR-based data in the MS-depression example) where appointment days vary by individual.

$$Y_{it} = I_i + S_i \tau_{it} + \delta_{it} \tag{2}$$

I_i is the latent intercept for subject i, S_i is the latent slope for subject i, and δ_{it} is the disturbance term for subject i at time t. The normality of I_i, S_i, and δ_{it} and independence of I_i and S_i from δ_{it} are assumed. We can extend the latent growth model to include a quadratic term (and cubic term) for nonlinear growth. In equation (3) we extend the growth model to include a quadratic growth term Q_i for subject i.

$$Y_{it} = I_i + S_i \tau_{it} + Q_i \tau_{it}^{\,2} + \delta_{it} \tag{3}$$

Without any further constraints, there is an identifiability problem for a latent growth model for an *unobserved* latent construct η (see Appendix A for proof).

Mediation model

(Figure 3) A general form of the structural equations for a single mediator model for independent variable x_i, mediator z_i and outcome y_i with unknown parameters $\beta_y, \beta_{zy}, \gamma_{xy}, \beta_z$, and β_{xz} [1,3,39]:

$$\begin{aligned} y_i &= \beta_y + \beta_{zy} z_i + \gamma_{xy} x_i + \zeta_i^y, \\ z_i &= \beta_z + \beta_{xz} x_i + \zeta_i^z \end{aligned} \tag{4}$$

Here we discuss a form that is applicable to cross-sectional data for subject i, $i = 1,\ldots, n$. ζ_i^y and ζ_i^z are random error terms associated with y_i and z_i respectively.

We typically in the context of mediation assume $cor(\zeta_i^y, \zeta_i^z) = 0$, where *cor* is the correlation, an important assumption for causal inference in performing mediation analysis [40]. Assuming multivariate normality for the error terms is a necessary underlying condition for a more straightforward definition of direct, indirect and total effects [2]. The *direct effect* is the pathway from the exogenous variable to the outcome while controlling for the mediator.

Therefore, in our mediation model and Figure 2, γ_{xy} is the direct effect. The *indirect effect* describes the pathway from the exogenous variable to the outcome through the mediator. This path is represented by the product $\beta_{xz}\beta_{zy}$. Finally, the *total effect* is the sum of the direct and indirect effects of the exogenous variable on the outcome, $\gamma_{xy} + \beta_{xz}\beta_{zy}$. Definitions of these type of effects under other distributional assumptions are discussed in the mediation literature [41].

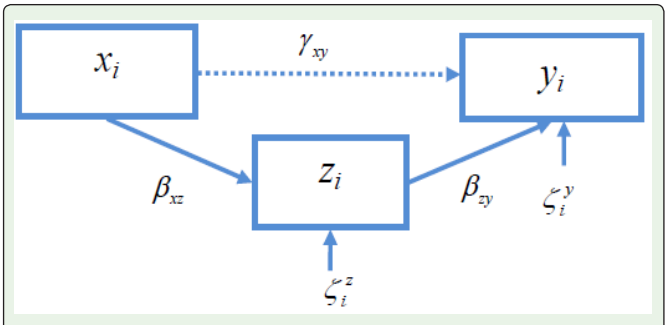

Figure 3: Path diagram for the hypothesized single mediation model.

The primary hypothesis of interest in a mediation analysis is to see whether the effect of the independent variable on the outcome can be mediated by a change in the mediating variable [2]. In a full mediation process, the effect is 100% mediated by the mediator, that is, in the presence of the mediator, the pathway connecting x_i to y_i is completely broken so that x_i has no direct effect on y_i. In most applications, however, partial mediation is more common, in which case the mediator z_i only mediates part of the effect of x_i on y_i, that is, x_i has some residual direct effect even after the mediator is introduced into the model.

Longitudinal SEM accounting for overlapping symptoms of co-occurring conditions

The structural equations corresponding to the path diagram from Figure 2 can be expressed in the following form:

$$\begin{aligned} Item\,1_{it} &= \mu_1 + \lambda_1\eta_{it} + \gamma_y z_{it} + e_i^{Item\,1} + \varepsilon_{it}^{Item\,1}, \\ Item\,2_{it} &= \mu_2 + \lambda_2\eta_{it} + e_i^{Item\,2} + \varepsilon_{it}^{Item\,2}, \\ Item\,3_{it} &= \mu_3 + \lambda_3\eta_{it} + e_i^{Item\,3} + \varepsilon_{it}^{Item\,3}, \end{aligned} \tag{5}$$

$$\begin{aligned} \eta_{it} &= I_i^{\eta} + S_i^{\eta}\tau_{it} + \delta_{it}^{\eta}, \\ z_{it} &= I_i^{z} + S_i^{z}\tau_{it} + \delta_{it}^{z}, \end{aligned} \tag{6}$$

$$\begin{aligned} S_i^{\eta} &= \beta_{\eta} + \beta_{z\eta}S_i^{z} + \gamma_{x\eta}x_i + \beta_{z\eta}^{I}I_i^{z} + \zeta_i^{\eta}, \\ S_i^{z} &= \beta_{z} + \beta_{xz}x_i + \beta_{\eta z}^{I}I_i^{\eta} + \zeta_i^{z}, \\ I_i^{z} &= \beta_{I^z} + \beta_{xI^z}x_i + \zeta_i^{I^z}, \\ I_i^{\eta} &= \beta_{xI^{\eta}}x_i + \beta_{I^zI^{\eta}}I_i^{z} + \zeta_i^{I^{\eta}} \end{aligned} \tag{7}$$

Note that there is no intercept for the structural equation with I_i^{η} as the outcome due to the identifiability problem in the latent growth model for the latent outcome (see Appendix A). Also note that since Item 1 overlaps with z_{it}, the structural equation for Item 1 includes the DIF term for the parameter for the DIF effect γ_y. We have assumed that this measurement structure and DIF relationship is time invariant, otherwise we might have different equations in (5) at each time point. For notation not yet discussed, the equation for *Item 1* includes a time invariant intercept μ_1, a time invariant factor loading λ_1 and corresponding random disturbance terms for each individual $e_i^{Item\,1}$ and time $\varepsilon_{it}^{Item\,1}$. Similar notation is seen in the equations for *Item 2* and *Item 3*. This model can be viewed as a multilevel model clustering on the individual, where the latent intercepts and slopes only live on the between-level. To implement this model in software such as MPlus, we describe how to re-express the model in a similar multilevel long format (Appendix B).

Given our parametric assumptions, we can use a maximum likelihood estimator with robust standard errors such as the MLR option in MPlus [34] to estimate the free model parameters. We use the more robust estimator given that our theory is developed for handling scales such as the PHQ-9 that have the potential for skewness. MPlus effectively handles ignorable missing data dependent on the data in hand (i.e., following a "missing at random" assumption) via Full Information Maximum Likelihood (FIML). As a result, respondents with missing data can still be included in the trajectory analyses for unbiased inference [11,12].

Although our example is for linear structural equations, these equations could be easily extended for categorical variables. For example, if Items 1, 2 and 3 are categorical, an appropriate estimator could be used such as weighted least squares [34,37].

These structural equations can be subdivided into three parts: (5) measurement model with an observed covariate and DIF path, (6) latent growth models and (7) mediation model. We assume linear growth in these equations for simplicity, while extending these equations for nonlinear growth will be shown in our MS-depression application. In addition, given the independence assumptions discussed in the section Latent growth modeling.

$$\begin{aligned} \text{var}(\eta_{it}) &= var(I_i^{\eta} + S_i^{\eta}\tau_{it} + \delta_{it}^{\eta}) \\ &= var(I_i^{\eta}) + \tau_{it}^2 var(S_i^{\eta}) + var(\delta_{it}^{\eta}) + 2\tau_{it}\,\text{cov}(I_i^{\eta}, S_i^{\eta}) \\ \text{var}(z_{it}) &= var(I_i^{z}) + \tau_{it}^2 var(S_i^{z}) + var(\delta_{it}^{z}) + 2\tau_{it}\,\text{cov}(I_i^{z}, S_i^{z}) \end{aligned} \tag{8}$$

where var and cov refer to the asymptotic covariance matrix of the parameters.

Assessing mediation effect in longitudinal SEM accounting for overlapping symptoms of co-occurring conditions

In assessing mediation for co-occurring conditions, primary interest may be in testing if growth in the slope in the mediator influences growth in the slope of the outcome [3,4]. As recommended by Baron and Kenny [1], formal testing of a mediation hypothesis involves checking if the estimates for both paths of an indirect effect, corresponding to the estimates $\hat{\beta}_{xz}$ and $\hat{\beta}_{z\eta}$ of β_{xz} and $\beta_{z\eta}$ in equation (7), are statistically significant. Given that these causal paths are statistically significant, a formal test for the longitudinal mediated effect of the estimated parameters can be performed [3,4,42,43]. A currently popular approach to assessing mediation is to bootstrap confidence intervals (percentile, bias-corrected, and bias-corrected and accelerated) for total and specific indirect effects [44]. Given the potential complexities of the proposed modeling approach for complex longitudinal clinical data (i.e. individually time varying repeated measures in a multilevel model with random effects), we chose to use the Monte Carlo confidence interval method instead of the bootstrap method. The Monte Carlo approach has been shown to perform comparably to other widely accepted methods of interval construction such as bootstrapping [43].

Here we provide a sketch of the Monte Carlo confidence interval method for calculating the indirect effect and confidence intervals for assessing mediation [43]. If we assume joint normality for the estimates, $\hat{\beta}_{xz}$ and $\hat{\beta}_{z\eta}$ and given the covariance between $\hat{\beta}_{xz}$ and $\hat{\beta}_{z\eta}$ will be nonzero in this latent variable model [3], then the sampling distribution for

$$\begin{bmatrix} \beta_{xz}^{*} \\ \beta_{z\eta}^{*} \end{bmatrix} \sim MVN\left(\begin{bmatrix} \hat{\beta}_{xz} \\ \hat{\beta}_{z\eta} \end{bmatrix}, \begin{bmatrix} \text{var}\left(\hat{\beta}_{xz}\right) & \text{cov}\left(\hat{\beta}_{xz}, \hat{\beta}_{z\eta}\right) \\ \text{cov}\left(\hat{\beta}_{xz}, \hat{\beta}_{z\eta}\right) & \text{var}\left(\hat{\beta}_{z\eta}\right) \end{bmatrix} \right), \tag{9}$$

where var and cov refer to the asymptotic covariance matrix of the parameters [43]. A sampling distribution of the indirect effect $\hat{\beta}_{xz}\hat{\beta}_{z\eta}$

is formed by repeatedly generating β^*_{xz} and $\beta^*_{z\eta}$ *(i.e. 10000000 times)* and computing their product. Values for $\hat{\beta}_{xz}$ and $\hat{\beta}_{z\eta}$ can be generated easily by writing a function in a statistical software program such as R [36] for the algorithm using pseudorandom number generation. Large sample assumptions are invoked for the distribution of $\hat{\beta}_{xz}$ and $\hat{\beta}_{z\eta}$, but no assumptions are made about the distribution of $\hat{\beta}_{xz}\hat{\beta}_{z\eta}$ conditional on $\hat{\beta}_{xz}$ and $\hat{\beta}_{z\eta}$. Percentiles of this sampling distribution (i.e. 5th and 95th) are identified to serve as limits for a 100(1 - α) % asymmetric confidence interval about the sample $\hat{\beta}_{xz}\hat{\beta}_{z\eta}$.

Model fit

As mentioned in the section for Longitudinal SEM accounting for overlapping symptoms of co-occurring conditions, our model parameters for the proposed approach can be estimated in a standard way using maximum likelihood estimator with robust standard errors such as the MLR option in MPlus [34]. However, standard definitions do not apply for the saturated and null models due to the overall model complexity in our proposed approach (i.e. individually varying follow up times, irregular follow up, missingness, multiple latent variables and latent growth models). As a result, many useful model fit indices, such as Root Mean Square Error of Approximation (RMSEA) and Comparative Fit Index (CFI) [45,46]), won't be output directly as in a simpler SEM in MPlus. MPlus output will provide for this type of model the number of free parameters, log-likelihood statistic (with a MLR correction factor) and information criteria (i.e. BIC, AIC, sample size adjusted BIC). Thus, we propose defining the appropriate saturated and null models given our proposed model and then checking model fit [9,10]. The null model, in our sense, includes all the latent intercepts and slopes along with all observed and latent variables (i.e. x_i, z_{it} and η_{it} in Figure 1), DIF effects and influence of the exogenous variables or covariates (i.e. x_i) on the latent intercepts and slopes, but no correlations or causal paths between latent intercepts and slopes. The saturated model extends the null model to include all pairwise correlations between all latent intercepts and slopes. To derive the chi-square statistic for the proposed model, we can use the log-likelihood statistic, MLR correction factor and degrees of freedom for both the proposed model and saturated model to perform a chi-square difference test [34]. Similarly, we can do such a chi-square difference test between the null and saturated model to derive the chi-square statistic for the null model. Thus, we can use these chi-square statistics along with associated degrees of freedom to calculate the RMSEA and CFI and assess model fit [45,46].

Results

Overlap of a depression screening scale and MS-related disability measures

We now apply the theory developed in the Methods section on the MS-depression example. The theory was presented for the simplified case corresponding to figure 2 for ease of comprehension. However, this type of problem, with a real clinical application in mind, typically will be much more complex. Thus, we showcase how to use the flexibility of this SEM framework to draw statistical inference and test mediation hypotheses for a much more complex example (i.e. multiple predictors, DIF paths, observed mediators and outcomes, nonlinearity).

PHQ-9 items for sleep problems, fatigue, poor concentration and psychomotor problems have been previously theorized [16,17,32] and found to overlap with symptoms described by MS disability scales [6,7]. The measurement properties of a unidimensional depression screening scale latent construct with some correlated residuals from the PHQ-9 were established within this MS population using prior expert-derived theory, model fit criteria, expected parameter changes and modification indices.

Corresponding to the standardized estimates for the cross-sectional analysis for Figure 4, the magnitude of the overlap was found to be large for fatigue (A = 0.21, B = 0.52), medium for cognitive impairment (C = 0.38, D = 0.23) and small for physical disability (E = 0.11, F = 0.00) at baseline [6,7]. These findings are the basis for hypothesizing the overlapping symptoms in longitudinal analysis. Note that the estimates on Figure 4 will be different than the standardized estimates for the cross-sectional analysis in previous work, since these are estimates of our longitudinal SEM investigating how MS-patient specific disease characteristics leads to depression as discussed in this paper.

The path diagram in Figure 4 shows the DIF paths to model the overlap of depressive symptoms with other symptoms for MS patients. The latent construct *depression*, can be interpreted as depressive symptoms adjusted for the overlapping symptoms of both depression and MS. Treating the PHQ-9 items as continuous more readily allows for correlated residuals, which is a key part of the model specification involving the items of overlap (sleep, fatigue, concentration and psychomotor) with MS symptoms. Further, due to sample size and since each of the PHQ-9 items has at least four ordinal categories, the measures can be viewed as approximating an interval scale [47-50]. Treating scales such as the PHQ-9 items as categorical in EFA also introduces threshold parameters (3 for every item for the PHQ-9) which won't be as straightforward to interpret. Thus, we treat the PHQ-9 items as continuous, as was commonly done in prior studies involving the PHQ-9 [47,48]. Note that to examine the potential for bias based on variable distributions we verified results using a mean and variance adjusted weighted least squares estimator

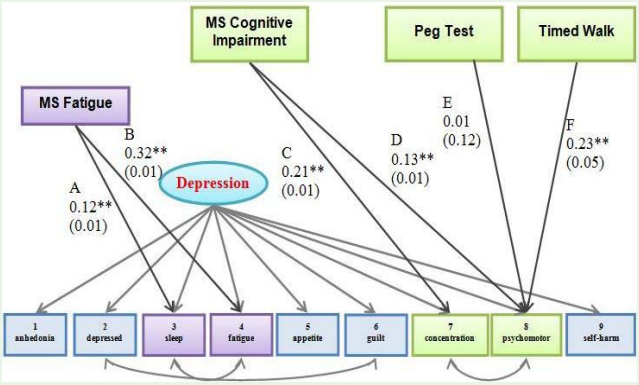

Figure 4: Path diagram for the hypothesized overlap of depressive symptoms with other symptoms for MS patients.
Estimates of the DIF paths on this figure are of the constrained to be equal within-subject level raw point estimate (standard error) and between-subject level raw point estimate (standard error).
Covariates (fatigue, cognitive impairment, timed walk and peg test) are all MS-related measures. In our study, the two correlated MS objective performance measures timed 25-foot walk and 9-hole peg test form a composite of physical disability.
**p ≤ 0.001.

(WLSMV option in MPlus) treating our outcomes (PHQ-9 items) as ordinal categorical measures. We also made the assumption that the measurement model and DIF relationships as shown in Figure 2 were time invariant. This assumption in this type of model with individually time-varying time points and multilevel effects relies mostly on clinical theory. In our example, the measurement model and DIF relationships should not change much over the four year period [6,51-53].

Conceptual model for longitudinal SEM investigating how MS-patient specific disease characteristics leads to depression

The relationships specified for our longitudinal analyses were derived from a priori theory from MS specialists and prior studies [6,32,51,54]. As mentioned above, due to unknown onset of both depression and MS symptoms in our population, we cannot make causal or temporal assumptions about the relationship between changes in MS disability and changes in depression. However, we do assume that the two co-occurring conditions are parallel processes, with the growth of MS disability influencing the growth of depression [4]. Hypothesized relationships using LGM between MS Type, disease duration, changes in functional limitation and changes in depression were evaluated in a prior study [51]. It was found that only functional limitations influenced the growth in depression. In our study we have divided functional limitations into fatigue, cognitive impairment and physical impairment components. Our models included an association between depression and fatigue [54]. Growth in MS-related fatigue are hypothesized to be correlated with growth in depression, but establishing causal and temporal ordering between the two is not clear due to the physical, emotional and mental aspects of both MS-related fatigue and depression [32]. Further, due to these physical aspects of MS-related fatigue, growth in physical impairment can lead to growth in both MS-related fatigue as well as depression. However, there does not seem to be clear prior evidence for making causal assumptions about the relationship between growth in MS-related cognitive impairment and MS-related fatigue. Thus, we omit this relationship. While we were primarily interested in the direct and indirect relationships affecting depression, we simultaneously controlled for the other outcome, fatigue.

Given the sample under study of only MS patients and longitudinal data available for temporal ordering of our growth models we performed confirmatory analyses of our a priori theory, drawing causal inference. We controlled for the effects of age, sex and race in our longitudinal analyses and made an ignorable assumption based on the a priori theory that we did not exclude any important covariate, pathway or interaction effect in our analyses.

While the models were developed for a generalizable theory of the longitudinal relationship between MS disability and depression, the data collected for these analyses occurred over a four year period. Thus, we recognize that this might not be a sufficient time period to observe critical illness periods in the population averages in our measures. Some prior studies over similar time ranges did not observe substantial depression trajectory changes in MS patients using the Center for Epidemiological Studies Depression Scale (CES-D) [51-53].

The analyses using these data, however, can help a clinician understand changes in these relationships over this four year time period. Patients in the sample had been diagnosed with MS for varying degrees of times before the first recorded Mellen Center visit in this database (see Table 1). Thus, for some of the patients in this database observations may represent only one segment of a longer period of overall MS disease progression (Figure 5).

In this longitudinal model, we used individually time varying repeated measures, which ranged from baseline only to 14, (we collapsed any visits for a patient less than 1 month apart) and set each first visit to baseline time zero. We considered the influence of MS patient-specific disease characteristics (baseline time since symptom onset and MS type) on the growth of MS disability (timed walk, peg test and cognitive impairment) and then the influence of the growth of MS disability on the growth of *depression*, while accounting for the role of the growth of fatigue (Figure 5). We stated eight longitudinal hypotheses for confirmatory analyses. Due to the simultaneous nature of the pathways from the MS patient-specific disease characteristics to the growth of *depression*, we did not address any one hypothesis without adjusting for the relationships expressed in the other hypotheses. Since we used *depression* as an outcome rather than the PHQ-9, these causal paths are adjusted for the overlap between depressive symptoms with other symptoms for MS patients as described in the section Overlap of a depression screening scale and MS-related disability measures.

(I) A longer baseline duration since symptom onset and a progressive MS type directly leads to increased growth of *depression*.

(II) A longer baseline duration since symptom onset and a progressive MS type directly leads to increased growth of MS disability, in terms of more physical disability and cognitive impairment.

(III) Greater MS disability directly leads to increased growth of *depression*.

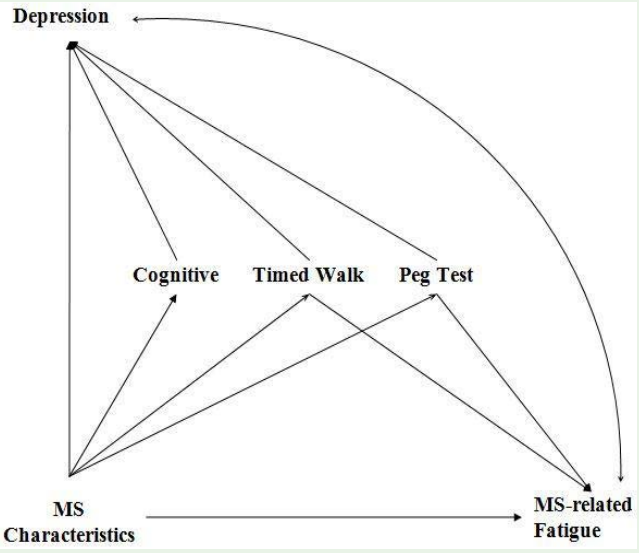

Figure 5: Conceptual model for investigating how MS patient-specific disease characteristics leads to *Depression.*
MS Characteristics consists of two separate measures representing patient-specific disease characteristics: MS type (relapsing or progressive) and baseline time since symptom onset.
While left out of this figure, we specify pairwise correlations between cognitive impairment, timed 25- foot walk and 9-hole peg test.

(IV) A longer baseline duration since symptom onset and a progressive MS type lead to increased growth of *depression* via the mechanism of growth of MS disability. In other words, indirectly, a longer baseline duration since symptom onset and a progressive MS type lead to an increased growth of MS disability which leads to an increased growth of *depression*.

(V) Growth of *depression* is highly correlated with growth of fatigue.

Hypotheses (VI), (VII) and (VIII) are the same as (I), (III) and (IV) but replace *depression* as the outcome measure with fatigue. Similarly, MS disability in (VI), (VII) and (VIII) is only in terms of more physical disability. We were interested in hypothesis (V) in order to understand how the two outcomes relate to each other.

Longitudinal SEM investigating how MS-patient specific disease characteristics leads to depression

Due to complexity of this MS-depression analyses as described above we expressed the hypothesized model (Appendix C) and coded the model (see Appendix D for MPlus code) using the multilevel modeling framework in MPlus. Appendix B explains in a simpler scenario why the coded model is similar to the type of model of equations (5), (6) and (7) . We used a maximum likelihood estimator with robust standard errors via the MLR option in MPlus [34] to estimate the free model parameters.

Before combining growth models for different variables and assessing the direct and indirect effects in our longitudinal model, we made adjustments to our model as necessary. We were able to assess the nonlinearity of our measures through examining trajectories and relationships between measures in descriptive plots and running preliminary models as suggested by Cheong *et al.* [4]. First, we examined individual trajectories over time using spaghetti plots on MS disability measures and individual items of the PHQ-9. Further, we evaluated nonparametric smoothing spline plots on these measures to get a sense of the average trajectory over time. A logarithmic transformation was appropriate for the timed walk and peg test because the data is strongly right skewed. We re-ran the model in MPlus after making the appropriate transformations. Then, we tested for nonlinearity of the outcomes and cognitive impairment through evaluating the quadratic growth terms for each of the latent growth models specified for each individual variable [4]. The quadratic growth term was appropriate for depression, and thus we include it in our model. This quadratic growth term for depression in a MS population is also consistent with what previous studies have shown [51].

We centered all measures by grand-mean, for ease of interpretation of the intercepts of the latent slopes in our models as the average overall effect (i.e. average overall effect of *depression* for the latent slope of *depression*). We are controlling for covariates in our mediation model.

Model Fit, DIF analyses and the trajectories of MS disability and latent dimension for depression

The overall model fit for our proposed model was excellent by our criteria (chi-square statistic = 0.026, df = 4, p-value >0.999; RMSEA ≈ 0.00; CFI ≈ 1.00). Since our model is nearly equivalent to the saturated model (see Kline [10] for more on *model equivalence* and *near equivalence*), these model fit results were not surprising.

Table 2: MPlus estimates of the main parameters of interest for the longitudinal SEM for MS-depression example.

	Modeling Overlap (DIF Paths)			Not Modeling Overlap		
	Estimate	SE	P	Estimate	SE	P
Q^D **ON**						
S^C	0.414	0.378	0.273	0.215	0.680	0.752
S^{TW}	-4.273	1.797	0.017	-1.438	4.301	0.738
S^{PT}	10.745	1.742	<0.001	5.781	16.890	0.732
S^D **ON**						
S^C	-1.706	1.163	0.142	-0.781	0.680	0.085
S^{TW}	19.385	5.630	0.001	7.253	4.169	0.082
S^{PT}	-48.373	3.768	<0.001	-28.127	10.969	0.010
S^C **ON**						
Depression^B	0.060	0.025	0.018	0.133	0.025	<0.001
S^{TW} **ON**						
Depression^B	-0.018	0.017	0.307	-0.002	0.012	0.893
S^{PT} **ON**						
Depression^B	0.011	0.020	0.576	0.017	0.018	0.333
S^F **ON**						
S^{TW}	8.285	3.622	0.022	3.633	2.696	0.178
S^{PT}	-18.918	10.911	0.083	-11.087	10.280	0.281
Q^D **ON**						
Duration	0.000	0.007	0.319	0.001	0.011	0.943
Type	0.041	0.107	0.698	0.043	0.136	0.752
I^C	0.077	0.060	0.184	0.073	0.268	0.785
I^{TW}	0.088	0.179	0.626	-0.028	0.096	0.773
I^{PT}	-0.079	0.278	0.778	-0.094	0.431	0.828
S^D **ON**						
Duration	0.002	0.016	0.884	-0.004	0.010	0.717
Type	-0.156	0.480	0.745	-0.175	0.279	0.530
I^C	-0.334	0.193	0.084	-0.324	0.219	0.138
I^{TW}	-0.415	0.769	0.589	0.128	0.529	0.809
I^{PT}	0.254	1.211	0.834	0.314	0.783	0.688
S^C **ON**						
Duration	-0.001	0.010	0.893	-0.002	0.006	0.728
Type	0.033	0.027	0.228	0.032	0.028	0.254
S^{TW} **ON**						
Duration	-0.001	0.025	0.984	0.000	0.015	0.976
Type	0.028	0.020	0.177	0.025	0.013	0.054
I^F	0.007	0.009	0.419	0.000	0.006	0.990
S^{PT} **ON**						
Duration	0.000	0.056	0.999	0.000	0.049	0.998
Type	0.006	0.042	0.892	0.005	0.009	0.603
I^F	-0.016	0.010	0.107	-0.014	0.007	0.047
S^F **ON**						
Duration	0.000	0.007	0.973	-0.003	0.005	0.579
Type	-0.039	0.192	0.839	-0.014	0.122	0.910
I^{TW}	-0.200	0.283	0.479	0.005	0.230	0.982
I^{PT}	-0.131	0.451	0.771	-0.189	0.245	0.440

Depression[B] ON						
Duration	-0.008	0.002	<0.001	-0.007	0.003	0.008
Type	-0.086	0.049	0.082	-0.069	0.048	0.152
I^{C}	0.251	0.018	<0.001	0.229	0.016	<0.001
I^{TW}	0.112	0.069	0.107	0.103	0.058	0.076
I^{PT}	0.767	0.171	<0.001	0.716	0.150	<0.001
I^{F} ON						
Duration	-0.004	0.004	0.382	-0.004	0.004	0.374
Type	-0.334	0.101	0.001	-0.331	0.110	0.002
I^{TW}	0.288	0.147	0.050	0.279	0.137	0.042
I^{PT}	2.060	0.365	<0.001	2.153	0.349	<0.001
I^{C} ON						
Duration	0.009	0.005	0.043	0.010	0.004	0.015
Type	-0.026	0.075	0.724	-0.036	0.077	0.638
I^{TW} ON						
Duration	0.007	0.002	0.001	0.007	0.001	<0.001
Type	0.430	0.033	<0.001	0.434	0.033	<0.001
I^{PT} ON						
Duration	0.004	0.017	0.811	0.004	0.014	0.765
Type	0.195	0.021	<0.001	0.194	0.021	<0.001
Mean						
PHQ-9						
Anhedonia	-2.171	0.436	<0.001	-1.969	0.393	<0.001
Depressed	-1.950	0.389	<0.001	-1.732	0.347	<0.001
Sleep	-1.238	0.360	0.001	-1.241	0.348	<0.001
Fatigue	-0.972	0.366	0.008	-1.231	0.388	0.002
Appetite	-1.749	0.363	<0.001	-1.580	0.327	<0.001
Guilt	-1.985	0.392	<0.001	-1.768	0.349	<0.001
Concentration	-1.698	0.363	<0.001	-1.706	0.349	<0.001
Psychomotor	-1.374	0.298	<0.001	-1.375	0.287	<0.001
Self-Harm	-0.810	0.146	<0.001	-0.719	0.129	<0.001
I^{F}	-4.920	0.934	<0.001	-5.193	0.925	<0.001
I^{C}	1.248	0.020	<0.001	1.248	0.020	<0.001
I^{TW}	1.868	0.008	<0.001	1.868	0.008	<0.001
I^{PT}	3.148	0.006	<0.001	3.146	0.006	<0.001
S^{D}	0.387	3.628	0.915	-0.688	2.213	0.756
Q^{D}	-0.012	0.811	0.988	0.227	1.002	0.821
S^{F}	0.765	1.390	0.582	0.620	0.759	0.414
S^{C}	-0.152	0.071	0.032	-0.333	0.072	<0.001
S^{TW}	0.067	0.035	0.058	0.035	0.026	0.179
S^{PT}	0.014	0.048	0.773	-0.002	0.043	0.961

There were significant DIF relationships between the PHQ-9 items for sleep problems, fatigue, poor concentration and psychomotor problems and symptoms described by our MS disability scales (see Figure 4) with the exception of DIF effect E (peg test → psychomotor). These results indicated that there was important, non-ignorable overlap between depressive symptoms and other symptoms for MS patients over time that we were adjusting for in our longitudinal analyses.

For comparison, Table 2 presents the estimates of the same longitudinal SEM both with and without modeling the overlapping symptoms. To fit the model without specifying overlapping symptoms, we removed all DIF paths from Figure 4. That is, in Figure 4 and Appendix C the corresponding DIF effects A through F were constrained to zero. As shown in Table 2, modeling the symptom overlap made a difference in our estimates of the coefficients. Specifically, the coefficients for the paths which involved depression as an outcome were changed by modeling the overlap, while the coefficients which did not involve depression remained relatively similar.

In the model with DIF paths, the mean of the slope for cognitive impairment ($p = 0.032$) was significantly decreasing, while the timed walk ($p = 0.058$) showed an increasing trend (see Table 2). However, in the model without DIF paths, the only increase/decrease for the mean of the slope was for a decrease in cognitive impairment ($p < 0.001$). Thus, clinicians could have obtained different trajectory information through modeling the overlap.

Given that we defined our clusters at the individual level in both models, we would expect the Intraclass Correlation Coefficients (ICCs) to be reasonably high. Indeed, the ICCs ranged from 0.463 (for PHQ-9 Item for Self-Harm) to 0.878 for the timed walk. Further our multilevel models included an individual level random intercept and slope to model this within-subject level variation. Thus, we were not concerned with further modeling autocorrelation within these measures. There was no significant association between the growth processes of the two outcomes or between baseline MS-related fatigue and baseline *depression* in either model.

MS-depression longitudinal mediation analyses

Since depression is modeled using quadratic growth, there were potentially two indirect effects (paths to the linear and quadratic outcome terms) of interest for testing if growth in the slope in each of the mediators influences growth in the slope of depression. The quadratic slope for depression was linear in terms of the parameters of the mediators and independent variables. Therefore, we could assume a joint distribution in the form of (9) and use the Monte Carlo simulation approach to formally assess any of the indirect effects for mediation.

First, however, we examined for clinically significant relationships of interest before performing mediation analyses. An increase in the growth of the 25 foot-timed walk and decrease in the growth of the peg test lead to an increase in the linear slope of depression (Table 2). A decrease in the growth of the timed walk and increase in the growth of the peg test lead to an increase in the quadratic slope of depression (Table 2). The effects of MS Type and Duration on the linear and quadratic slopes of depression as well as the slope of the timed walk and peg test were not significant (Table 2).

Prior studies have found significant relationships between increasing physical impairment and increasing levels of depression in MS patients [18,51]. However, in prior studies, where overlapping symptoms were not accounted for, the mobility aspect of physical impairment did not lead to changes in depression [18]. Similarly, without modeling the DIF paths in our data, $S^{TW} \rightarrow S^{D}$ was not significant ($p=0.082$). In contrast, including the DIF path lead to a statistically significant relationship ($p=0.001$). Further, without the

Table 3: Assessment of potential mediation effects of interest for MS-Depression example using the Monte Carlo simulation method in model with overlapping symptoms of MS and depression.

Mediation Pathway	Indirect Effect	95% Lower CI	95% Upper CI	p
Duration→*Depression*^{B} →S^{C}	-0.0005	-0.0010	-0.0002	0.018
Duration→I^{C}→*Depression*^{B}	0.0023	0.0000	0.0040	0.050
Duration→I^{TW}→I^{F}	0.0020	0.0000	0.0039	0.051
Type→I^{PT}→*Depression*^{B}	0.1495	0.0810	0.2043	<0.001
Type→I^{PT}→I^{F}	0.4014	0.2500	0.5235	<0.001
Type→I^{TW}→I^{F}	0.1237	0.0002	0.2157	0.050

DIF paths neither of the quadratic terms for growth in depression on growth of the timed walk or peg test was significant.

There were other inconsistencies in comparing between the models with and without DIF paths. When modeling overlapping symptoms, an increase in the growth of the 25 foot-timed walk lead to an increase in the growth of MS-related fatigue (S^{TW}→S^{F}), but this path was not significant in the model without the DIF paths.

In Table 3 we evaluated the indirect effects in which both individual causal paths involved in the point estimate were significant for our model including the DIF paths (Table 2). While it was nearly a significant mediation effect for Type→I^{TW}→I^{F} after accounting for overlapping symptoms (p = 0.051), this mediation effect was significant in the model not accounting for overlapping symptoms (Indirect Effect [IE] = 0.002, 95% Confidence Interval [CI] = 0.000, 0.003, p = 0.042). Also, in the model not accounting for overlapping symptoms an additional indirect effect was identified where both causal paths were significant for Type→I^{F}→S^{PT}. However, this mediation effect was only nearly significant (IE = 0.004, 95% CI = 0.000, 0.008, p = 0.056).

There were mediation effects as listed in Table 3 that were statistically significant. A longer duration led to increased depression at baseline which led to increased growth in cognitive impairment. The direct effect was not significant in this case. On average, patients had been diagnosed with MS in this population for more than 10 years at baseline. Thus, the cross-sectional information contained in the intercepts may hold important information for clinicians about the relationship between MS progression and depression. A longer duration led to increased cognitive impairment at baseline which led to increased depression at baseline. Also, directly, a longer duration led to increased depression at baseline. More progressive forms of MS lead to higher scores of the peg test at baseline. A higher peg test score leads to higher depression and fatigue at baseline. More progressive forms of MS also led to a longer timed walk at baseline which led to higher fatigue at baseline. The direct effect for MS type at baseline to fatigue at baseline was significant. However, the direct effect indicated that relapsing forms of MS lead to higher fatigue at baseline.

Discussion

Clinical implications of the MS-depression longitudinal mediation analyses

In a clinical sense, understanding the results of our SEM-based analyses may provide the mechanisms of changes in these relationships over a four year period. Clinicians would potentially draw different conclusions when using such a longitudinal SEM approach if they do not account for overlapping symptoms. All our DIF paths under study were statistically significant.

We reported that growth in mobility and hand function influenced growth in depression over this time period. However, given the nonlinear nature of depression in our model, we should further evaluate these relationships over a longer period of time. Growth in leg function also lead to an increase in the growth of MS-related fatigue.

Baseline measures of the mediators were potential mechanisms of change for baseline measures of the depression and MS-related fatigue. Clinicians might consider high levels of these measures (cognitive and physical impairment) when tailoring care in patients with high levels of depressive symptoms or fatigue.

Patients in this population on average did not show mean trajectory changes in depressive symptoms and MS-related fatigue over this four year time period whether the overlap was modeled or not. This result is consistent with prior studies over similar time periods [51,52]. Thus, the mediation analysis results are important for clinicians to consider in determining a treatment plan for managing MS and depressive symptoms simultaneously for more individualized care.

The cognitive impairment findings warrant further validation since they involve self-report cognitive impairment, which may fundamentally be associated with self-report depression. For example, the Paced Auditory Serial Addition Test, representing cognitive function [28-31], was not available in our database. However, using such a measure could help internally validate our cognitive impairment findings.

These longitudinal findings over a four year time period are surely a reflection of the unpredictable and heterogeneous nature of disability accumulation in MS. The field of MS is lacking simple objective measures of disease severity which might more accurately segregate patients with earlier but more severe disease from those with later but milder disease.

There was no direct relationship between MS patient-specific disease characteristics (i.e. baseline time since symptom onset and MS type) and the slopes of depressive symptoms (quadratic or linear) or MS-related fatigue. This was also consistent with prior studies [51]. Since depression and fatigue may occur early in and throughout MS, this finding may be in line with the nature of the coincident conditions.

Conclusions

A longitudinal structural equation modeling approach was developed to perform confirmatory analyses of a priori theory to model multiple growth trajectories of multiple measures of co-occurring conditions simultaneously. The overlap of symptoms of the co-occurring conditions has been accounted for in the model removing that source of bias [7]. In the MS-depression application it was demonstrated that not adjusting for this overlap can lead to different results. We developed the model for only one latent outcome, with all other variables observed. Extension to multiple latent variables or for categorical measures can be done using SEM theory but was beyond the scope of this paper.

Other approaches to mediation analysis, such as latent difference score or autoregressive modeling [3], could have been used depending on the temporal assumptions between the co-occurring conditions. We discussed an extension of the parallel process LGM approach (combining features of MIMIC modeling) for co-occurring conditions when the time of the onset of particular symptoms was not clear. Our approach incorporated features of a complex longitudinal clinical data, such as accounting for individually varying follow-up visit times, irregular follow-up, missingness and the measurement error involved with modeling patient reported outcomes and time-varying covariates, for real time clinical use. A limitation, however, of the proposed approach is due to the overall model complexity and random effects interpretation of model estimates of DIF effects and model fit indices is not as straightforward as a simpler SEM.

The modeling approach is extremely flexible and was motivated by the real clinical application to disentangle some of the complexities of disease symptom etiology, given co-occurring conditions, and support treatment decision making. In the MS-depression application, we evaluated the potential mediating roles of MS disability measures. Depressive symptoms overlap with other symptoms of MS patients such as fatigue, cognitive impairment and physical disability.

The MIMIC modeling type of approach for modeling overlap has several advantages within the context of our study [15]. We can express *depression* as a latent variable, which is likely more accurate for this MS population than the summed PHQ-9 and adjusted for the overlap with other symptoms for MS. Further, within *depression*, multiple factors or correlated items are easily modeled. A more robust estimator, MLR, is readily available in MPlus for valid inference with the skewed PHQ-9 items within these analyses [34]. Covariates, such as the PS domains and objective performance measures in this study, may be treated as continuous.

While the longitudinal model used in this MS-depression study example has been developed via a priori theory for a general MS population, we make an ignorability assumption within our sample under study in drawing causal inference. Given the assumption holds, the study may still have limited external validity outside the Mellen Center population. In future work, our applied models will require further validation using other populations with varied symptom patters and perhaps data from alternate measures and scales such as the Hamilton Rating Scale for Depression [55].

We note that even if we did have available other measures related to depression that are already validated for use in MS, validation of measures in general would be complex and is the subject of future research. To the best of our knowledge, no prior study besides Gunzler *et al.* [6,7] has identified any scale or diagnoses related to depression that corrects for overlap with other symptoms for MS patients. The PHQ-9 had previously been validated for use in MS patients [56]. Along this line, any measure previously considered a valid measure for depression screening in MS should be evaluated in the future for differential item functioning before considering it valid in MS patients. Alternatively, a new study involving clinical diagnostic interview of MS patients by psychiatric experts familiar with MS symptom presentation would provide another benchmark for diagnostic comparison of the validity of self-report scoring approaches. Perhaps most importantly, the proposed flexible modeling framework for LGM-DIF is potentially very useful for understanding symptom overlap across a wide variety of other diseases. This modeling technique would be well-suited for those studies that also involve the use of patient self-report measures to make diagnoses and monitor disease progression and response to treatment.

In future work, we will explore the predictive value of *depression* by linking response to particular anti-depressants, and subgroups (multiple heterogeneous trajectories) of MS patients identified via Growth Mixture Modeling (GMM) [57]. Additional studies will be necessary to explore the implementation of these type of models (such as adding a column for latent subgroup in an EHR data base, to be considered along with expert judgment for patient care) for real time clinical use.

Authors' Contribution

DG conceived of the study and participated in its design, developed the methods, performed the data analyses and drafted the manuscript. NM developed the methods, performed the data analyses and helped draft the manuscript. AP participated in the study design, helped develop the methods and helped draft the manuscript. DM participated in the study design and helped with the clinical interpretation of the data analyses for multiple sclerosis patients. SL helped with data cleaning and study design. RB also participated in the study design and helped with the clinical interpretation of the data analyses for multiple sclerosis patients.

Acknowledgment

Financial support for this study was provided by a grant from NIH/NCRR CTSA KL2TR000440 and by a grant from Novartis. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report. We appreciate contributions from Drs. Randall Cebul, Thomas Love, Irene Katzan, and Neal Dawson at the Center for Health Care Research and Policy, Drs. Alex Rae-Grant and Francois Bethoux at the Mellen Center, and Dr. Martha Sajatovic, Departments of Psychiatry and Neurology at Case Western Reserve University School of Medicine.

References

1. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. J Pers Soc Psychol. 1986. 51: 1173-1182.
2. Gunzler D, Chen T, Wu P, Zhang H. Introduction to mediation analysis with structural equation modeling. Shanghai Arch Psychiatry. 2013; 25: 390-394.
3. MacKinnon DP. Introduction to statistical mediation analysis. 2008: Routledge.
4. Cheong J, MacKinnon DP, Khoo ST. Investigation of mediational processes using parallel process latent growth curve modeling. Struct Equ Modeling. 2003. 10: 238-262.
5. Messick S. Validity, in Educational Measurement R.L. Linn, Editor. American Council on Education/Macmillan: New York. 1989; p. 13-103.
6. Gunzler DD, Perzynski A, Morris N, Bermel R, Lewis S, Miller D. Disentangling Multiple Sclerosis and depression: an adjusted depression screening score for patient-centered care. J Behav Med. 2015. 38: 237-250.
7. Gunzler DD, Morris N. A tutorial on structural equation modeling for analysis of overlapping symptoms in co-occurring conditions using MPlus. Stat Med. 2015; 34: 3246-80.
8. Alemayehu D, Cappelleri JC. Conceptual and Analytical Considerations toward the Use of Patient-Reported Outcomes in Personalized Medicine. Am Health Drug Benefits. 2012; 5: 310-317.
9. Bollen KA. Structural equations with latent variables. New York, NY: Wiley. 1989.

10. Kline RB. Principles and practice of structural equation modeling. Guilford press. 2011.

11. Duncan TE, Duncan SC, Strycker LA. An introduction to latent variable growth curve modeling: Concepts, issues, and application. Routledge Academic. 2013.

12. Preacher KJ. Latent growth curve modeling. Sage. 2008.

13. Asparouhov T, Muthen B. Exploratory structural equation modeling. Structural Equation Modeling: A Multidisciplinary Journal. 2009; 16: 397-438.

14. Brown TA. Confirmatory factor analysis for applied research. Guilford Press. 2012.

15. Woods CM, Oltmanns TF, Turkheimer E. Illustration of MIMIC-model DIF testing with the Schedule for Nonadaptive and Adaptive Personality. J Psychopathol Behav Assess. 2009; 31: 320-330.

16. Goldman Consensus Group. The Goldman Consensus statement on depression in multiple sclerosis. Mult Scler. 2005; 11: 328-337.

17. Siegert RJ, Abernethy DA. Depression in multiple sclerosis: a review. J Neurol Neurosurg Psychiatry. 2005; 76: 469-75.

18. Wallin MT, Wilken JA, Turner AP, Williams RM, Kane R. Depression and multiple sclerosis: Review of a lethal combination. J Rehabil Res Dev. 2006; 43: 45-62.

19. Ferrando SJ, Samton J, Mor N, Nicora S, Findler M, Apatoff B. Patient health questionnaire-9 to screen for depression in outpatients with multiple sclerosis. International Journal of MS Care. 2007; 9: 99-103.

20. Cleveland Clinic's Knowledge Program (KP), 2008-2013.

21. Mellen Center for Multiple Sclerosis Treatment and Research, Cleveland Clinic, Neurological Institute. 2013.

22. Blacker D. Psychiatric rating scales. In: Kaplan and Sadock's Comprehensive Textbook of Psychiatry, 8th ed. Edited by Sadock BJ, Sadock VA. Philadelphia, Lippincott Williams & Wilkins. 2005; p. 929-955.

23. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001; 16: 606-613.

24. Pinto-Meza A, Serrano-Blanco A, Peñarrubia MT, Blanco E, Haro JM. Assessing depression in primary care with the PHQ-9: can it be carried out over the telephone? J Gen Intern Med. 2005; 20: 738-742.

25. Marrie RA, Goldman M. Validity of performance scales for disability assessment in multiple sclerosis. Mult Scler. 2007; 13: 1176-1182.

26. Schwartz CE, Vollmer T, Lee H. Reliability and validity of two self-report measures of impairment and disability for MS. North American Research Consortium on Multiple Sclerosis. Neurology. 1999; 52: 63-70.

27. Chamot E, Kister I, Cutter GR. Item response theory-based measure of global disability in multiple sclerosis derived from the Performance Scales and related items. BMC Neurol. 2014; 14: 192.

28. Polman CH, Rudick RA. The Multiple Sclerosis Functional Composite: A clinically meaningful measure of disability. Neurology. 2010; 74: S8-S15.

29. Fischer JS, Rudick RA, Cutter GR, Reingold SC. The Multiple Sclerosis Functional Composite Measure (MSFC): an integrated approach to MS clinical outcome assessment. National MS Society Clinical Outcomes Assessment Task Force. Mult sclera. 1999; 5: 244-250.

30. Rudick R, Antel J, Confavreux C, Cutter G, Ellison G, Fischer J, et al. Clinical outcomes assessment in multiple sclerosis. Ann Neurol. 1996; 40: 469-479.

31. Whitaker JN, McFarland HF, Rudge P, Reingold SC. Outcomes assessment in multiple sclerosis clinical trials: a critical analysis. Mult Scler. 1995; 1: 37-47.

32. Krupp LB. Fatigue in multiple sclerosis: a guide to diagnosis and management. Demos Medical Publishing. 2004.

33. Poser CM, Brinar VV. The nature of multiple sclerosis. Clin Neurol Neurosurg. 2004; 106: 159-171.

34. Muthen LK, Muthen BO. Mplus. The comprehensive modelling program for applied researchers: user's guide, 2012. 5.

35. SAS Institute Inc. SAS/STAT 9.2 user's guide. Cary, NC: SAS Institute Inc. 2008.

36. Venables WN, Smith DM, the R Core Team. An introduction to R. 2002, Network Theory Ltd.

37. Muthen B. A general structural equation model with dichotomous, ordered categorical and continuous latent variable indicators. Psychometrika. 1984; 49: 115-132.

38. Muthen BO. Beyond SEM: General latent variable modeling. Behaviormetrika. 2002; 29: 81-117.

39. Kowalski J, Tu XM. Modern applied U-statistics. John Wiley & Sons. 2008.

40. Gunzler D, Tang W, Lu N, Wu P, Tu XM. A class of distribution-free models for longitudinal mediation analysis. Psychometrika. 2014; 79: 543-568.

41. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. Psychol Methods. 2010; 15: 309-334.

42. Sobel ME. Asymptotic confidence intervals for indirect effects in structural equation models. Sociol Methodol. 1982; 13: 290-312.

43. Preacher KJ, Selig JP. Advantages of Monte Carlo confidence intervals for indirect effects. Communication Methods and Measures. 2012; 6: 77-98.

44. Preacher KJ, Hayes AF. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behavior Research Methods. 2008; 40: 879-891.

45. Kenny DA. Terminology and Basics of SEM. 2011.

46. O'Boyle EH, Williams LJ. Decomposing model fit: Measurement vs. theory in organizational research using latent variables. J Appl Psychol. 2011; 96: 1-12.

47. Hansson M, Chotai J, Nordstöm A, Bodlund O. Comparison of two self-rating scales to detect depression: HADS and PHQ-9. Br J Gen Pract. 2009; 59: e283-288.

48. Huang FY, Chung H, Kroenke K, Delucchi KL, Spitzer RL. Using the Patient Health Questionnaire-9 to Measure Depression among Racially and Ethnically Diverse Primary Care Patients. J Gen Intern Med. 2006; 21: 547-552.

49. Johnson DR, Creech JC. Ordinal measures in multiple indicator models: A simulation study of categorization error. American Sociological Review. 1983; 48; 398-407.

50. Zumbo BD, Zimmerman DW. Is the selection of statistical methods governed by level of measurement? Canadian Psychology. 1993; 34: 390-400.

51. Beal CC, Stuifbergen AK, Brown A. Depression in multiple sclerosis: a longitudinal analysis. Arch Psychiatr Nurs. 2007; 21: 181-191.

52. Koch MW, Patten S, Berzins S, Zhornitsky S, Greenfield J, Wall W, et al. Depression in multiple sclerosis: a long-term longitudinal study. Mult Scler. 2015; 21: 76-82.

53. Radloff LS. The CES-D scale a self-report depression scale for research in the general population. Applied Psychological Measurement. 1977; 1: 385-401.

54. Brown RF, Valpiani EM, Tennant CC, Dunn SM, Sharrock M, Hodgkinson S, et al. Longitudinal assessment of anxiety, depression, and fatigue in people with multiple sclerosis. Psychol Psychother. 2009; 82: 41-56.

55. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry. 1960; 23: 56-62.

56. Sjonnesen K, Berzins S, Fiest KM, M Bulloch AG, Metz LM, Thombs BD, et al. Evaluation of the 9-item Patient Health Questionnaire (PHQ-9) as an assessment instrument for symptoms of depression in patients with multiple sclerosis. Postgrad Med. 2012; 124: 69-77.

57. Muthen B, Muthen LK. Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. Alcohol Clin Exp Res. 2000; 24: 882-891.

Citation: Gunzler DD, Morris N, Perzynski A, Miller D, Lewis S and Bermel RA. Mediation Analysis of Co-occurring Conditions for Complex Longitudinal Clinical Data. SM J Biometrics Biostat. 2017; 2(1): 1004.