



SERVIZIO SANITARIO REGIONALE
EMILIA-ROMAGNA
Azienda Unità Sanitaria Locale di Bologna

Istituto delle Scienze Neurologiche
Istituto di Ricovero e Cura a Carattere Scientifico

Multivariate statistics for secondary analyses

Fabio La Porta, MD PhD

Neuroriabilitazione "Casa dei Risvegli"

UO Medicina Riabilitativa e Neuroriabilitazione (SC)

IRCCS Istituto delle Scienze Neurologiche

AUSL Bologna

1

JPND call for proposals: Understanding the mechanisms of non-pharmacological interventions

SYNOPSIS

Full proposal application

Title:	Steps against the burden of Parkinson's Disease
Acronym:	StepuP
Duration:	36 months
Total funding:	€1,589,883

- **Efficacy of SDTT: hypothesized mechanism**

Decreased beta band oscillations in sensorimotor areas

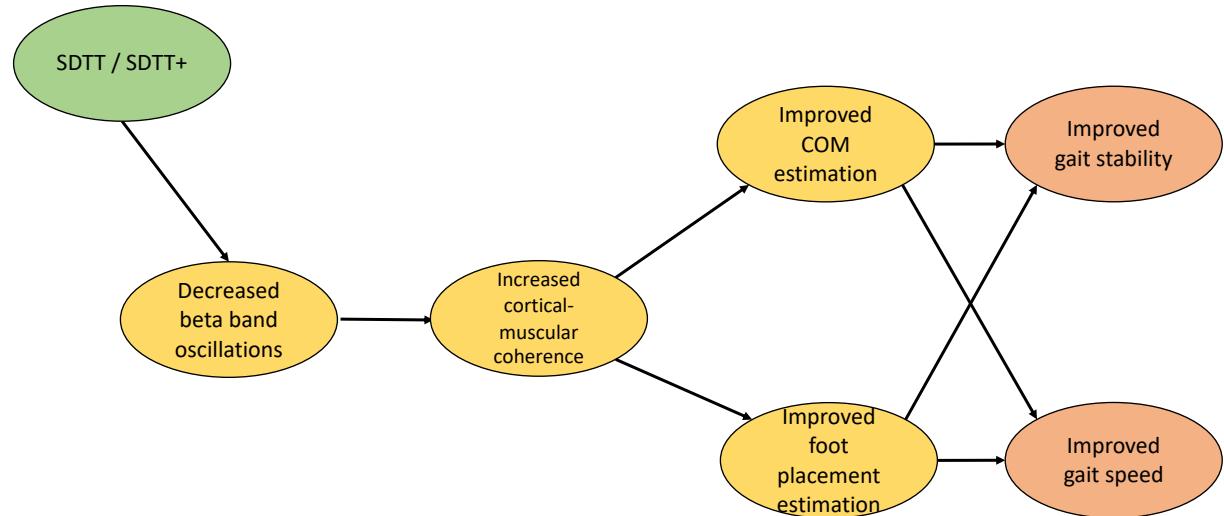
Increased cortical-muscular coherence

Improved center of mass state estimation

Improved foot placement coordination

2

Our starting mechanistic model



How can we demonstrate this hypothesized complex mechanism?

3

Agenda



Standard regression techniques to establish causal nexus

Problems with standard regression techniques

Path analysis and Structural Equation Modeling

How to deal with predictors to save on sample size

Conclusions

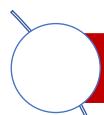
4

A few provisos

- Assessment of causality (mechanisms in the StepuP) can be tricky
- The idea is to provide some suggestions for secondary analyses
- Disclaimer: I am not a statistician!
- I have experience with the statistical and psychometric techniques presented
- No conflicts of interests declared

5

Agenda



Standard regression techniques to establish causal nexus

Problems with standard regression techniques

Path analysis and Structural Equation Modeling

How to deal with predictors to save on sample size



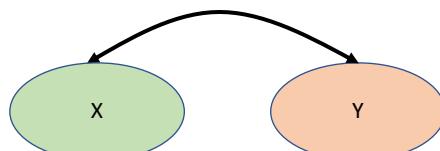
Conclusions

6

Criteria for establishing a causal nexus between two variables

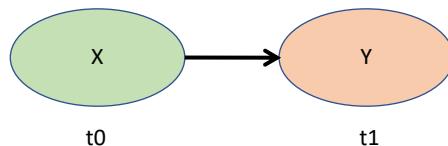
1. Association

There must be a demonstrated association between the putative cause and its effect



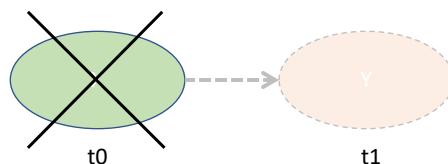
2. Temporality

The putative causal factor must precede the effect



3. Manipulability

The removal of the putative causal factor should remove the effect



7

Regression techniques

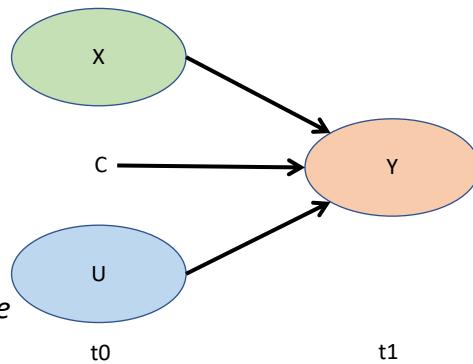
- Set of techniques to estimate the relationship between variables

- Independent variable(s)

*The putative causal factor (x: i.e., the predictor or covariate)
A constant (c)*

- Dependent variable

The effect (y: i.e., the outcome)



- Most common regression models

Linear regression: Y is a linear outcome variable

Logistic regression: Y is a binary outcome variable

8

Regression techniques in RCTs

- Used to assess causal nexus between intervention and outcome

> *Contemp Clin Trials Commun.* 2018 Mar 28:10:80-85. doi: 10.1016/j.conctc.2018.03.008.
eCollection 2018 Jun.

- Logistic regression**

If outcome is binary (e.g., fallen vs. not fallen)

Intervention/control is a covariate

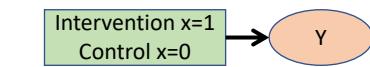
Other covariates may be added to adjust for pretest differences

Different ways to estimate treatment effects in randomised controlled trials

Twisk J¹, Bosman L¹, Hoekstra T^{1,2}, Rijnhart J³, Welten M¹, Heymans M¹

- ANCOVA (Analysis of Covariance)**

If outcome is linear (e.g., balance)



t0

t1

Combines regression with ANOVA on posttest outcome variable (Y_{t1})

Intervention/control is a covariate

Pretest outcome variable (Y_{t0}) is added to the model to adjust for baseline differences

Other covariates may be added to adjust for pretest differences

9

Regression techniques in RCTs: The PRE.C.I.S.A. RCT

> *Front Neurol.* 2022 Sep 1:13:943918. doi: 10.3389/fneur.2022.943918. eCollection 2022.

Efficacy of a multiple-component and multifactorial personalized fall prevention program in a mixed population of community-dwelling older adults with stroke, Parkinson's Disease, or frailty compared to usual care: The PRE.C.I.S.A. randomized controlled trial

Fabio La Porta¹, Giada Lullini¹, Serena Caselli², Franco Valzania³, Chiara Mussi², Claudio Tedeschi³, Giulio Pioli³, Massimo Bondavalli³, Marco Bertolotti², Federico Banchelli^{2,4}, Roberto D'Amico^{2,4}, Roberto Vicini^{2,4}, Silvia Puglisi², Pierina Viviana Clerici², Lorenzo Chiarì^{5,6}; PRECISA Group

- Population (N=403)**

Community-dwelling older adults with frailty, PD, or stroke

Two centers

- Intervention (N=203)**

Multicomponent and multifactorial personalized intervention to prevent falls (11 weeks)

- Comparator (N=200)**

Usual care

- Outcome**

Falls at one year

10

Regression techniques in RCTs: The PRE.C.I.S.A. RCT

Endpoint	Statistic	Estimate	CI 95%	p-value
Number of falls	IRR	0.94	0.69	1.29
Fall probability (1 fall)	RR	0.94	0.79	1.12
Multi fall probability (≥ 2 falls)	RR	0.89	0.67	1.17
Multi fall probability (≥ 3 falls)	RR	0.68	0.45	1.01
Time to the first fall	HR	0.89	0.69	1.16

CI95%, Confidence Interval at 95% level; IRR, Incidence Rate Ratio;
RR, Relative Risk; HR, Hazard Ratio.

11

Regression techniques in RCTs: The PRE.C.I.S.A. RCT

Variable	Control group			Intervention group			Comparison		Effect size	
	mean	sd	n	mean	sd	n	MD (CI95%)	p	Cohen's d (CI 95%)	
FROP-Com	Pre	-0.5	0.5	200	-0.5	0.5	203			
	Post	-0.9	0.6	132	-0.9	0.6	153			
	Post-pre	-0.3	0.4	132	-0.3	0.4	153	-0.03 (-0.13; 0.07)	0.543	-0.08 (-0.32; 0.15)
BBS	Pre	2.2	1.6	200	2.1	1.6	203			
	Post	2.5	1.8	156	2.6	1.8	184			
	Post-pre	0.1	2.7	156	0.5	2.3	184	0.15 (-0.23; 0.53)	0.445	0.14 (-0.08; 0.35)
POMA	Pre	2.6	1.8	200	2.6	1.9	203			
	Post	3.0	1.9	156	3.1	1.9	182			
	Post-pre	0.2	1.2	156	0.3	1.3	182	0.12 (-0.14; 0.37)	0.363	0.10 (-0.11; 0.31)
MBT	Pre	0.6	2.3	198	0.4	2.4	200			
	Post	1.0	2.5	156	1.2	2.7	182			
	Post-pre	0.2	1.8	155	0.6	2.0	179	0.42 (0.03; 0.81)	0.035	0.26 (0.04; 0.48)
FABS	Pre	0.4	1.2	200	0.4	1.2	203			
	Post	0.4	1.2	156	0.4	1.3	184			
	Post-pre	0.1	0.7	156	0.3	0.7	184	0.21 (0.06; 0.36)	0.006	0.32 (0.10; 0.53)

12

Regression techniques in RCTs

> *Contemp Clin Trials Commun.* 2018 Mar 28:10:80-85. doi: 10.1016/j.conctc.2018.03.008.
eCollection 2018 Jun.

Different ways to estimate treatment effects in randomised controlled trials

Twisk J¹, Bosman L¹, Hoekstra T^{1 2}, Rijnhart J¹, Welten M¹, Heymans M¹

Descriptive information^a regarding the example dataset.

	Baseline	T1	T2
Treatment	126.5 (12.5); n = 68	122.6 (11.5); n = 63	121.6 (12.3); n = 59
Control	130.7 (17.6); n = 71	130.1 (17.0); n = 67	127.2 (14.4); n = 60

^aMean systolic blood pressure and SD between brackets.

- Three methods compared to estimate treatment effect in RCT

ANCOVA

Repeated Measures (RM)

Analysis of Change (AoC)

Equation	Method	Overall treatment effect
(1a)	Longitudinal analysis of covariance	-3.7 (-6.8 to -0.6) ^a
(2a)	Repeated measures ^b	-6.2 (-10.7 to -1.7) ^a
(2c)	Repeated measures without treatment ^b	-3.5 (-6.6 to -0.3) ^a
(3a)	Analysis of changes (not adjusted)	-1.9 (-5.5 to 1.8)
(3b)	Analysis of changes (adjusted)	-3.7 (-6.8 to -0.6) ^a

^aStatistically significant at $\alpha = 0.05$.

13

Regression techniques in RCTs

> *Contemp Clin Trials Commun.* 2018 Mar 28:10:80-85. doi: 10.1016/j.conctc.2018.03.008.
eCollection 2018 Jun.

Different ways to estimate treatment effects in randomised controlled trials

Twisk J¹, Bosman L¹, Hoekstra T^{1 2}, Rijnhart J¹, Welten M¹, Heymans M¹

> *J Clin Epidemiol.* 2006 Sep;59(9):920-5. doi: 10.1016/j.jclinepi.2006.02.007. Epub 2006 Jun 23.

ANCOVA versus change from baseline: more power in randomized studies, more bias in nonrandomized studies [corrected]

Gerard J P Van Breukelen¹

- ANOVA gave biased results

Standard ANOVA: overestimation of treatment effect

ANOVA without treatment variable: underestimation of treatment effect

- ANCOVA and adjusted Analysis of Change gave comparable results

- ANCOVA has more power than Analysis of Change

14

Regression techniques: summary

- Set of techniques to estimate the relationship between variables
- Within RCTs to assess causal nexus between intervention and outcome
 - For binary outcomes: Logistic regression
 - For linear outcomes: ANCOVA or Analysis of Change with baseline adjustment
 - Both inform on **WHAT** makes the difference (efficacy)

But...

15

Agenda



Standard regression techniques to establish causal nexus

Problems with standard regression techniques

Path analysis and Structural Equation Modeling

How to deal with predictors to save on sample size

Conclusions

16

Regression techniques: summary

- Set of techniques to estimate the relationship between variables
- Within RCTs to assess causal nexus between intervention and outcome
 - For binary outcomes: Logistic regression
 - For linear outcomes: ANCOVA or Analysis of Change with baseline adjustment
 - Both inform on **WHAT** makes the difference (efficacy)
- **But they do not inform on WHY and HOW there is a difference (mechanism)**

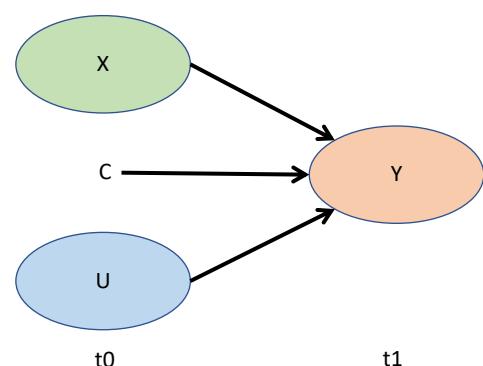


- ***Let's go back to the beginning (regression techniques outside RCTs)***

17

Regression techniques to establish mechanism: step 1

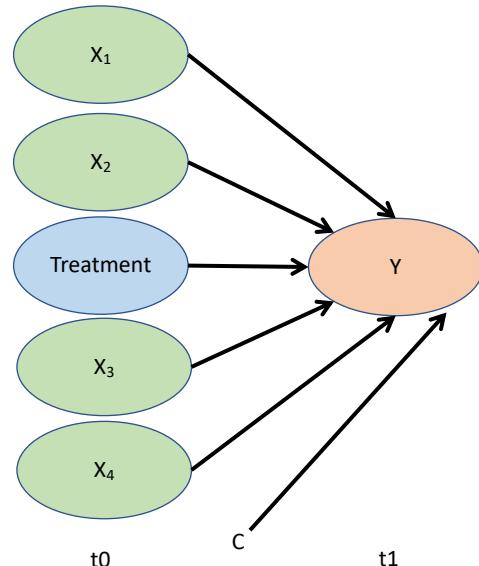
- **Simple linear regression**
 - Dependent variable: *gait speed*
 - Independent variable(s): a putative predictor
- **Simple logistic regression**
 - Dependent variable: *gait speed*
 - above or below a certain cut-off OR above or below a certain change from baseline*
 - Independent variable(s): a putative predictor



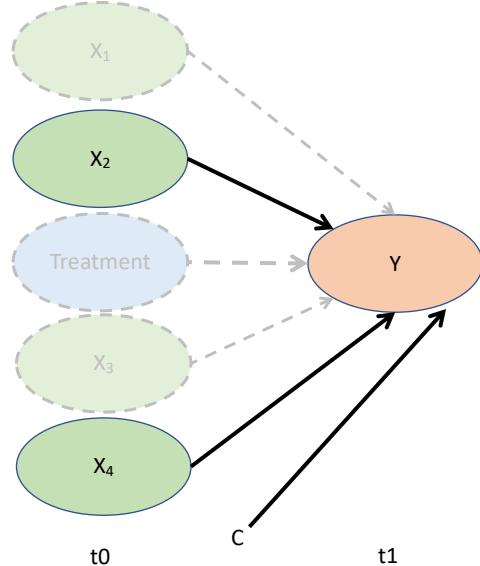
18

Regression techniques to establish mechanism: step 2

Significant predictors from simple regression



Multiple regression



19

Regression techniques to establish mechanism: summary

- **First step: simple regression (linear or logistic)**

Significant predictors chosen from a large set of putative factors are selected

- **Second step: multiple regression (linear or logistic)**

All significant predictors are included in the model

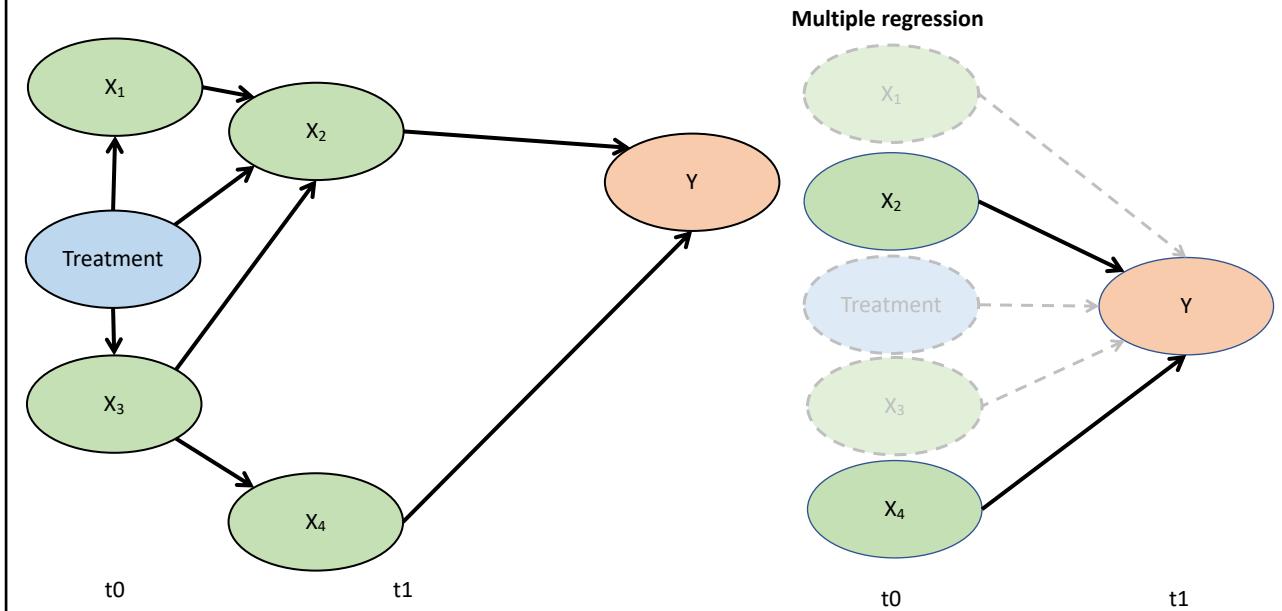
Treatment is one of the covariates

The model retains only the variables contributing the most to the model

But...

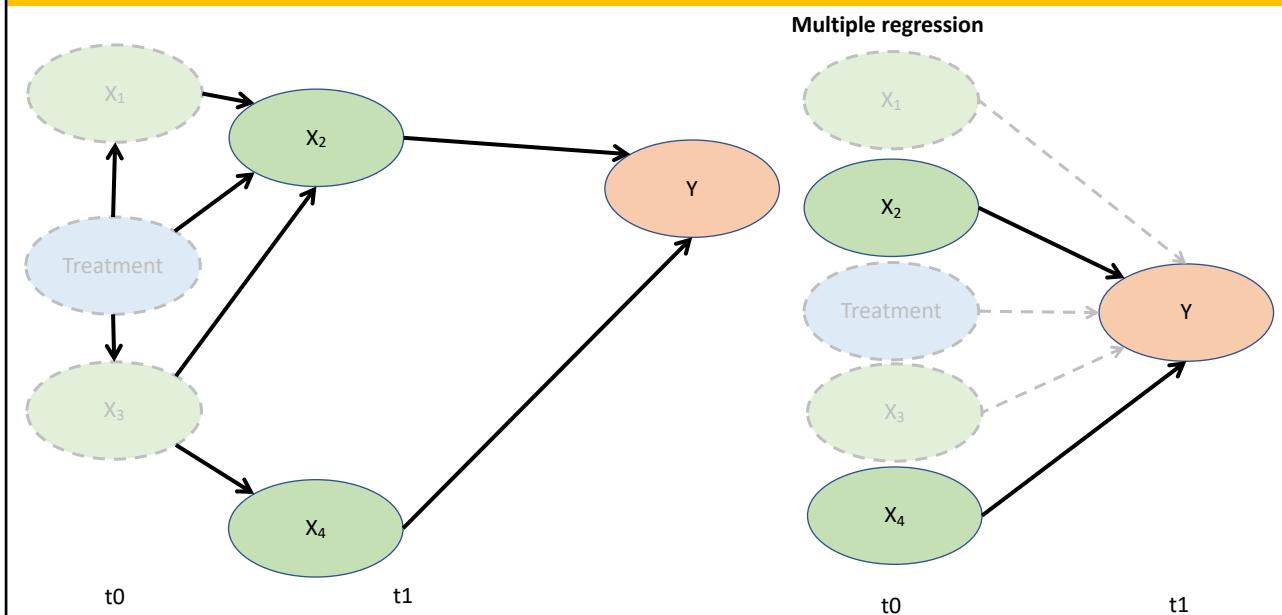
20

What results could we expect if the mechanism was as such?



21

What results could we expect if the mechanism was as such?



22

Regression techniques to establish mechanism: summary

- **First step: simple regression (linear or logistic)**
Significant predictors chosen from putative factors are selected
- **Second step: multiple regression (linear or logistic)**
All significant predictors are included in the model
Treatment is one of the covariates
The model retains only the variables most contributing to the model
- **But significant indirect predictors may be excluded from the model**
- **How can we address this issue?**

23

Agenda



Standard regression techniques to establish causal nexus

Problems with standard regression techniques

Path analysis and Structural Equation Modeling

How to deal with predictors to save on sample size

Conclusions

24

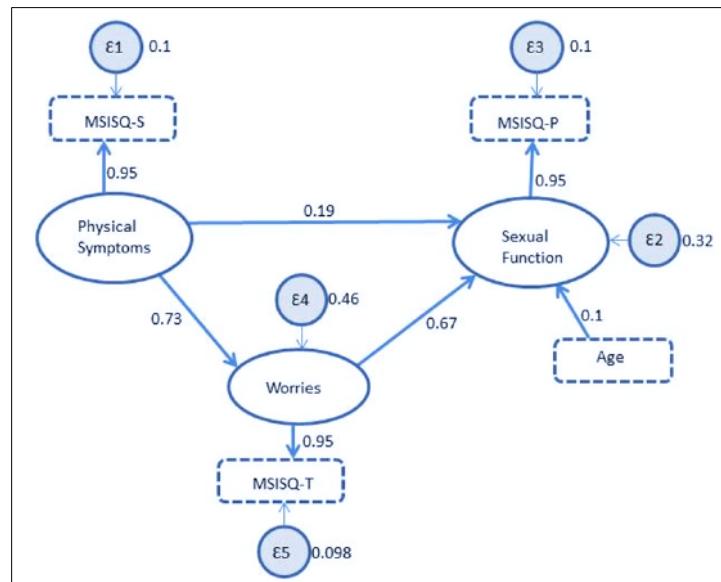
Regression techniques to establish mechanism: summary

- First step: simple regression (linear or logistic)
Significant predictors chosen from putative factors are selected
- Second step: multiple regression (linear or logistic)
All significant predictors are included in the model
Treatment is one of the covariates
The model retains only the variables most contributing to the model
- But significant indirect predictors may be excluded by the model
- How can we address this issue?
- Path analysis and Structural Equation Modeling (SEM)!

25

Path analysis: an example

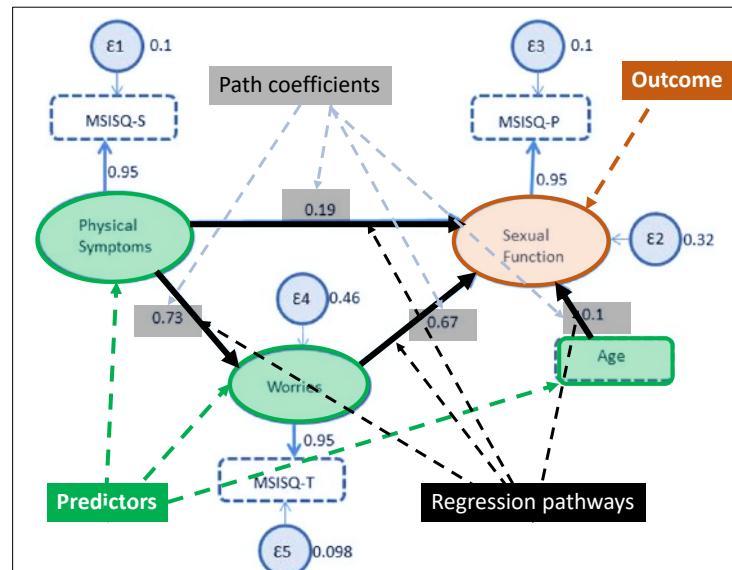
> Mult Scler. 2017 Aug 1;23(9):1268-1275. doi: 10.1177/1352458516675749. Epub 2016 Nov 1.
**Sexual functioning in multiple sclerosis:
Relationships with depression, fatigue and physical function**
 Carolyn A Young¹, Alan Tennant², TONIC Study Group



26

Path analysis: an example

- Path analysis
- Extension of regression model
Simple and multiple regressions are special cases of path analysis
- Direct and indirect effects
- Path coefficients
Standardized regression coefficient (beta)
- Model fit indices are provided
- This model accounts for 96% of the variance of sexual functioning in MS!



27

Structural Equation Modeling: a closer example to StepuP

> Neuroimage. 2023 Apr 15:270:119942. doi: 10.1016/j.neuroimage.2023.119942.
Epub 2023 Feb 14.

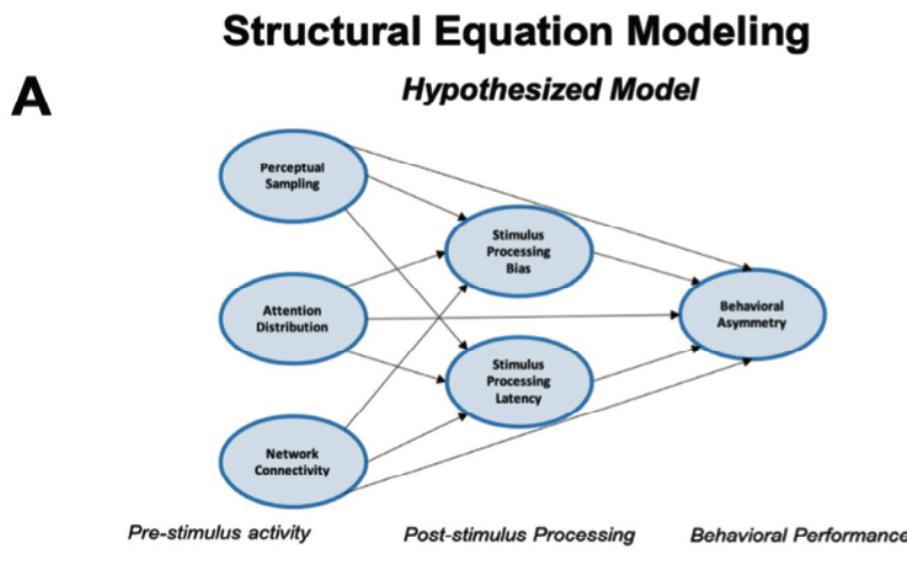
Hierarchical psychophysiological pathways subtend perceptual asymmetries in Neglect

Francesco Di Gregorio ¹, Valeria Petrone ², Emanuela Casanova ², Giada Lullini ²,
Vincenzo Romei ³, Roberto Piperno ², Fabio La Porta ⁴

- Aim
To develop a causative neurophysiological model for Neglect (LHN) in stroke patients
- Step 1
To identify neurophysiological measures (NPhM) discriminating between LHN and controls
- Step 2
To hypothesize a Structural Equation Model linking NPhM to LHN behavior
- Step 3
To test (and refine) the hypothesized model

28

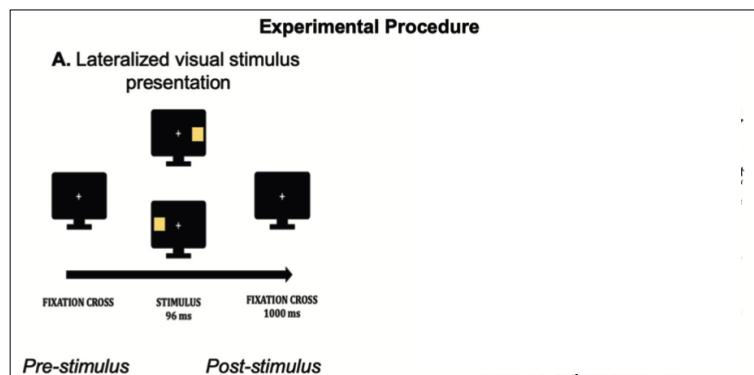
The hypothesized mechanistic model



29

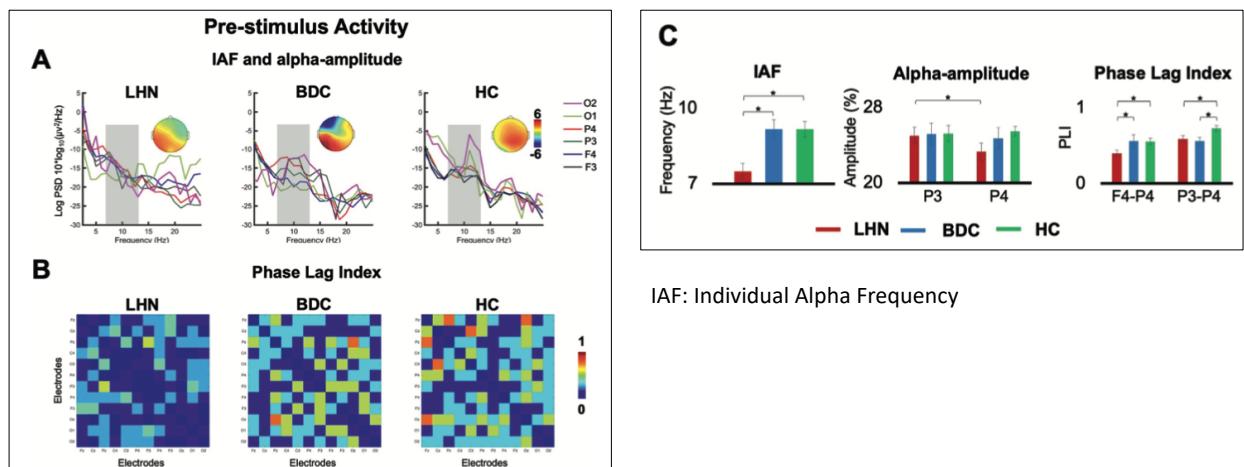
Step 1: experimental paradigm to select the NPh measures

- **Populations**
 - Stroke patients with LHN
 - Stroke patients without LHN
 - Healthy controls
- **Pre-stimulus features**
 - 12-channels EEG
- **Post-stimulus features**
 - Visual evoked potentials
- **Behavioral performance**
 - Bells test



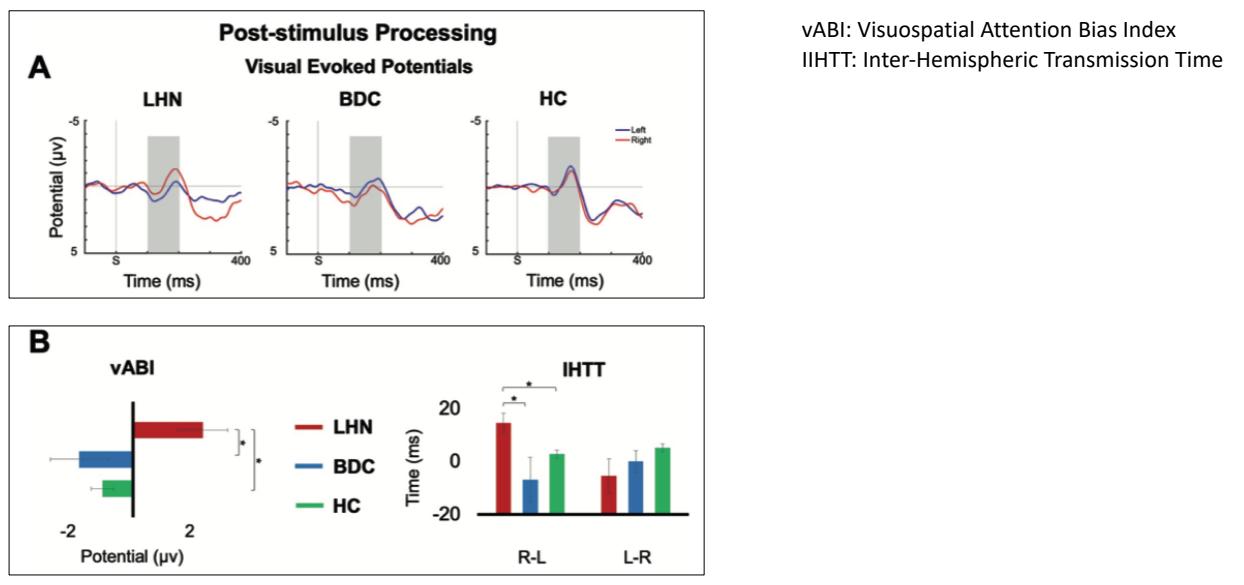
30

Step 1: pre-stimulus results (EEG)



31

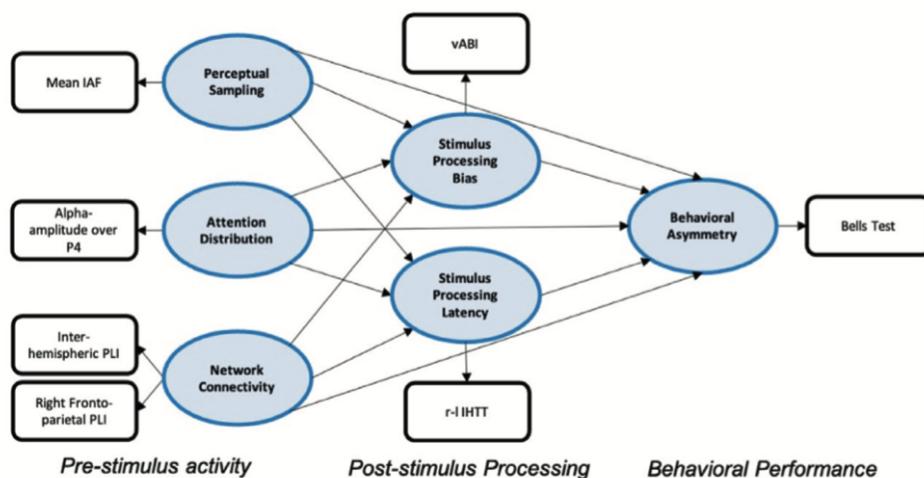
Step 1: post-stimulus results (VEPs)



32

Step 2: hypothesized Structural Equation Model

B

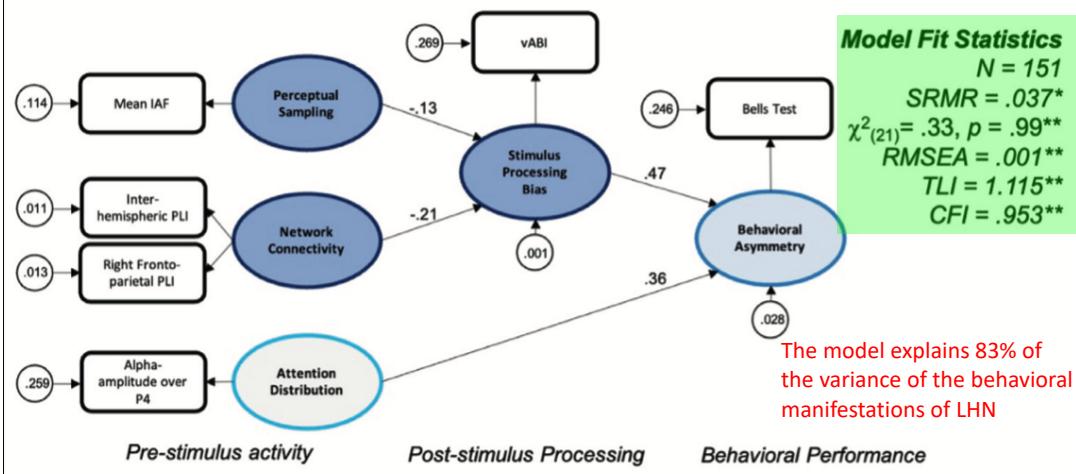


33

Step 3: The refined Structural Equation Model

C

Final Model based on Results of Aim 2



34

Path analysis and Structural Equation Modeling: summary

- **Path analysis**

Extension of the regression model

Just observable (manifest) variables

- **Structural Equation Modeling (SEM)**

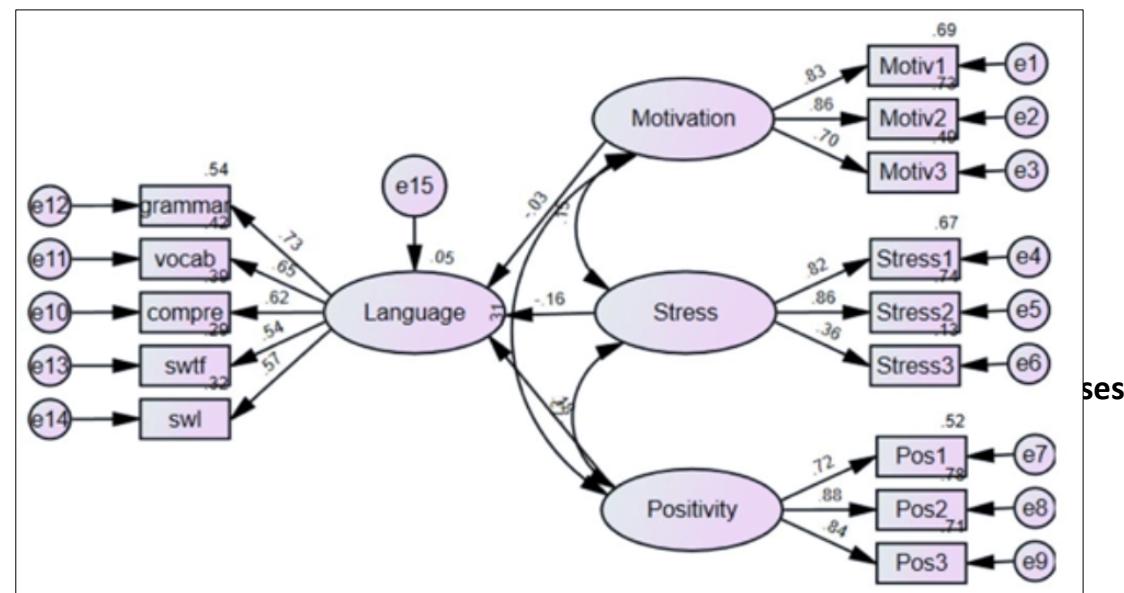
Include path analysis and a measurement model (Confirmatory Factor Analysis)

SEM is used to test the relationship between latent and manifest variables

(both the presented model were SEM)

35

Path analysis and Structural Equation Modeling: summary



36

Path analysis and Structural Equation Modeling: summary

- **Path analysis**

Extension of the regression model
Just observable (manifest) variables

- **Structural Equation Modeling (SEM)**

Include path analysis and a measurement model (Confirmatory Factor Analysis)
SEM is used to test the relationship between latent and manifest variables
(both the presented model were SEM)

- **Both are powerful multivariate techniques to confirm causality hypotheses**

But...

37

Agenda



Standard regression techniques to establish causal nexus

Problems with standard regression techniques

Path analysis and Structural Equation Modeling

How to deal with predictors to save on sample size

Conclusions

38

Path analysis and Structural Equation Modeling: summary

- **Path analysis**

Extension of the regression model
Just observable (manifest variables)

- **Structural Equation Modeling (SEM)**

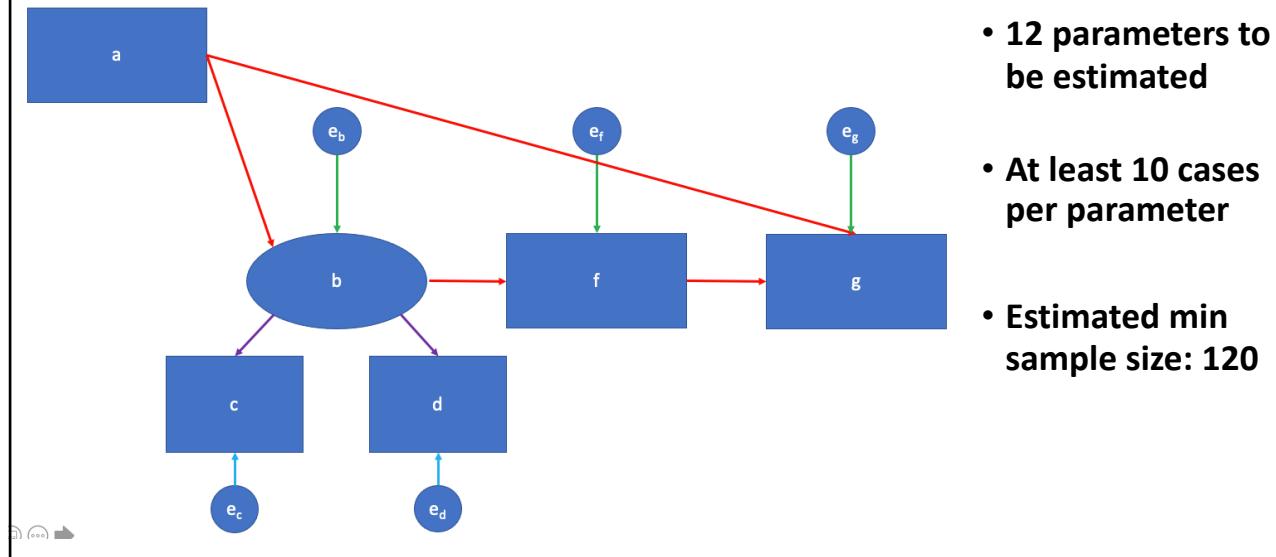
Include path analysis and a measurement model (Confirmatory Factor Analysis)
SEM is used to test the relationship between latent and manifest variables
(*both the presented model were SEM*)

- Both are powerful multivariate techniques to confirm causality hypotheses

- **But how to deal with a potential tons of putative predictors within Stepup?**

39

Parameters to be estimated and sample size



40

Strategy one: start from well-grounded hypotheses

- Path analysis and SEM are confirmatory techniques
- Thus, we need to build models starting from well-grounded hypotheses
 - Literature and prior exploratory analyses
- Efficacy of SDTT: hypothesized mechanisms
 - Decreased beta band oscillations in sensorimotor areas
 - Increased cortical-muscular coherence
 - Improved center of mass state estimation
 - Improved foot placement coordination
- Post-hoc modifications are possible
 - As suggested within the analyses
 - Need to make sense from a substantive point of view

41

Strategy two: reduce the number of predictors

- Principal Component Analysis (PCA)
 - To reduce a large set of variables into a smaller one still containing most of the information (variance) of the larger set
 - Applicable only to linear variables
- May be dangerous
 - Does not require unidimensionality of the data
 - Reduction made on exploratory statistical grounds
 - Sample dependency

42

Strategy three: to aggregate variables into unidimensional sets

- Linear variables**

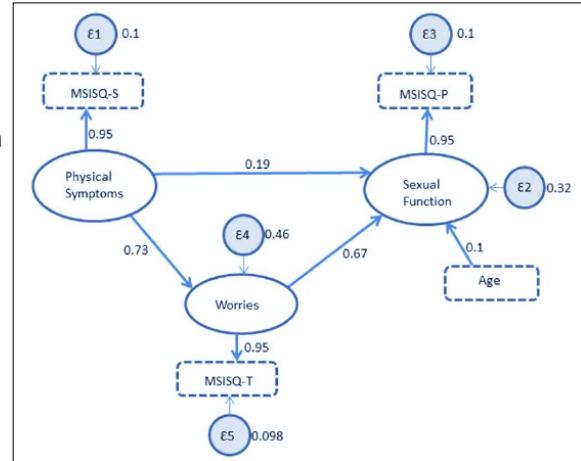
Exploratory / Confirmatory Factor Analysis
 Try to link manifest variables to latent variables
 Provide evidence of unidimensionality of the data
 Both parametric and non-parametric

- Categorical variables**

Rasch analysis
 Unidimensional model
 Invariant measurement
 Very powerful

- Mixed approach**

Transform linear variables into categorical
 Then perform Rasch analysis



43

How to deal with predictors to save on sample size: summary

- Strategy one: start from well-grounded hypotheses**

Model modifications to improve fit must make sense from a substantive point of view

- Strategy two: reduce the number of variables with PCA**

No dimensionality assumed, reduction on statistical grounds
 Potentially dangerous

- Strategy three: aggregate variables into unidimensional sets**

Unidimensionality assumed and verifiable
 Less likely to make choices only on statistical grounds (confirmatory methods)
 Confirmatory Factor analysis
 Rasch analysis

44

Agenda



- Standard regression techniques to establish causal nexus
- Problems with standard regression techniques
- Path analysis and Structural Equation Modeling
- How to deal with predictors to save on sample size
- Conclusions

45

Overall summary: analysis of efficacy

- **Analysis of efficacy is really needed within the StepuP**
- **Need to demonstrate that SDTT and SDTT+ both work before anything else**
i.e., to assess 'the WHAT' first
Besides, this will make methodologists and clinicians happy ☺
- **Repeated measures ANOVA may give biased results**
- **ANCOVA or adjusted Analysis of Change are preferable**
both give unbiased estimates of treatment effect

46

Overall summary: Simple and multiple regressions (logistic and linear)

- **Can be used to assess causal relationships**
First the WHAT, then the WHY-HOW
- **Simple regression to identify significant predictors**
- **Multiple regression to identify an overall causal model**
- **No distinction between direct and indirect effect**
- **May exclude significant indirect predictors**

47

Overall summary: Structural Equation Modeling and path analysis

- Both are powerful multivariate techniques to confirm causality hypotheses
- Make distinctions between direct and indirect effects
- May require larger sample sizes
depending on the number of parameters to be estimated

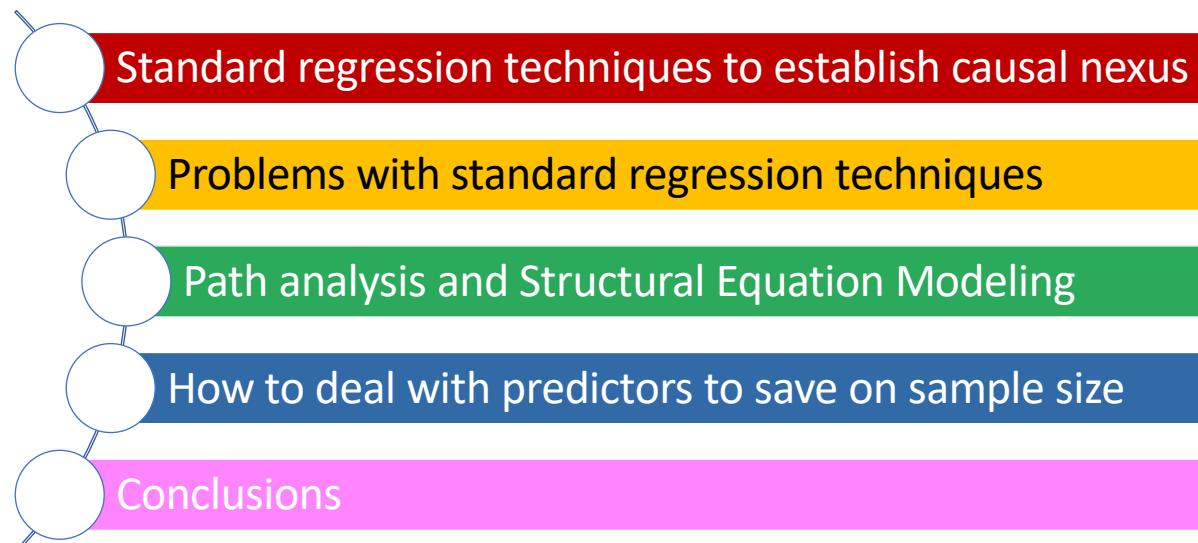
48

Overall summary: strategies to save on the sample size

- **Need to keep the number of parameters low**
To avoid too large sample sizes
- **Start from well-grounded hypotheses**
We are there!
- **Reduce the number of parameters**
Principal Component Analysis
- **Aggregate variables into unidimensional latent variables**
Confirmatory Factor Analysis and Rasch analysis

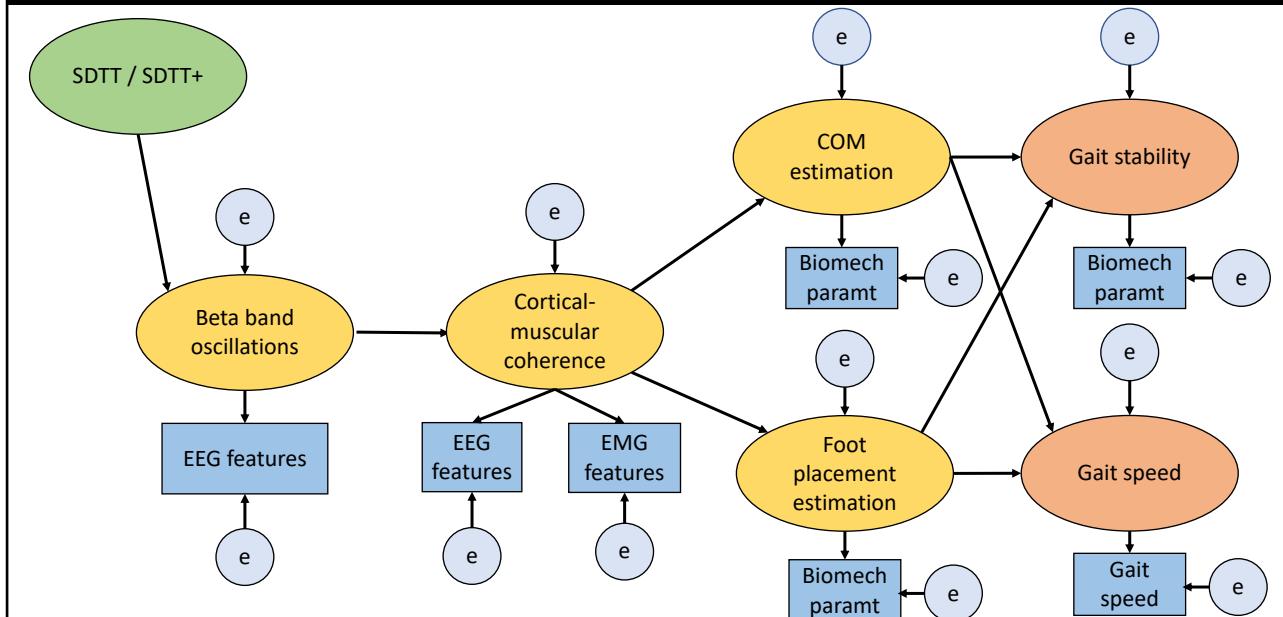
49

Agenda

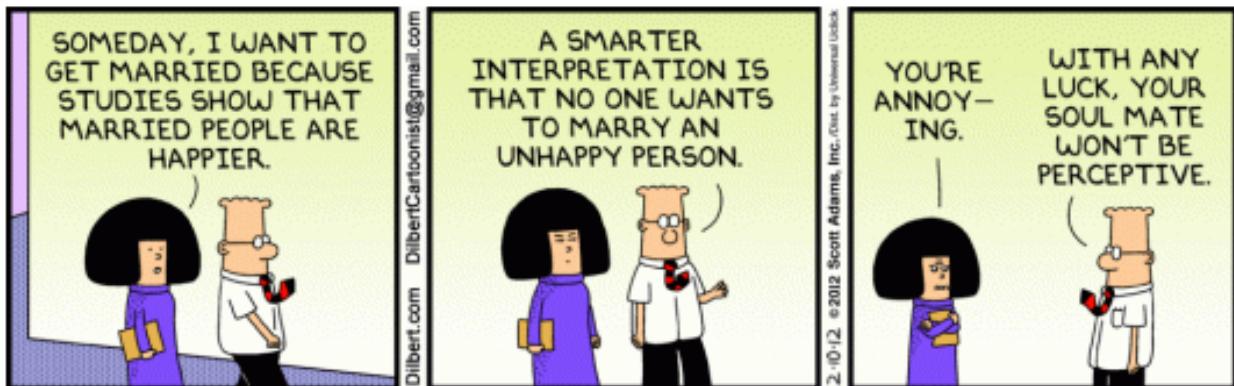


50

Our starting SEM model?



51



Thank you for listening

52