An Introduction to Dynamic Structural Equation Models Using Stan: A Practical Guide For Cognitive Researchers

ESCoP - 9 September, 2023

Jean-Paul Snijder[a], Valentin Pratz[a], and Anna-Lena Schubert[b]

[a]University of Heidelberg, [b]University of Mainz

1 Dynamic Structural Equation Modeling (DSEM)

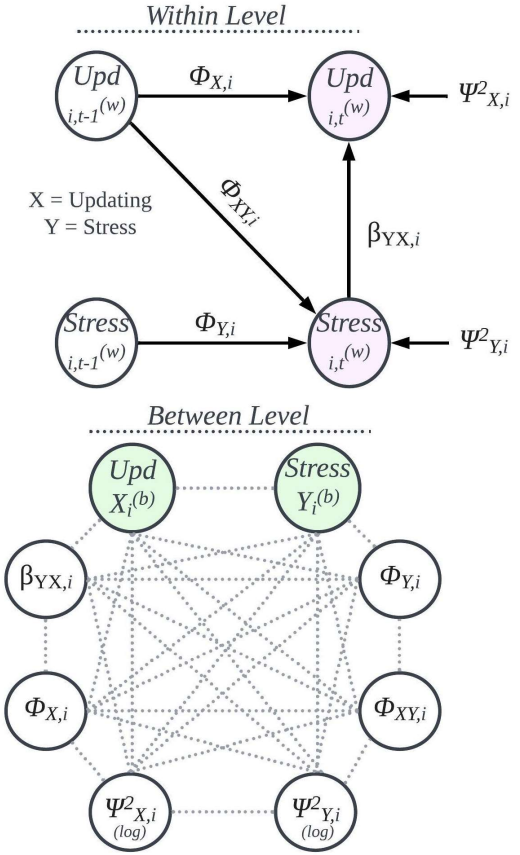
Introduction

- Technological advancements are increasing availability of Intensive Longitudinal Data (ILD) from:
  - Experience Sampling Methods (ESM, EMA, AA)
  - Electro-EncephaloGram (EEG)
  - Wearables
- ILD are densely spaced repeated measures data collected from large samples
- Need for models that allow to examine dynamic changes over time
- Computational models are developed and adapted to match this growing demand

Dynamic Structural Equation Modeling

Combines: [1]

- **Time-series modeling**
  - *allows lagged relationships*
- **Multilevel modeling**
  - *allows modeling of nested data structures*
- **Structural equation modeling**
  - *allows latent variable/path analysis*



1. Hamaker et al. (2023)

DSEM in Mplus

Pros

- Widely used
- online user and program support
- Considered user-friendly
- Low computational time
  - Gibbs sampler with conjugate priors (Normal $\Leftrightarrow$ Inverse Wishart)

Cons

- Not fully customizable
- currently doesn't support some model extensions and specifications
- limited prior options and access to sampler settings
  - *i.e., no LKJ distribution*
- Limited options for missing data
- License costs money

Stan

Pros

- Free
- Fully customizable
- Open Code & Reproducible Science
- Online community support
- Hamiltonian Monte Carlo
  - Efficient general-purpose MCMC sampler

Cons

- Programming can pose a barrier
  - *Fully code-based*
  - *No GUI*
- Higher computational time
  - *but reasonable (minutes to hours)*
  - *not optimized for a specific model family*

Our project

- Stan tutorial using DSEM framework as example
  1. Introducing DSEM
  2. Improving the accessibility to Stan

- **6 model archetypes**[1]
  1. Bivariate, Single Case
  2. **Bivariate, Multilevel**
  3. Model 2 + predictor variable
  4. Model 2 + latent variable
  5. Model 3 + outcome variable
  6. Model 4 + mediation

- **For each archetype in Stan:**
  1. *Simple:* tutorial model
  2. *Reparam:* reparameterized model
  3. *Full:* missing data model

1. Hamaker et al. (2023)

Handling missing data

- Missing data is unknown
- Parameters are unknown

treat missing data like parameters

- Preserves uncertainty (unlike mean imputation etc.)

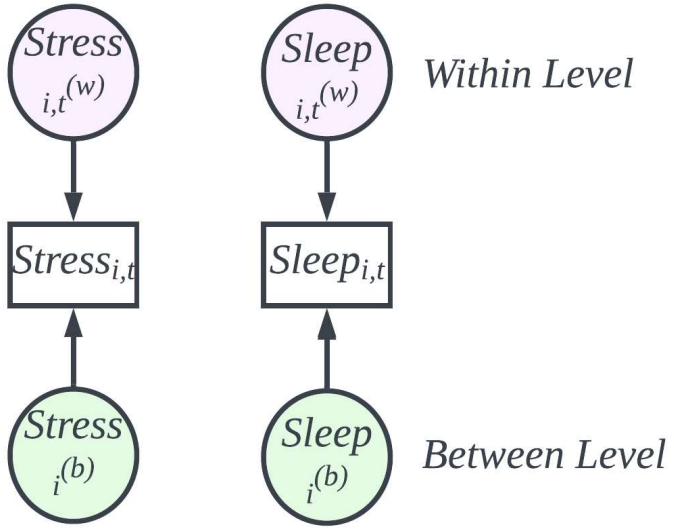
2 DSEM

M2: Two Variable, Multilevel Model

- Model 2: two variables + multilevel
  - Stress
  - Sleep

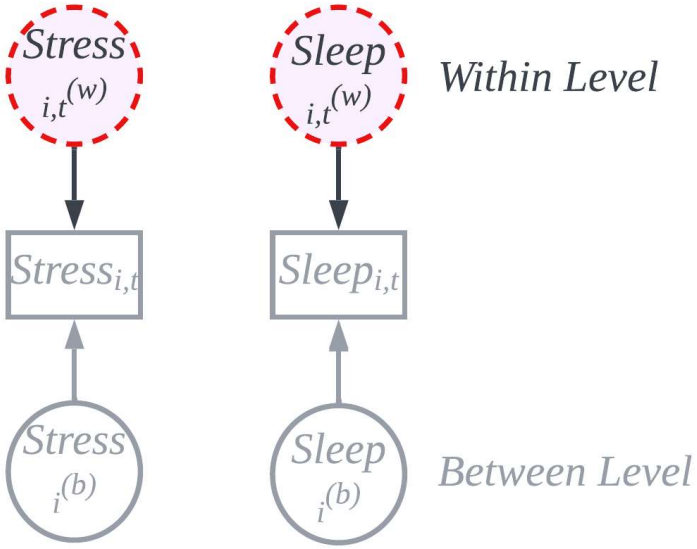
M2: Two Variable, Multilevel Model

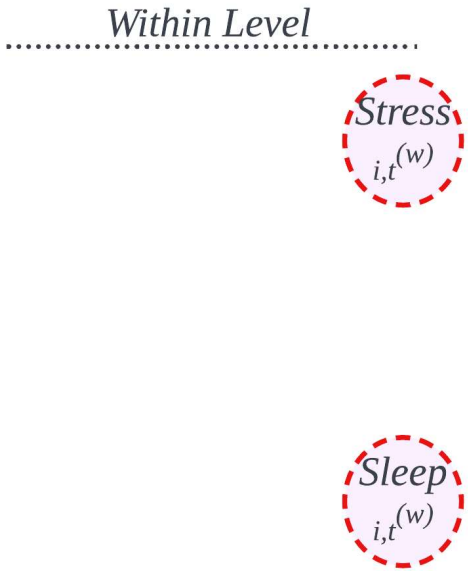
- Model 2: two variables + multilevel
  - Stress
  - Sleep
- Within- & between-person decomposition
  - *Between: time-**insensitive** mean of subject*
  - *Within: time-**sensitive** deviation from that mean*
- Allows for specifying time-*dynamics* in within-person model



M2: Within-person Model I

- The decomposed within-person variables are the start of the within-person model

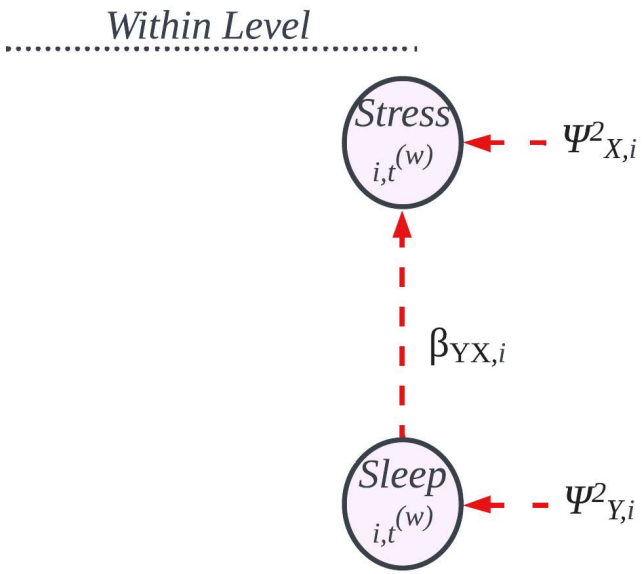




M2: Within-person Model II

Relationships & Parameters:

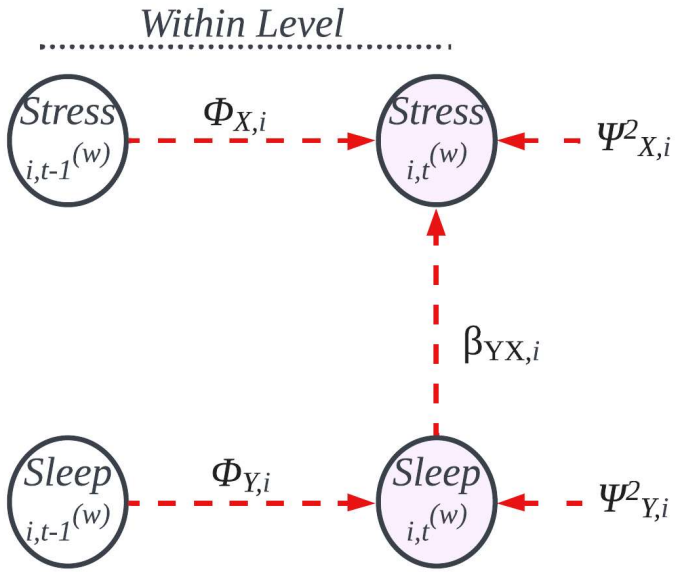
- Regression:
  - γ_X = $Stress_t$ regressed on $Sleep_t$



M2: Within-person Model II

Relationships & Parameters:

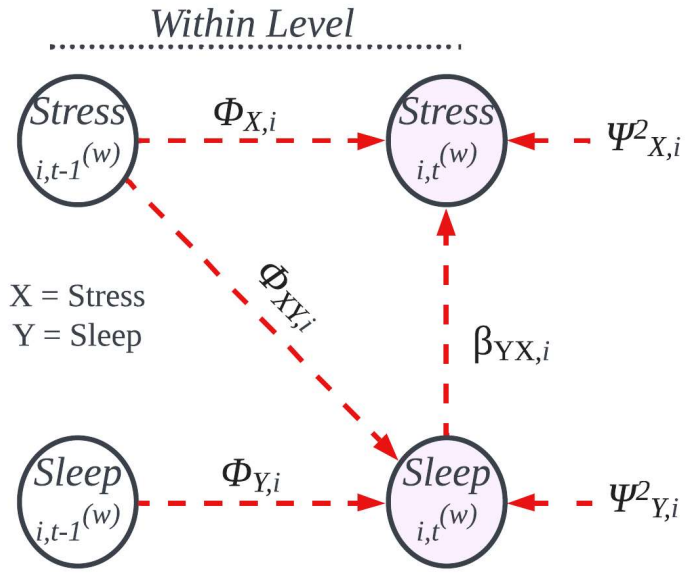
- Regression:
  - γ_{YX} = $Stress_t$ regressed on $Sleep_t$
- Time Dynamic Regressions:
  - $\gamma_{X,i}$ = auto-regressive parameter Stress
  - $\gamma_{Y,i}$ = auto-regressive parameter Sleep
    - $Stress_{i,t\text{-}1}^{(w)}$ and $Sleep_{i,t\text{-}1}^{(w)}$ are lag(1) variables
    - E.g., *if* $_t$ = *observation 9* $_{t\text{-}1}$ = *observation 8* …



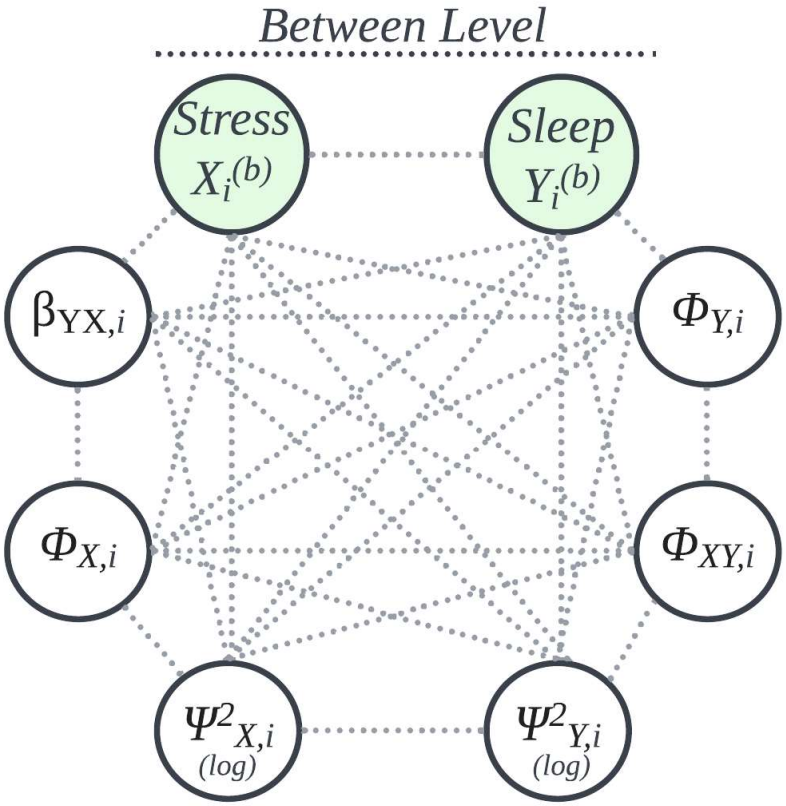
M2: Within-person Model II

Relationships & Parameters:

- Regression:
  - γ_{YX} = $Stress_t$ regressed on $Sleep_t$
- Time Dynamic Regressions:
  - $\chi_{,i}$ = auto-regressive parameter Stress
  - $\gamma_{,i}$ = auto-regressive parameter Sleep
  - $\chi\gamma_{,i}$ = cross-regressive parameter $Sleep_{i,t-1}$ onto $Stress_{i,t}$
- Residual variances:
  - $^2\chi_{,i}$ and $^2\gamma_{,i}$



M2: Between-person Model

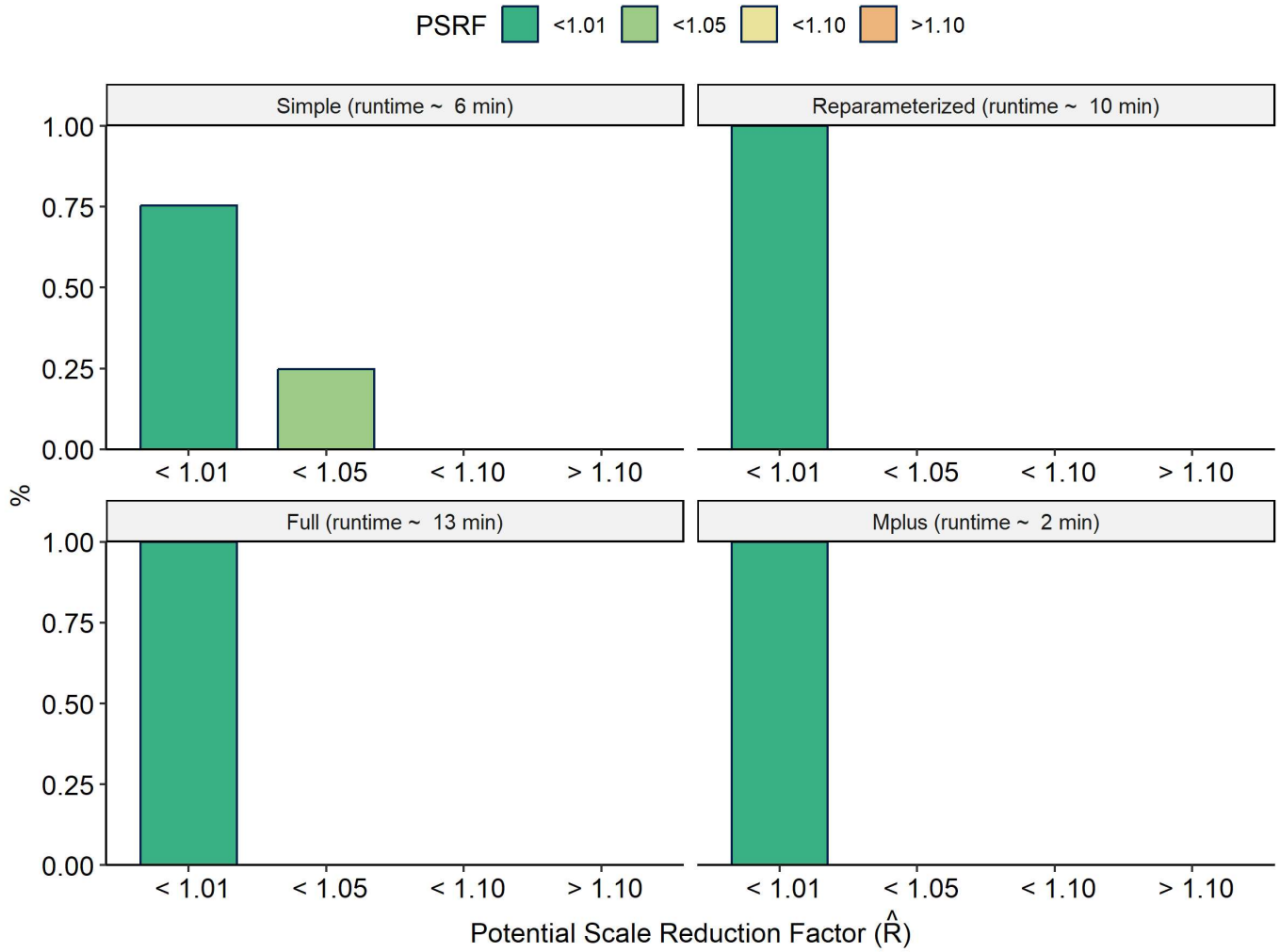


3 Simulation

Results with simulated data

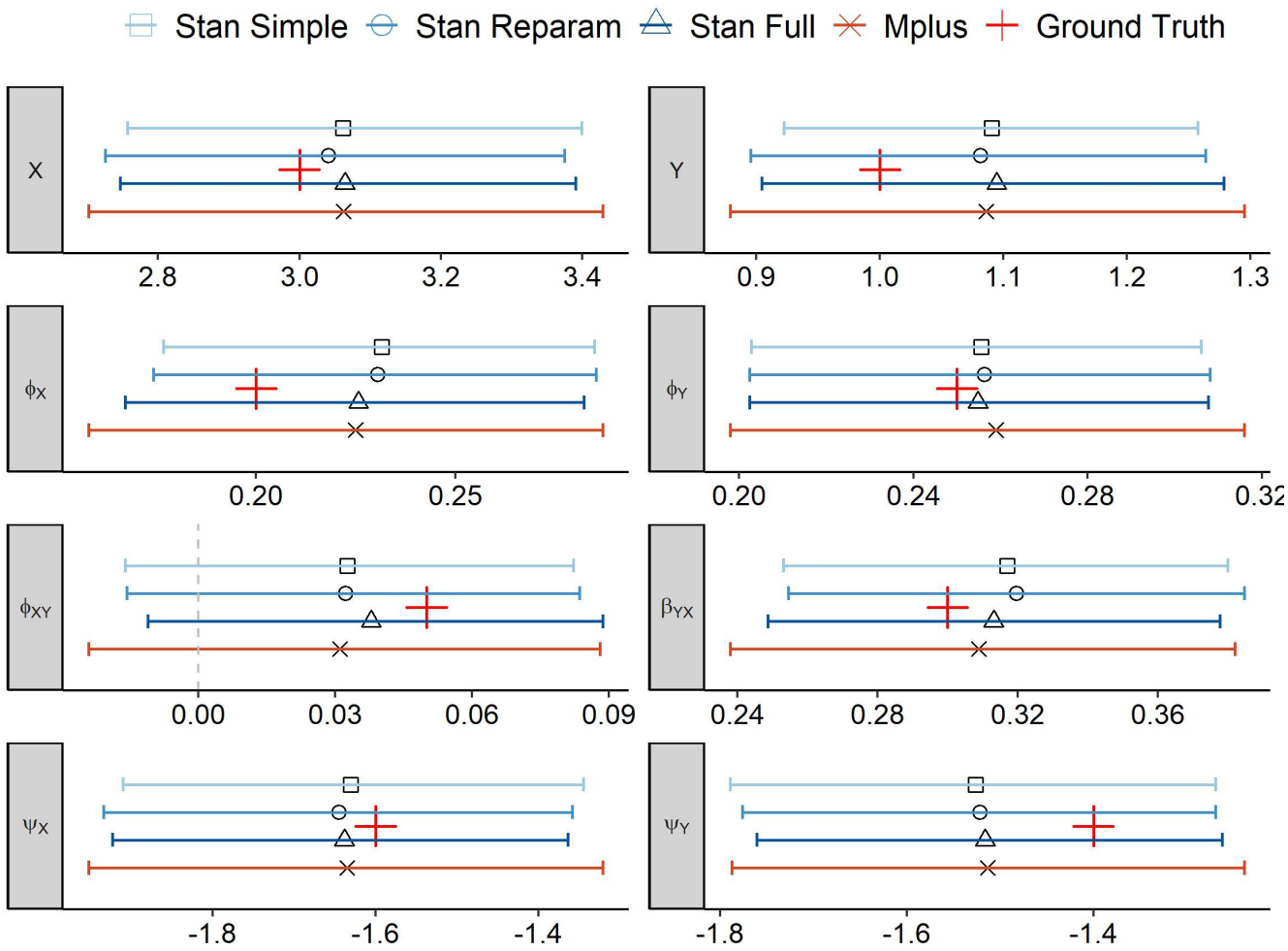
- 100 subjects
- 100 observations
- relevant parameter ranges for sleep and stress
- for missing data model: 5% missingness
- no model misspecification
- Sampler:
  - 500 warmup/3500 sampling iterations
  - 4 chains, 16 cores

Model convergence



Model 2, simulated data convergence.

Parameter recovery



Model 2, simulated data. Errorbars: 95% CI.

4 Real Data Results

COGITO Data

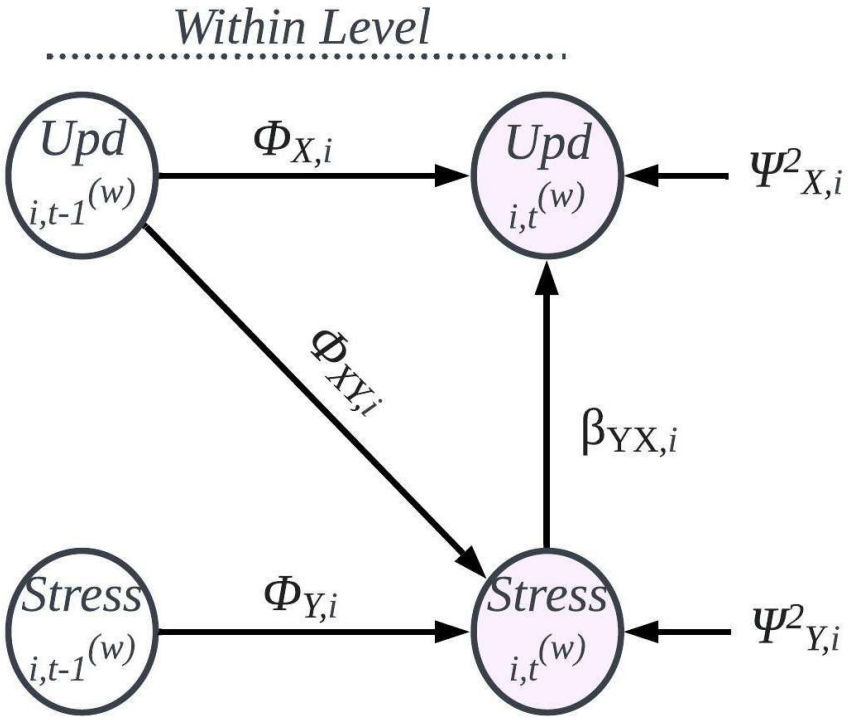
COGITO Study [1]

- Cognitive function and self-report measures
- 101 younger adults & 103 older adults
- Daily measures ~ 100 days

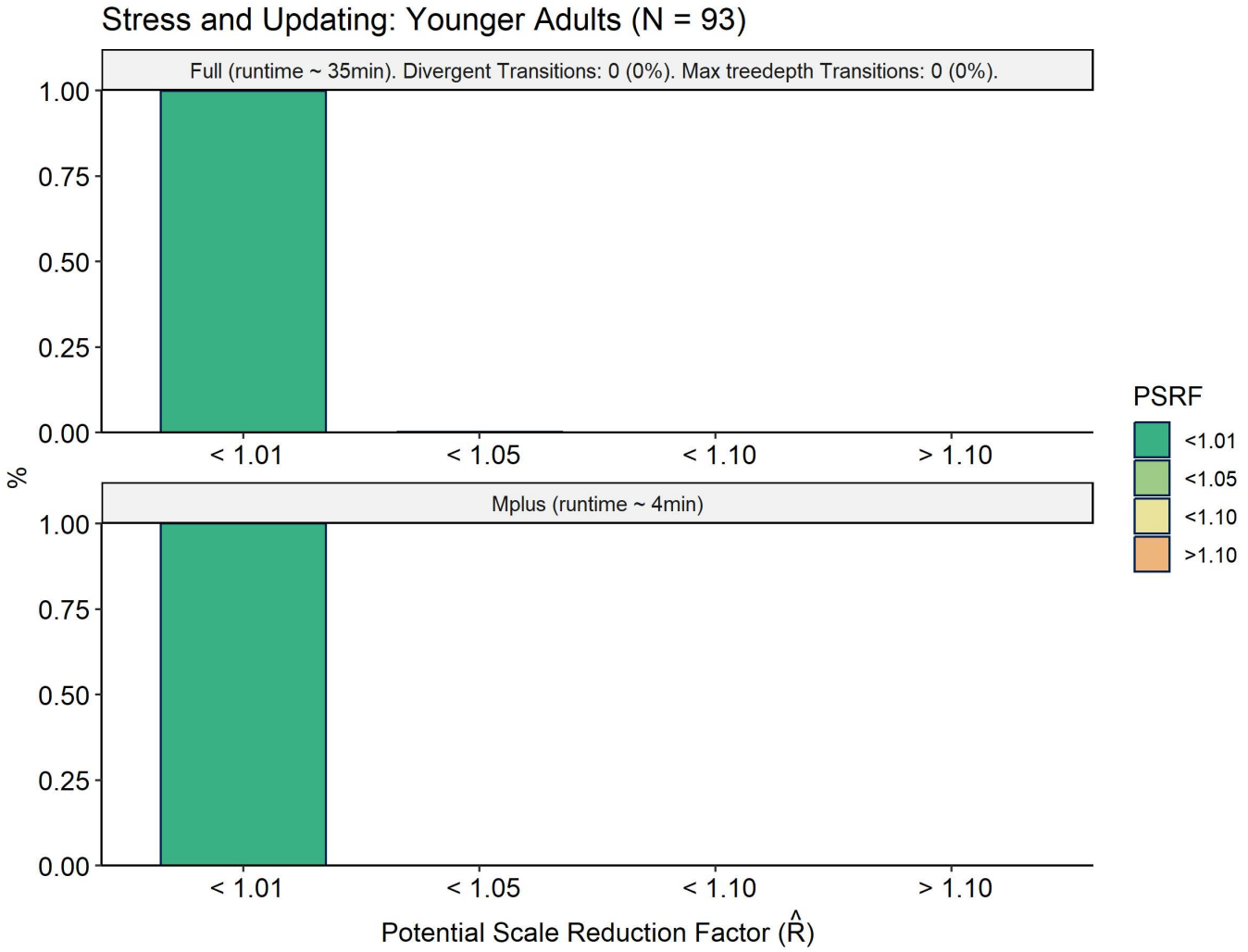
1. Schmiedek et al. (2010)

Results with real data

- COGITO data
  - ~ 100 subjects per group
    - 93 younger & 95 older adults
  - 100 daily observations
    - Stress
    - Memory Updating
  - missing at random: < 1% of observations
  - 500/3500 iterations
  - 4 chains, 16 cores

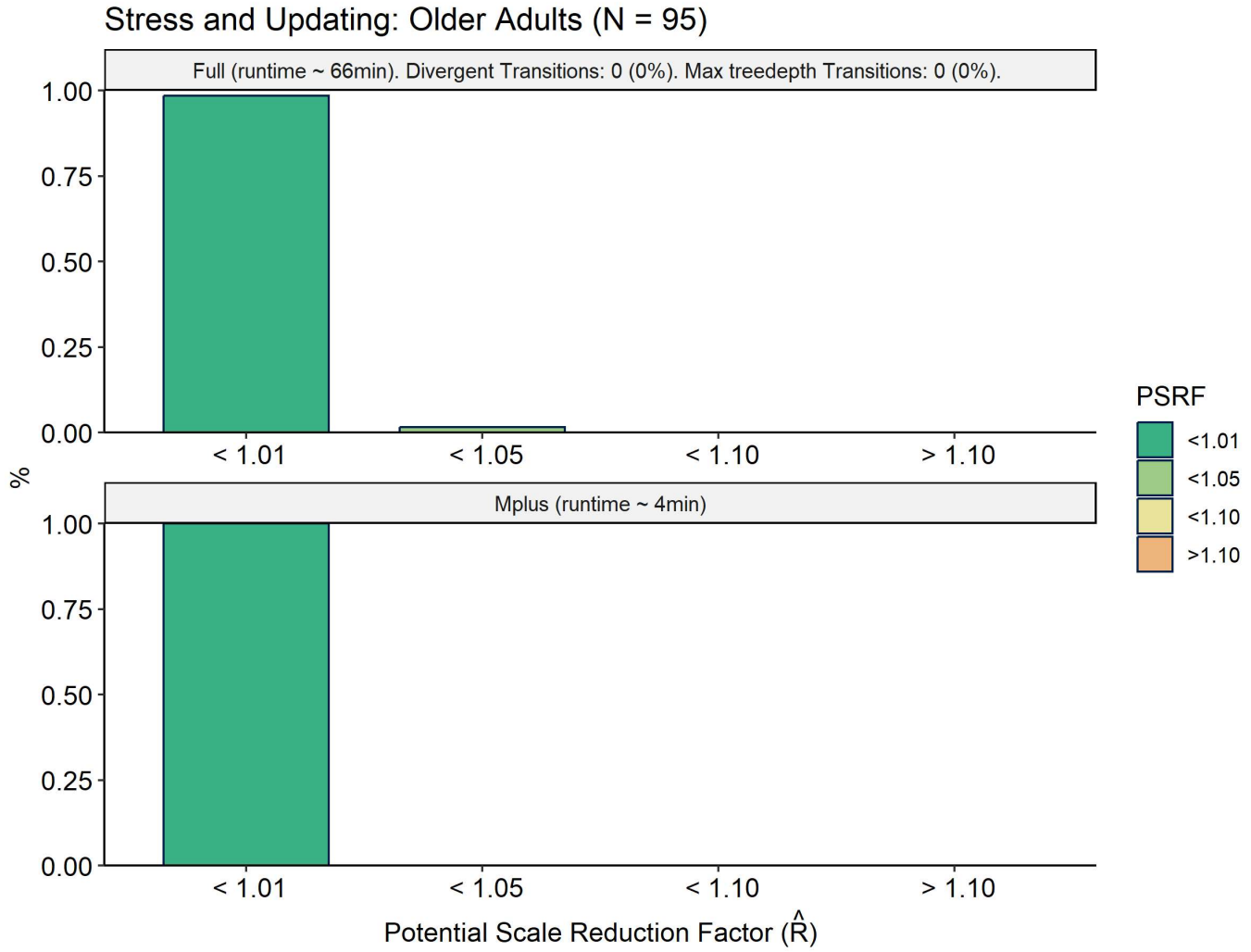


Model convergence (Younger adults)



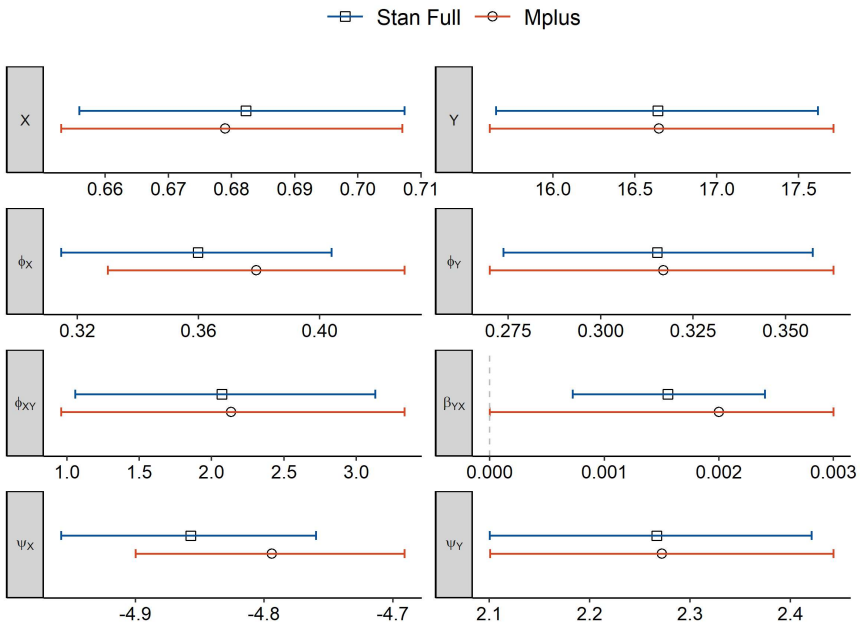
Younger adults.

Model convergence (Older adults)

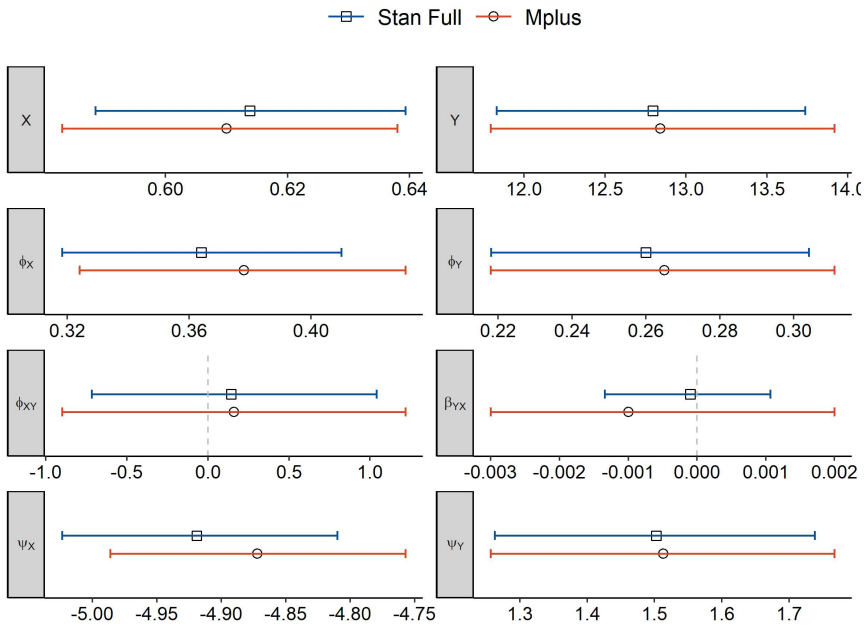


Older adults.

Parameters (age comparison)

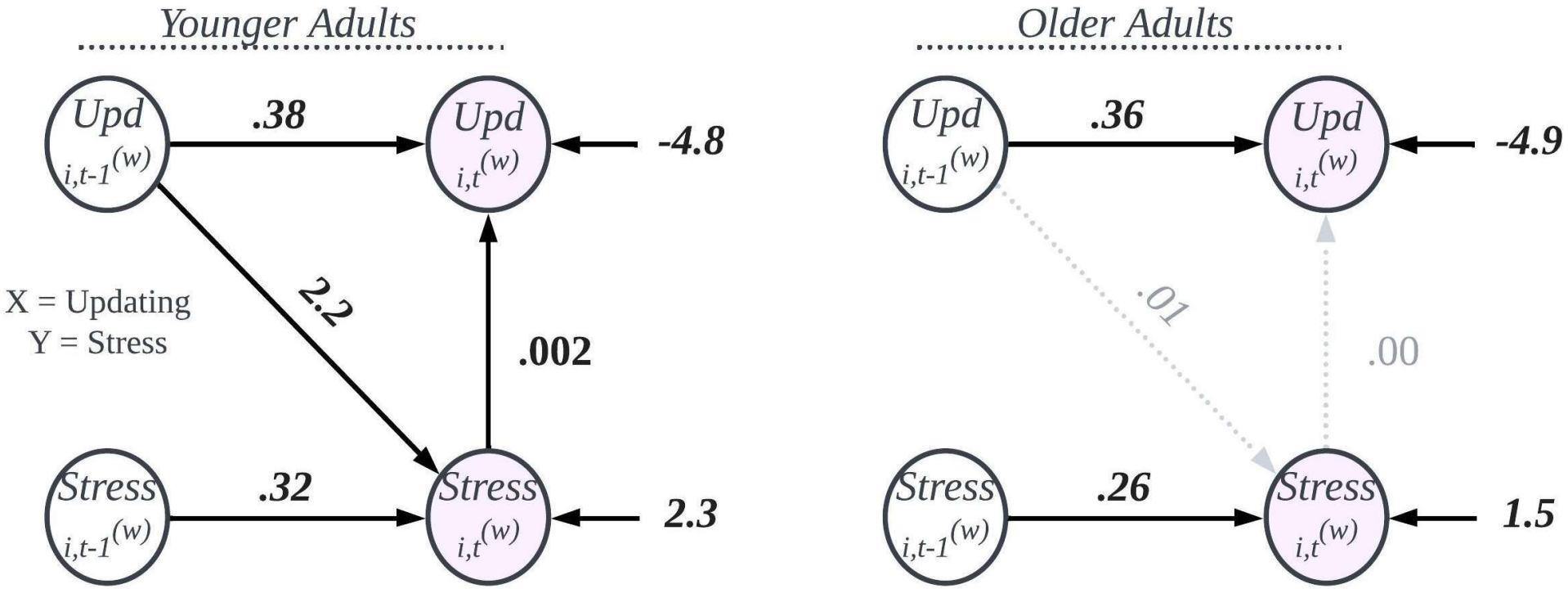


Younger adults.

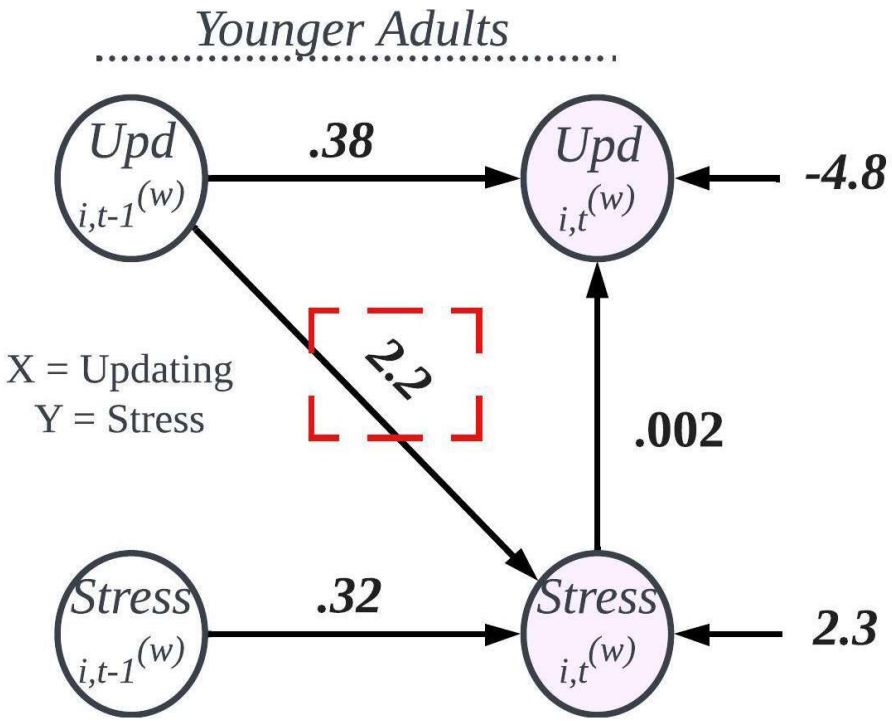


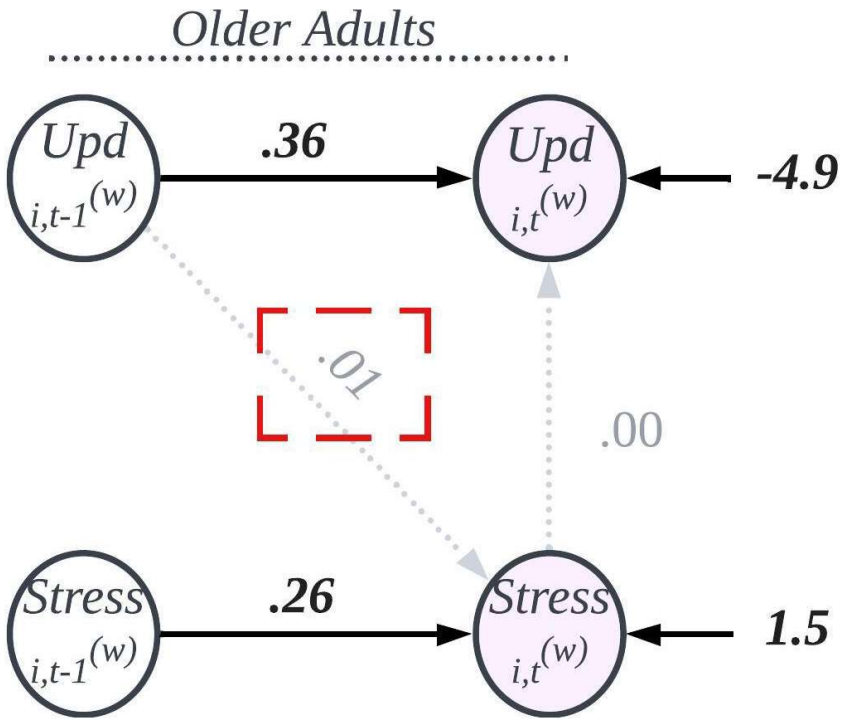
Older adults.

Model Plots


Younger Adults
Upd i,t-1 (w)
.38
Upd i,t (w)
-4.8
X = Updating
Y = Stress
2.2
.002
Stress i,t-1 (w)
.32
Stress i,t (w)
2.3
Older Adults
Upd i,t-1 (w)
.36
Upd i,t (w)
-4.9
.01
.00
Stress i,t-1 (w)
.26
Stress i,t (w)
1.5

Model Plots


Younger Adults
Upd i,t-1 (w)
.38
Upd i,t (w)
-4.8
X = Updating
Y = Stress
2.2
.002
Stress i,t-1 (w)
.32
Stress i,t (w)
2.3


Older Adults
Upd i,t-1 (w)
.36
Upd i,t (w)
-4.9
.01
.00
Stress i,t-1 (w)
.26
Stress i,t (w)
1.5

5 Discussion

Future

- *current:* Simulations
  - relevant parameter ranges
  - prior calibration
- *near:* Standardized estimates
- *near:* Model implementation for cognitive behavioral tasks
- *far:* R Package with Stan as back-end

Thanks to

- **Ellen Hamaker** for instructional material on DSEM
- **Mauricio Garnier-Villarreal** (blavaan) for sharing his Stan knowledge online
- The **Stan community** for educational material on reparameterization and other tricks

Thank you

Questions?

Github repo with presentation + reproducable model 2 example

References

Hamaker, E. L., Asparouhov, T., & Muthén, B. (2023). Dynamic Structural Equation Modeling as a Combination of Time Series Modeling, Multilevel Modeling, and Structural Equation Modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (2nd ed., pp. 576–597). New York: Guilford Press.

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2010). Hundred Days of Cognitive Training Enhance Broad Cognitive Abilities in Adulthood: Findings from the COGITO Study. *Frontiers in Aging Neuroscience*, *2*, 27. https://doi.org/10.3389/fnagi.2010.00027

–>