STRUCTURAL EQUATION MODELING IN CLINICAL RESEARCH: NEW METHODS AND APPLICATIONS FOR EHR DATA, MULTIPLE SCLEROSIS AND DEPRESSION

DOUGLAS D. GUNZLER, PH.D.
CENTER FOR HEALTH CARE RESEARCH & POLICY
FRIDAY, OCTOBER 28, 2011

Outline

1. Smoking Prevention Program Mediation Process Example
2. Overview of Structural Equation Models
3. Disentangling Multiple Sclerosis & Depression
4. A New Class of Distribution-free Models for Longitudinal Mediation Analysis
5. Conclusion

Smoking Prevention Program

- A small study finds that a cost-effective tobacco prevention program reduces cigarette smoking.

- How does the tobacco prevention program reduce cigarette smoking?

- A cost-effective tobacco prevention program **changes social norms for tobacco use and this change in social norms** leads to a reduction of cigarette smoking.

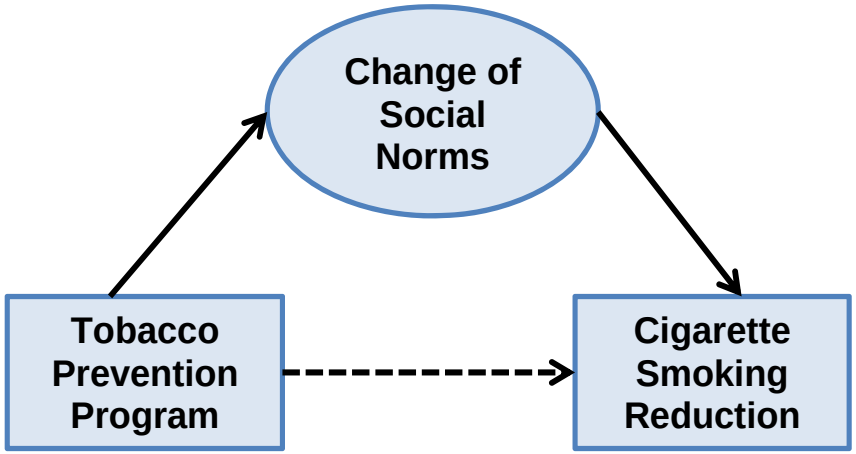
Mediation Analysis

- Evaluates a hypothesized mediation process where an intervening (or middle) measure helps explain the mechanism by which a predictor (such as an intervention) influences an outcome.
- Helps determine focus for future intervention research.
- Can be used to help find surrogate outcomes.
- How does X affect Y?

Mediation Analysis

- Baron and Kenny (1986)
- Causal sequence of events
- **Tobacco prevention program→change of social norms→cigarette smoking reduction**

Path Diagram showing the pathway of a mediation process for tobacco prevention



- Direct vs. indirect effect
- Multiple mediators or outcomes, confounders and latent constructs

Structural Equation Modeling (SEM)

- Very general technique combining complex path analysis with latent (unobserved) constructs.
- simple ANOVA to complex models with observed and unobserved measures, reciprocal relationships and correlations.
- Uses a system of linked regression-style models.
- *Measurement* and *Structural* Components.
- General form for Structural Equation Models in matrix notation.
- Well-suited for analyzing:

1) latent (unobserved) constructs
2) causal relationships between *endogenous* and *exogenous* measures.

SEM "Assumptions"

1) Analytic Component
Statement 1: MS *is associated with* Depression
Statement 2: MS *causes* Depression

Association does not imply causation:
As ice cream sales increase, the rate of drowning deaths increases sharply. Therefore, ice cream causes drowning.

Giving meaning to a grouping of items in a latent construct.
i.e. high stress and low stress groups.

2) Empirical Component
$\text{Depression} = \alpha_0 + \beta_1 \text{MS} + \varepsilon_{\text{Depression}}$

Factor Analysis for Uncovering Latent Constructs

- variations in observed measures reflect the variations in fewer such latent (unobserved) constructs.
- finds simple patterns and connections between multiple measures, compensates for random error, and disentangles complex interrelationships.
- Exploratory Factor Analysis (EFA)
- Confirmatory Factor Analysis (CFA)
- Used in measurement component of SEM framework usually CFA with a unidimensional scale

Test of a Clinical Model of Drinking and Suicidal Risk

- Project MATCH multi-site, randomized clinical trial.
- 1,726 participants of alcohol dependent individuals (ADIs) including 24% women and a mean age of 40.2±11.0 years.
- Pre-treatment interview (baseline) and three reassessments at 3-, 9-, and 15 months.
- Suicidal thoughts and behavior is known to occur during intense drinking bouts.
- Role of drinking in suicidal thoughts and behavior requires the understanding of the interrelationship of drinking with other risk factors.

Test of a Clinical Model of Drinking and Suicidal Risk

Measures Used in Analyses:

- Drinking measures were latent constructs, consisting of three months of drinking data.
- Beck Depression Inventory (BDI) 20-item self-report measure of depression severity .
- Suicidal Ideation was dichotomous (yes or no) measured from BDI and Addiction Severity Index (ASI) scales separately.
- Confounders included age, sex, race (white, nonwhite), treatment assignment (cognitive behavioral treatment, motivational enhancement and 12-step facilitation), and study arm (aftercare, outpatient).

Statistical Analyses:

- Structural Equation Model using a Probit Link with Confirmatory Factor Analysis (CFA) implemented in Mplus.
- Constrained each of the three time points to have the same estimates.

Test of a Clinical Model of Drinking and Suicidal Risk

- Hypothesis of Interest: The direct path from drinking to suicidal ideation is broken by depression (or H_0: direct effect = 0).

- Strategy for Mediation Analysis:

1) Regression model to describe the relationship between drinking and suicidal ideation without mediation.

2) If the relationship changes (as assessed by a substantial change in the magnitude of the standardized slope coefficient relating drinking to SI) after the addition of the mediator, depression, in a structural model, then depression is a mechanism of change.

Test of a Clinical Model of Drinking and Suicidal Risk

- **Results:**

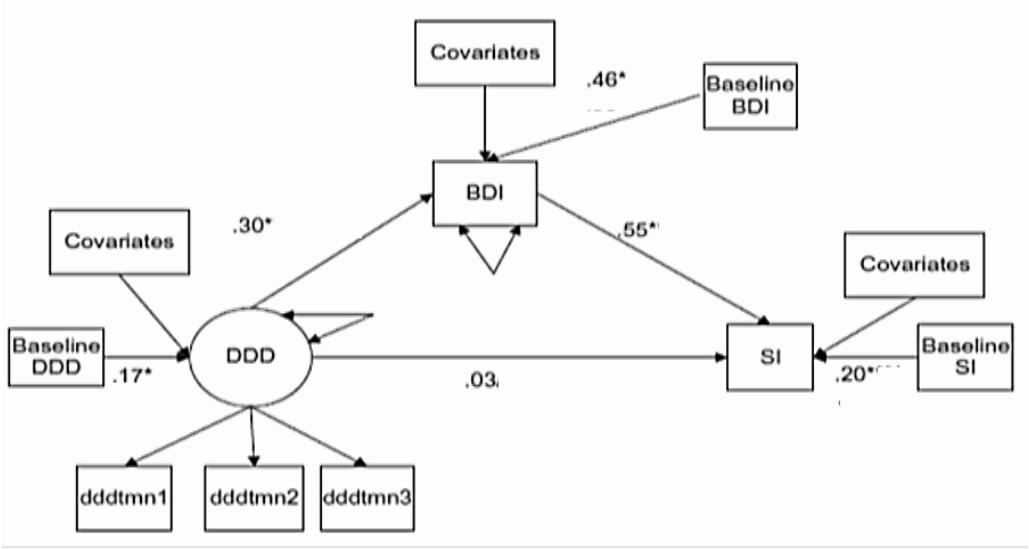
Both higher average drinking intensity (DDD) and greater drinking frequency (PDA) predicted greater likelihood of suicidal thoughts.

BDI-SI outcome resulted in a *full mediation* (p-value in regression model $\leq .001$, mediation model $> .05$)

ASI-SI outcome resulted in a *partial mediation* (p-value in regression model $\leq .001$, mediation model $< .01$).

Test of a Clinical Model of Drinking and Suicidal Risk

For Drinking Intensity → Depression → BDI Measure of Suicidal Ideation



Reported standardized estimates rather than raw estimates (* = significant).

Model fit indices:
Chi-Squared Test p-value < .001
Root Mean Square Error of Approximation = 0.042
Comparative fit index = 0.947
Tucker-Lewis index = 0.933

Knowledge Program (KP)

- At the Cleveland Clinic's Neurological Institute.
- links patient-reported Patient Health Questionnaire (PHQ-9) responses to the EPIC Electronic Health Records (EHR).
- powerful opportunity to study and improve patient care and clinical research.
- Within this neurological population, the KP can be used to assess illness severity and treatment efficacy over time, across a volume of over 140,000 neurological and 1,000 new multiple sclerosis (MS) patients per year.

Multiple Sclerosis

- MS is the most common progressive neurological disease of young adults and affects approximately 400,000 persons in the United States.
- Average age of clinical onset is about 30 years of age.
- The ratios of women to men and white to non-white are both approximately 2:1.
- Depression is the most frequent psychiatric diagnosis in MS patients, with lifetime risk estimated at 50%.
- Patients with MS show increased severity of depressive symptoms compared to patients with other chronic neurological conditions.

PHQ-9

- primary care patients self-report on their health.
- shorter depression scale is used to diagnose, classify and monitor depression and select or modify treatments
- patient specifies the frequency over the past two weeks (0 = not at all to 3 = every day) of nine depressive symptoms, yielding a summated (range: 0–27) total score.
- Scores of 5, 10, 15, and 20 describe validated thresholds for mild, moderate, moderately severe and severe depression.
- PHQ-9 score of 10 or higher has displayed 88% sensitivity and 88% specificity for major depression.
- PHQ-9 has been validated for multiple modes of administration, several clinical (including neurological) populations, and across ethnicities

Common symptoms of MS and Depression

- initial study of its use with MS patients could not reliably relate the PHQ-9 to MS severity without further formal validation.

- roles of MS and depression in patient-reported fatigue and functional disability is not clear.

Can lead to:

1) inappropriate clinical decisions (medication selection, intensification, etc.)
2) false-positive and false-negative inferences about the effectiveness of new treatments.

Disentangling Multiple Sclerosis & Depression

Clinical Research Overview

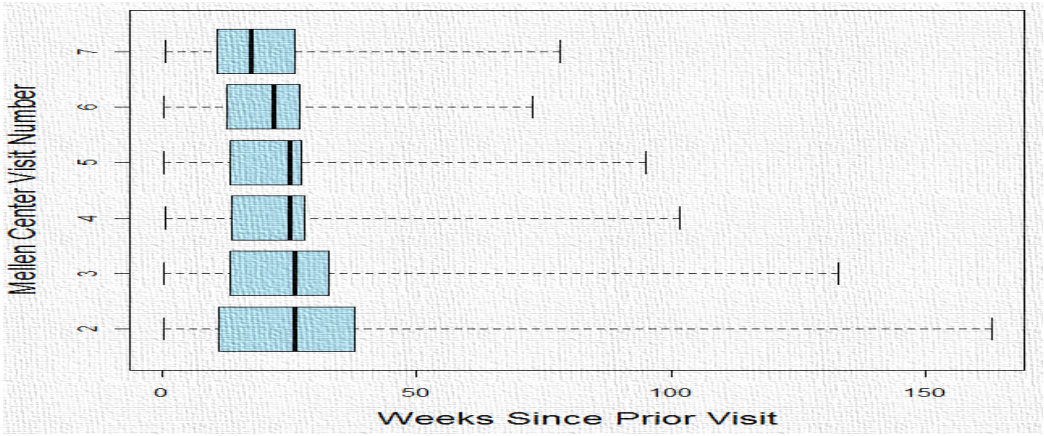
Use the detailed 4-year KP data base to untangle the overlap of fatigue in MS and depression and create a new, MS-adjusted baseline PHQ-9 score. Then, use this new scale to describe causal pathways over time between MS and depression.

Study Aims

1. Correct the PHQ-9 for common symptoms of MS and depression.
2. Causal path for MS leading to depression.
3. Identify subgroups of patients who respond to certain anti-depressants via growth mixture modeling (GMM).

SEM well-suited for the following issues:

- Identify the overlap in symptoms and re-weight the PHQ-9, using a latent variable approach.
- Test hypothesized causal paths for MS leading to depression.
- Irregular follow-up visits in KP data base.



Boxplot of Weeks Since Last Visit

Room for New SEM Methodology for KP data base

- Nearly three-quarters (72%) of patients returned for a second visit in the available data window. A similar drop off pattern was seen through the first seven visits.
- patient visit number, address with zip code of primary residence, location of patient's referring provider and whether the patient had previously seen any physician at the Neurological Institute.
- Missing data reasonably based on observed information. "Missing at Random (MAR)" assumption.
- New SEM methodology necessary to provide unbiased inference under MAR assumption under departures from normality in the context of mediation analysis.
- Comprehensive inference for binary outcomes and mediators using the logit link.

Study Design

- Retrospective Cohort
- Limit study to MS patients
- **Main Sample:** KP subjects with at least one visit to the Mellen Center
- Data are available for 5,671 MS patients from June 2, 2008 up to July 19, 2011 (approximately 95% of these have 1+ PHQ-9 score)

1. Develop MS-adjusted PHQ-9 via SEM

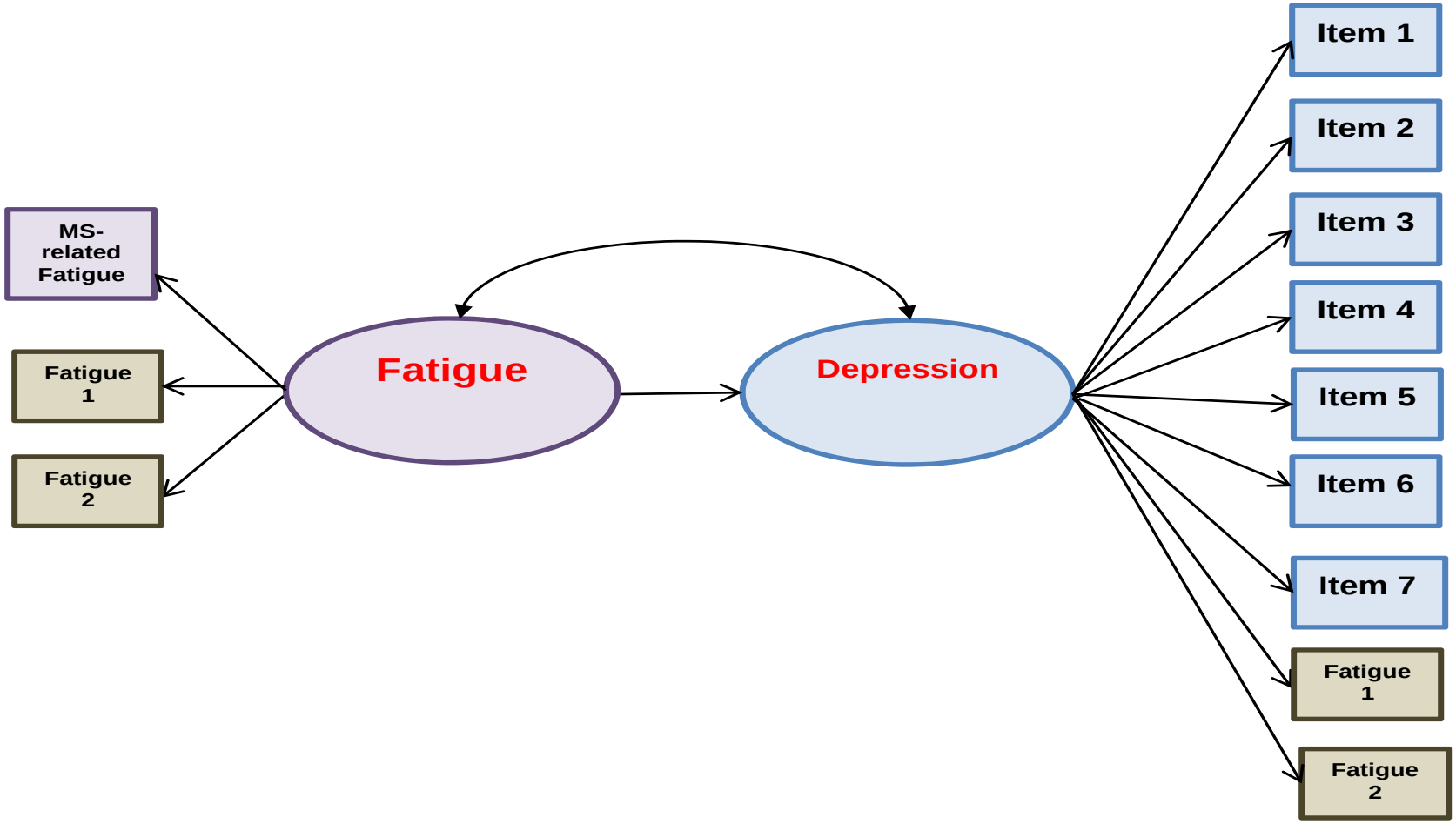
- Accounting for symptoms that are common to depression and MS will result in a significant difference between PHQ-9 scores before and after our SEM-based adjustment.

- Baseline Study

Strategy:

1) Use CFA to create latent variables for Depression and Fatigue, both with two fatigue items from PHQ-9.
2) Re-weight PHQ-9 based on factor loadings, which will reflect the MS-adjusted depression score.

Finding Overlap in MS and PHQ-9 Fatigue



2. Causal Path of MS leading to Depression

- MS variables must account for two concerns:

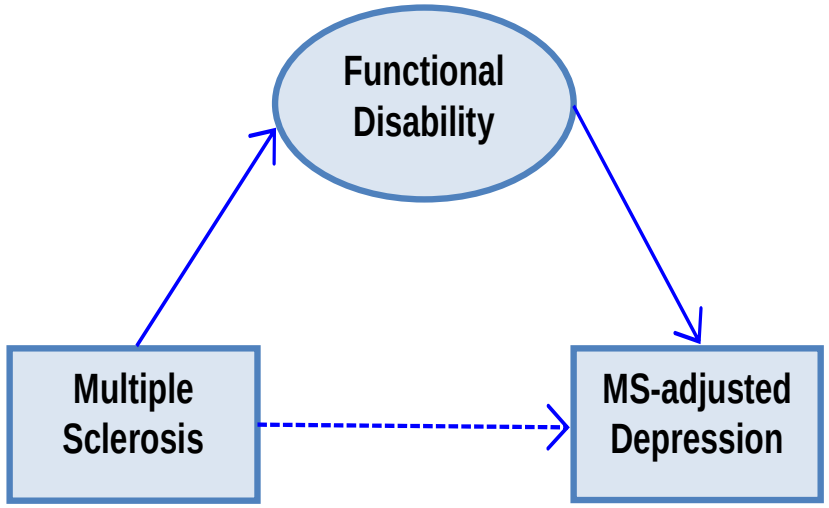
1) baseline time since diagnosis.
2) MS type -- relapsing-remitting, primary progressive, progressive relapsing, clinically isolated syndrome, other categories.

- Functional disability measures:
  timed 25-foot walk, 9-hole peg test, MS performance test

- Longitudinal Study

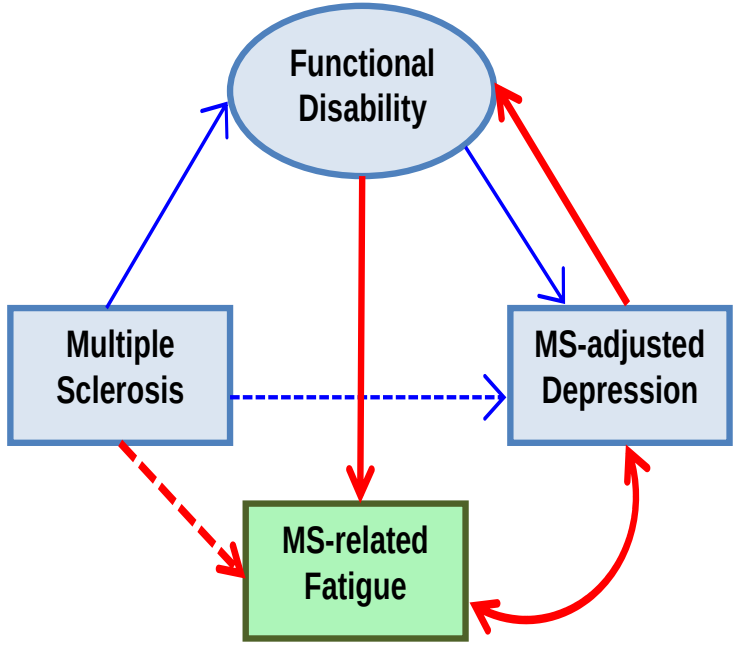
Causal Path of MS leading to Depression

Figure 1. (*Model 1*) Simplified Mediation Model



Oval = latent variable
Rectangle = observed variable

Figure 2. (*Model 2*) Simplified Causal Path adding Fatigue and Reciprocal Relationship



3. Identify subgroups of patients who respond to certain anti-depressants via GMM

Assess and explore the predictive value of the re-weighted PHQ-9 by determining the effectiveness of those anti-depressants in MS patients.

- Latent growth modeling is a longitudinal analysis technique used in the SEM framework to estimate growth over a period of time (i.e. anti-depressant response for depression over the visits).

- Flexibility to use latent constructs and account for irregular follow-up.

- Extension of latent growth modeling in which unobserved heterogeneity can be captured using latent categorical variables (latent classes).

GMM

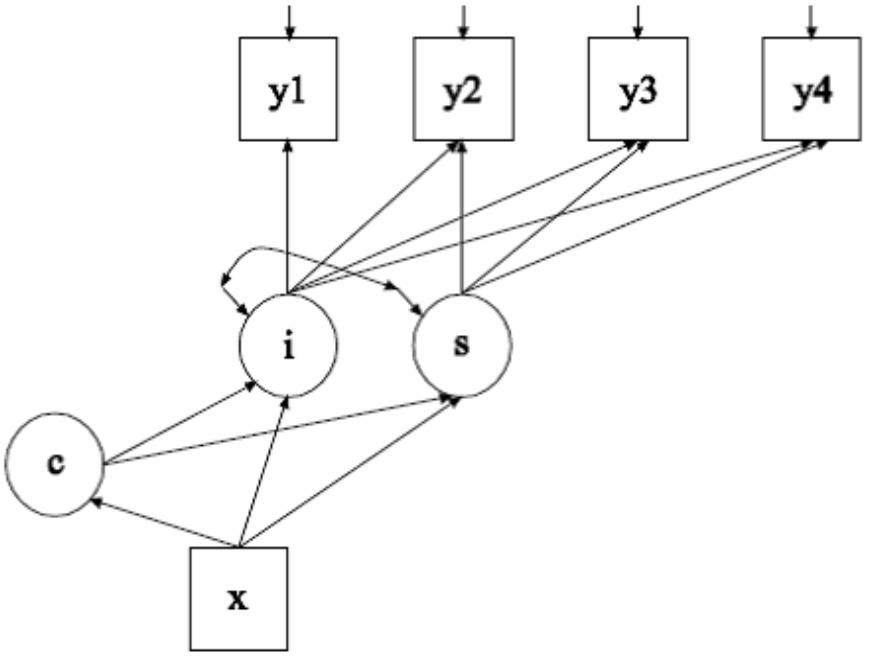
- Latent classes represent sets of endorsed items of the adjusted PHQ-9 within each path.

- To determine the most appropriate number of latent classes:
  Bayesian information criterion (BIC), entropy, percentage of individuals in each class, and uniqueness of trajectory path.

- We then compare trajectories across anti-depressants in each latent class over subsequent visits, to assess changes in our MS-adjusted PHQ-9.

GMM

- y1=MS-adjusted Depression at visit 1 … y4=MS-adjusted Depression at visit 4
- x=anti-depressant medication
- i=intercept
- s=slope
- c=latent classes



Limitations

- Fatigue measure only detailed among patients seen at the spasticity clinic (n=586 at present), where we have a fatigue severity scale.
- EPIC EHR provides only imprecise indicators as to the clinical motivation and starting date for each patient's anti-depressant medications. i.e. anti-depressants are given for smoking or diabetes-related neuropathy.
- Potential selection bias: patients more satisfied with their care or more severely disabled.
- limited external validity outside the Mellen Center population.

Extensions

- Analogous SEM-based methods should transfer cleanly to other subgroups and measures.
  Can also apply to complete KP database i.e. spine disease.

- Explore whether the effectiveness of MS or fatigue interventions are compromised by the overlap of disability and fatigue with depression via under-medication for MS.

- What if fatigue and functional disability appear to play only a minor role in the overlap of MS and depression?
  Other measures of overlap in the KP data may be considered, including pain, unemployment and quality of life.

- Possible RO3?

A New Class of Distribution-free Models for Longitudinal Mediation Analysis

- Joint-normality widely assumed for data in SEM-based analyses.

- Must address missing data to ensure valid inference in the presence of incomplete observations.

- Distribution-free method using SEM framework within the context of longitudinal mediation analysis:

1) Unbiased estimation of parameters under MAR under departures from normality.
2) Comprehensive inference for non-continuous mediators or responses.

Standard methods of Estimation

- Maximum Likelihood (ML) assumes joint normality.
- Generalized Least Squares (GLS) is distribution-free?!?
- Weighted Least Squares (WLS) is distribution-free.

- Must estimate at least the variance-covariance matrix to identify all parameters.

- Only ML doesn't listwise delete missing observations, but biased estimates under MAR assumption under departures from normality.

Functional Response Model (FRM)

1) Provides a single distribution-free framework for modeling first and higher moments with multiple outcomes and extends such models to longitudinal data settings.
2) Addresses the inherent missing data issue under a unified paradigm.

Inference looks similar to Generalized Estimating Equations II (GEE II) and for longitudinal studies with missing data Weighted Generalized Estimating Equations II (WGEE II).

Don't have to estimate nuisance parameters (i.e. variance-covariance matrix) as in ML, GLS and WLS.

Non-continuous Mediators and Outcomes

- Straightforward to extend FRM approach for a comprehensive model.
- Binary outcomes or mediators are very common in real world problems (i.e. cardiovascular disease=yes or no).
- Logit link provides ease of interpretation
- ML estimation requires very complex numerical integration, which is often carried out using ad-hoc methods such as Monte Carlo integration, yielding quite unstable model estimates.
- Three-stage generalized least squares procedure along with variations of the procedure use a probit link.

Simulation Studies Continuous Case

Estimates, standard errors and type I errors based on normal error under complete data with 1000 MC replications						
	Estimate			Standard error		
♦	Method			Method		
	FRM	ML	WLS	FRM	ML	WLS
		GLS			GLS	
Sample size ■50						
☹$_{0}$ ■1	1.014	1.014	1.000	0.192	0.198	0.196
☹$_{xy}$ ■1	0.998	0.998	0.997	0.136	0.142	0.135
☹$_{xy}$ **■1**	1.007	1.007	1.000	0.190	0.200	0.193
☹$_{0}$ ■1	0.997	0.997	1.012	0.137	0.138	0.140
☹$_{xz}$ ■1	0.996	0.996	1.005	0.135	0.141	0.135
☺for H_0: ☹$_{xy}$ ■1				0.050	0.052	0.074
Sample size of 500						
☹$_{0}$ ■1	0.997	0.997	0.997	0.063	0.063	0.063
☹$_{xy}$ ■1	1.000	1.000	1.000	0.044	0.045	0.045
☹$_{xy}$ **■1**	0.998	0.998	1.001	0.063	0.063	0.063
☹$_{0}$ ■1	1.002	1.002	1.002	0.045	0.045	0.045
☹$_{xz}$ ■1	1.001	1.001	0.999	0.044	0.045	0.044
☺for H_0: ☹$_{xy}$ ■1				0.048	0.054	0.049

Missing data (15%/30% at time 2/3) with correlated normal error terms with 1,000 MC replications				
	Estimate		Standard error	
♦	Method		Method	
	FRM	ML	FRM	ML
Sample size ■50				
☹$_{0}$ ■1	1.105	1.557	0.621	0.629
☹$_{xy}$ ■1	1.015	1.053	0.169	0.174
☹$_{xy}$ **■1**	0.968	0.925	0.405	0.410
☹$_{0}$ ■1	1.024	1.228	0.547	0.543
☹$_{xz}$ ■1	1.001	1.050	0.319	0.327
☺for H_0: ☹$_{xy}$ ■1			0.054	0.072
Sample size of 500				
☹$_{0}$ ■1	1.003	1.531	0.235	0.201
☹$_{xy}$ ■1	1.001	1.054	0.067	0.055
☹$_{xy}$ **■1**	1.004	0.946	0.161	0.129
☹$_{0}$ ■1	0.997	1.215	0.181	0.175
☹$_{xz}$ ■1	1.004	1.059	0.110	0.104
☺for H_0: ☹$_{xy}$ ■1			0.045	0.077

Simulation Studies Binary Case

Logit Model for FRM				
with complete data				
	Estimate		Standard error	
♦	Method		Method	
	FRM	ML	FRM	ML
Sample size 50				
0 984	1.156	1.157	0.672	0.691
1	1.190	1.190	0.511	0.535
1	1.204	1.206	0.652	0.688
1	0.994	0.996	0.138	0.139
1	0.996	0.995	0.136	0.142
for H_0: 1			0.064	0.038
Sample size 500				
1	1.016	1.018	0.186	0.186
1	1.016	1.016	0.141	0.141
1	1.010	1.009	0.185	0.186
1	1.002	1.002	0.045	0.045
1	1.000	1.000	0.044	0.045
for H_0: 1			0.049	0.046

Simulation Studies

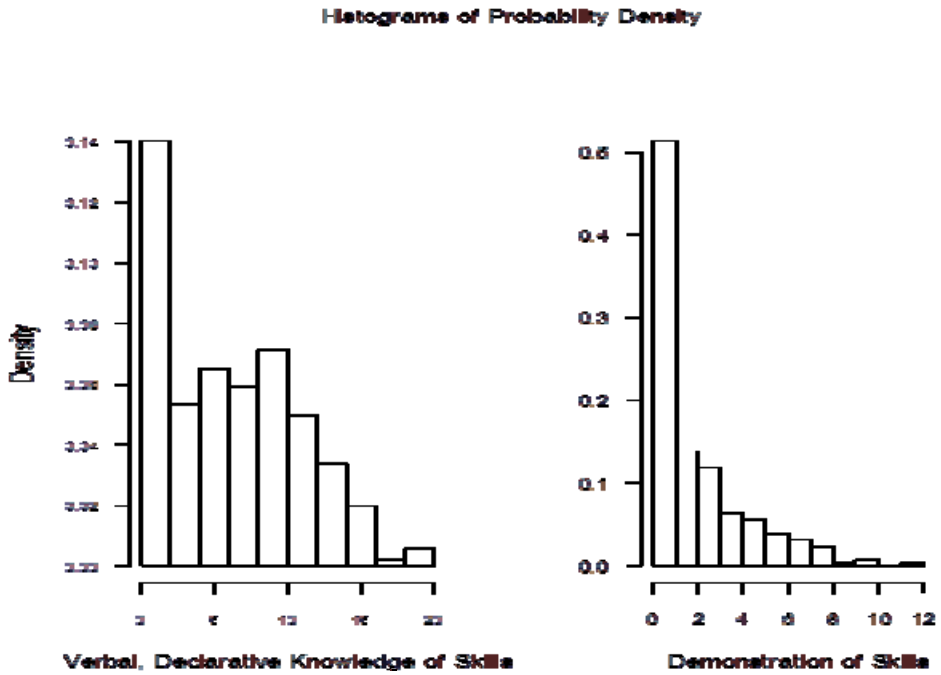
- 1,000 Monte Carlo replications under sample sizes of 50, 100, and 500.
- FRM programmed into R.
- Imported exact same data set into Mplus for ML, GLS and WLS, so results are directly comparable.
- Continuous case: FRM performed comparably well to ML, GLS and WLS with complete data under normal error terms.
- Missing data under MAR, with correlated non-normal errors: FRM performed better than ML.
- Binary outcome, continuous mediator: FRM performed comparably well to ML with complete data under normal error terms.
- Currently, working on simulation for binary outcome, continuous mediator with missing data under MAR.

Child Resilience Project

- 401 students from first up to third grade in five Rochester City School District elementary schools.
- Examines how children with a higher risk of developing behavioral problems with a mentor socially improve compared to the control and lower risk children over periods of 6 and 18 months.
- Both the mediator and response were focused on helping children to manage challenging emotions- emotion self-regulation.
- X=treatment at baseline (mentor or no mentor)
- Z=self-reported verbal, declarative knowledge of the skills the child is learning in the Resilience Project at 6 months
- Y=child's self-initiated demonstration of skills he/she is learning, at 18 months.

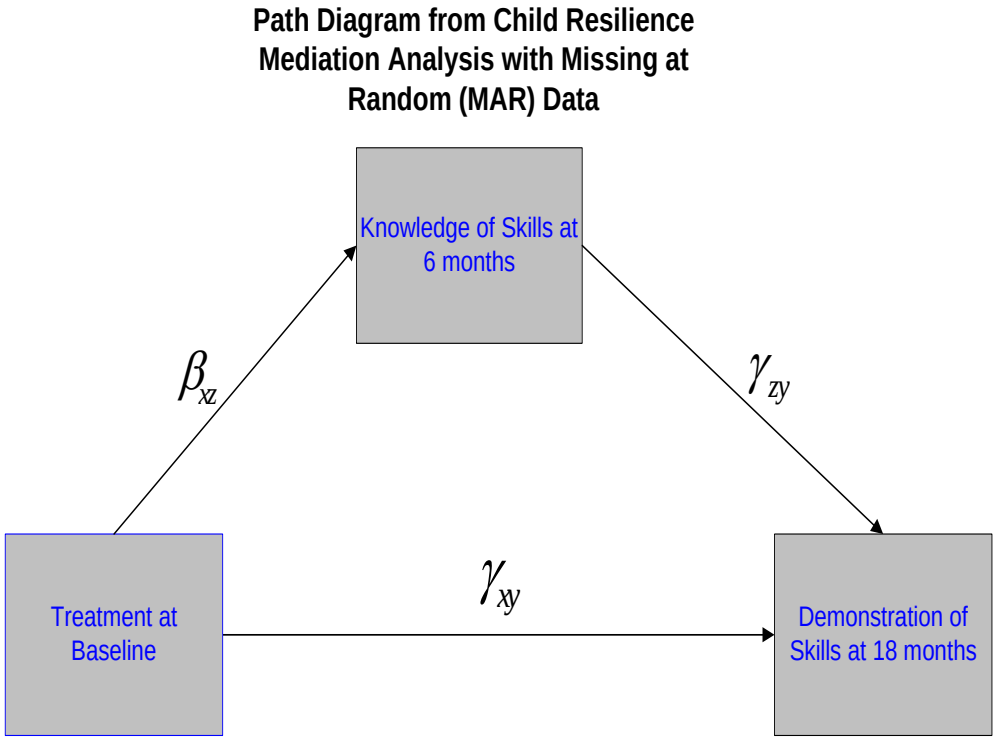
Child Resilience Project

- Highly skewed Z and Y.



- Missing data at 6 months (37%) and 18 months (59%).
- Hypothesize: children receiving a mentor improved their social skills over time given an increase in the mediator.

Child Resilience Project

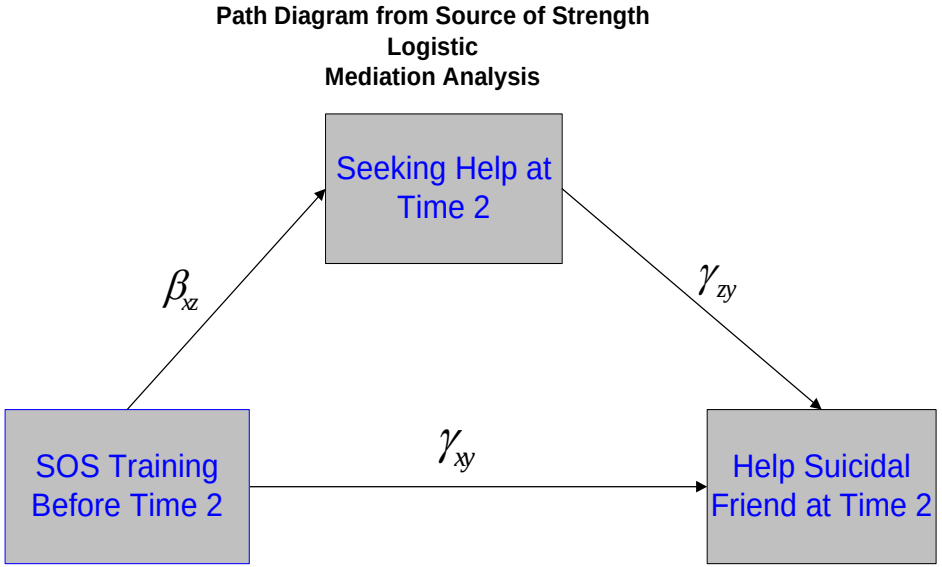


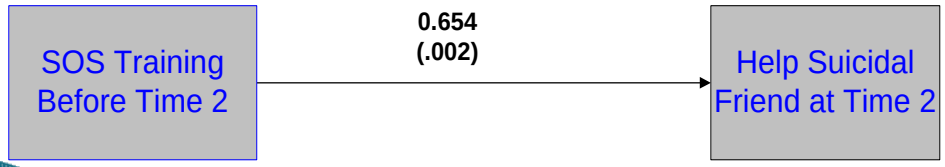
Estimates, standard errors and type I errors				
Child Resilience Study example				
under missing data (37%/59%)				
	Estimate		Standard error	
♦	Method		Method	
	FRM	ML	FRM	ML
Sample size = 401				
γ_0	1.812	1.810	0.278	0.352
γ_{zy}	-0.042	-0.039	0.053	0.050
γ_{xy}	2.330	2.283	0.503	0.480
γ_0	3.429	3.429	0.370	0.374
γ_{xz}	4.390	4.390	0.528	0.529
Type I for H_0: γ_{xy} = 0			.001	.001

Source of Strength Peer Leaders Study with Binary Outcome

- Anonymous survey from 2433 students from 9th through 12th grade with 12 participating rural schools, 8 in New York and 4 in North Dakota, collected over the 2008-2009 school year.
- Examined what type of mentor and available help decreases the risk of suicidal ideation in the students.
- Peer leaders were selected by a team of counselors, adult advisors and teachers within each school.
- Peer leaders were surveyed to assess changes in behavior and attitudes following training and participation in Sources of Strength (SOS) activities occurring between baseline and time 2, approximately 4 months later.
- X=Training in Source of Strength activities before time 2
- Z= Seeking help from adults at time 2
- Y=Telling a suicidal friend to get help from an adult at time 2 (binary indicator with Pr(Y=Yes)=0.286)
- 510 peer leaders with these three measures forming a complete dataset with 321 of the students receiving Source of Strength training before time 2.
- In MPlus used Weighted Least Squares Mean and Variance Adjusted Estimator (WLSMV)

Source of Strength Peer Leaders Study with Binary Outcome





Estimates, standard errors and type I errors				
Source of Strength example under complete data				
	Estimate		Standard error	
♦	Method		Method	
	FRM	WLSMV	FRM	WLSMV
Sample size = 510				
α_0	2.290	1.383	0.547	0.303
α_{zy}	0.318	0.193	0.179	0.098
α_{xy}	0.562	0.329	0.225	0.129
τ_0	2.925	2.925	0.046	0.041
τ_{xz}	0.307	0.307	0.056	0.053
for H_0: α_{xy} = 0			0.013	0.011

Current and Future Research

- Current Research:

Extending methodology for binary outcomes and mediators to missing data.

- Future Research:

Multiple Mediators (Very Straightforward)

Double Robust Method (Robins et al., 1995)

Latent Variables

Not Missing at Random (NMAR)

General SEM

Conclusion

- Structural Equation Models are a general and flexible technique for handling latent constructs and causal pathways, such as a mediation process, in clinical research studies.
- SEM is well-suited to disentangle the relationship between depression and MS using the KP data base.
- A new distribution-free SEM-based approach within the context of longitudinal mediation analysis provides valid inference under MAR and comprehensive inference with a logit link for binary outcomes.

THANK YOU!