

Methods: Mind the Gap
Webinar Series

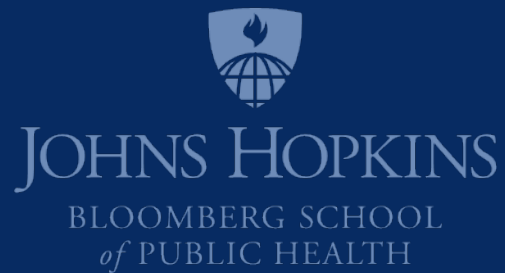
Integrated Data Analysis in Prevention Science



Presented by:

Rashelle J. Musci, Ph.D.
Johns Hopkins Bloomberg School of Public Health

INTEGRATED DATA ANALYSIS IN PREVENTION SCIENCE



Rashelle J. Musci, Ph.D.
Department of Mental
Health

ACKNOWLEDGMENTS



Land Acknowledgement

- This presentation was developed on stolen Nanticoke and Piscataway land.
- <https://native-land.ca/>

Redlining Acknowledgment

- This presentation was developed in a community impacted by redlining and gentrification.
- <https://dsl.richmond.edu/panorama/redlining>

GOALS IN PREVENTION SCIENCE



- Prevention social, physical, mental health, and academic problems.
- Promote health and well-being
- Research, support, and implement evidenced based programming

CURRENT CHALLENGES IN PREVENTION RESEARCH



- Large scale RCTs are expensive and resource heavy, typically designed and powered for main outcome analyses
- Increasingly, the literature is pointing toward an emphasis on mechanisms and subgroups
- Researchers, practitioners, and policy makers are interested in rarer outcomes and exposures

WHAT IS INTEGRATIVE DATA ANALYSIS?



- A set a techniques used to fit models to pooled or harmonized data
 - Allows us to capitalize on between-study heterogeneity
 - Increases power for mediation and moderation analyses
 - Allows for longitudinal models to be used across developmental periods
- To date, most IDA studies within mental health research have focused either on cohort studies or data from clinical trials with psychiatric medication.

TALK OVERVIEW



- Ultimate goal:
 - Integrative data analysis is an important methodological tool in your prevention research toolbox.
- Our motivating example:
 - Early childhood prevention programming and suicidal behaviors

BACKGROUND



- Suicide is a significant public health concern, particularly among black youth (13 and below) who are more likely to die from suicide compared to their white peers.
- Early prevention programming has shown significant impacts on suicidal behaviors across the life course.
- We do expect that universal programming to have some effect heterogeneity.
 - Exploring this can be challenging within study, particularly with rare outcomes

INTRODUCING THE DATA



- Through the use of integrative data analysis (IDA) we harmonize extant school-based cluster randomized prevention trials with longitudinal follow up
 - JHU Center for prevention & early intervention, 1st generation trial
 - JHU Center for prevention & early intervention, 2nd generation trial
 - JHU Center for prevention & early intervention, Paths to PAX trial
 - Schools and Families Educating (SAFE) Children Study
 - Fast Track Project
 - Linking the Interests of Families and Teachers (LIFT) study

DATASET INFORMATION



Table 1. Basic Information about included Datasets

	PIRC G1	PIRC G2	P2P	Fast Track	LIFT	SAFE Children
Prevention Type	Universal	Universal	Universal	Universal + Indicated	Universal	Universal
Intervention type*	Individual Strategies, Skills Training	Individual Strategies, Parent Training	Individual Strategies, Social Emotional Learning	Cognitive Training, Individual Skills Training, Emotional Learning	Behavioral School Strategies, Social Learning, Parent Training	Individual Strategies, Social Emotional Learning, Parent Training
Randomization Level	School Classroom &	Classroom	School	School	School	Classroom
Sample size	2,311	678	5,611	1,199	671	424
Male (%)	50%	53%	51%	69%	49%	51%
White (%)	33%	13%	4%	47%	90%	0%
African Amer (%)	65%	87%	90%	51%	1%	43%
Hispanic (%)	<1%	0%	4%	0%	4%	58%
Other (%)	0%	0%	2%	2%	5%	0%

*Intervention type classification captured based on classifications from Blueprints for Healthy Youth Development

ADDRESSING CURRENT CHALLENGES



- Larger sample size allows us to
 - Study cross-over effects
 - Have power to detect impacts on rare outcomes
 - Explore subgroup analyses not previously possible
- Rich measures collected across the life-course allows us to explore proximal prevention targets to understand mechanisms

A LIFE COURSE PERSPECTIVE



Table 2. Assessments completed for each trial (age in years)

Age	PIRC G1	PIRC G2	P2P	Fast Track	LIFT	SAFE Children
6-7	X	X	X	X	X	X
8-10	X		X	X	X	X
11-14	X	X		X	X	X
15		X		X	X	X
16-18		X		X	X	
19-20	X	X		X	X	
21-22	X	X			X	
23-24		X			X	
25		X		X	X	
30+	X			X		

- Each dataset varies in the number of waves of data collected and timing of the waves

THE CHALLENGE



- Harmonizing data from disparate trials is challenging, particularly with complex study designs and both multilevel and longitudinal modeling

Data Discovery

- Obtaining codebooks
- Collect datasets

Pre-statistical Harmonization

- Create concordance table
- Merge and clean data, confirm response scales
- Identify common data domains

Statistical harmonization

- Harmonize response scales
- Create covariate informed factor scores

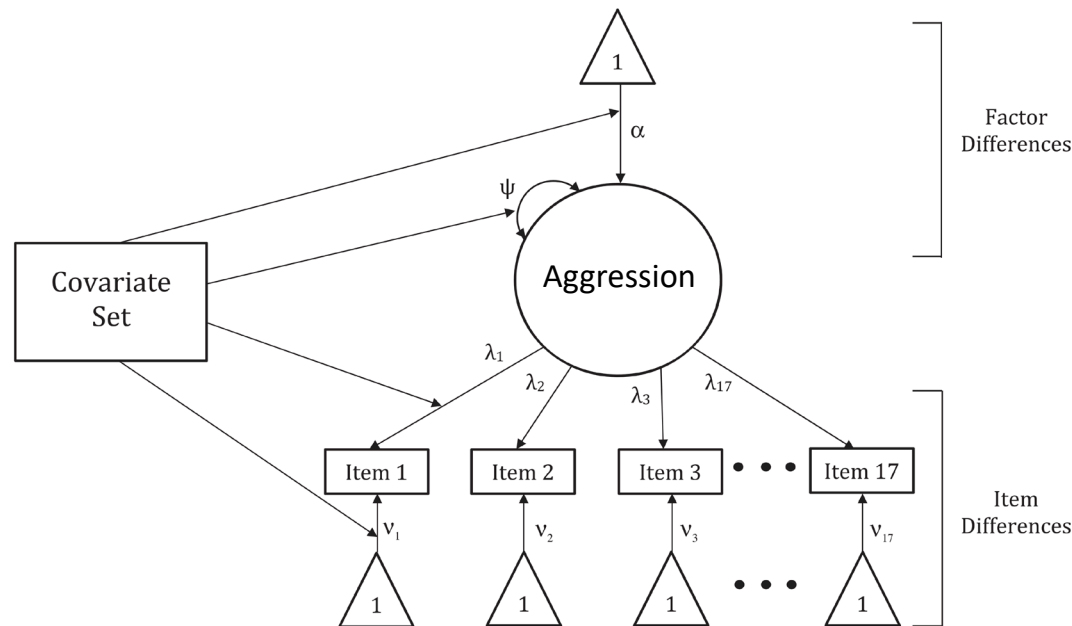
PRE-ANALYTICAL HARMONIZATION

MEASUREMENT INVARIANCE



- Ensure the hypothesized factor being measured is the same under different conditions or across different groups
 - First, a baseline two-group model is established
 - Progressively specify more stringent constraints
 - Conduct a likelihood ratio difference test; If $p > .05$, MI is established
- Typically, when using multi-sample approaches, (a) discrete groups are (b) tested in succession
 - treatment vs. control, old vs. young, females, vs. males
 - Why not tested simultaneously?!

MODERATED NON-LINEAR FACTOR ANALYSIS



- α_j = latent factor mean for individual j
- ψ_j = latent factor variance for individual j
- v_{ij} = intercept for item I for individual j
- λ_{ij} = factor loading for item I for individual j

- Extends the generalized factor analysis framework by allowing the four model parameters to vary as a function of covariates

STEPS OF MNLFA



- Step 1: Graphical and descriptive analyses of items
- Step 2: Dimensionality Testing
- Step 3: Testing for factor and item differences
- Step 4: Estimate scale scores in pooled dataset

STEPS OF MNLFA

Step 1: Graphical and descriptive analyses of items

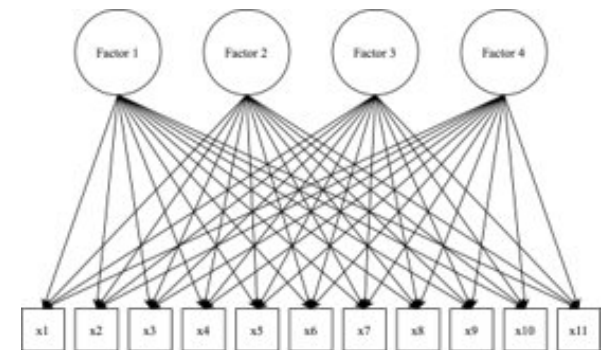
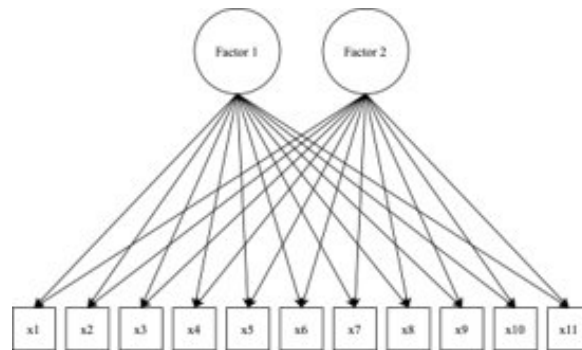
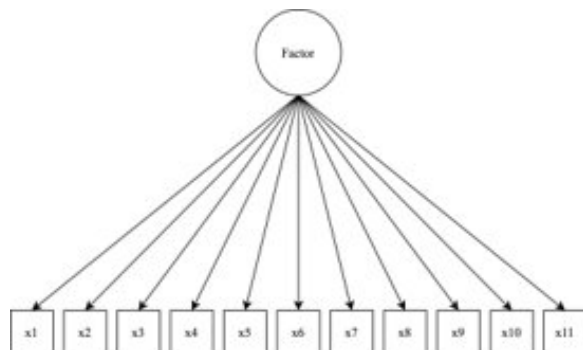
- Considering ordinal responses as binary
 - Item "Yells at others" originally scored along a 6-point Likert scale

Teacher-Yells at others	Freq.	Percent	Cum.
Almost Never	4,504	47.41	47.41
Rarely	2,080	21.89	69.31
Sometimes	1,634	17.20	86.51
Often	618	6.51	93.01
Very Often	421	4.43	97.44
Always	243	2.56	100.00
Total	9,500	100.00	

Teacher-Yells at others	Freq.	Percent	Cum.
No	4,504	47.41	47.41
Yes	4,996	52.59	100.00
Total	9,500	100.00	

STEPS OF MNLFA

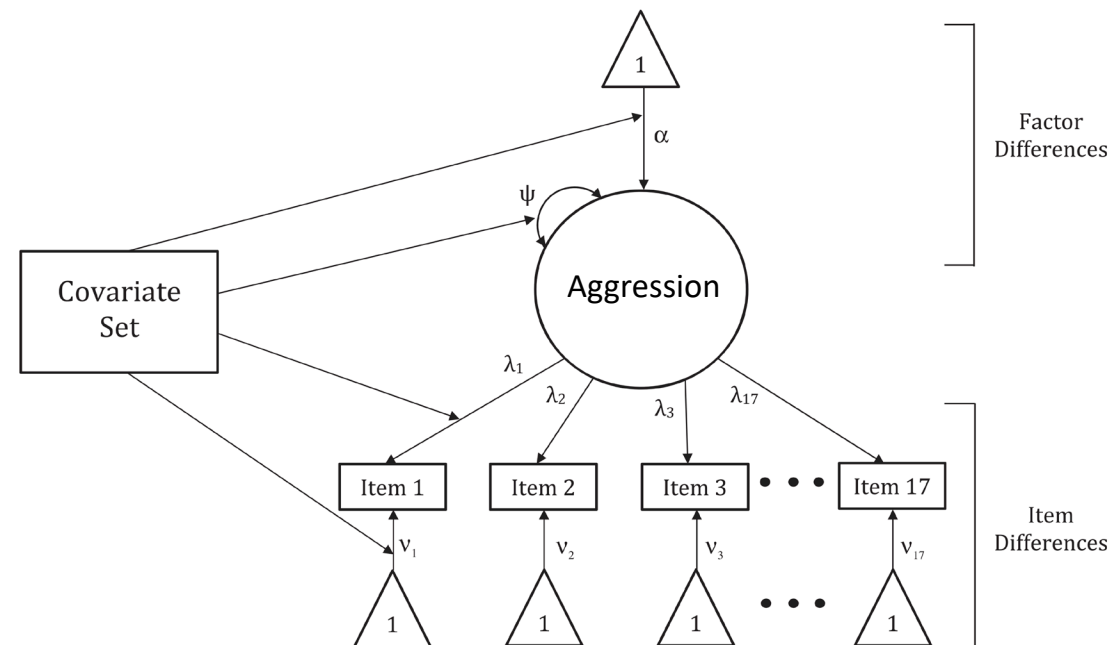
- Step 2: Dimensionality Testing
 - Conduct EFA to determine appropriate number of factors to retain
- Tested 1, 2, 3, and 4 factors



STEPS OF MNLFA



■ Step 3: Testing for factor and item differences



TESTING FOR FACTOR AND ITEM DIFFERENCES



Final MNLFA Model Examining Covariate Effects on Factor Mean and Variance

Covariate Effect	Estimate	<i>SE</i>	<i>p</i>
Factor Mean			
Study	-0.205	0.016	< .001
Sex	-0.579	0.036	< .001
Race	0.586	0.048	< .001
Factor Variance			
Study	0.107	0.018	< .001
Sex	0.081	0.045	.072
Race	0.116	0.060	.052

Note: *SE* = standard error.

TESTING FOR FACTOR AND ITEM DIFFERENCES

Examining DIF in Items Using a Sequential Modelling Building Approach

Model	Covariate	# of Parameters	Log likelihood	Scaling factor	Factor Mean		Factor Variance		Item Intercept		Item Loading	
					Est	<i>p</i>	Est	<i>p</i>	Est	<i>p</i>	Est	<i>p</i>
Baseline		27	-39166.368	1.0401								
	Study				-0.358	< .001	0.186	< .001				
	Sex				-0.912	< .001	0.341	< .001				
	Race				0.778	< .001	0.253	< .001				
X1		34	-38978.743	0.9970								
	Study				-0.181	< .001	0.051	.002	-0.213	< .001	-0.109	.042
	Sex				-0.476	< .001	-0.047	.275	-0.544	.004	0.006	.977
	Race				0.418	< .001	0.117	.047	-0.803	.006	-0.530	.037
X2		33	-39086.742	1.0330								
	Study				-0.357	< .001	0.191	< .001	-0.432	< .001	-0.136	< .001
	Sex				-0.972	< .001	0.356	< .001	0.319	.099	-0.094	.194
	Race				0.754	< .001	0.258	< .001	0.647	.002	0.035	.673
X3		33	-39121.400	1.0566								
	Study				-0.364	< .001	0.205	< .001	-0.586	< .001	-0.213	< .001
	Sex				-0.908	< .001	0.320	< .001	0.114	.479	0.058	.431
	Race				0.786	< .001	0.206	< .001	1.114	< .001	0.348	< .001
X4		33	-39132.380	1.0455								
	Study				-0.368	< .001	0.186	< .001	-0.118	.087	-0.163	.001
	Sex				-0.905	< .001	0.351	< .001	-0.447	.008	-0.043	.706
	Race				0.820	< .001	0.290	< .001	0.932	< .001	-0.313	.047

Note: Baseline model includes moderation on the factor mean and variance only (i.e., no item DIF). Moderation was investigated using three covariates, including study (an indicator variable for the study in which an individual is from), sex (0 = Female, 1 = Male, and race (0 = non-Black, 1 = Black).

STEPS OF MNLFA



- Step 4: Estimate scale scores in pooled dataset
 - Factor scores were estimating using
 - maximum likelihood estimation with robust standard errors and a logit link function.
 - Estimated factor scores can then be used in subsequent analyses

OUTCOME MODELING



$$\text{logit}(\text{Attempt Suicide}) = \beta_0 + \beta_1 \times \text{Male} + \beta_2 \times \text{Black} + \beta_3 \times \text{Intervention} + \beta_4 \times \text{Aggressive} + \beta_5 \times \text{Intervention} * \text{Aggressive}$$

- Those with higher 'aggressive behavior' factor scores were significantly more likely to have attempted suicide (OR = 2.39, $p = .001$)

LIMITATIONS & NEXT STEPS



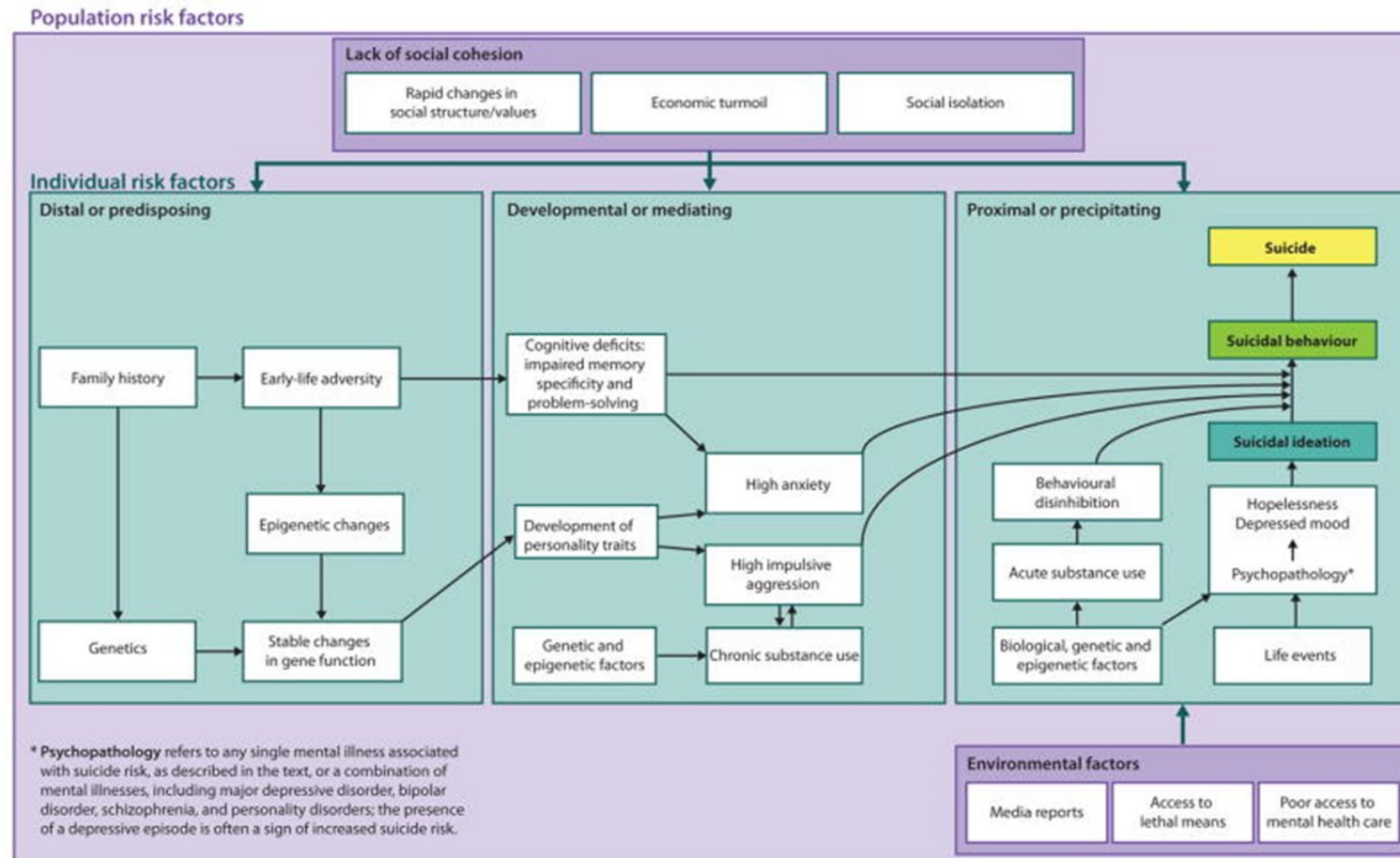
- Inclusion of multi-level model– allowing the intervention to be at level 2, modeling intervention specific effects
- Inclusion of longitudinal modeling
- Improvement of the measurement model around exposure, baseline covariates, and outcomes

CHALLENGES TO IDA



- Principal Investigator participatory research
 - PI's of the original studies should be involved at each step of the data harmonization process
- Just because the data is harmonized doesn't mean its perfect, there are still limitations
- Sometimes there is too much data!
 - Work needs to be theoretically driven

MODEL FOR SUICIDE RISK (TURECKI & BRENT, 2016)

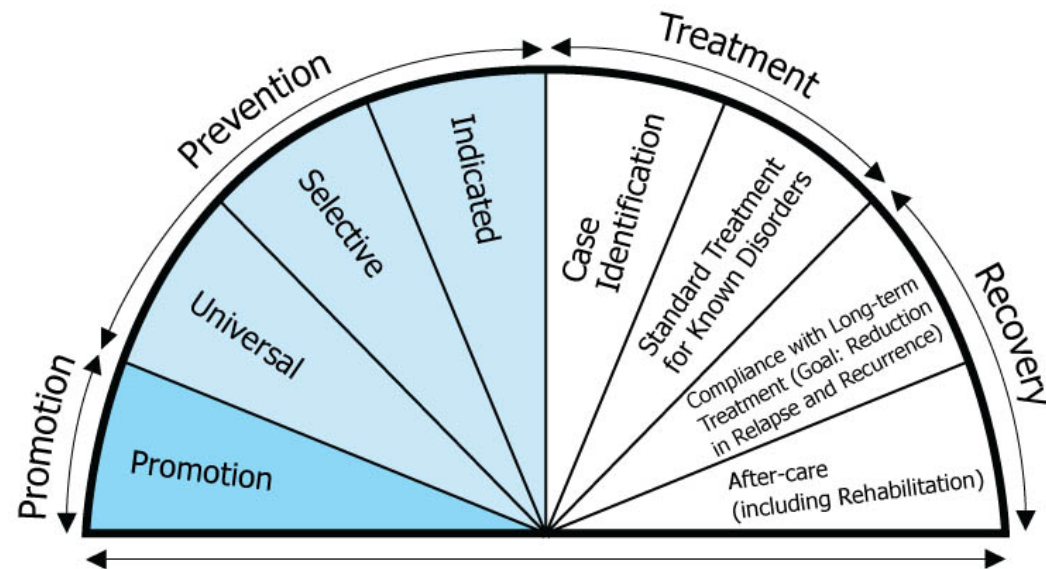


MEDIATORS & MODERATORS



- Aggressive & Disruptive behavior
- Deviant peer affiliation
- Substance misuse
- Sex
- Racial/ethnic group
- Sexual orientation

IMPLICATIONS



- Long-term cross over effects are important to estimate for early childhood prevention programming
- Advances in IDA methodology to account for heterogeneity in behavior and heterogeneity in program impact

ACKNOWLEDGMENTS



- **Holly Wilcox, PhD (co-PI)**
- Ryoko Susukida
- Masoumeh Aminehmaeili
- Katherine Masyn
- Joe Kush

- NIMH ---- MH122214
- Grant support for included datasets:
 - PIRC G1: MH42968, MH38725, MH40859, DA09897, DA004392
 - PIRC G2: MH38725, MH57005, DA11796
 - P2P: R305A080326
 - Fast Track Project: R18MH48043, R18MH50951, R18MH50952, R18MH50953, K05MH00797, and K05MH01027, DA016903, K05DA15226, RC1DA028248, P30DA023026, S184U30002
 - LIFT: MH054248
 - SAFEChildren: MH048248, R01DA020829

PREVENTION SCIENCE SPECIAL ISSUE



- Innovations and Applications of Integrative Data Analysis (IDA) and Related Data Harmonization Procedures in Prevention Science
- Co-editors:
 - Antonio A. Morgan-López (RTI International),
 - Rashelle J. Musci (Johns Hopkins University), &
 - Catherine P. Bradshaw (University of Virginia & Johns Hopkins University)
- LOI due September 1, 2021