Dynamic Structural Equation Modeling of Intensive Longitudinal Data Using Mplus Version 8 Part 8

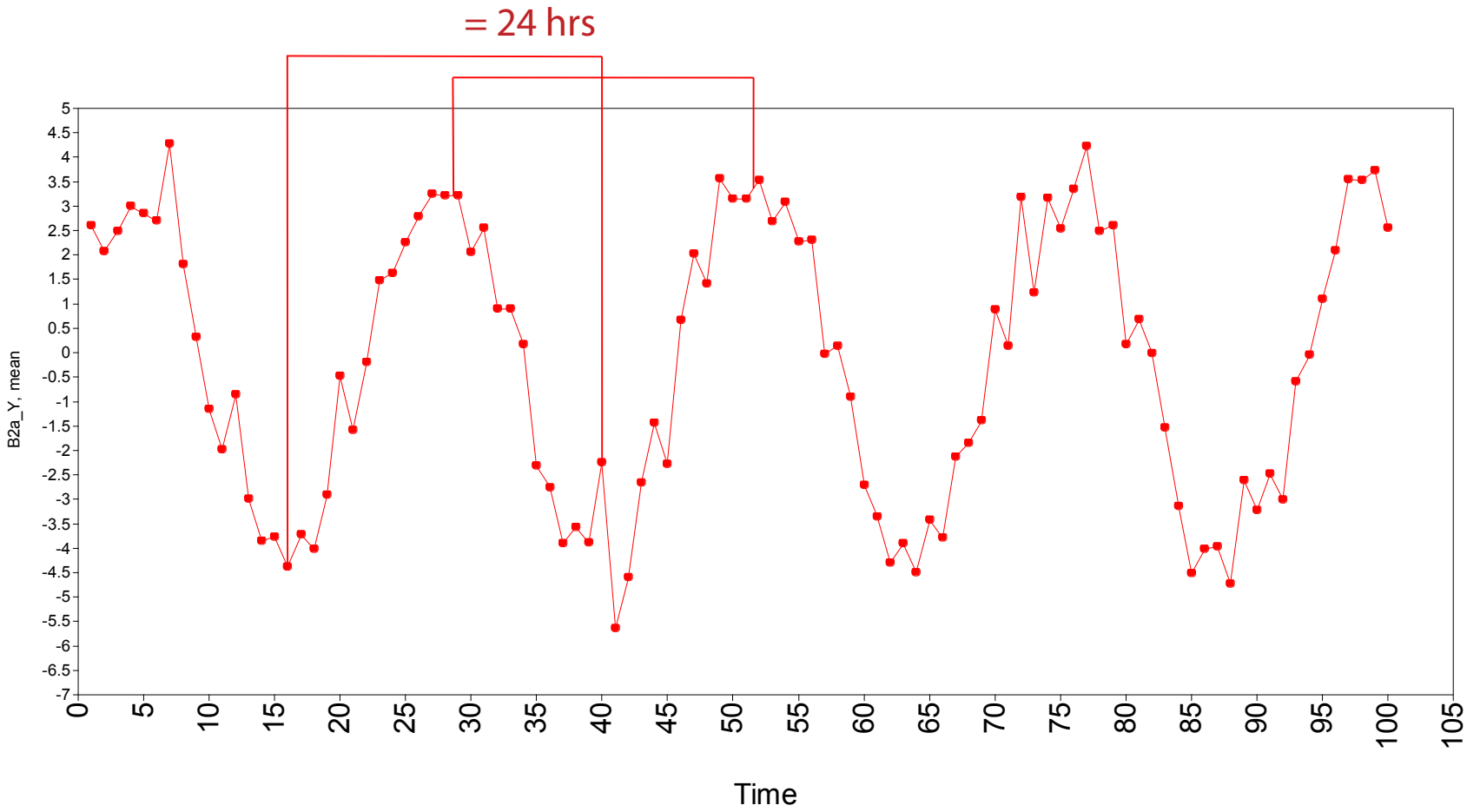
Bengt Muthén

I thank Noah Hastings for excellent assistance

Outline

- DSEM with cycles (circadian rythm)
- Two-part DSEM
- Multilevel time series analysis is in the air. Report on our survey

Modeling Cycles



24-hour cycles: Circadian rhythm such as heart rate.

The picture corresponds to ibi (time in between heart beats) with lows around 4PM (time = 16) and highs around 4AM (time = 28).

Cyclic Formulas

Shumway & Stoffer (2011) Time Series Analysis And Its Applications, pp. 175-177 and Ching Ting Fok & Ramsay (2006) in Walls & Schafer (eds.) Models for ILD, p. 188:

$$f(t) = \beta_1 \, x_1 + \beta_2 \, x_2, \tag{1}$$

where β_1, β_2 determine the amplitude and phase and

$$x_1 = sin(2 \, \pi \, \omega \, t) \tag{2}$$

$$x_2 = cos(2 \, \pi \, \omega \, t), \tag{3}$$

where ω is a frequency index defined as cycles per unit. Using $\omega = 1/24 = 0.04167$ gives 24-hour cycles.

Spectral analysis - finding the components of the cycles

Generating the Cyclic Data: Step 1
Cross-Classified Generation with AR(1) - No Cycles Yet

```
MONTECARLO:          NAMES = y;
                     NOBSERVATIONS = 20000;
                     NREPS = 1;
                     CSIZES = 200[100(1)];! 200 subjects (2b), 100 time points (2a)
                     NCSIZE = 1[1];
                     LAGGED = y(1);
                     SAVE = step1AR.dat; ! saves the time and subject variables too
ANALYSIS:            TYPE = CROSS RANDOM;
                     ESTIMATOR = BAYES;
                     PROCESSORS = 2;
                     BITERATIONS = (500);
MODEL POPULATION:
                     %WITHIN%
                     y ON y&1*0.3;
                     y*1;
                     %BETWEEN level2a% ! across time variation
                     y*.5;
                     %BETWEEN level2b% ! across subjects variation
                     y*2; [y*0];
```

MODEL command same as MODEL POPULATION

Generating Cycles: Step 2
Adding the Cycles And Analyzing Using the Same Model

```
DATA:       FILE = step1AR.dat;
VARIABLE:   NAMES = y time subject;
            USEVARIABLES = y;
            CLUSTER = subject time;
            LAGGED = y(1);
DEFINE:     y = 3*sin(6.28*(1/24)*time) + 2* cos(6.28*(1/24)*time) + y;
            ! 6.28 = 2 π
ANALYSIS:   TYPE = CROSSCLASSIFIED;
            ESTIMATOR = BAYES;
            PROCESSORS = 2;
            BITERATIONS = (10000);
MODEL:      %WITHIN%
            y on y&1;
            %BETWEEN subject%
            y;
            %BETWEEN time%
            y;
OUTPUT:     TECH8;
PLOT:       TYPE = PLOT3;
            FACTORS = ALL(50);
```

Slow convergence, but a time series plot of the between time y factor scores shows the cycles clearly early on.

Two-Level Analysis Modeling the Cyclic Trend

```
DATA:       FILE = step1AR.dat;
VARIABLE:   NAMES = y time subject;
            USEVARIABLES = y x1 x2;
            CLUSTER = subject;
            WITHIN = x1 x2;
            LAGGED = y(1);
DEFINE:     y = 3*sin(6.28*(1/24)*time) + 2* cos(6.28*(1/24)*time) + y;
            x1 = sin(6.28*(1/24)*time);
            x2 = cos(6.28*(1/24)*time);
ANALYSIS:   TYPE = TWOLEVEL;
            ESTIMATOR = BAYES;
            PROCESSORS = 2;
            BITERATIONS = (1000);
MODEL:      %WITHIN%
            y ON y&1;
            y ON x1 x2;
            %BETWEEN%
            y;
OUTPUT:     TECH8;
PLOT:       TYPE = PLOT3;
            FACTORS = ALL(50);
```

The auto-correlation and the x1, x2 slopes are well estimated. Converges in 17 seconds.

Example: Houtveen Heart Beat Data

- N = 162, T = 38 (hourly measures used here)
- Outcome: ibi (time in between heart beats - high is good)
- Covariates: Gender, smoking, sports

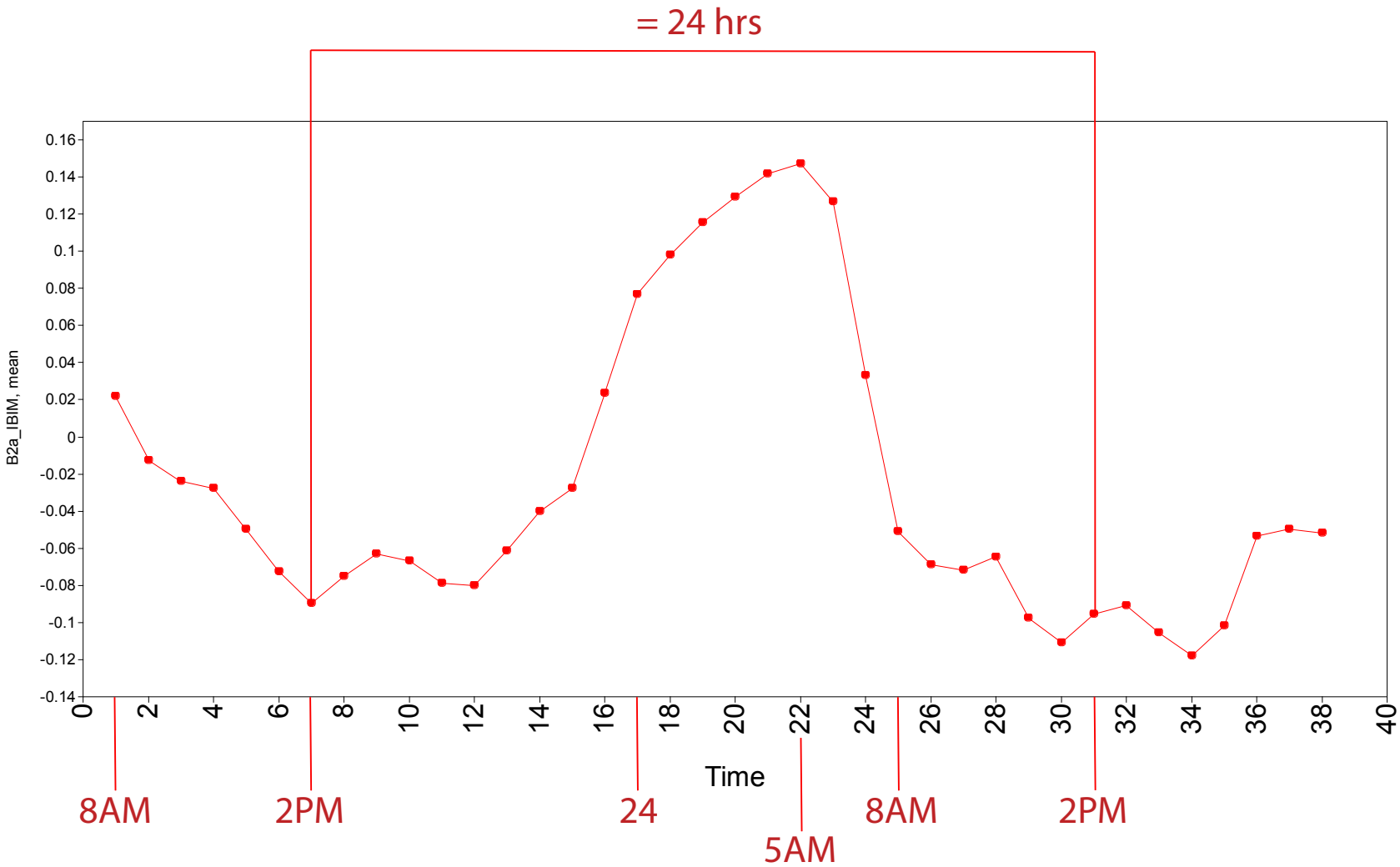
Background:

Houtveen, Hamaker, van Doornen (2010). Using multilevel path analysis in analyzing 24-h ambulatory physiological recordings applied to medically unexplained symptoms. Psychophysiology, 47, 570-578.

Input for Circadian Analysis of Heart Rate Data

```
DATA:       FILE = dataset_for_Ellenandmuthen_TimeAndAverages.csv;
VARIABLE:   NAMES = subject hour body sex age height weight sports smoker ibi-
            stat ibisel time ibim ibiselm;
            USEVARIABLES = subject time ibim sex sports smoker x1 x2;
            CLUSTER = subject time;
            MISSING = ALL(-999);
            LAGGED = ibim(1);
            TINTERVAL = time(1);
            BETWEEN = (subject) sex sports smoker (time) x1 x2;
DEFINE:     x1 = sin(6.28*(1/24)*time);
            x2 = cos(6.28*(1/24)*time);
ANALYSIS:   TYPE = CROSSCLASSIFIED RANDOM;
            ESTIMATOR = BAYES;
            PROCESSORS = 2; BITERATIONS = (2000);
MODEL:      %WITHIN%
            phi | ibim ON ibim&1;
            logv | ibim;
            %BETWEEN SUBJECT%
            ibim phi logv ON sex sports smoker;
            sex sports smoker;
            %BETWEEN TIME%
            ibim ON x1 x2;
OUTPUT:     TECH1 TECH8 FSCOMPARISON;
PLOT:       TYPE = PLOT3;
            FACTORS = ALL;
```

Estimated IBI Time Level Factor Scores



First time point is 8 o'clock in the morning

Model Estimates

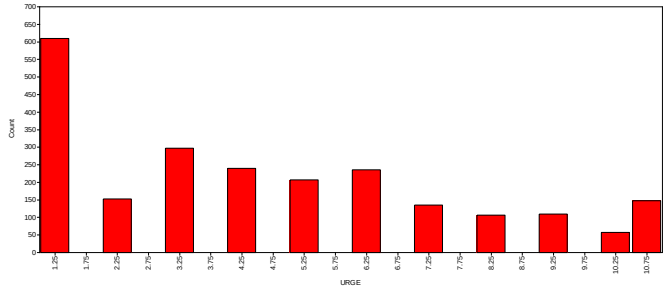
	Estimate	Posterior S.D.	95% C.I. Lower 2.5%	95% C.I. Upper 2.5%	Significance
Within Level					
Between TIME Level					
ibim ON					
x1	-0.084	0.010	-0.104	-0.063	*
x2	0.055	0.009	0.037	0.074	*
Variances					
phi	72.767	26.953	36.735	136.396	*
Residual Variances					
ibim	0.001	0.000	0.001	0.002	*
Between SUBJECT Level					
ibim ON					
sex	-0.076	0.015	-0.105	-0.046	*
sports	0.029	0.015	0.001	0.058	*
smoker	-0.046	0.017	-0.079	-0.011	*
phi ON					
sex	-0.039	0.036	-0.106	0.034	
sports	0.020	0.033	-0.044	0.085	
smoker	0.067	0.037	0.002	0.148	*
logv ON					
sex	-0.162	0.101	-0.362	0.035	
sports	0.243	0.098	0.044	0.430	*
smoker	-0.280	0.109	-0.503	-0.071	*

Outline

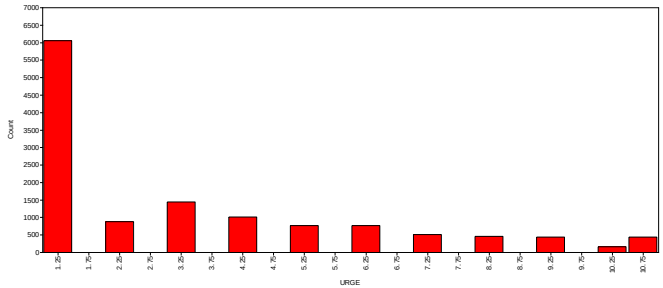
- DSEM with cycles (circadian rythm)
- **Two-part DSEM**
- Multilevel time series analysis is in the air. Report on our survey

Smoking Urge Has a Strong Floor Effect

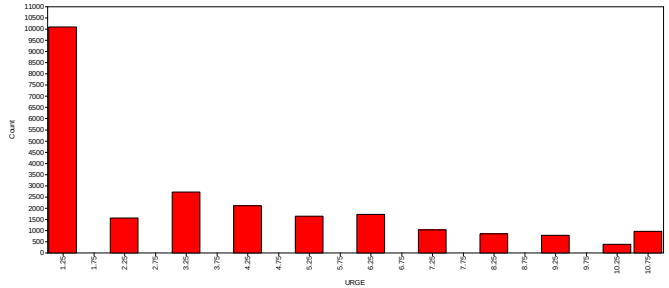
Early: 27% at the floor value



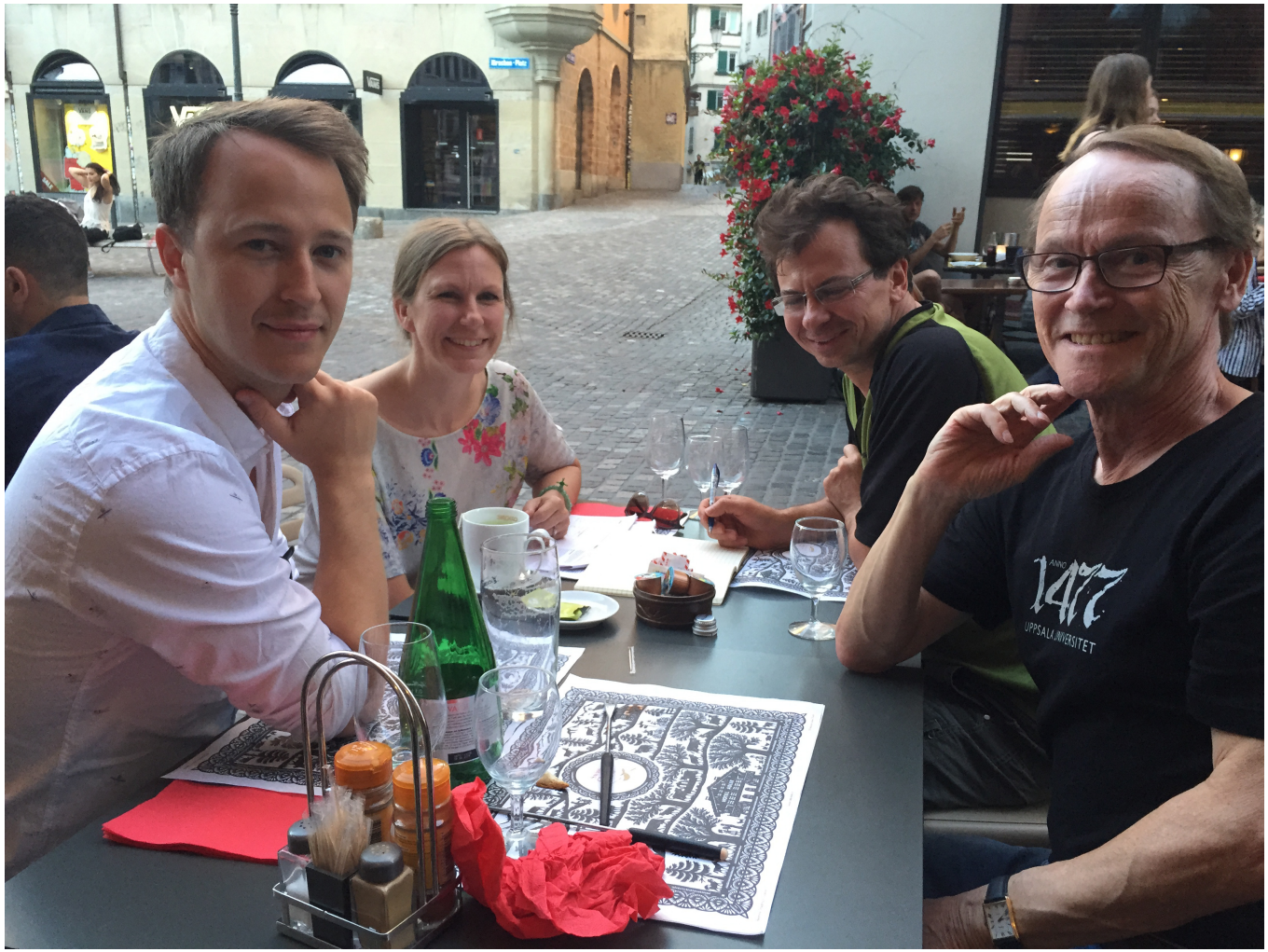
Late: 47% at the floor value



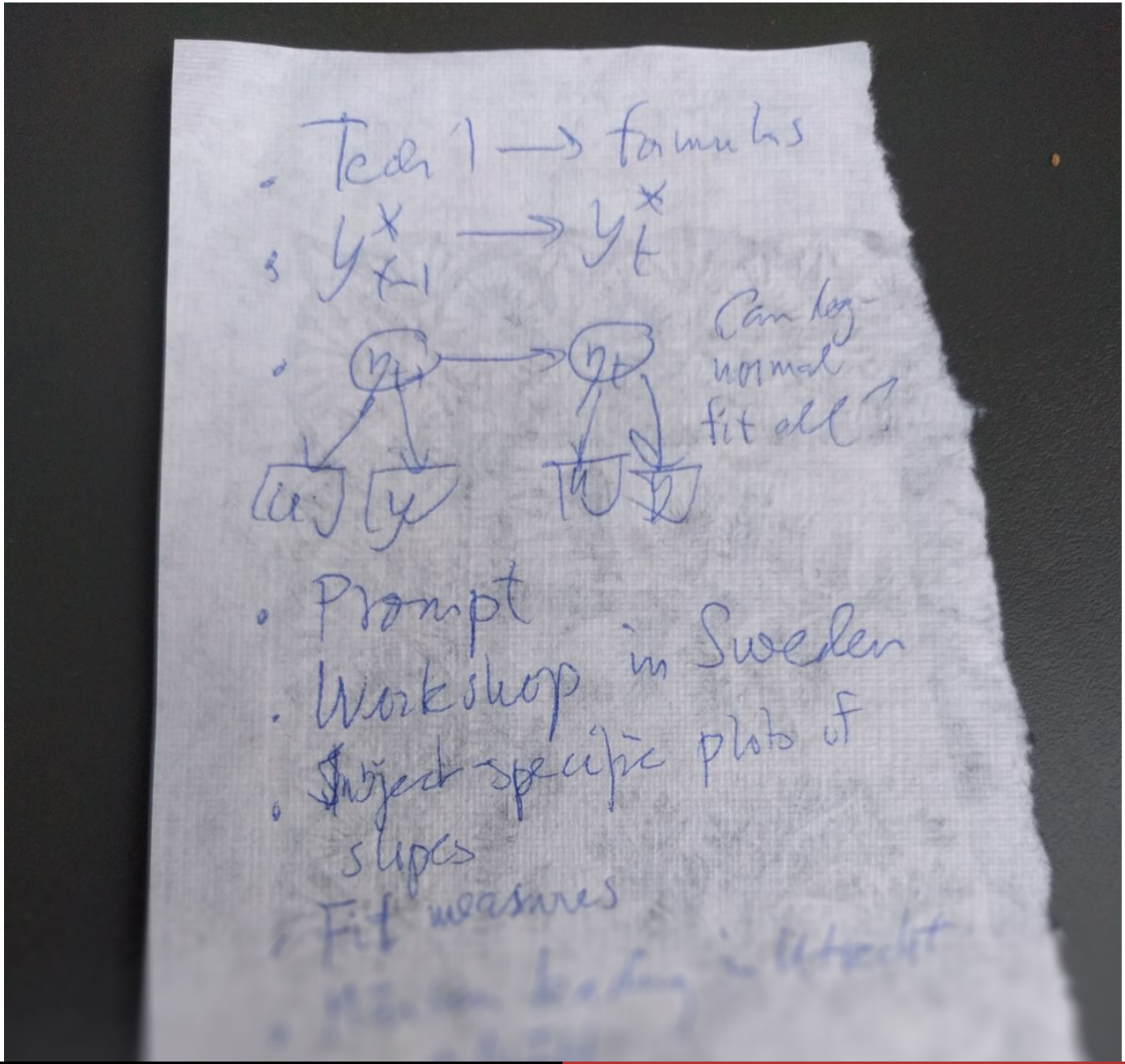
Overall: 42% at the floor value



Swiss Fondue at IMPS July 2017



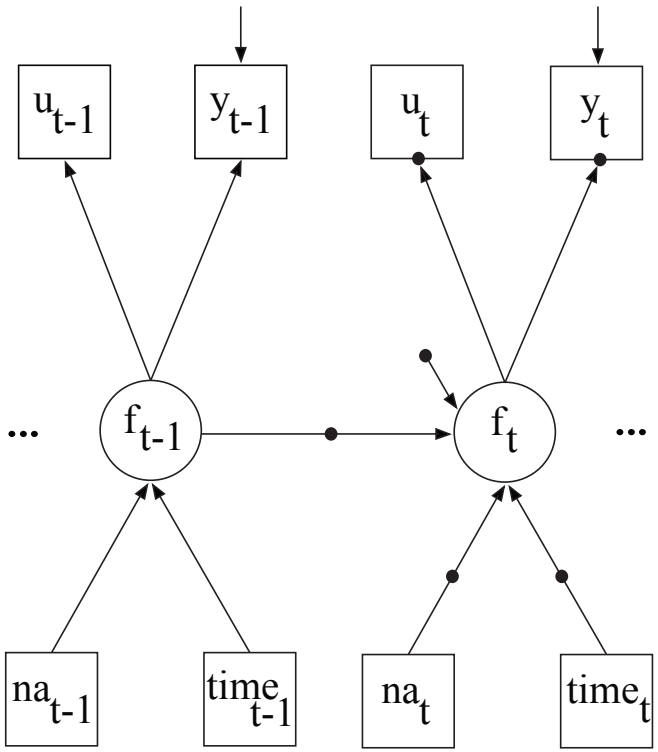
Ideas on a Napkin



Two-Part Modeling

- Transform the variable into 2 variables:
  - - a binary u and a continuous y (DATA TWOPART)
- u = 0 if at the floor: y is missing
- u = 1 if not at the floor: y is observed
- Probit model for u
- Log normal model for y

Within-level model:



Two-Part DSEM Input

```
DATA:       FILE = twopart.dat; ! DATA TWOPART done in a previous run
VARIABLE:   NAMES = subject timeqd urge negaff age gender quit u positive;
            USEVARIABLES = quit u positive negaff age female;
            CATEGORICAL = quit u;
            CLUSTER = subject;
            BETWEEN = quit age female;
            WITHIN = negaff;
            MISSING = *;
            TINTERVAL = timeqd(0.08);
DEFINE:     female = gender - 1;
            age = (age-44.3)/10.1;
ANALYSIS:   TYPE = TWOLEVEL RANDOM;
            ESTIMATOR = BAYES;
            PROCESSORS = 2;
            BITERATIONS = (20000);
            THIN = 10;
MODEL:      %Within%
            f BY u positive (&1);
            phi | f ON f&1;
            syx | f ON negaff;
            logv | f;
            negaff;
```

Two-Part DSEM Input Continued

```
                    %Between%
                    positive u phi syx logv ON female age;
                    positive u phi syx logv WITH positive u phi syx logv;
                    quit ON positive u phi syx logv female age;
OUTPUT:             TECH1 TECH8 TECH4 RESIDUAL STANDARDIZED
                    FSCOMPARISON;
PLOT:               TYPE = PLOT3
                    FACTORS = ALL;
```

Two-Part DSEM Between-Level Results

The run takes a very long time

Outline

- Two-part DSEM
- DSEM with cycles (circadian rythm)
- **Multilevel time series analysis is in the air**

Multilevel Time Series Analysis of ILD: We Built It - Will They Come?

Here are some positive signs...

New U.S. National Institutes of Health (NIH) Funding Opportunities Spring 2017

Funding Opportunity Title

Intensive Longitudinal Analysis of Health Behaviors: Leveraging New Technologies to Understand Health Behaviors (U01)

National Institutes of Health (NIH)

Office of Behavioral and Social Sciences Research (OBSSR)
National Cancer Institute (NCI)
National Institute on Alcohol Abuse and Alcoholism (NIAAA)
National Institute on Drug Abuse (NIDA)
National Institute of Mental Health (NIMH)

Purpose:

This (FOA) is intended to provide funding to encourage research projects that seek to explain underlying mechanisms and predict health behaviors within individuals over time utilizing intensive longitudinal, within-person protocols that leverage recent advances in mobile and wireless sensor technologies and big data analytics. The research projects will collect and analyze data, disseminate project findings, and work collaboratively with each other and the research coordinating center (supported under RFA-OD-17-005).

In contrast, a within-person approach to health behavior theory research seeks to explain why a given individual engages in healthy or risky behaviors at one time versus another. Within-person analysis of intensive longitudinal data is likely to provide insight into the dynamic factors in the physical, social, and/or built environment that facilitate or hinder engaging in certain behaviors

Biostatistics (2017) **18**, 3, *pp.* 403–404
doi:10.1093/biostatistics/kxx018
Advance Access publication on 12 June 2017

Discussion: The FDA is Unprepared for Personalized Medicine

ALEX TABARROK

Department of Economics, George Mason University, Fairfax, VA, USA

Tabarrok@gmu.edu

The Food and Drug Administration (FDA) is unprepared for the new world of personalized medicine. Consider what is possible now that nearly everyone carries with them the processing power of a 1990s Cray supercomputer. Smartphones equipped with sensors can monitor blood pressure, perform ECGs, even analyze DNA. Other devices being developed or available now include contact lens that can track glucose levels and eye pressure, real-time gait analysis, and head-bands that monitor and even adjust your brain waves.

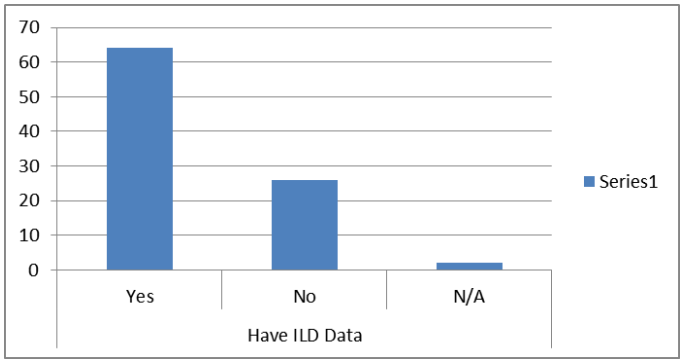
High Registration Numbers for Our Summer Workshops

- Utrecht DSEM workshop July 13:
  - 130 registrants
- Johns Hopkins Baltimore DSEM workshop Aug. 17-18 (videotaped):
  - 233 registrants
    - Students: 80
    - Post-docs: 31
    - Others: 122
- - And strong interest in time series analysis at the International Meeting of the Psychometric Society in July

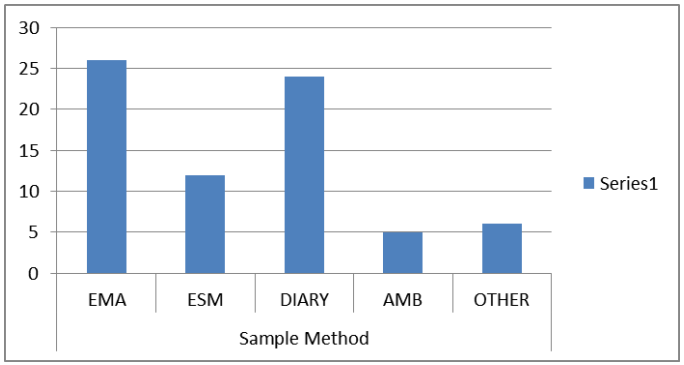
Registrant Survey

- Do you have or plan to collect intensive longitudinal data (ILD)? If so,
- What is your data collection method (EMA, ESM, Diary, Ambulatory, etc)?
- What sample size (N) and number of time points (T) do you have?
- How many time points per day?
- Are the time points fixed or random?
- What is the substantive area of your study?
- What is your primary outcome variable?
- Do you expect trends, cycles?
- Do you have experience with analyzing ILD?

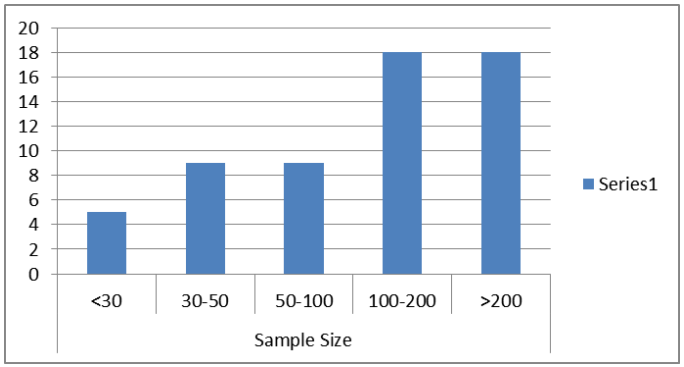
Survey Results



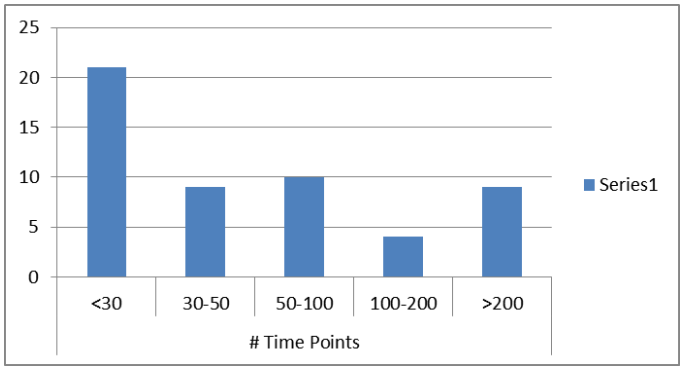
(a) Have/Plan to Collect ILD Data



(b) Sampling Method (EMA, etc)



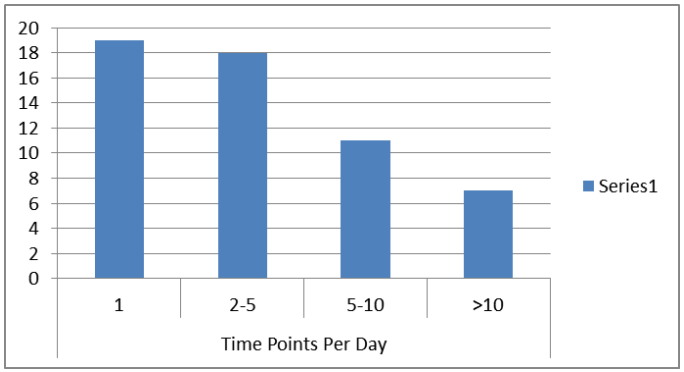
(c) Sample Size



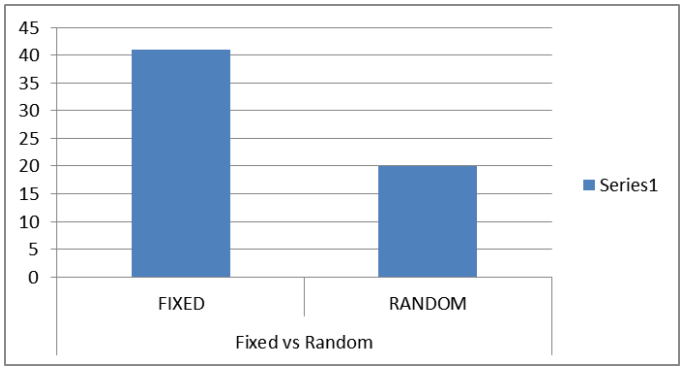
(d) Number of Time Points

Figure : Survey answers, part 1

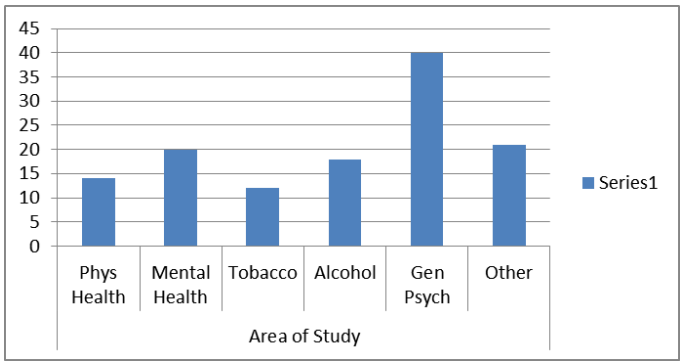
Survey Results Continued



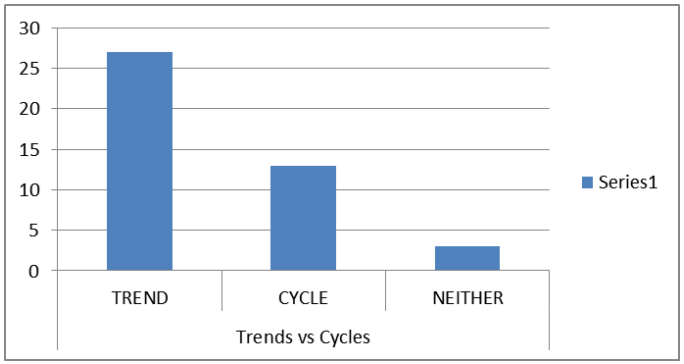
(a) Time Points Per Day



(b) Fixed vs Random Time Points



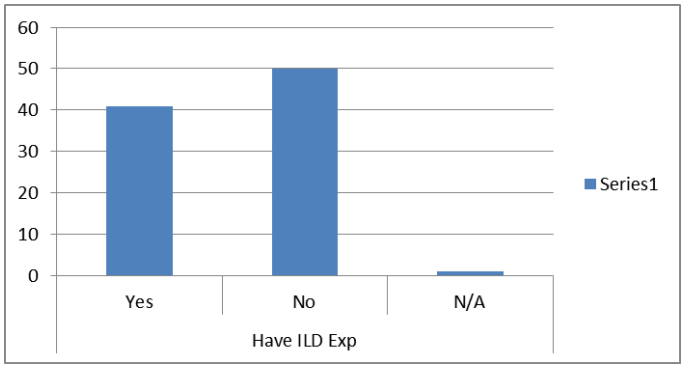
(c) Area of Study



(d) Trends vs Cycles

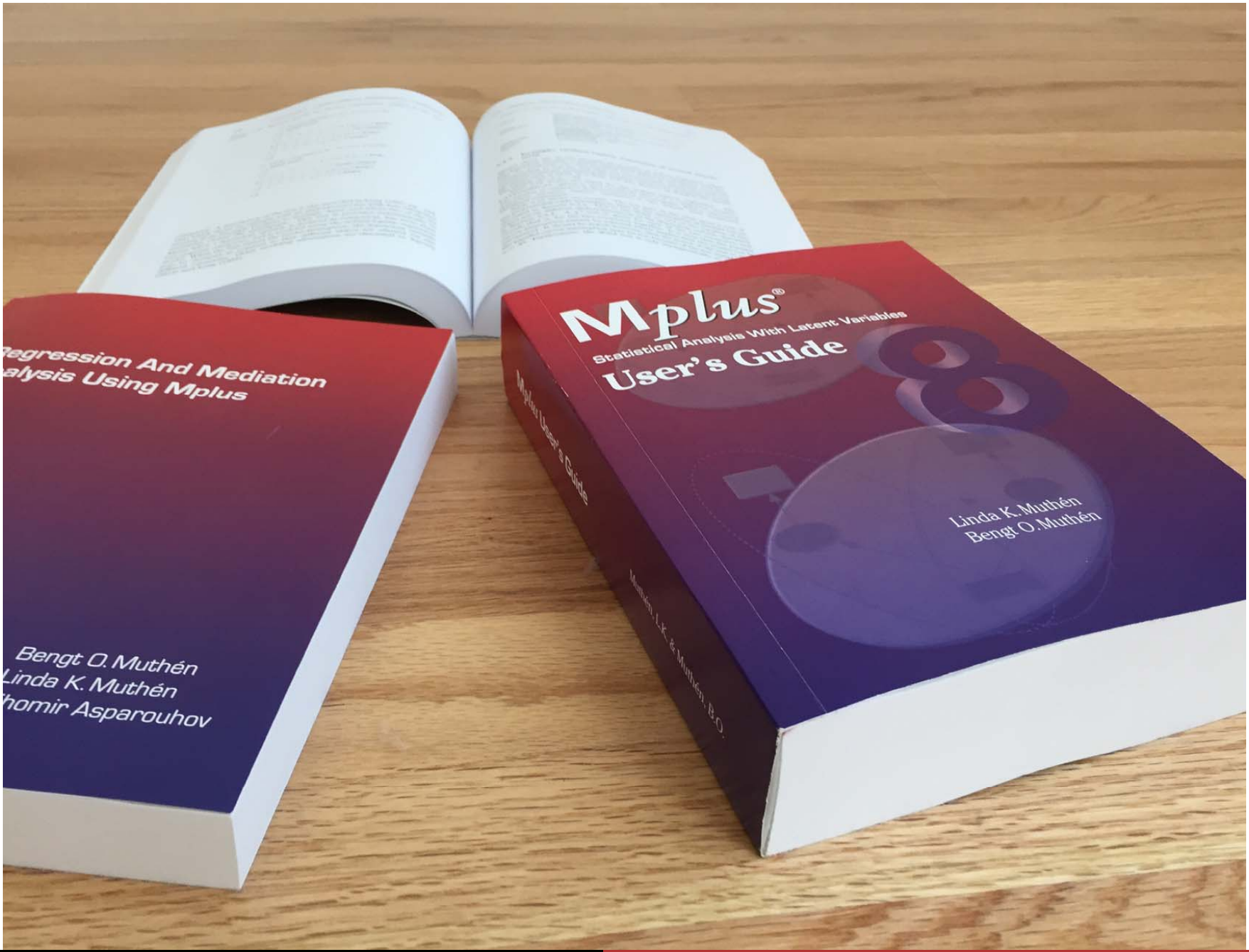
Figure : Survey answers, part 2

Final Survey Result



(a) ILD Experience

Figure : We have work to do!


Mplus®
Statistical Analysis With Latent Variables
User's Guide
Linda K. Muthén
Bengt O. Muthén
Bengt O. Muthén
Linda K. Muthén