LONGITUDINAL MODELING AND MISSING DATA HANDLING IN BLIMP

Craig K. Enders, UCLA

WWW.APPLIEDMISSINGDATA.COM/BLIMP

BLIMP 3.0

Blimp 3 offers powerful latent variable modeling and imputation for incomplete data sets with up to three levels. Blimp's unique Bayesian computational architecture allows easy specification of complex analyses that are difficult or impossible to fit in other software packages.

Download Now

User's Guide

WWW.APPLIEDMISSINGDATA.COM/BLIMP-PAPERS

APPLIED MISSING DATA

home analysis examples blimp **blimp papers** centerstat workshop quantitude podcast Contact Me

Workshops and Training

Enders, C. K. (2023, March). Longitudinal modeling and missing data handling in Blimp. Workshop presented at the Advanced Techniques for Longitudinal Data Analysis in Social Science (ATLASS) conference. Bielefeld, Germany.

Download

ACKNOWLEDGEMENTS

Brian Keller & Han Du

IES Institute of Education Sciences

WORKSHOP OUTLINE

- Linear growth
- Nonlinear change
- Incomplete predictors
- Interaction effects
- Heterogeneous within-cluster variation
- Structural modeling
- Missing not at random processes
- Nonnormal data

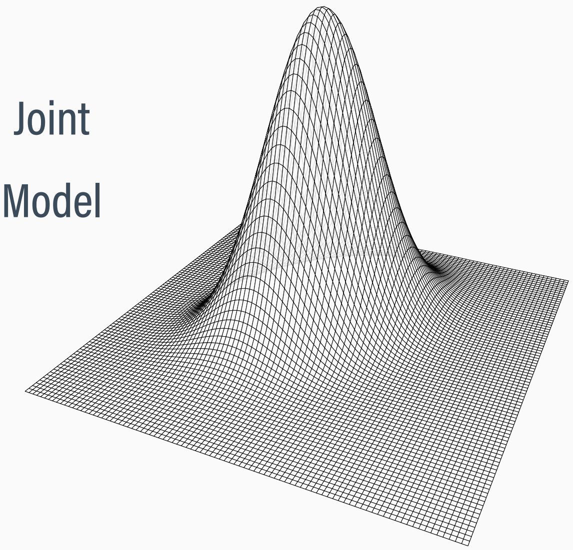
THE BLIMP MODELING FRAMEWORK

MODELING FRAMEWORKS

Multivariate modeling

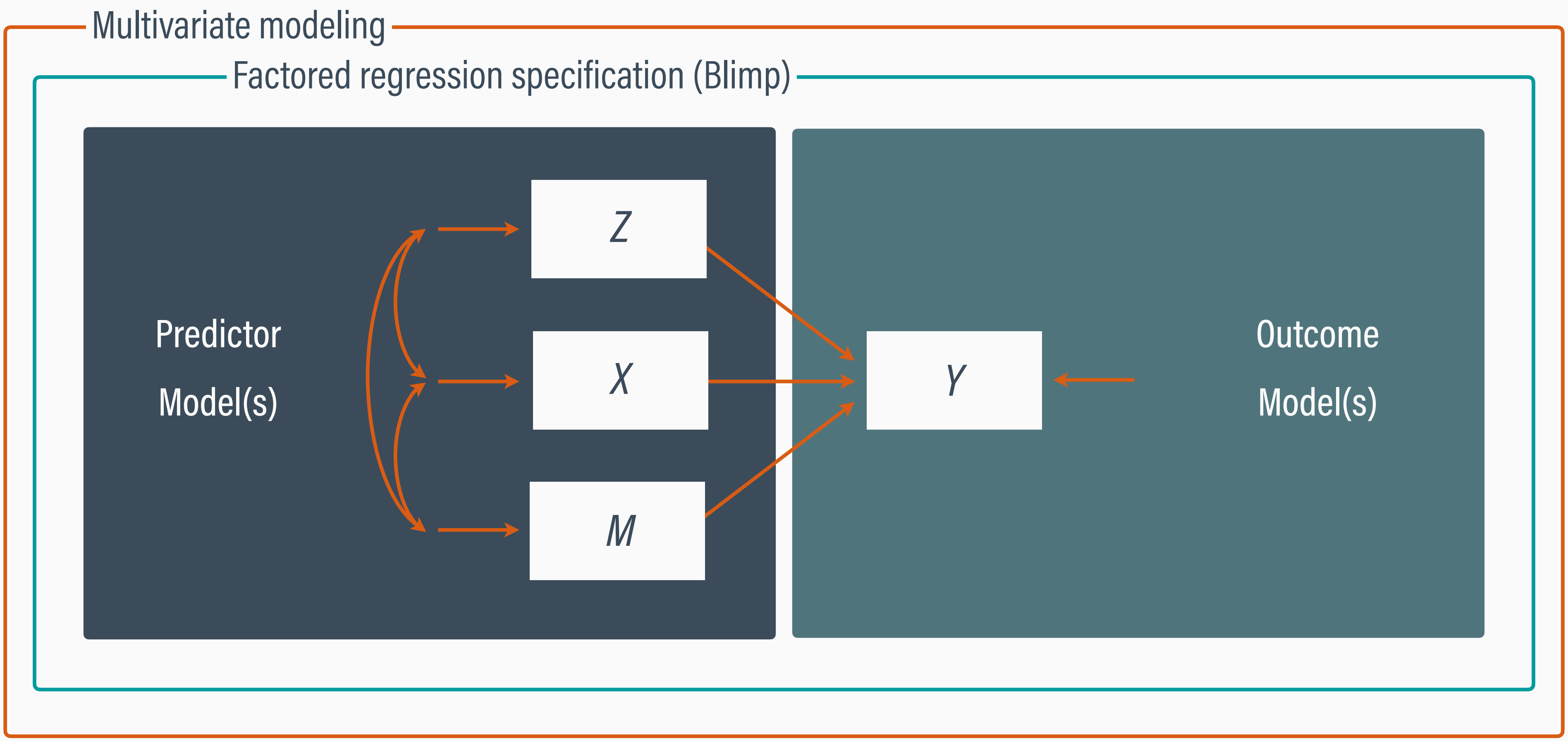
Joint

Model

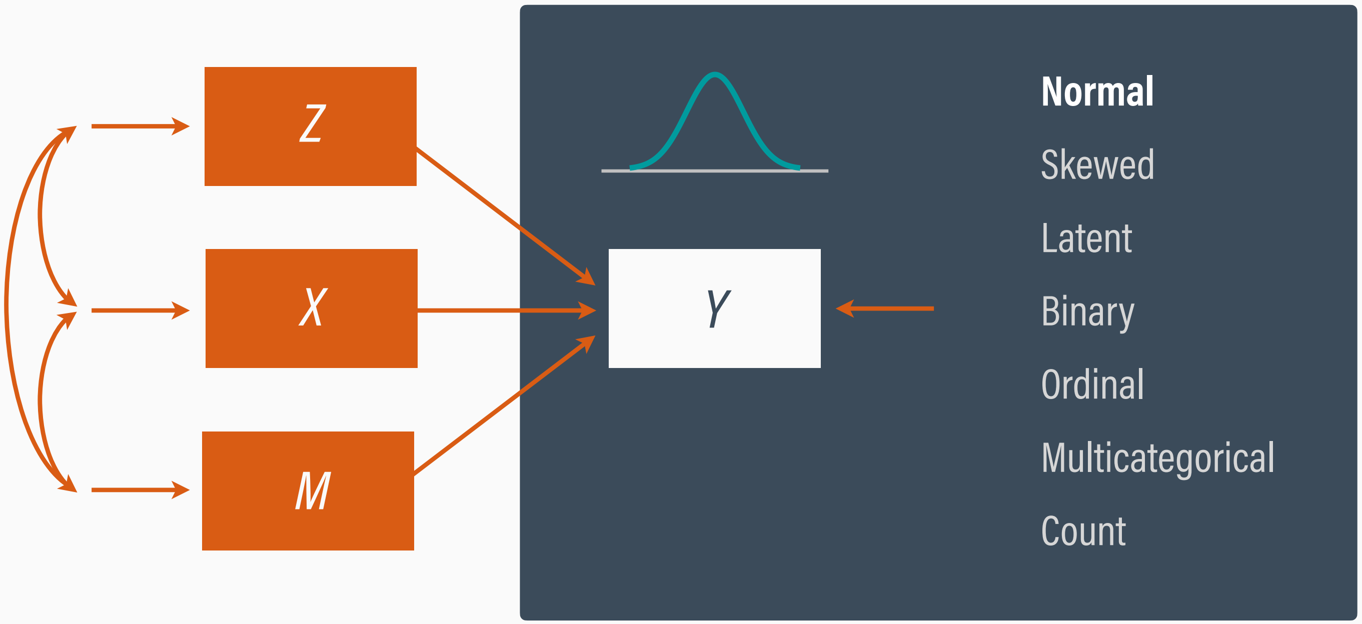


- Classic methods often assume a multivariate distribution (e.g., normal)
- e.g., FIML and joint model multiple imputation

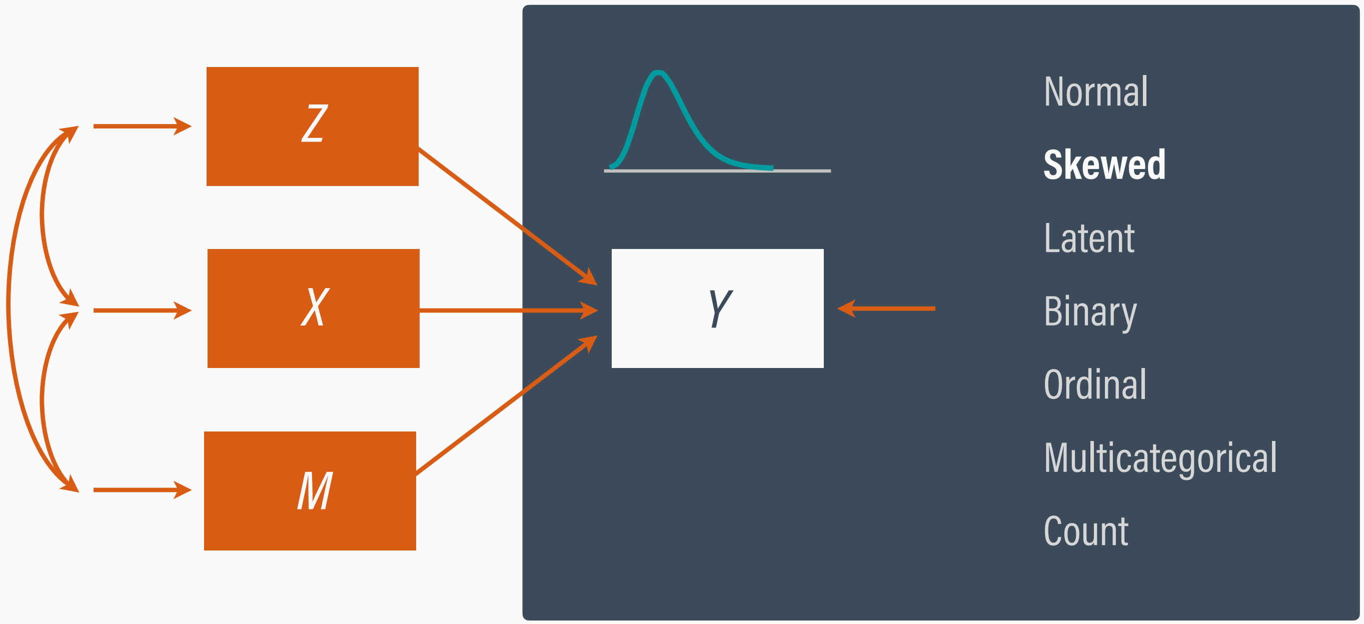
MODELING FRAMEWORKS



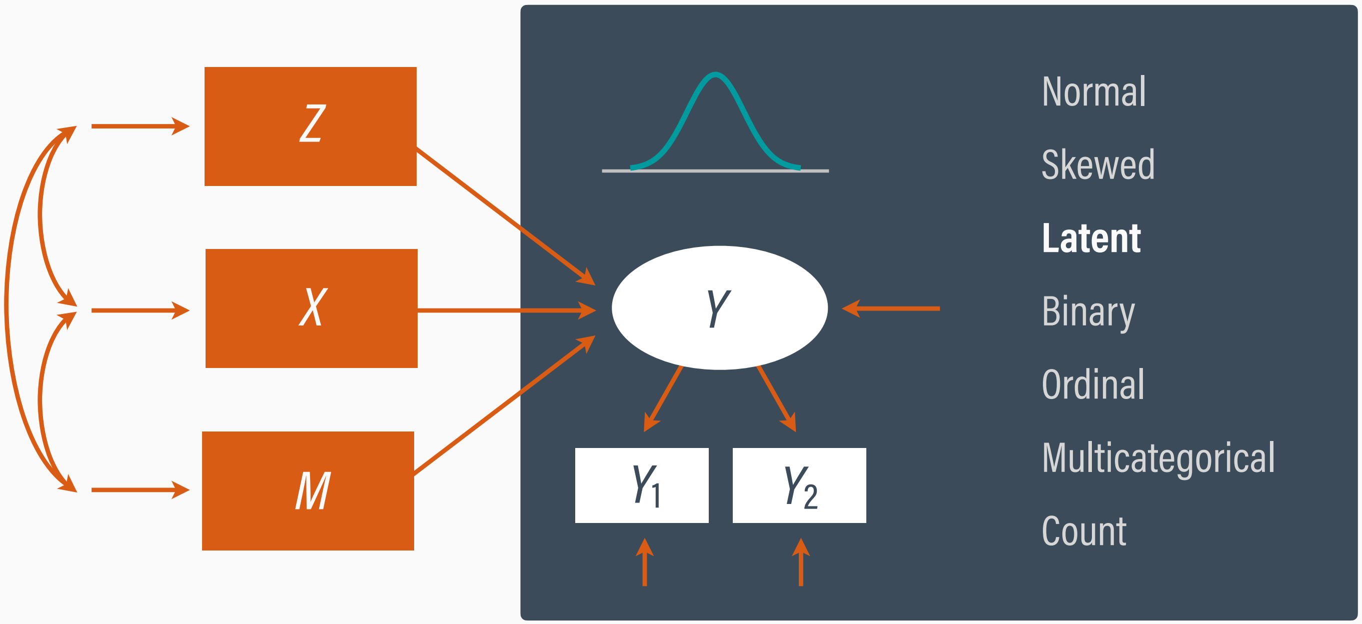
OUTCOME VARIABLE TYPES



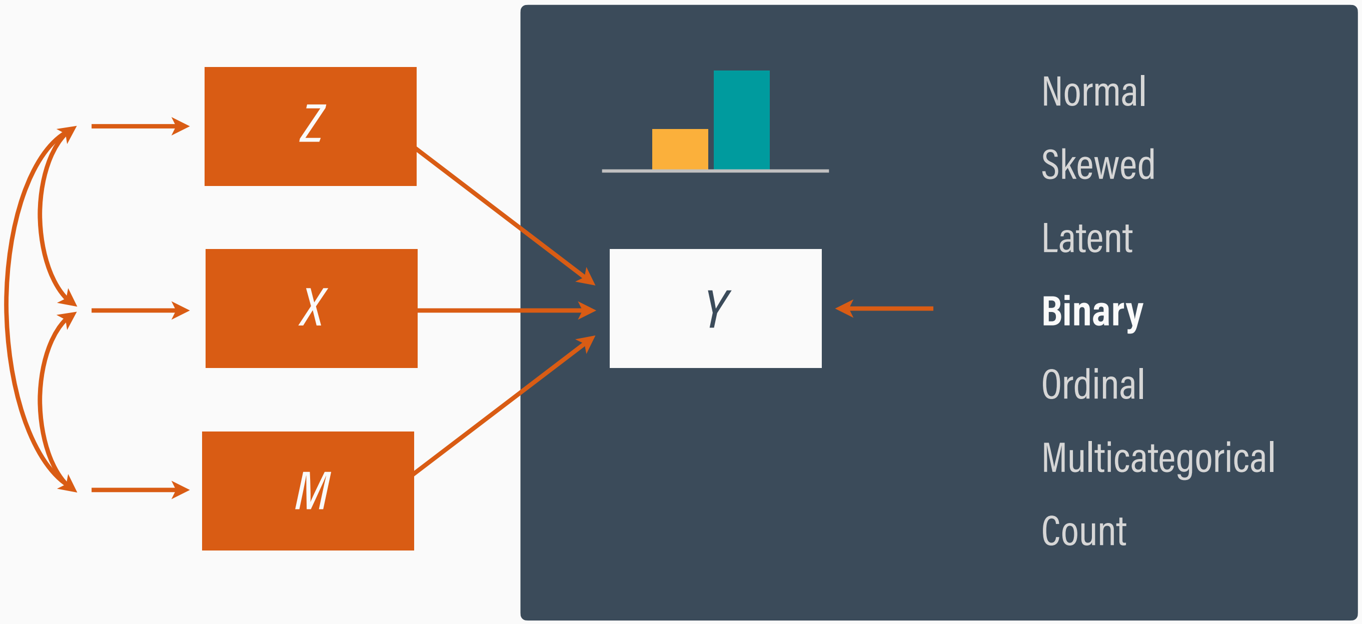
OUTCOME VARIABLE TYPES



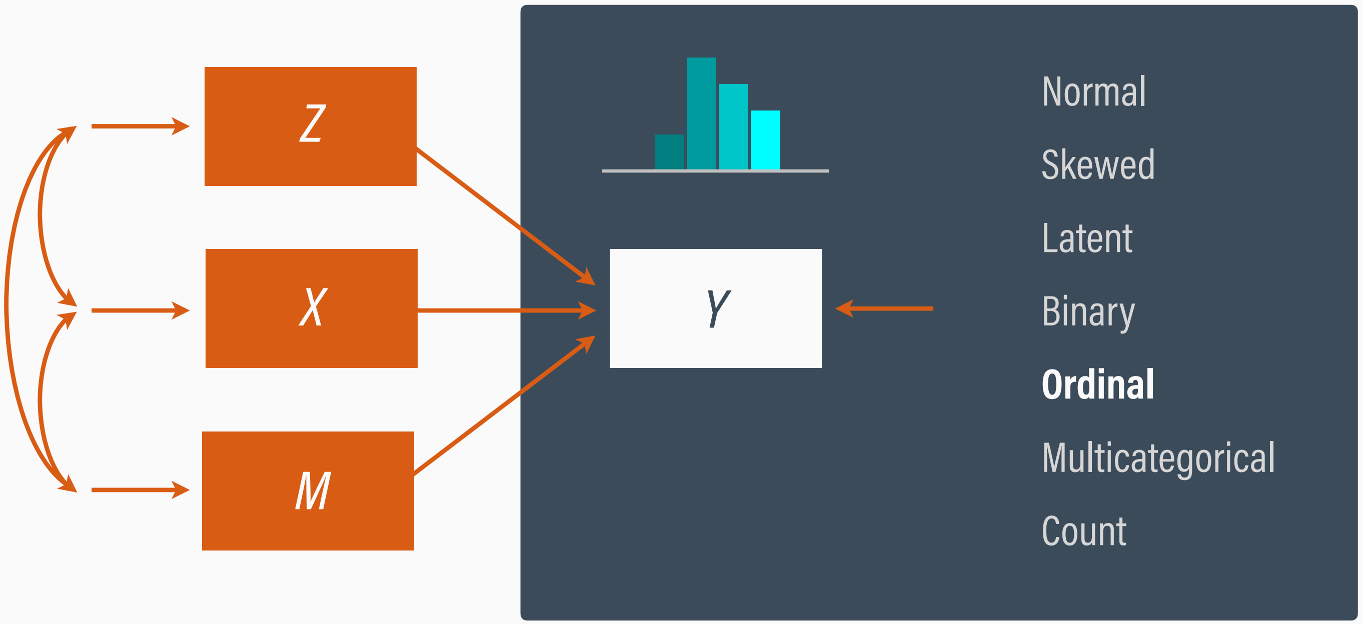
OUTCOME VARIABLE TYPES



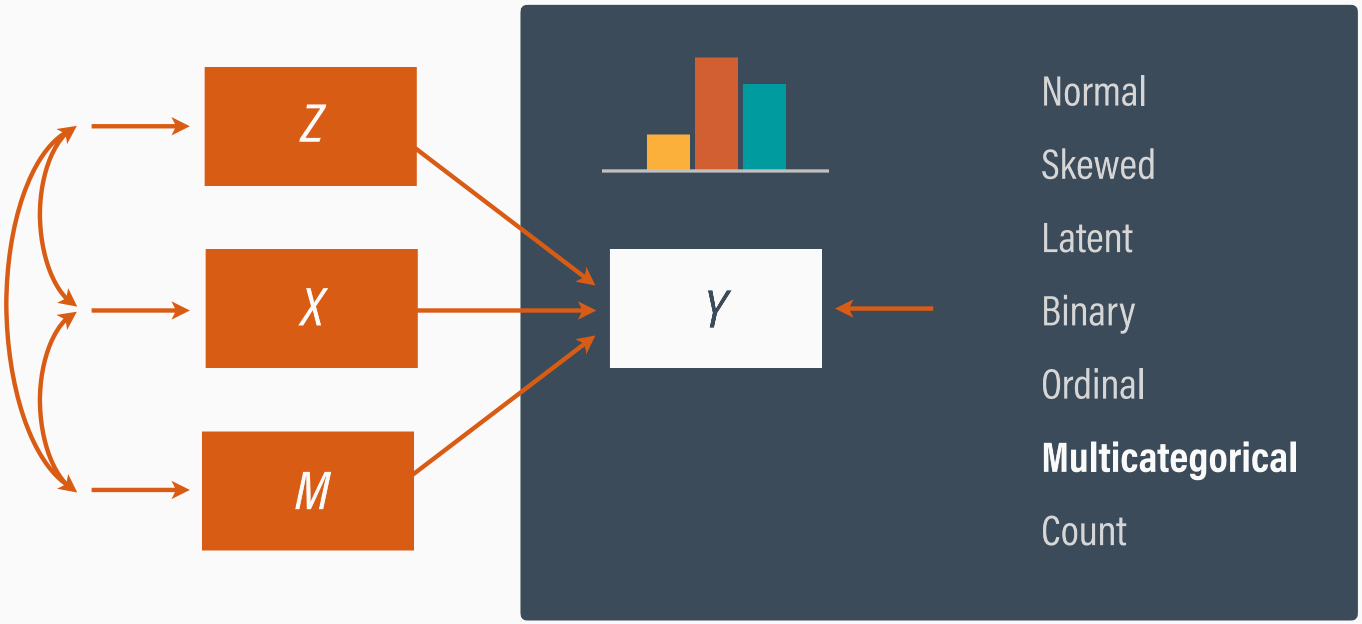
OUTCOME VARIABLE TYPES



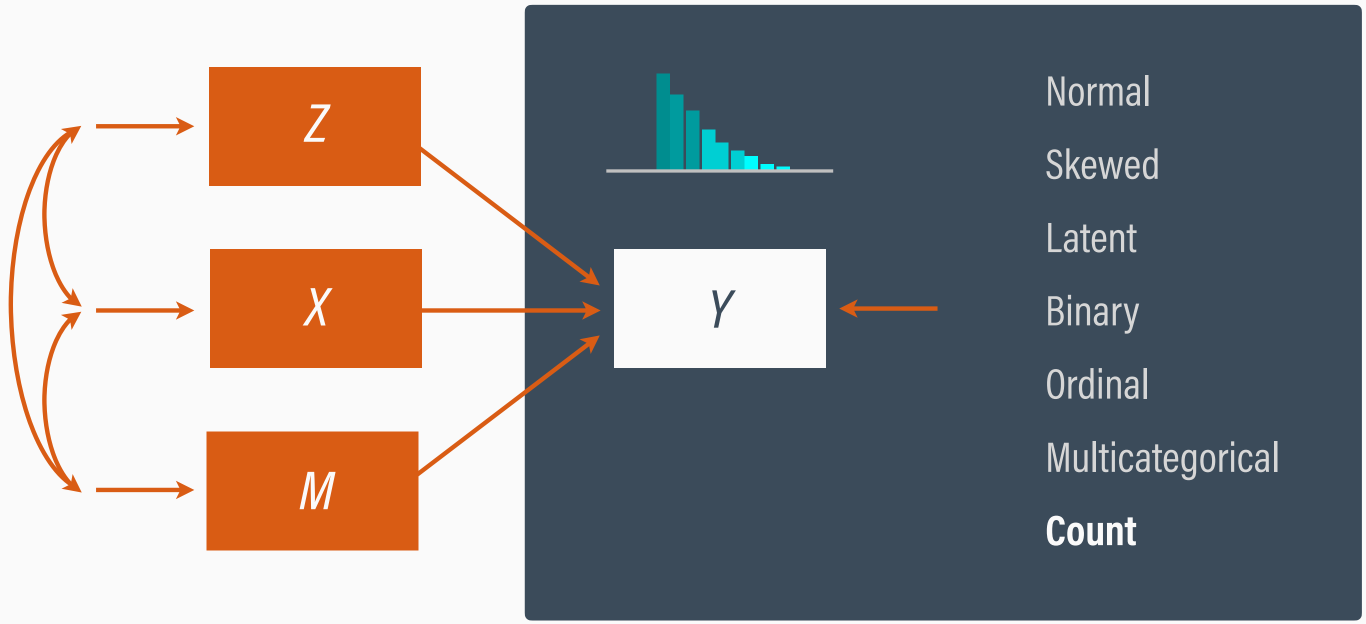
OUTCOME VARIABLE TYPES



OUTCOME VARIABLE TYPES

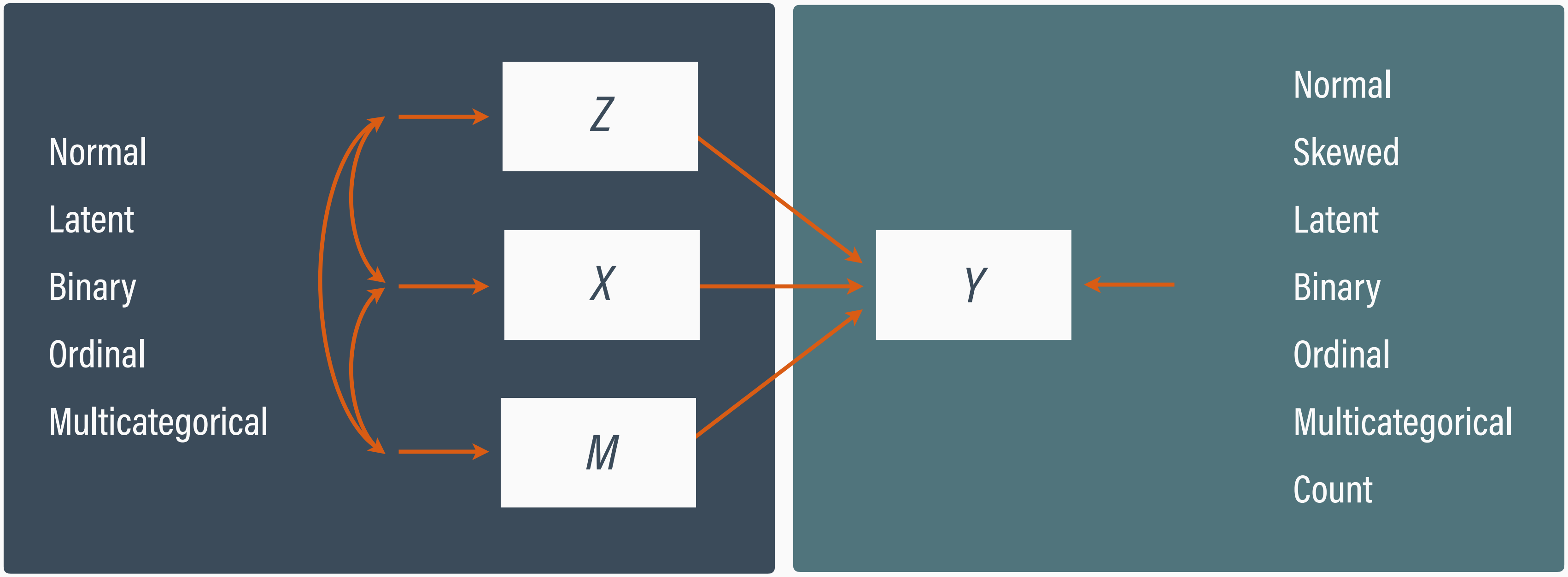


OUTCOME VARIABLE TYPES

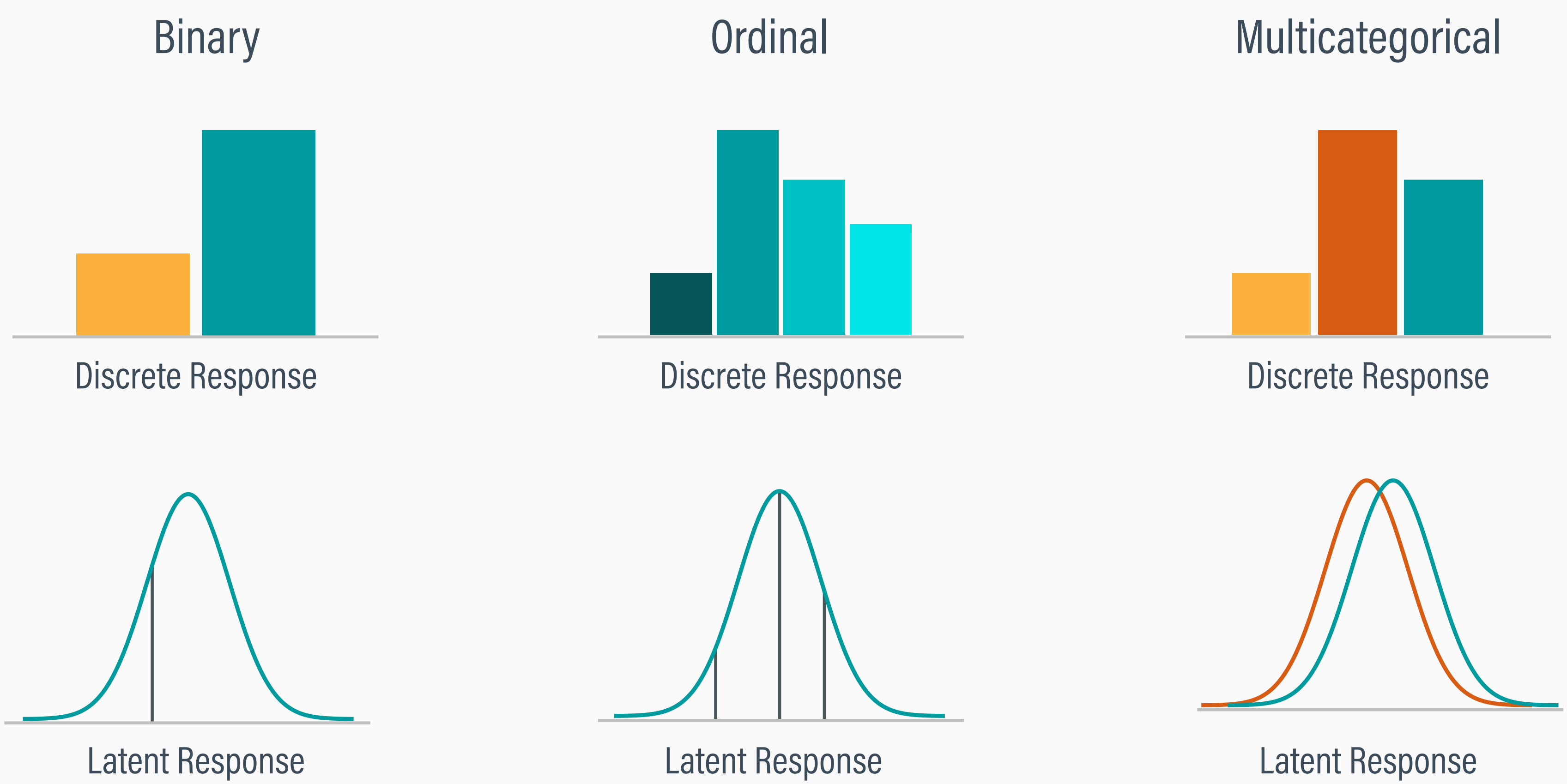


PREDICTOR VARIABLE TYPES

Multivariate Distribution = Univariate Outcome Model × Multivariate Predictor Model

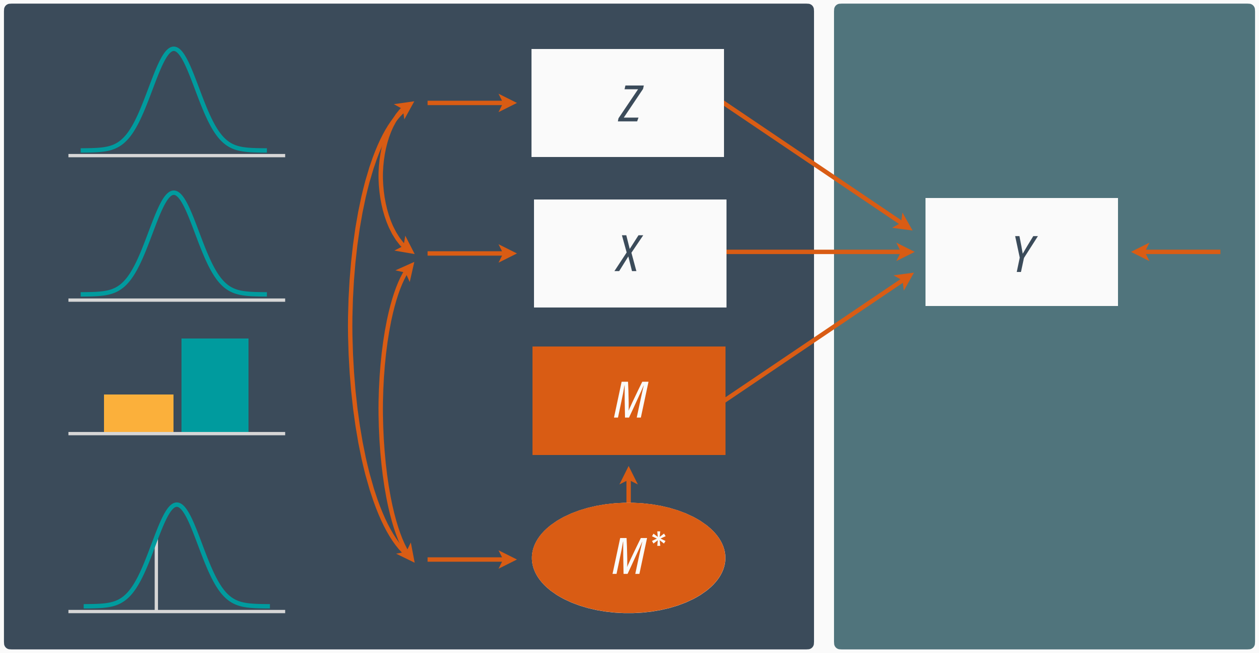


LATENT RESPONSE FORMULATION



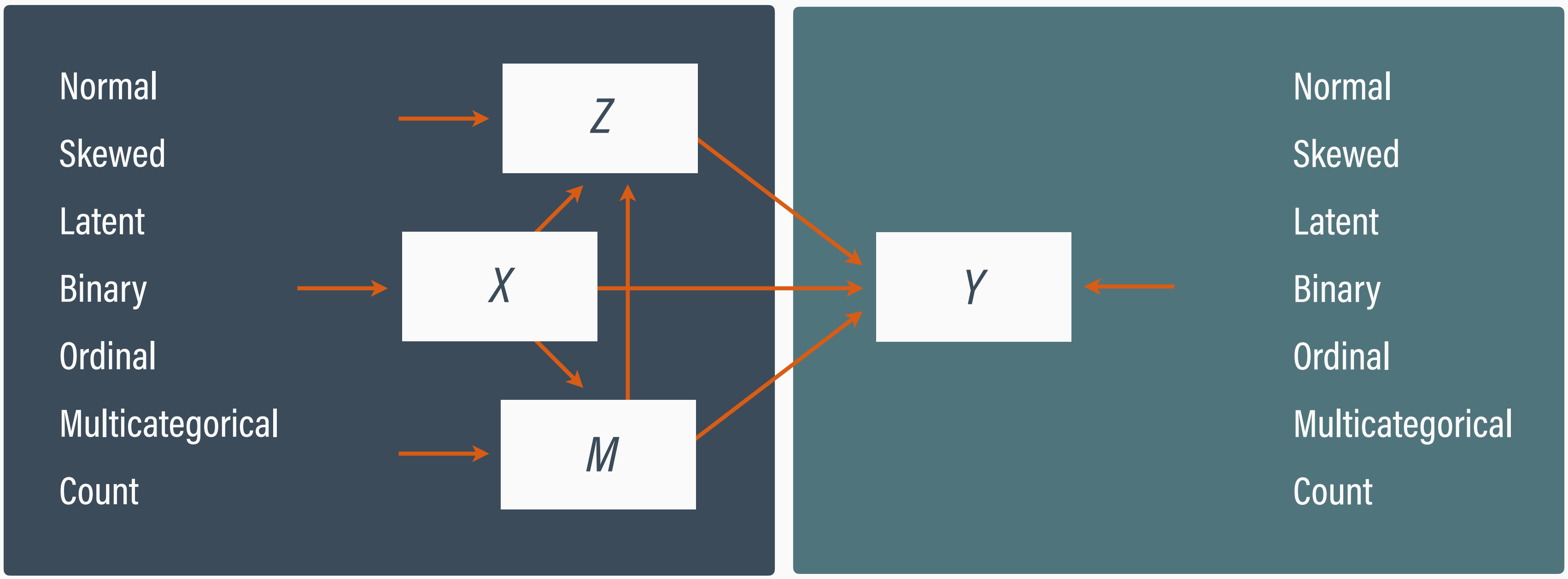
CATEGORICAL PREDICTOR

Multivariate Distribution = Univariate Outcome Model × Multivariate Predictor Model



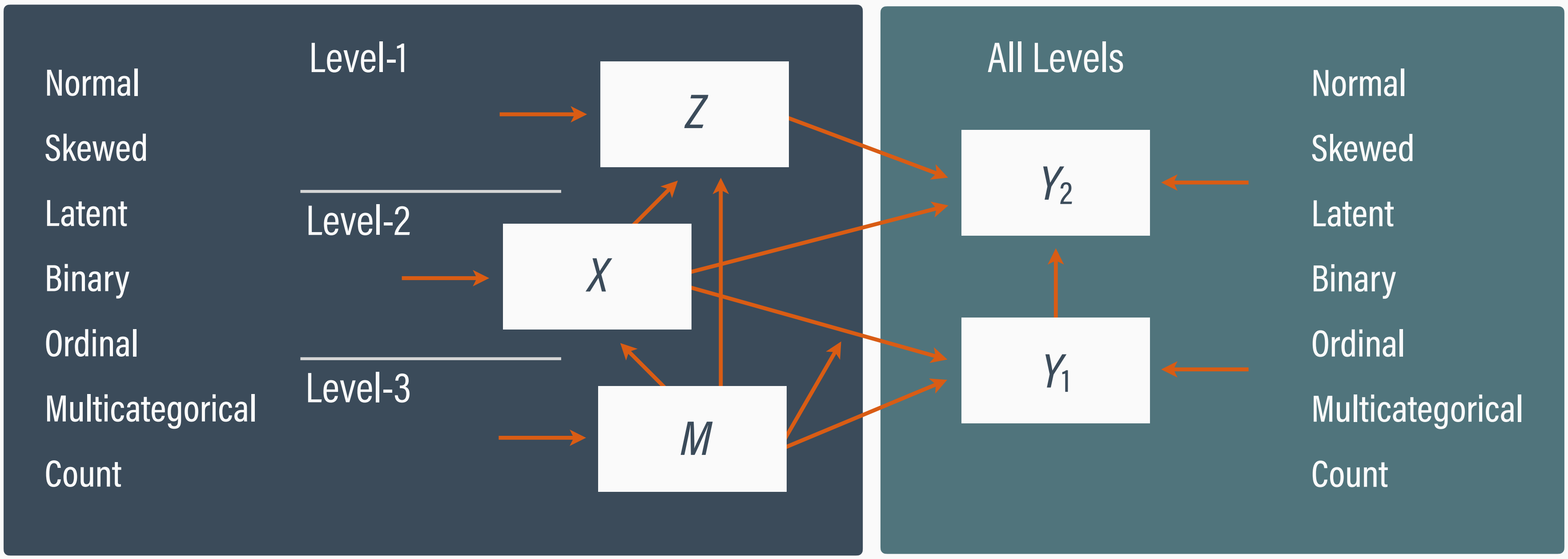
SEQUENTIAL PREDICTOR TYPES

Multivariate Distribution = Univariate Outcome Model × Univariate Predictor Models



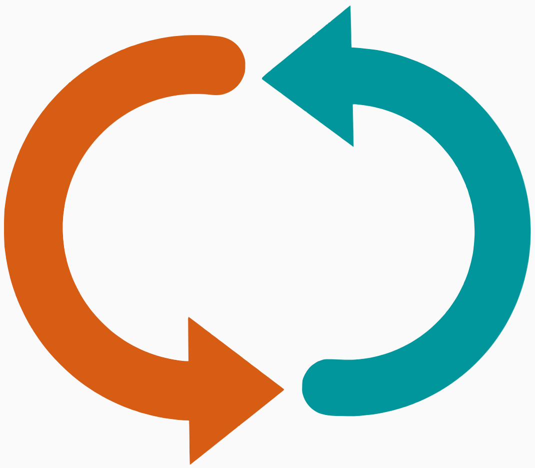
MULTILEVEL MODELING

Multivariate Distribution = Univariate Outcome Model × Multivariate or Univariate Predictor Models



MCMC ESTIMATION

Estimate regression models

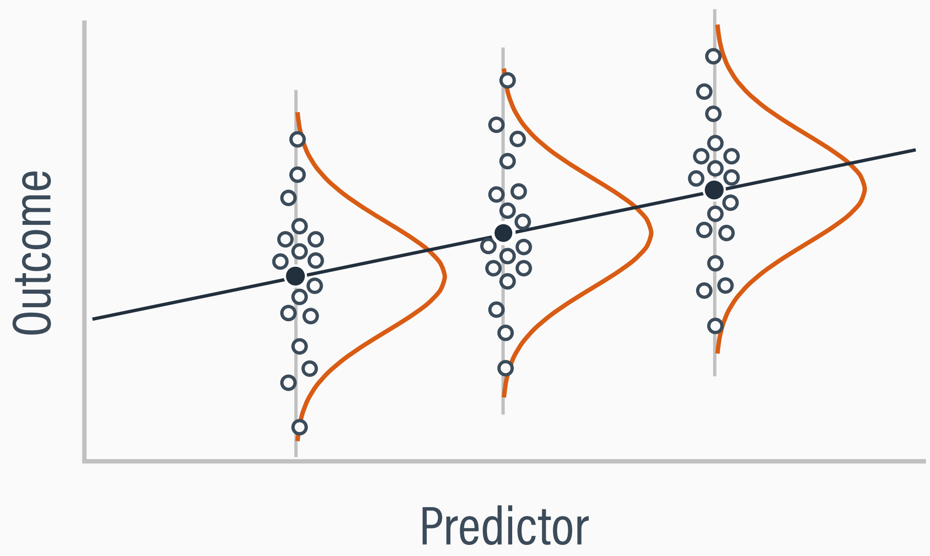


Impute missing values

- Do for $t = 1$ to T iterations
  - Estimate focal model parameters, conditional on filled-in data
  - Estimate predictor model parameters, conditional on filled-in data
  - Impute outcome scores, conditional on the focal model parameters
  - Impute predictors, conditional on the focal and predictor model parameters
- Repeat

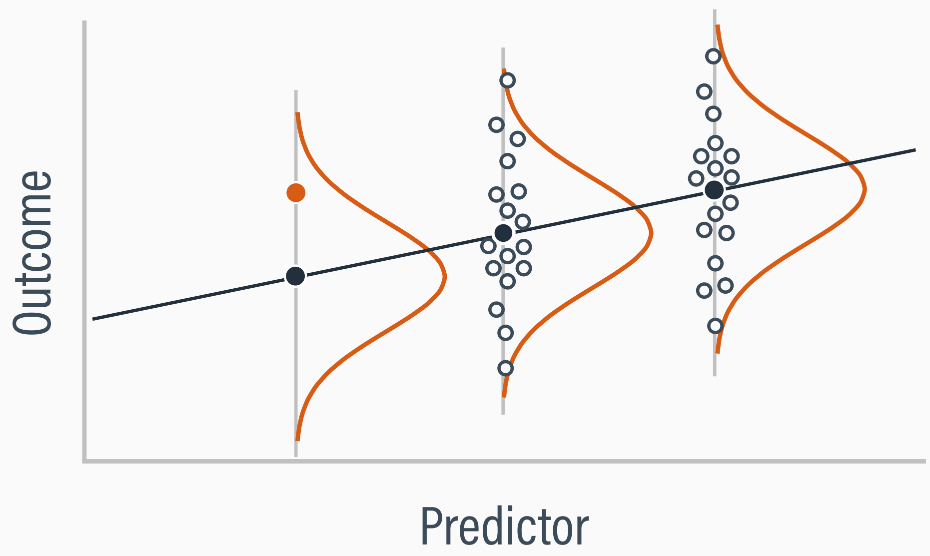
MISSING DATA IMPUTATION

- Imputations are sampled at random from distributions of plausible values
- One or more sets of model parameters define the center and spread



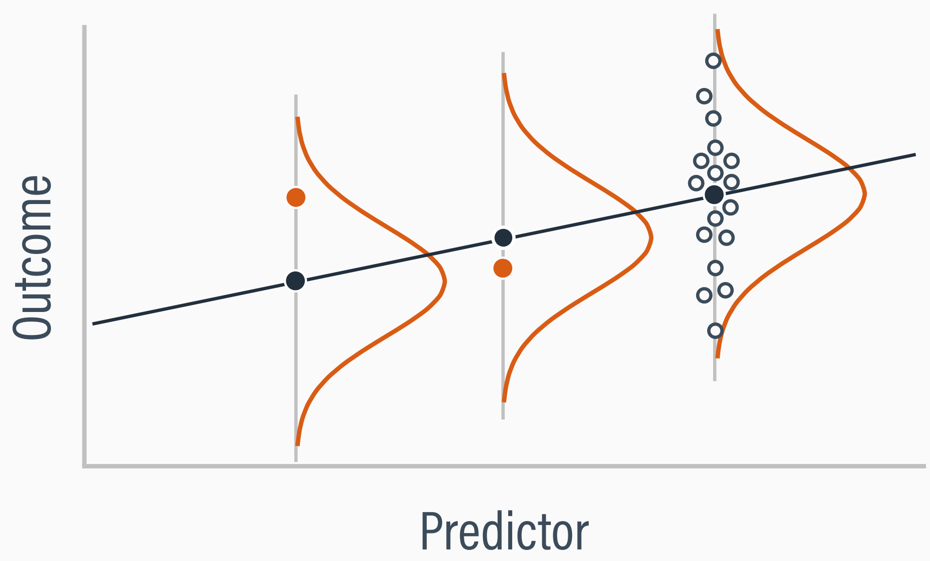
MISSING DATA IMPUTATION

- Imputations are sampled at random from distributions of plausible values
- One or more sets of model parameters define the center and spread



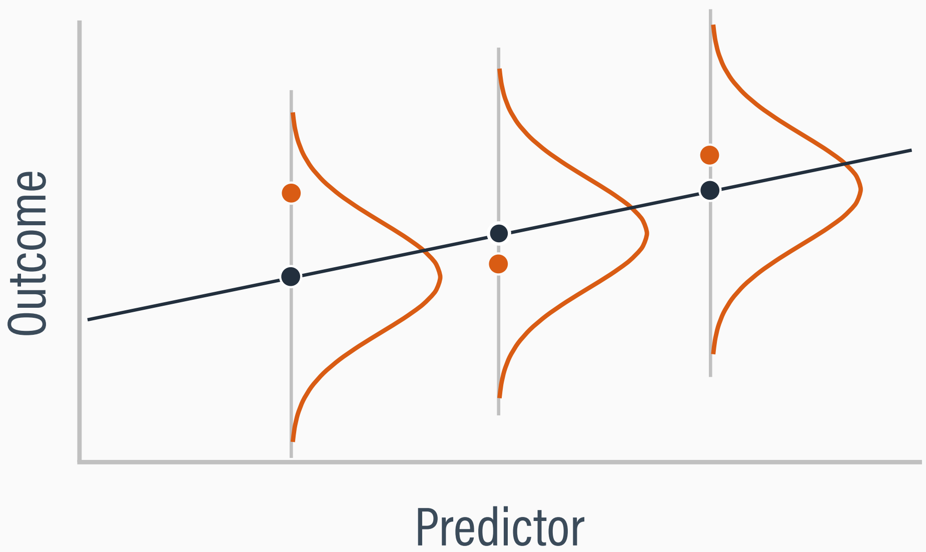
MISSING DATA IMPUTATION

- Imputations are sampled at random from distributions of plausible values
- One or more sets of model parameters define the center and spread



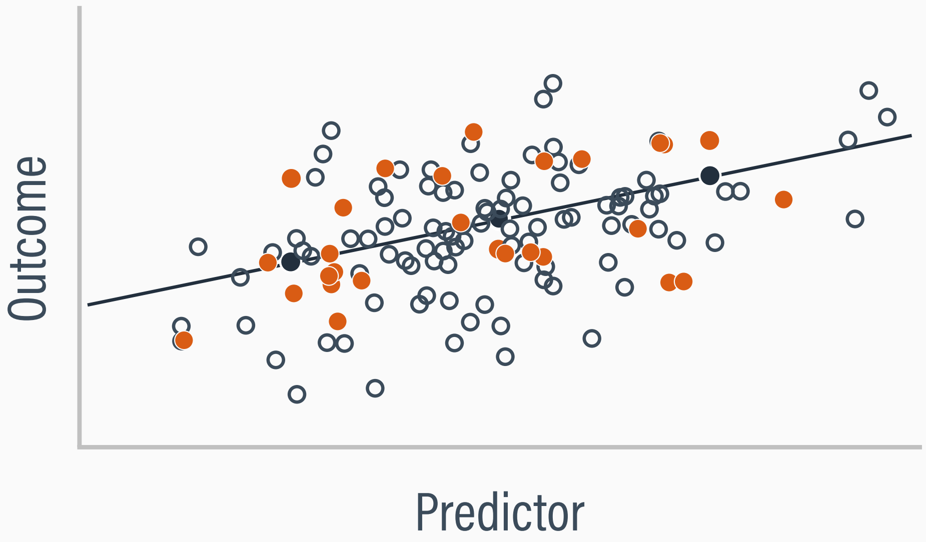
MISSING DATA IMPUTATION

- Imputations are sampled at random from distributions of plausible values
- One or more sets of model parameters define the center and spread



MISSING DATA IMPUTATION

- Imputations are sampled at random from distributions of plausible values
- One or more sets of model parameters define the center and spread



MISSING DATA MECHANISMS

PARTITIONING THE DATA

Complete = Observed + Missing | Indicators

Y_1	Y_2	Y_3		Y_1	Y_2	Y_3		Y_1	Y_2	Y_3	M_1	M_2	M_3
4	4	3		4	4	3					0	0	0
3	**3**	5		3	**NA**	5			**3**		0	**1**	0
7	1	6		7	1	6					0	0	0
2	1	6		**NA**	1	6		**2**			**1**	0	0
5	9	3	=	5	9	3	+				0	0	0
3	**2**	**2**		3	**NA**	**NA**			**2**	**2**	0	**1**	**1**
1	6	7		1	6	7					0	0	0
9	4	9		9	4	9					0	0	0
2	**5**	6		2	**NA**	6			**5**		0	**1**	0

RUBIN'S MISSINGNESS MECHANISMS

- Missing data mechanisms describe different ways in which the pattern of 0s and 1s in M relate to the observed or missing data
- Missingness may be independent of the data, or it could relate to the observed or missing parts (or both)
- Mechanisms are essentially missingness models

MISSING COMPLETELY AT RANDOM

- The probability of missing values is completely unrelated to the data

$$f(\mathbf{M} = 1 | \mathbf{Y}_{\text{obs}}, \mathbf{Y}_{\text{mis}}, \phi) = f(\mathbf{M} = 1 | \phi)$$

- Purely haphazard missingness

Missingness Conditioning on the Data

M			Y_{obs}			Y_{mis}		
M_1	M_2	M_3	Y_1	Y_2	Y_3	Y_1	Y_2	Y_3
0	0	0	4	4	3			
0	1	0	3	NA	5		3	
0	0	0	7	1	6			
1	0	0	NA	1	6	2		
0	0	0	5	9	3			
0	1	1	3	NA	NA		2	2
0	0	0	1	6	7			
0	0	0	9	4	9			
0	1	0	2	NA	6		5	

(CONDITIONALLY) MISSING AT RANDOM

- The probability of missing values is unrelated to the missing (latent) data

$$f(\mathbf{M}=1|\mathbf{Y}_{\text{obs}},\mathbf{Y}_{\text{mis}},\phi)=f(\mathbf{M}=1|\mathbf{Y}_{\text{obs}},\phi)$$

- Missingness is haphazard after conditioning on observed data

Missingness Conditioning on the Data

M			Y_{obs}			Y_{mis}		
M_1	M_2	M_3	Y_1	Y_2	Y_3	Y_1	Y_2	Y_3
0	0	0	4	4	3			
0	1	0	3	NA	5		3	
0	0	0	7	1	6			
1	0	0	NA	1	6	2		
0	0	0	5	9	3			
0	1	1	3	NA	NA		2	2
0	0	0	1	6	7			
0	0	0	9	4	9			
0	1	0	2	NA	6		5	

MISSING NOT AT RANDOM

- The probability of missing values is related to the missing (latent) data

$$f(\mathbf{M}=1|\mathbf{Y}_{\text{obs}},\mathbf{Y}_{\text{mis}},\phi) \text{ or } f(\mathbf{M}=1|\mathbf{Y}_{\text{mis}},\phi)$$

- The observed data may or may not additionally determine missingness

Missingness Conditioning on the Data

M			Y_{obs}			Y_{mis}		
M_1	M_2	M_3	Y_1	Y_2	Y_3	Y_1	Y_2	Y_3
0	0	0	4	4	3			
0	1	0	3	NA	5		3	
0	0	0	7	1	6			
1	0	0	NA	1	6	2		
0	0	0	5	9	3			
0	1	1	3	NA	NA		2	2
0	0	0	1	6	7			
0	0	0	9	4	9			
0	1	0	2	NA	6		5	

CLINICAL TRIAL DATA

Predictors
Outcome

Variable	Definition	Missing %	Scale
PERSON	Person-level (level-2) identifier	0	Integer index
WAVE	Data collection wave	0	Integer index (1 to 4)
MONTH	Months relative to 4-week follow-up	0	Integer index (–1 to 2)
BREATHCO	Breath CO reading	23.94	Numeric (0 to 46)
DRINKS	Number of drinks per drinking day	15.00	Numeric (0 to 34)
CIGS	Number of cigarettes per smoking day	14.85	Numeric (0 to 50)
FEMALE	Gender dummy code	0	0 = Male, 1 = Female
CONDITION	Treatment condition	0	0 = Single-medication, 1 = Dual-medication
NICDEP	Nicotine dependence dummy code	6.67	0 = Very Low, to Low 1 = Moderate to Very High
QUIT16	Quit smoking at 4-month follow-up	33.94	0 = Did not quit, 1 = Quit smoking
DROPOUT	Dropout indicator	13.33	0 = In study 1 = Dropped out

LONGITUDINAL PROCESSES

- The conditional MAR process assumes that missingness at occasion *t* is explained by the observed values of other variables (covariates and repeated measurements)
- Two plausible MNAR processes: one's underlying growth trajectory is responsible for missing data, or the unseen value of the outcome at a particular occasion predicts nonresponse

MULTILEVEL GROWTH MODELS

MOTIVATING EXAMPLE

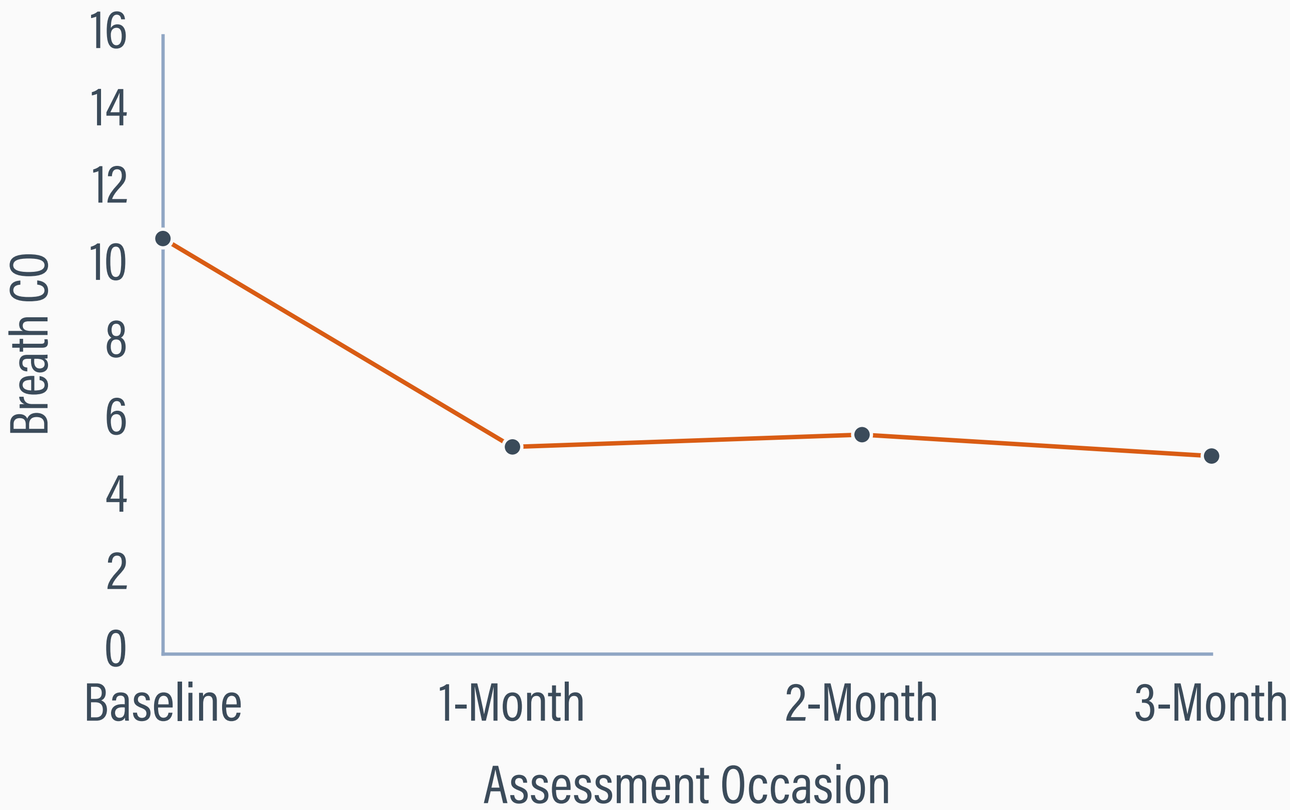
- Smoking cessation clinical trial with four repeated measurements nested in participants
- Two treatment arms: varenicline vs. varenicline + naltrexone
- Cigarette4 smoking (breath CO) assessed at baseline, 4-week, 8-week, and 12-week follow-ups

CLINICAL TRIAL DATA

Predictors
Outcome

Variable	Definition	Missing %	Scale
PERSON	Person-level (level-2) identifier	0	Integer index
WAVE	Data collection wave	0	Integer index (1 to 4)
MONTH	Months relative to 4-week follow-up	0	Integer index (–1 to 2)
BREATHCO	Breath CO reading	23.94	Numeric (0 to 46)
DRINKS	Number of drinks per drinking day	15.00	Numeric (0 to 34)
CIGS	Number of cigarettes per smoking day	14.85	Numeric (0 to 50)
FEMALE	Gender dummy code	0	0 = Male, 1 = Female
CONDITION	Treatment condition	0	0 = Single-medication, 1 = Dual-medication
NICDEP	Nicotine dependence dummy code	6.67	0 = Very Low, to Low 1 = Moderate to Very High
QUIT16	Quit smoking at 4-month follow-up	33.94	0 = Did not quit, 1 = Quit smoking
DROPOUT	Dropout indicator	13.33	0 = In study 1 = Dropped out

OBSERVED MEANS



GROWTH MODELING OVERVIEW

- Multilevel growth models express outcome changes as a function of a temporal predictor (e.g., change per month)
- The basic model features an average change trajectory (fixed effects) and person-specific trajectories (random effects)
- Person-specific trajectories can be outcomes and predictors

TIME SCORES

- The growth model requires a meaningful zero value for the temporal predictor (measurement occasion, passage of time)
- Substantive concerns should dictate centering, I use months since the 4-week follow-up assessment

Measurement	Baseline	4-week	8-week	12-week
Time Scores (MONTH)	–1	0	1	2

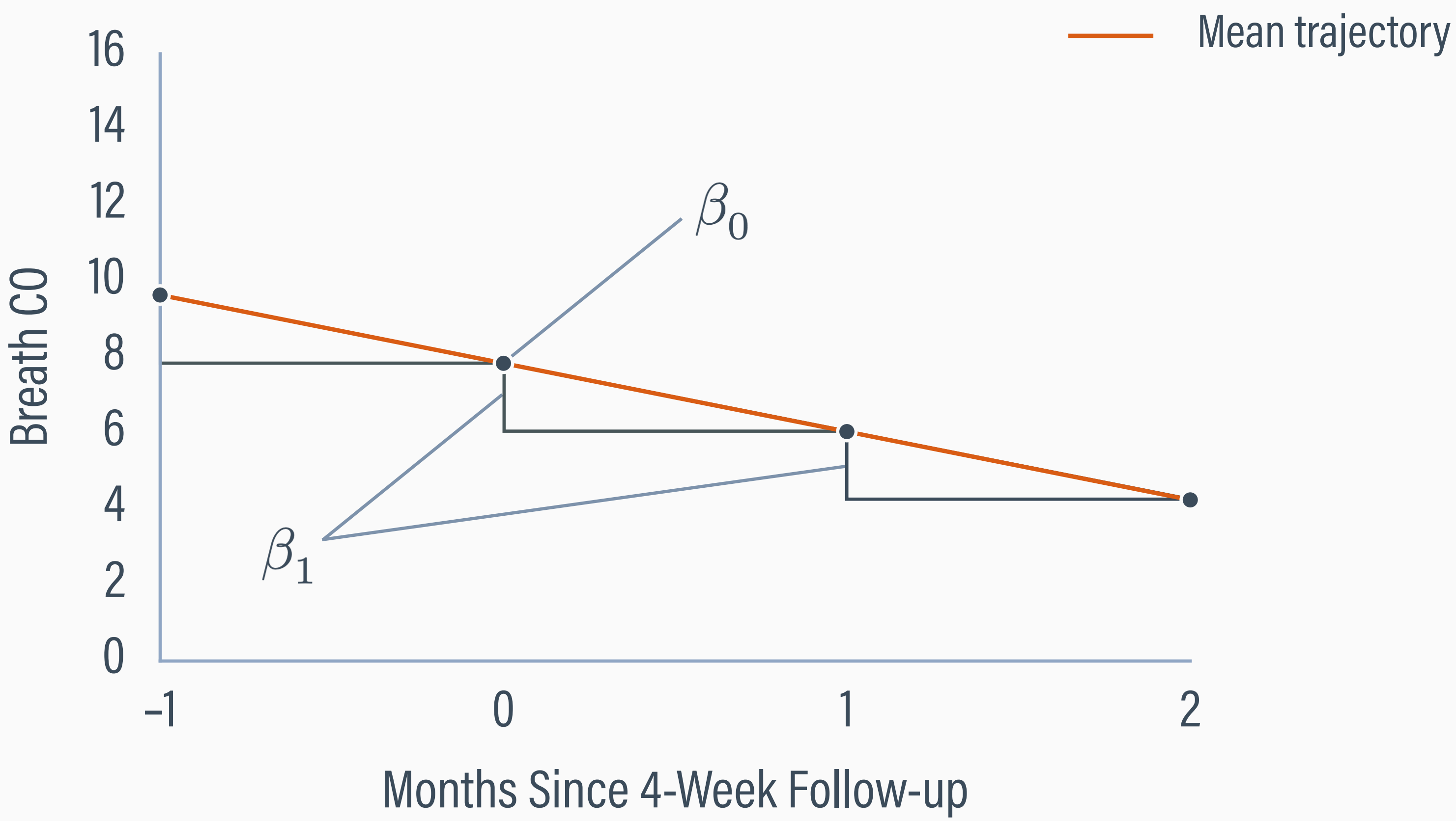
LINEAR GROWTH MODEL

- Individual intercepts and slopes are normally distributed latent variables, β_0 and β_1 define the mean trajectory, and time-specific residuals are normal with constant variation

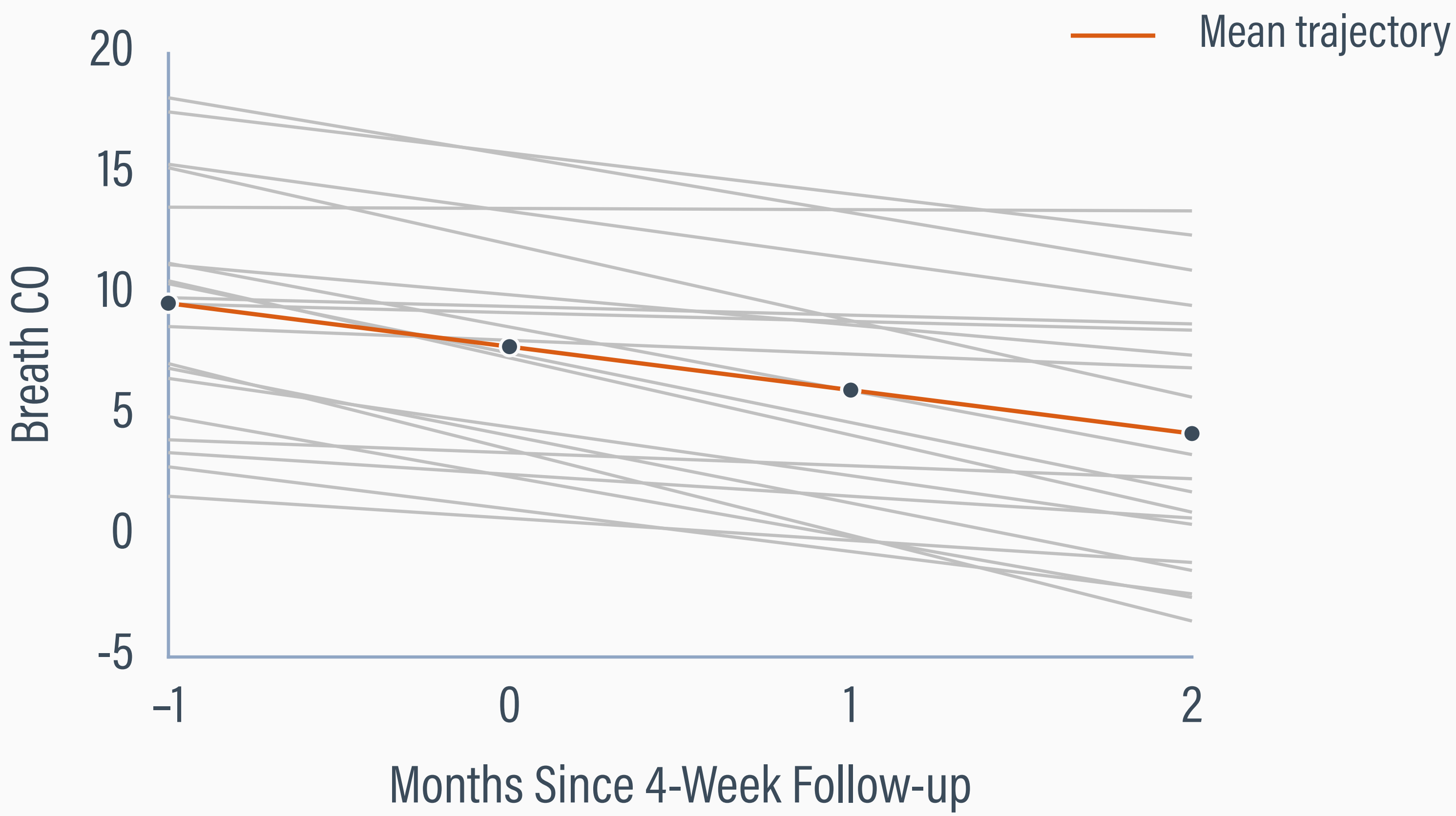
$$BREATHCO_{ti} = \beta_{0i} + \beta_{1i}(MONTH_{ti}) + \varepsilon_{ti}$$

$$\begin{pmatrix} \beta_{0i} \\ \beta_{1i} \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \Sigma_{\beta_i} \right) \qquad \varepsilon_{ti} \sim N(0, \sigma_\varepsilon^2)$$

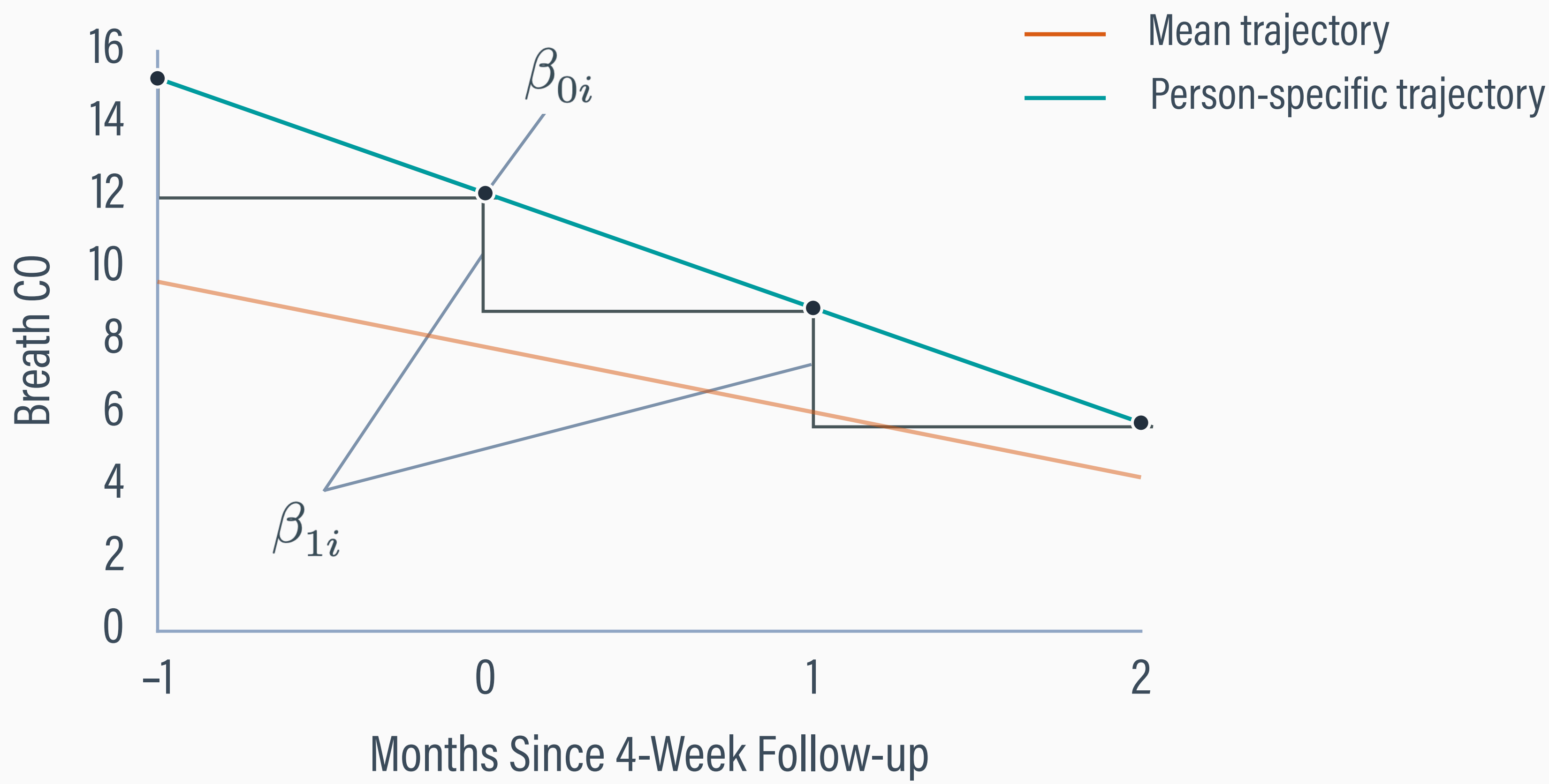
MEAN INTERCEPT AND SLOPE



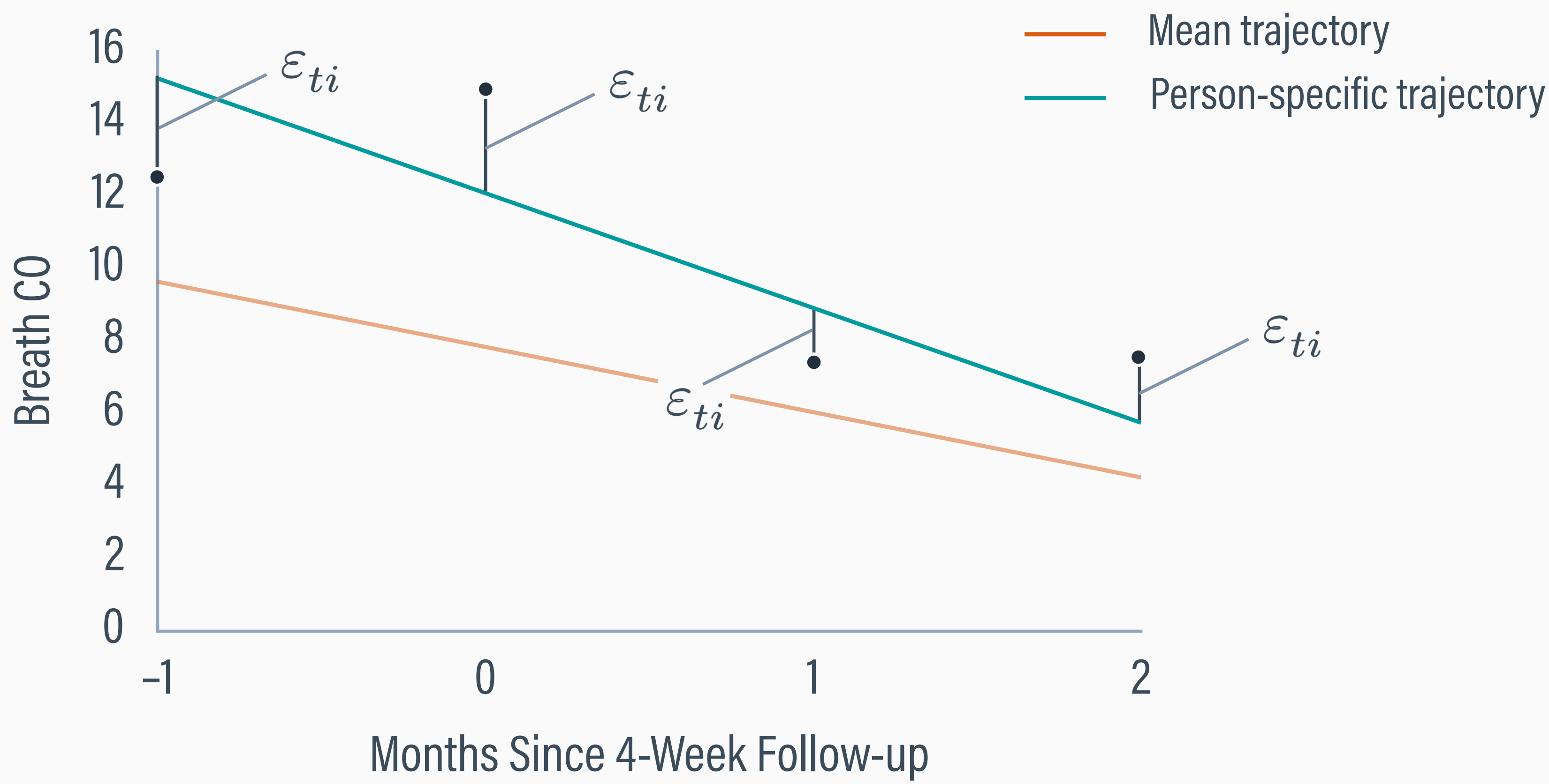
PERSON-SPECIFIC TRAJECTORIES



PERSON-SPECIFIC INTERCEPT AND SLOPE



TIME-SPECIFIC RESIDUALS



LINEAR GROWTH MODELS IN BLIMP

CLINICAL TRIAL DATA

Predictors

Outcome

Variable	Definition	Missing %	Scale
PERSON	Person-level (level-2) identifier	0	Integer index
WAVE	Data collection wave	0	Integer index (1 to 4)
MONTH	Months relative to 4-week follow-up	0	Integer index (–1 to 2)
BREATHCO	Breath CO reading	23.94	Numeric (0 to 46)
DRINKS	Number of drinks per drinking day	15.00	Numeric (0 to 34)
CIGS	Number of cigarettes per smoking day	14.85	Numeric (0 to 50)
FEMALE	Gender dummy code	0	0 = Male, 1 = Female
CONDITION	Treatment condition	0	0 = Single-medication, 1 = Dual-medication
NICDEP	Nicotine dependence dummy code	6.67	0 = Very Low, to Low 1 = Moderate to Very High
QUIT16	Quit smoking at 4-month follow-up	33.94	0 = Did not quit, 1 = Quit smoking
DROPOUT	Dropout indicator	13.33	0 = In study 1 = Dropped out

ANALYSIS MODEL

- Linear growth model with random intercepts and slopes (person-specific intercepts and growth rates)

$$BREATHCO_{ti} = \beta_{0i} + \beta_{1i}(MONTH_{ti}) + \varepsilon_{ti}$$

Measurement	Baseline	4-week	8-week	12-week
Time Scores (MONTH)	–1	0	1	2

BLIMP SCRIPT

```
DATA: clinicaltrial.dat;
VARIABLES: person wave month breathCO drinks cigs female condition nicdep quit16 dropout;
CLUSTERID: person;
MISSING: 999;
FIXED: month;
MODEL:
breathCO ~ month | month;
BURN: 5000;
ITERATIONS: 10000;
SEED: 90291;
```

DATA AND VARIABLES

```
DATA: clinicaltrial.dat;
VARIABLES: person wave month breathCO drinks cigs female condition nicdep quit16 dropout;
CLUSTERID: person;
MISSING: 999;
FIXED: month;
MODEL:
breathCO ~ month | month;
BURN: 5000;
ITERATIONS: 10000;
SEED: 90291;
```

MODEL DETAILS

```
DATA: clinicaltrial.dat;
VARIABLES: person wave month breathCO drinks cigs female condition nicdep quit16 dropout;
CLUSTERID: person;
MISSING: 999;
FIXED: month;
MODEL:
breathCO ~ month | month;
BURN: 5000;
ITERATIONS: 10000;
SEED: 90291;
```

COMPUTATIONAL DETAILS

```
DATA: clinicaltrial.dat;
VARIABLES: person wave month breathCO drinks cigs female condition nicdep quit16 dropout;
CLUSTERID: person;
MISSING: 999;
FIXED: month;
MODEL:
breathCO ~ month | month;
BURN: 5000;
ITERATIONS: 10000;
SEED: 90291;
```

DATA AND VARIABLES

```
DATA: clinicaltrial.dat;                                # ascii text data
VARIABLES: person wave month breathCO drinks cigs female
   condition nicdep quit16 dropout;                     # variable order
CLUSTERID: person;                                      # level-2 identifier variable
MISSING: 999;                                           # missing value code
```

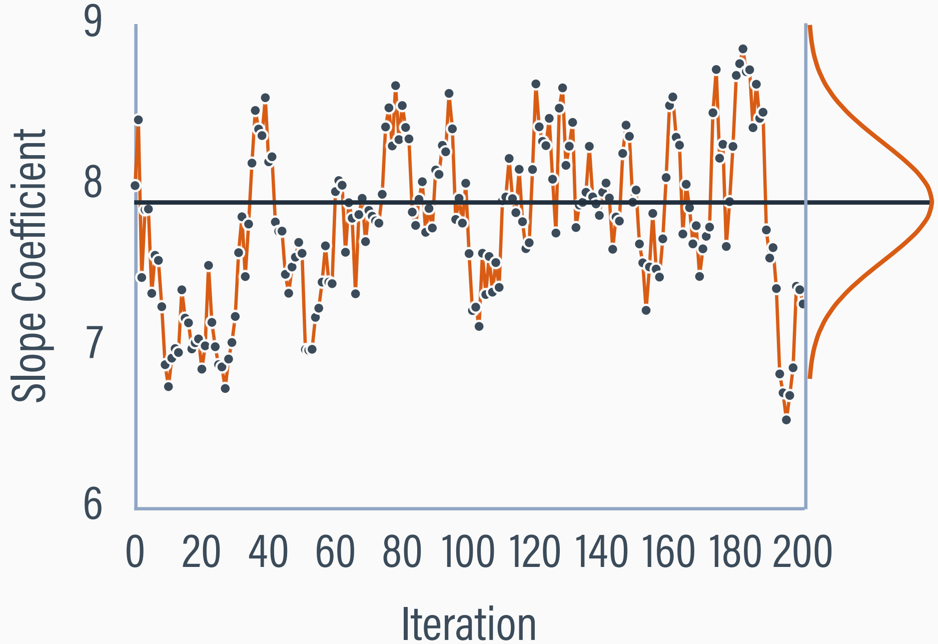
MODEL DETAILS

```
FIXED: month;                          # complete predictors
MODEL:                                 # regression equations
breathCO ~ month | month;              # DV ~ predictors | random predictors
```

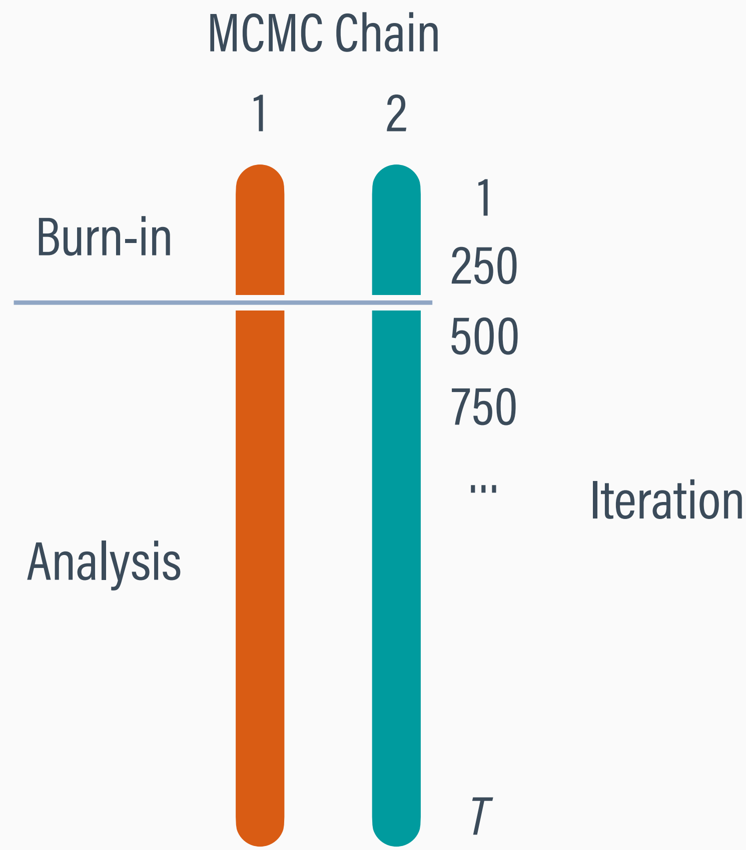
MCMC ESTIMATION

- MCMC uses computer simulation to "draw" or "sample" parameters from a distribution of plausible values

- Estimates continually vary across iterations in a random pattern

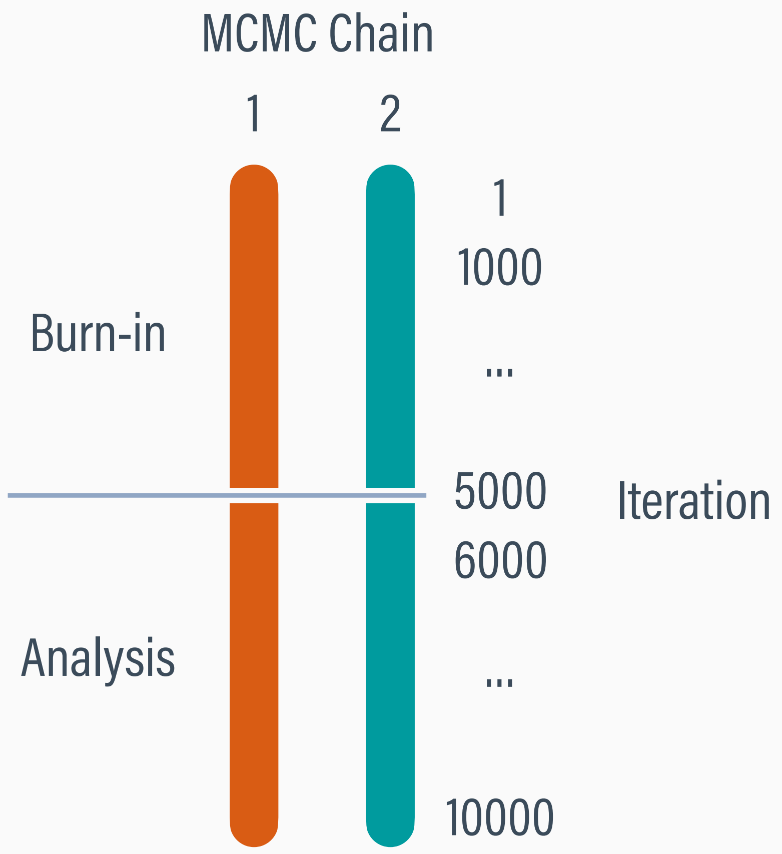


MCMC ALGORITHM



COMPUTATIONAL DETAILS

```
BURN: 5000;             # burn-in iterations
ITERATIONS: 10000;      # analysis iterations
SEED: 90291;            # random number seed
```



MISSING DATA IMPUTATION

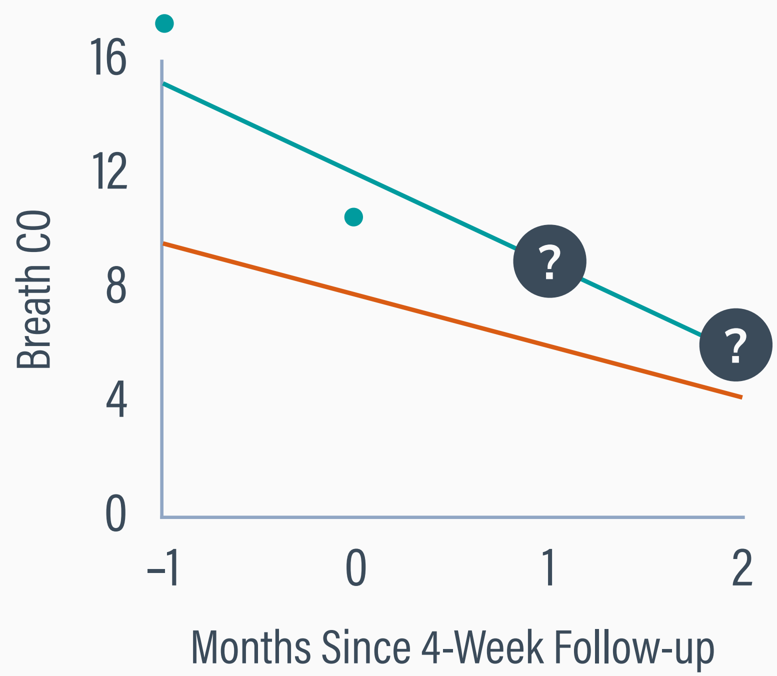
- Missing values are imputed at each MCMC iteration

$$BREATHCO_{ti(mis)} \sim N(\beta_{0i} + \beta_{1i}(MONTH_{ti}), \sigma^2_{\varepsilon})$$

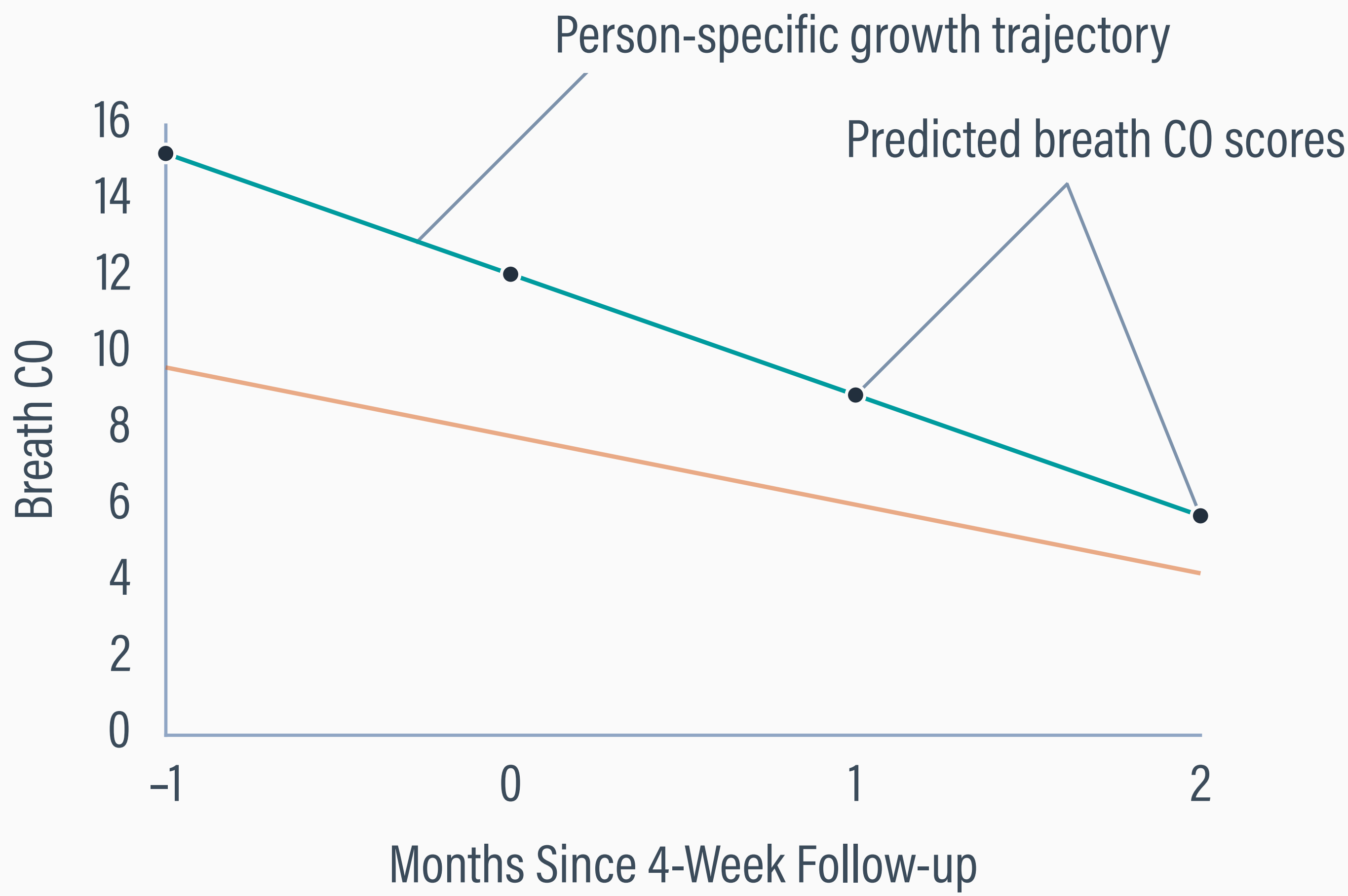
- The model parameters and random effects define normal distributions of imputations around person-specific trajectories

ILLUSTRATION

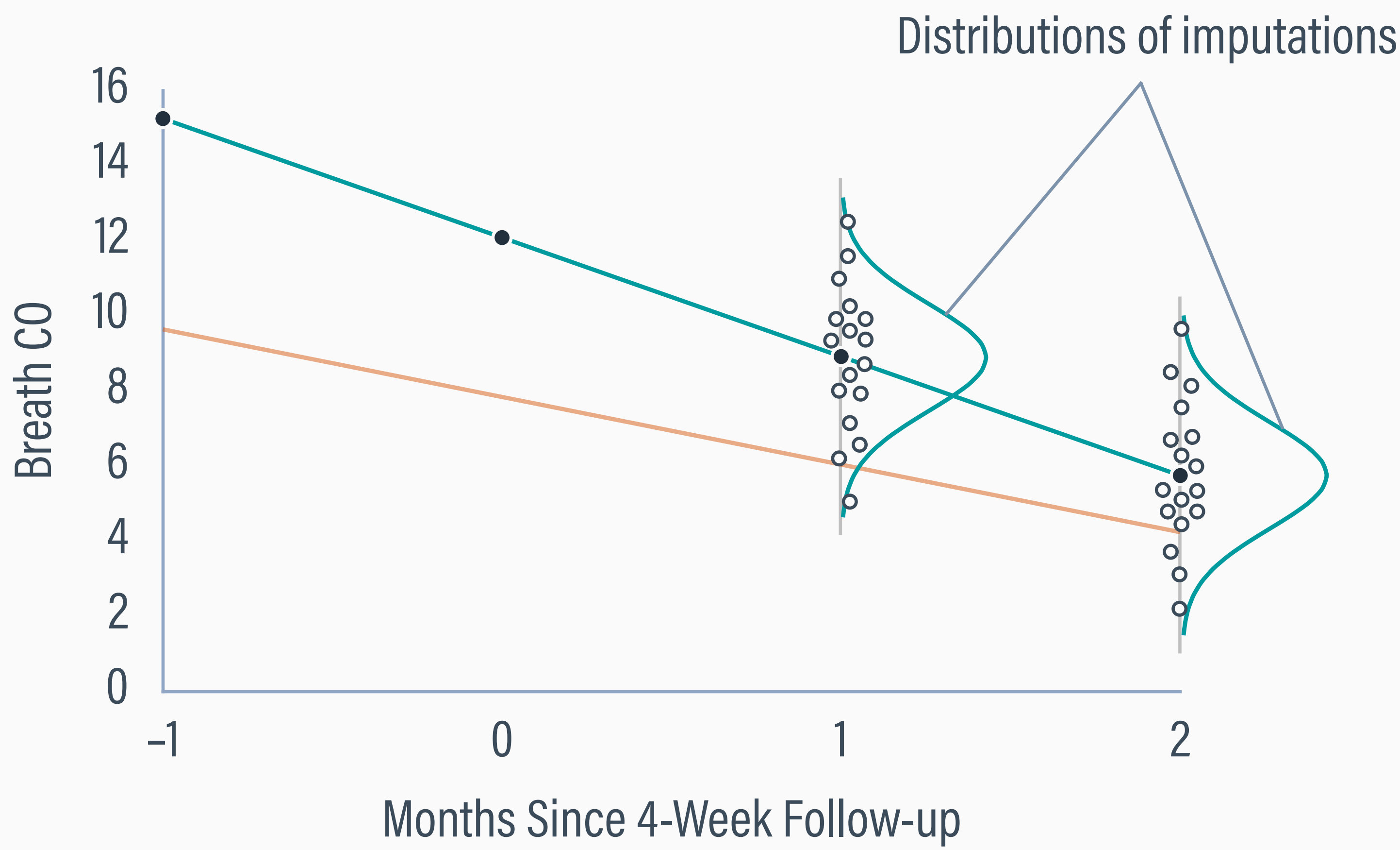
- A participant provides baseline and one-month follow-up scores then drops out
- Missing scores are imputed by sampling from a distribution of plausible values



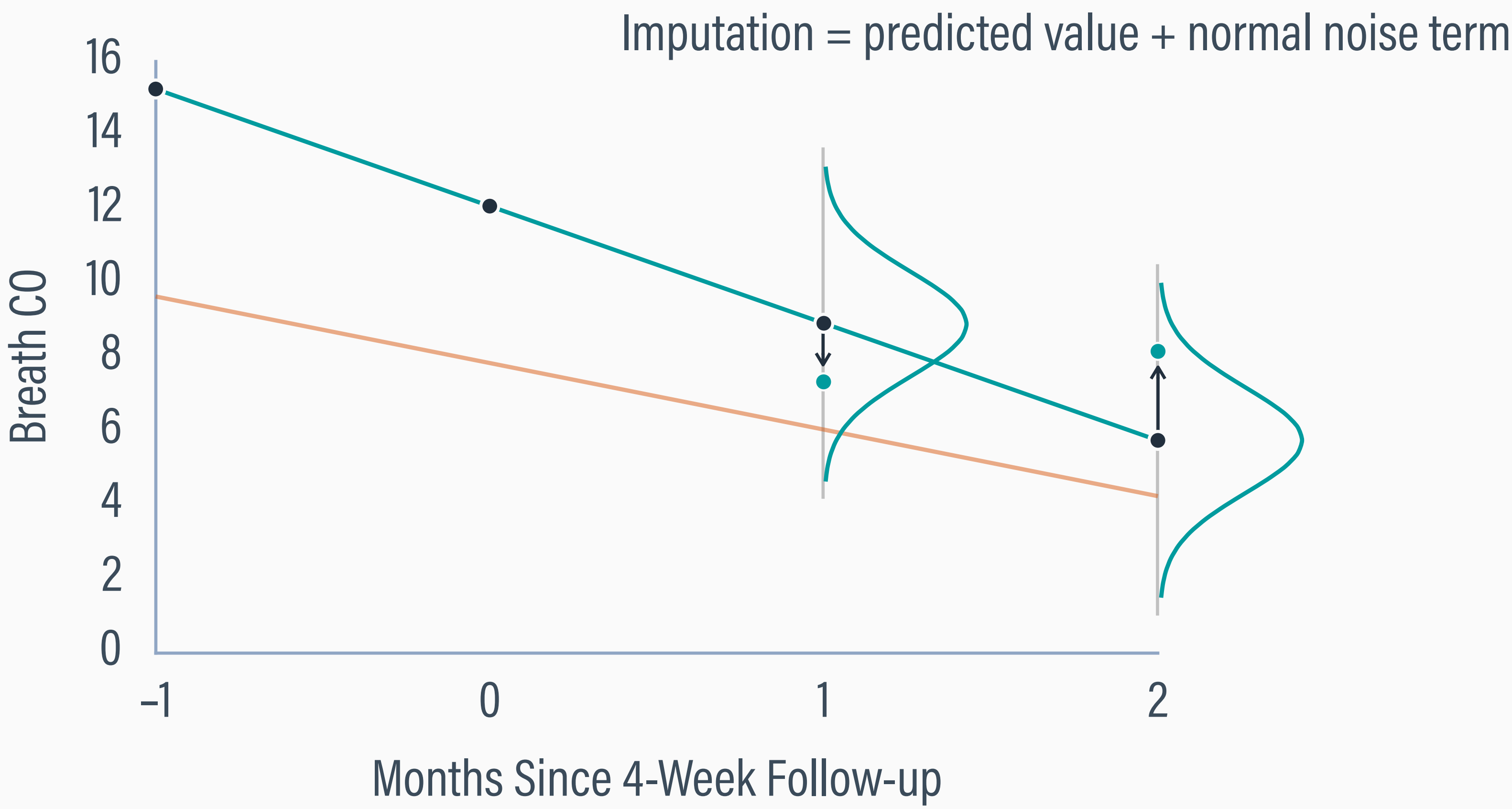
IMPUTATION AT 8- AND 12-WEEK FOLLOW-UP



IMPUTATION AT 8- AND 12-WEEK FOLLOW-UP



IMPUTATION AT 8- AND 12-WEEK FOLLOW-UP



UNDERSTANDING BLIMP OUTPUT

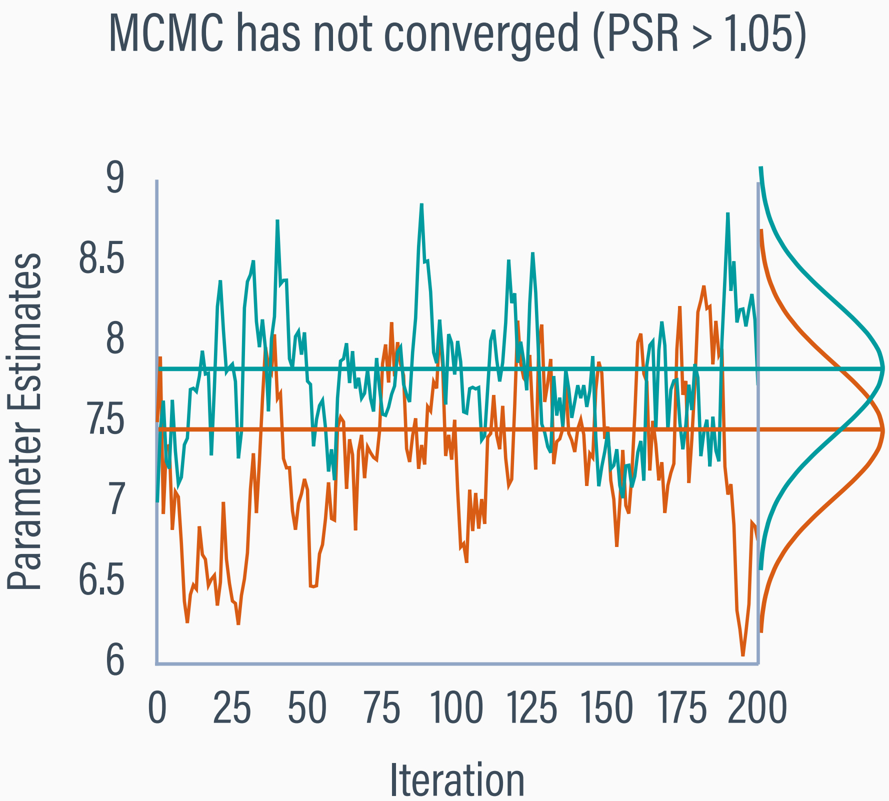
MCMC CONVERGENCE

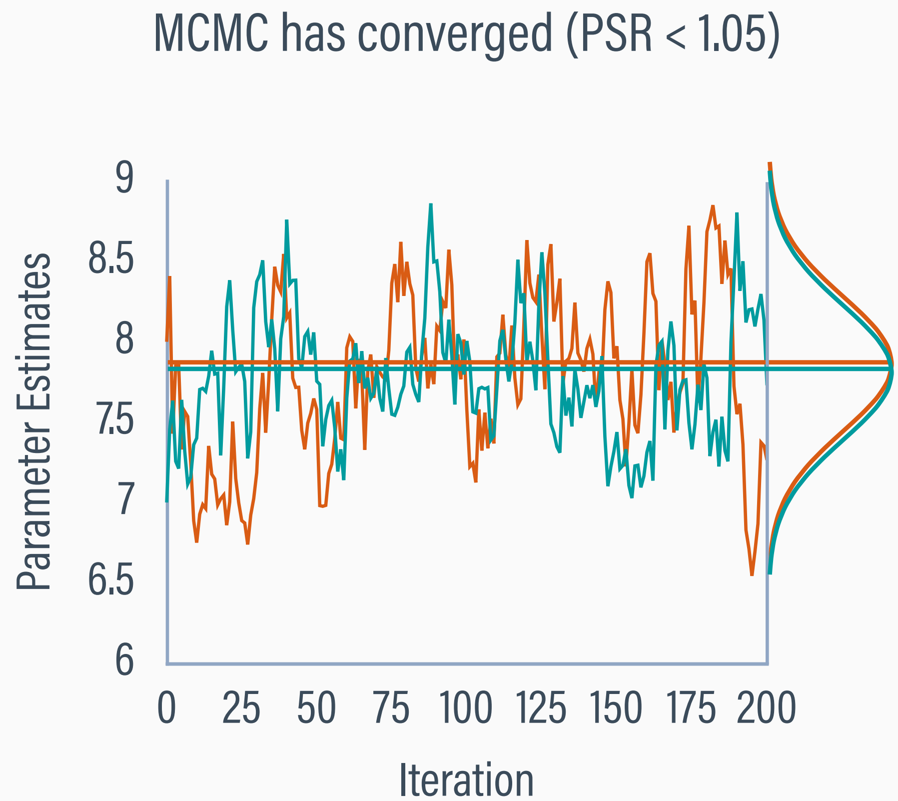
- MCMC converges when posterior distributions are stationary
- Parameter estimates oscillate around a stable mean, and variation doesn’t change with additional iterations
- Set burn-in cycles > number of iterations needed to converge

POTENTIAL SCALE REDUCTION

- The potential scale reduction (PSR) factor compares parameter distributions from two MCMC processes
- MCMC converges when two chains give estimates with same means and spread
- PSRs for all parameters should be < 1.05

PSR GRAPHIC


MCMC has not converged (PSR > 1.05)
Parameter Estimates
9
8.5
8
7.5
7
6.5
6
0
25
50
75
100
125
150
175
200
Iteration


MCMC has converged (PSR < 1.05)
Parameter Estimates
9
8.5
8
7.5
7
6.5
6
0
25
50
75
100
125
150
175
200
Iteration

PSR DIAGNOSTIC OUTPUT

```
BURN-IN POTENTIAL SCALE REDUCTION (PSR) OUTPUT:

  NOTE: Split chain PSR is being used. This splits each chain's
        iterations to create twice as many chains.

  Comparing iterations across 2 chains     Highest PSR   Parameter #
                          126 to 250             1.186            10
                          251 to 500             1.090             3
                          376 to 750             1.037             3
                          501 to 1000            1.033            10
                            ...     ...              ...              ...
                         2001 to 4000            1.019            10
                         2126 to 4250            1.010            10
                         2251 to 4500            1.010            10
                         2376 to 4750            1.005            10
                         2501 to 5000            1.010             2
```

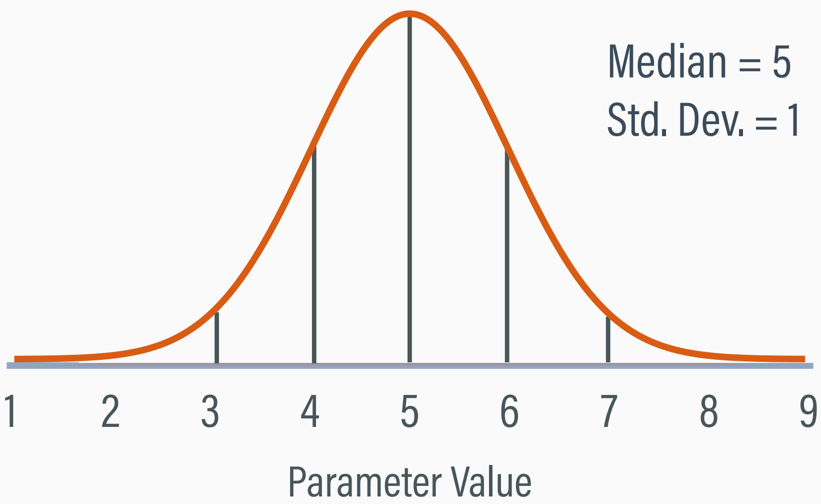
Worst PSR < 1.05

BAYESIAN POSTERIOR SUMMARIES

- Probability distributions are tools for expressing our knowledge about a parameter in a Bayesian analysis
- The posterior distribution describes plausible parameter values that are consistent with the data (no repeated sampling)
- We use summary statistics to describe parameter distributions

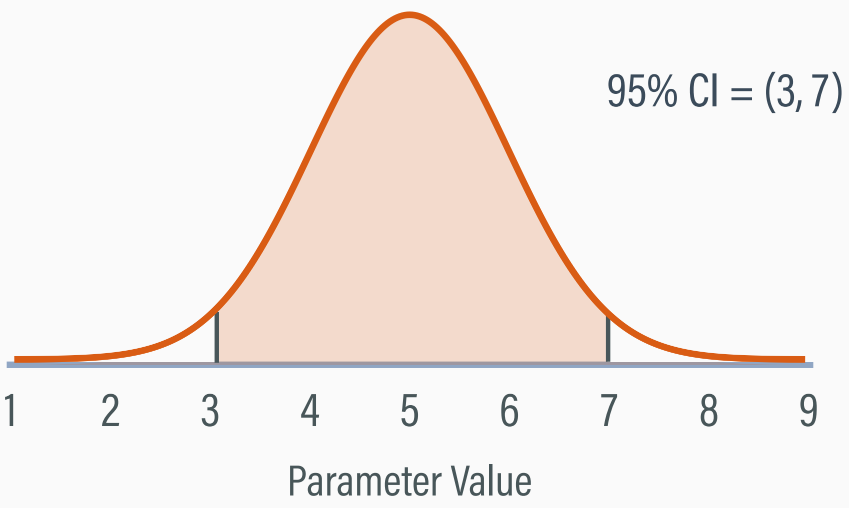
MEDIAN AND STANDARD DEVIATION

- The posterior median and standard deviation quantify the most likely parameter value and uncertainty
- Analogous to a point estimate and standard error, sans repeated sampling

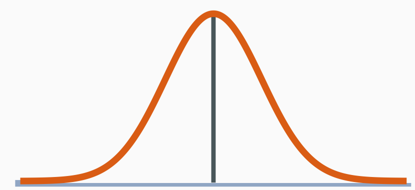


95% CREDIBLE INTERVALS

- The 95% credible interval gives limits spanning 95% of the parameter's range
- Akin to a confidence interval, but references a range of highly plausible parameter values



POINT ESTIMATES

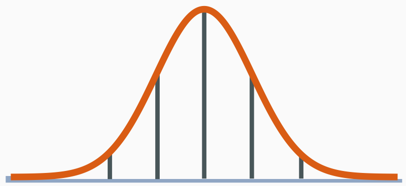


Parameter Values

OUTCOME MODEL ESTIMATES:

```
  Summaries based on 10000 iterations using 2 chains.

Outcome Variable:  breathCO

Parameters                            Median     StdDev       2.5%      97.5%        PSR      N_Eff
                                   -------------------------------------------------------------------

Variances:
  L2 : Var(Intercept)                 19.826      3.219     14.342     27.084      1.002   2214.658
  L2 : Cov(month,Intercept)           -1.155      1.162     -3.526      1.070      1.001    861.249
  L2 : Var(month)                      2.274      0.863      0.815      4.218      1.011    290.118
  Residual Var.                       18.381      1.746     15.316     22.168      1.003    910.242

Coefficients:
  Intercept                            7.837      0.409      7.052      8.645      1.003   1020.162
  month                               -1.800      0.220     -2.235     -1.367      1.001   2631.023

Standardized Coefficients:
  month                               -0.301      0.035     -0.369     -0.230      1.000   2595.367

Proportion Variance Explained
  by Coefficients                      0.091      0.021      0.053      0.136      1.000   2596.186
  by Level-2 Random Intercepts         0.431      0.046      0.340      0.520      1.001   1764.658
  by Level-2 Random Slopes             0.064      0.024      0.023      0.116      1.010    322.944
  by Level-1 Residual Variation        0.411      0.044      0.331      0.504      1.006    750.430

                                   -------------------------------------------------------------------
```

MEASURES OF UNCERTAINTY



Parameter Values

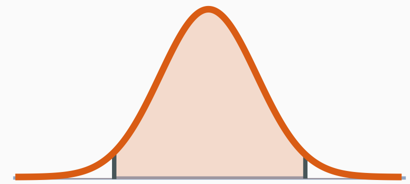
OUTCOME MODEL ESTIMATES:

Summaries based on 10000 iterations using 2 chains.

Outcome Variable: **breathCO**

Parameters	Median	StdDev	2.5%	97.5%	PSR	N_Eff
Variances:						
L2 : Var(Intercept)	19.826	3.219	14.342	27.084	1.002	2214.658
L2 : Cov(month,Intercept)	-1.155	1.162	-3.526	1.070	1.001	861.249
L2 : Var(month)	2.274	0.863	0.815	4.218	1.011	290.118
Residual Var.	18.381	1.746	15.316	22.168	1.003	910.242
Coefficients:						
Intercept	7.837	0.409	7.052	8.645	1.003	1020.162
month	-1.800	0.220	-2.235	-1.367	1.001	2631.023
Standardized Coefficients:						
month	-0.301	0.035	-0.369	-0.230	1.000	2595.367
Proportion Variance Explained						
by Coefficients	0.091	0.021	0.053	0.136	1.000	2596.186
by Level-2 Random Intercepts	0.431	0.046	0.340	0.520	1.001	1764.658
by Level-2 Random Slopes	0.064	0.024	0.023	0.116	1.010	322.944
by Level-1 Residual Variation	0.411	0.044	0.331	0.504	1.006	750.430

95% CREDIBLE INTERVALS



Parameter Values

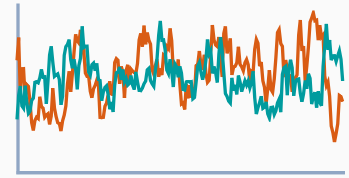
OUTCOME MODEL ESTIMATES:

Summaries based on 10000 iterations using 2 chains.

Outcome Variable: **breathCO**

Parameters	Median	StdDev	2.5%	97.5%	PSR	N_Eff
Variances:						
L2 : Var(Intercept)	19.826	3.219	14.342	27.084	1.002	2214.658
L2 : Cov(month,Intercept)	-1.155	1.162	-3.526	1.070	1.001	861.249
L2 : Var(month)	2.274	0.863	0.815	4.218	1.011	290.118
Residual Var.	18.381	1.746	15.316	22.168	1.003	910.242
Coefficients:						
Intercept	7.837	0.409	7.052	8.645	1.003	1020.162
month	-1.800	0.220	-2.235	-1.367	1.001	2631.023
Standardized Coefficients:						
month	-0.301	0.035	-0.369	-0.230	1.000	2595.367
Proportion Variance Explained						
by Coefficients	0.091	0.021	0.053	0.136	1.000	2596.186
by Level-2 Random Intercepts	0.431	0.046	0.340	0.520	1.001	1764.658
by Level-2 Random Slopes	0.064	0.024	0.023	0.116	1.010	322.944
by Level-1 Residual Variation	0.411	0.044	0.331	0.504	1.006	750.430

DIAGNOSTICS



Iteration

OUTCOME MODEL ESTIMATES:

Summaries based on 10000 iterations using 2 chains.

Outcome Variable: **breathCO**

Parameters	Median	StdDev	2.5%	97.5%	PSR	N_Eff
Variances:						
L2 : Var(Intercept)	19.826	3.219	14.342	27.084	1.002	2214.658
L2 : Cov(month,Intercept)	-1.155	1.162	-3.526	1.070	1.001	861.249
L2 : Var(month)	2.274	0.863	0.815	4.218	1.011	290.118
Residual Var.	18.381	1.746	15.316	22.168	1.003	910.242
Coefficients:						
Intercept	7.837	0.409	7.052	8.645	1.003	1020.162
month	-1.800	0.220	-2.235	-1.367	1.001	2631.023
Standardized Coefficients:						
month	-0.301	0.035	-0.369	-0.230	1.000	2595.367
Proportion Variance Explained						
by Coefficients	0.091	0.021	0.053	0.136	1.000	2596.186
by Level-2 Random Intercepts	0.431	0.046	0.340	0.520	1.001	1764.658
by Level-2 Random Slopes	0.064	0.024	0.023	0.116	1.010	322.944
by Level-1 Residual Variation	0.411	0.044	0.331	0.504	1.006	750.430

COEFFICIENTS

OUTCOME MODEL ESTIMATES:

```
  Summaries based on 10000 iterations using 2 chains.

Outcome Variable:  breathCO

Parameters                            Median     StdDev       2.5%      97.5%        PSR      N_Eff
                                 -------------------------------------------------------------------
Variances:
  L2 : Var(Intercept)                 19.826      3.219     14.342     27.084      1.002   2214.658
  L2 : Cov(month,Intercept)           -1.155      1.162     -3.526      1.070      1.001    861.249
  L2 : Var(month)                      2.274      0.863      0.815      4.218      1.011    290.118
  Residual Var.                       18.381      1.746     15.316     22.168      1.003    910.242

Coefficients:
  Intercept                            7.837      0.409      7.052      8.645      1.003   1020.162
  month                               -1.800      0.220     -2.235     -1.367      1.001   2631.023

Standardized Coefficients:
  month                               -0.301      0.035     -0.369     -0.230      1.000   2595.367

Proportion Variance Explained
  by Coefficients                      0.091      0.021      0.053      0.136      1.000   2596.186
  by Level-2 Random Intercepts         0.431      0.046      0.340      0.520      1.001   1764.658
  by Level-2 Random Slopes             0.064      0.024      0.023      0.116      1.010    322.944
  by Level-1 Residual Variation        0.411      0.044      0.331      0.504      1.006    750.430

                                 -------------------------------------------------------------------
```

STANDARDIZED COEFFICIENTS

OUTCOME MODEL ESTIMATES:

```
  Summaries based on 10000 iterations using 2 chains.

Outcome Variable:  breathCO

Parameters                          Median     StdDev       2.5%      97.5%        PSR      N_Eff
                                ------------------------------------------------------------------
Variances:
  L2 : Var(Intercept)               19.826      3.219     14.342     27.084      1.002   2214.658
  L2 : Cov(month,Intercept)         -1.155      1.162     -3.526      1.070      1.001    861.249
  L2 : Var(month)                    2.274      0.863      0.815      4.218      1.011    290.118
  Residual Var.                     18.381      1.746     15.316     22.168      1.003    910.242

Coefficients:
  Intercept                          7.837      0.409      7.052      8.645      1.003   1020.162
  month                             -1.800      0.220     -2.235     -1.367      1.001   2631.023

Standardized Coefficients:
  month                             -0.301      0.035     -0.369     -0.230      1.000   2595.367

Proportion Variance Explained
  by Coefficients                    0.091      0.021      0.053      0.136      1.000   2596.186
  by Level-2 Random Intercepts       0.431      0.046      0.340      0.520      1.001   1764.658
  by Level-2 Random Slopes           0.064      0.024      0.023      0.116      1.010    322.944
  by Level-1 Residual Variation      0.411      0.044      0.331      0.504      1.006    750.430

                                ------------------------------------------------------------------
```

VARIANCES AND COVARIANCES

OUTCOME MODEL ESTIMATES:

Summaries based on 10000 iterations using 2 chains.

Outcome Variable: **breathCO**

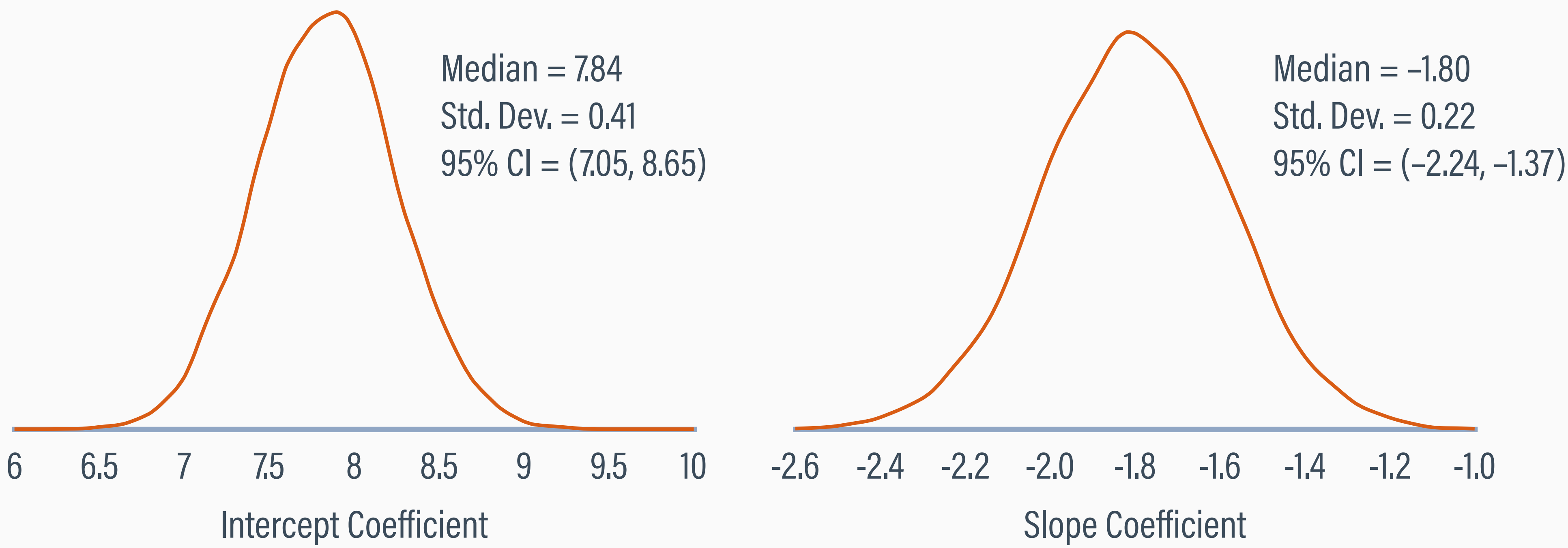
Parameters	Median	StdDev	2.5%	97.5%	PSR	N_Eff
Variances:						
L2 : Var(Intercept)	19.826	3.219	14.342	27.084	1.002	2214.658
L2 : Cov(month,Intercept)	-1.155	1.162	-3.526	1.070	1.001	861.249
L2 : Var(month)	2.274	0.863	0.815	4.218	1.011	290.118
Residual Var.	18.381	1.746	15.316	22.168	1.003	910.242
Coefficients:						
Intercept	7.837	0.409	7.052	8.645	1.003	1020.162
month	-1.800	0.220	-2.235	-1.367	1.001	2631.023
Standardized Coefficients:						
month	-0.301	0.035	-0.369	-0.230	1.000	2595.367
Proportion Variance Explained						
by Coefficients	0.091	0.021	0.053	0.136	1.000	2596.186
by Level-2 Random Intercepts	0.431	0.046	0.340	0.520	1.001	1764.658
by Level-2 Random Slopes	0.064	0.024	0.023	0.116	1.010	322.944
by Level-1 Residual Variation	0.411	0.044	0.331	0.504	1.006	750.430

EFFECT SIZE ESTIMATES

OUTCOME MODEL ESTIMATES:

```
  Summaries based on 10000 iterations using 2 chains.

Outcome Variable:  breathCO

Parameters                            Median     StdDev       2.5%      97.5%        PSR      N_Eff
                                ------------------------------------------------------------------
Variances:
  L2 : Var(Intercept)                 19.826      3.219     14.342     27.084      1.002   2214.658
  L2 : Cov(month,Intercept)           -1.155      1.162     -3.526      1.070      1.001    861.249
  L2 : Var(month)                      2.274      0.863      0.815      4.218      1.011    290.118
  Residual Var.                       18.381      1.746     15.316     22.168      1.003    910.242

Coefficients:
  Intercept                            7.837      0.409      7.052      8.645      1.003   1020.162
  month                               -1.800      0.220     -2.235     -1.367      1.001   2631.023

Standardized Coefficients:
  month                               -0.301      0.035     -0.369     -0.230      1.000   2595.367

Proportion Variance Explained
  by Coefficients                      0.091      0.021      0.053      0.136      1.000   2596.186
  by Level-2 Random Intercepts         0.431      0.046      0.340      0.520      1.001   1764.658
  by Level-2 Random Slopes             0.064      0.024      0.023      0.116      1.010    322.944
  by Level-1 Residual Variation        0.411      0.044      0.331      0.504      1.006    750.430

                                ------------------------------------------------------------------
```

POSTERIOR DISTRIBUTIONS



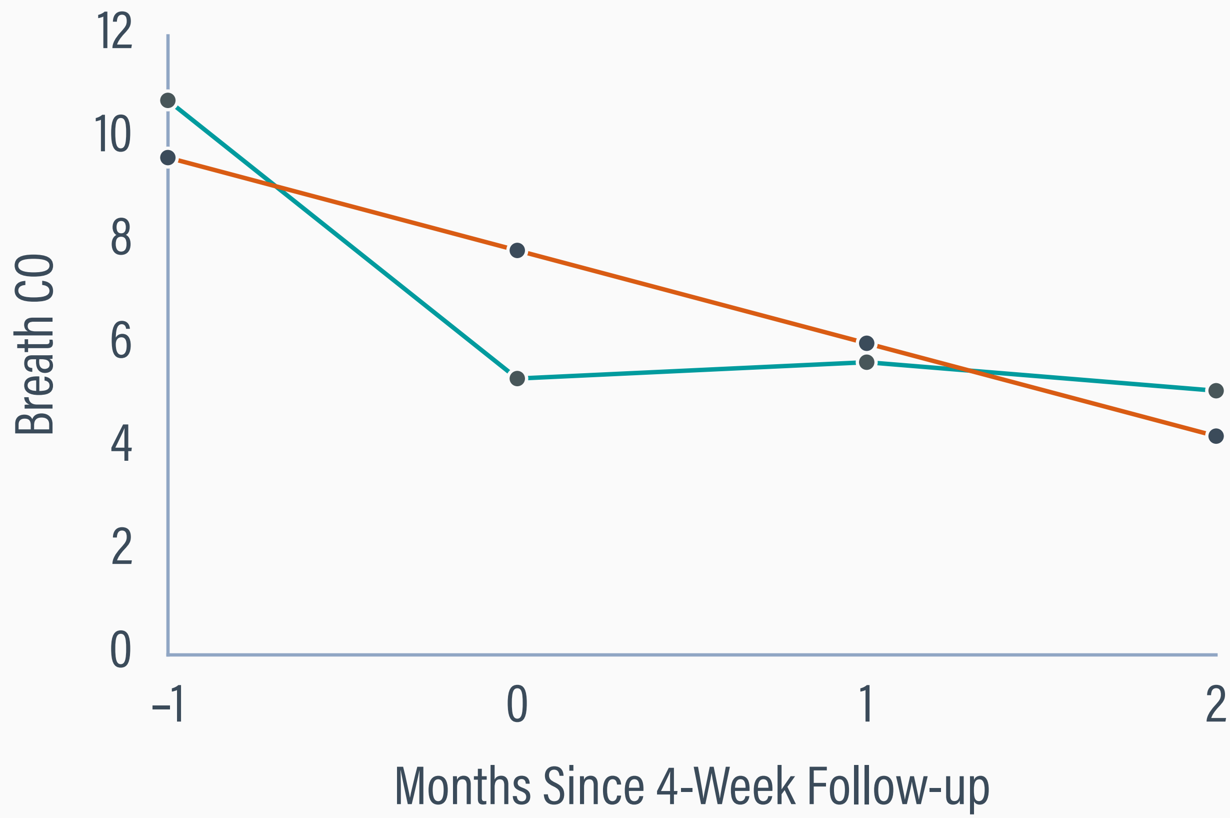
COEFFICIENT OUTPUT

```
OUTCOME MODEL ESTIMATES:

  Summaries based on 10000 iterations using 2 chains.

Outcome Variable:  breathCO

Parameters                            Median     StdDev       2.5%      97.5%        PSR      N_Eff

-------------------------------------------------------------------
Variances:
  ...

Coefficients:
  Intercept                            7.837      0.409      7.052      8.645      1.003   1020.162
  month                               -1.800      0.220     -2.235     -1.367      1.001   2631.023

Standardized Coefficients:
  ...
                              -------------------------------------------------------------------
```

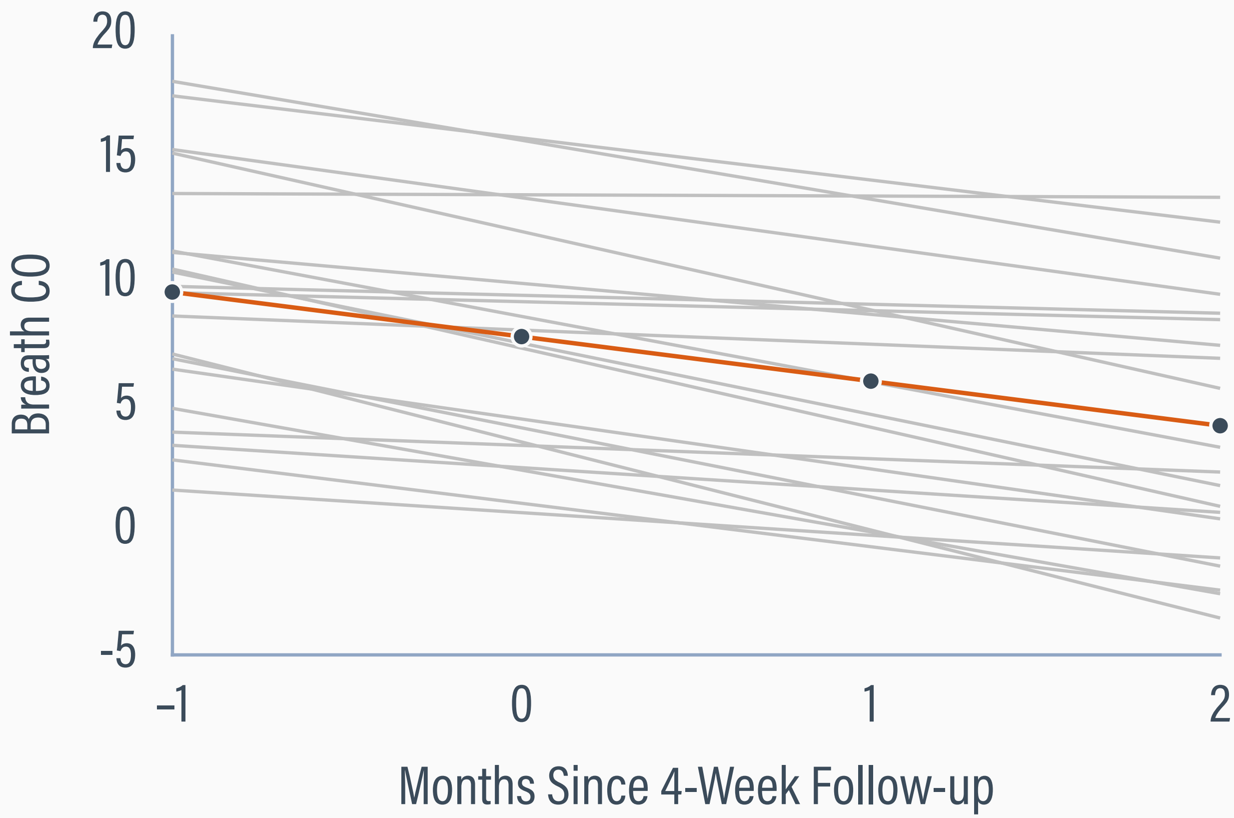
Average monthly change rate

Model-predicted mean at MONTH = 0

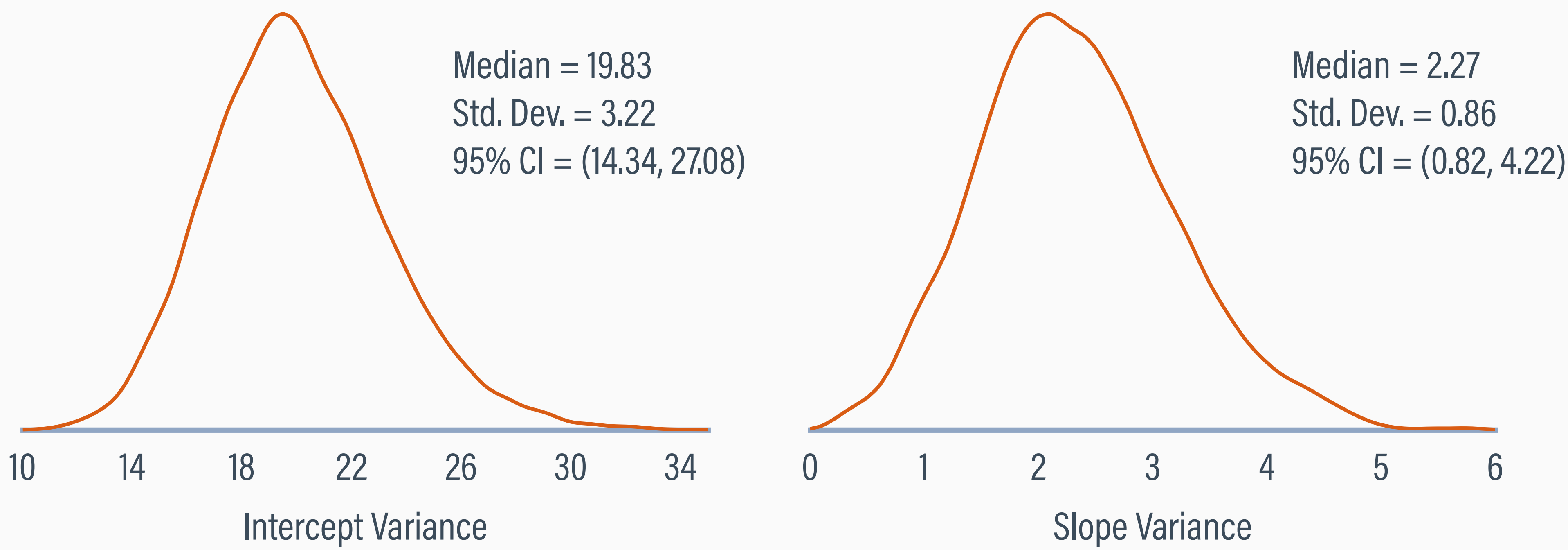
OBSERVED VS. PREDICTED MEANS



INDIVIDUAL TRAJECTORIES



POSTERIOR DISTRIBUTIONS



VARIANCE COMPONENT OUTPUT

```
OUTCOME MODEL ESTIMATES:

  Summaries based on 10000 iterations using 2 chains.

Outcome Variable:  breathCO

Parameters                          Median     StdDev       2.5%      97.5%        PSR      N_Eff
                              -------------------------------------------------------------------
Variances:
  L2 : Var(Intercept)               19.826      3.219     14.342     27.084      1.002   2214.658
  L2 : Cov(month,Intercept)         -1.155      1.162     -3.526      1.070      1.001    861.249
  L2 : Var(month)                    2.274      0.863      0.815      4.218      1.011    290.118
  Residual Var.                     18.381      1.746     15.316     22.168      1.003    910.242

Coefficients:
  ...
                           -------------------------------------------------------------------
```

Variance of monthly change rates

Covariance

Intercept variance at MONTH = 0

EFFECT SIZE OUTPUT

```
OUTCOME MODEL ESTIMATES:

  Summaries based on 10000 iterations using 2 chains.

Outcome Variable:  breathCO

Parameters                           Median     StdDev       2.5%      97.5%        PSR      N_Eff

-------------------------------------------------------------------
Variances:
  ...

Proportion Variance Explained
  by Coefficients                     0.091      0.021      0.053      0.136      1.000   2596.186
  by Level-2 Random Intercepts        0.431      0.046      0.340      0.520      1.001   1764.658
  by Level-2 Random Slopes            0.064      0.024      0.023      0.116      1.010    322.944
  by Level-1 Residual Variation       0.411      0.044      0.331      0.504      1.006    750.430

                               -------------------------------------------------------------------
```

» Rights & Sterba (2019)

HOW MANY ITERATIONS?

- MCMC estimates are autocorrelated across iterations
- The effective number of MCMC samples estimates the number of independent estimates after removing autocorrelation
- Gelman et al. (2014, p. 267) recommend at least 100 independent MCMC samples per parameter

EFFECTIVE NUMBER OF MCMC SAMPLES

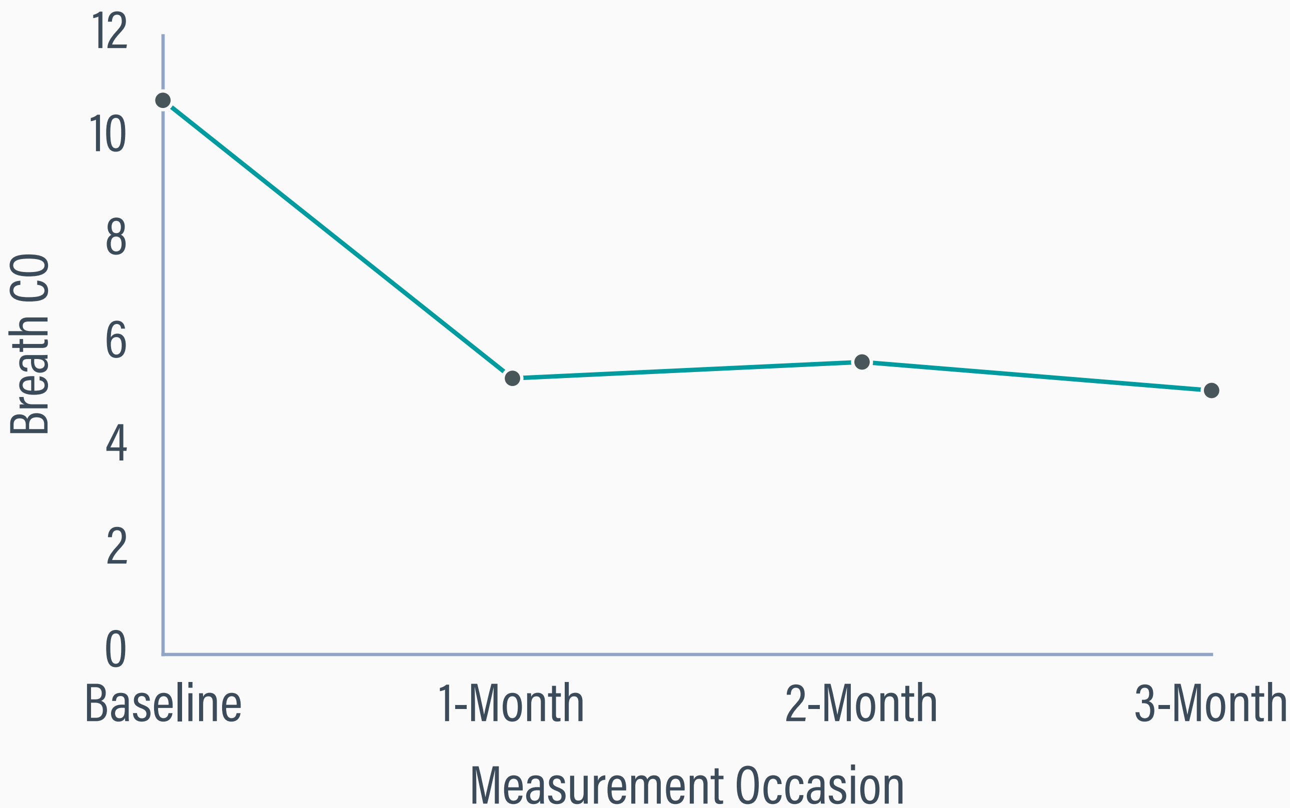
» Effective number of MCMC samples for every parameter is > 100

OUTCOME MODEL ESTIMATES:

```
  Summaries based on 10000 iterations using 2 chains.

Outcome Variable:  breathCO

Parameters                            Median     StdDev       2.5%      97.5%        PSR      N_Eff
                                -------------------------------------------------------------------

Variances:
  L2 : Var(Intercept)                 19.826      3.219     14.342     27.084      1.002   2214.658
  L2 : Cov(month,Intercept)           -1.155      1.162     -3.526      1.070      1.001    861.249
  L2 : Var(month)                      2.274      0.863      0.815      4.218      1.011    290.118
  Residual Var.                       18.381      1.746     15.316     22.168      1.003    910.242

Coefficients:
  Intercept                            7.837      0.409      7.052      8.645      1.003   1020.162
  month                               -1.800      0.220     -2.235     -1.367      1.001   2631.023

Standardized Coefficients:
  month                               -0.301      0.035     -0.369     -0.230      1.000   2595.367

Proportion Variance Explained
  by Coefficients                      0.091      0.021      0.053      0.136      1.000   2596.186
  by Level-2 Random Intercepts         0.431      0.046      0.340      0.520      1.001   1764.658
  by Level-2 Random Slopes             0.064      0.024      0.023      0.116      1.010    322.944
  by Level-1 Residual Variation        0.411      0.044      0.331      0.504      1.006    750.430

                                -------------------------------------------------------------------
```

NONLINEAR GROWTH MODELS IN BLIMP

OBSERVED MEANS



NONLINEAR MODELING OPTIONS

- Transform the time metric (e.g., square root of MONTHS)
- Piecewise growth model (multiple epochs of linear trends)
- Polynomial growth trajectories

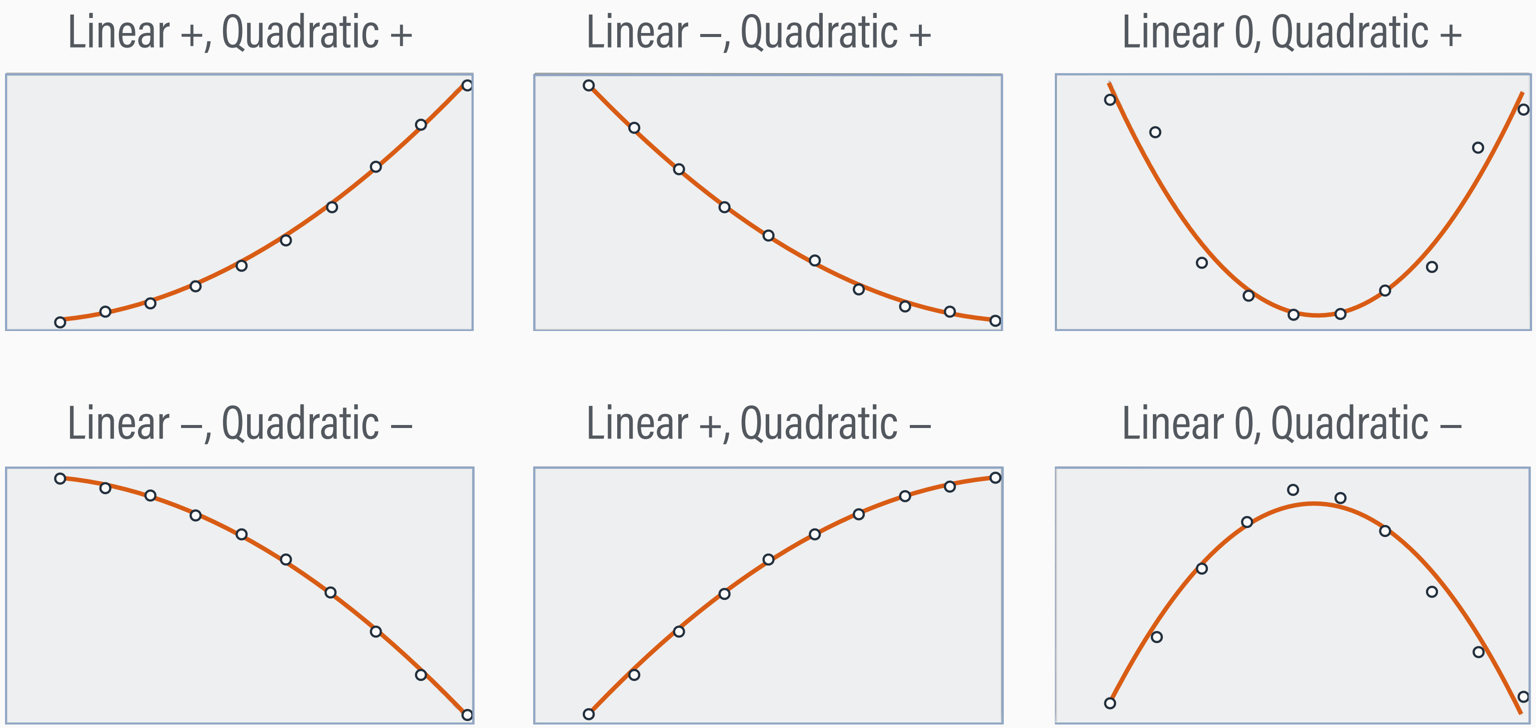
POLYNOMIAL REGRESSION

- A quadratic function uses a squared term as a predictor

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon$$

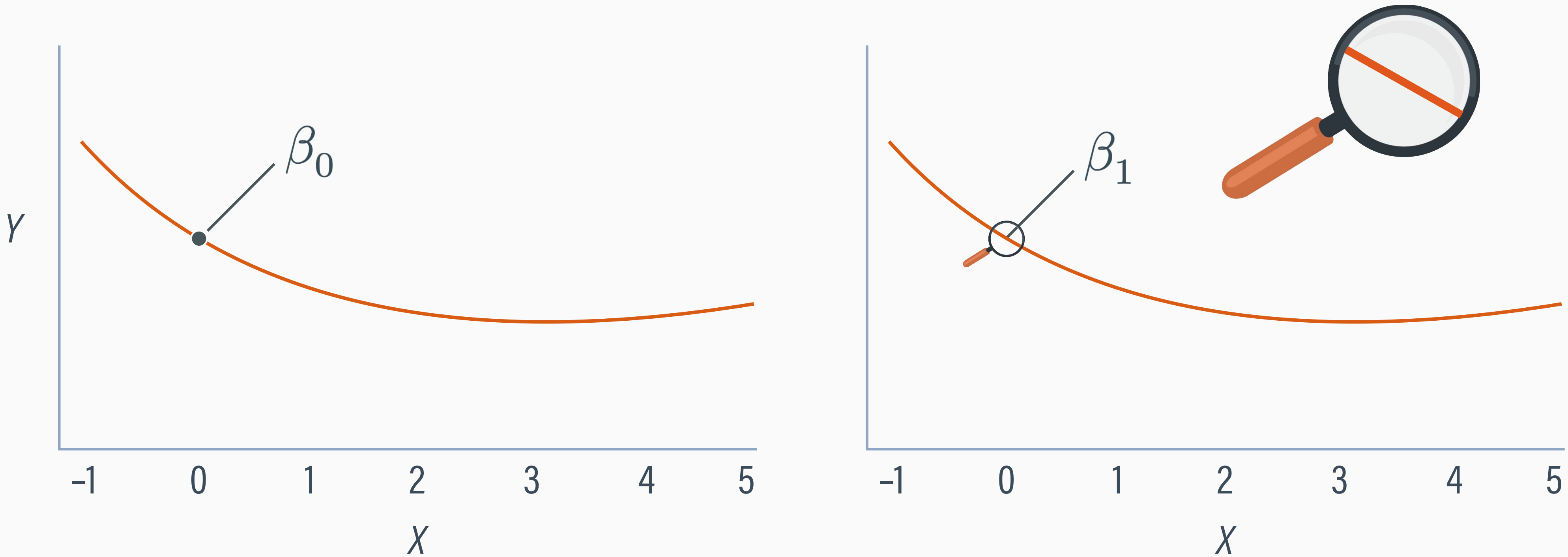
- β_0 and β_1 are the expected value and instantaneous linear change when X = 0, and β_2 captures curvature

POLYNOMIAL COMBINATIONS



LOWER-ORDER TERMS AND SCALING

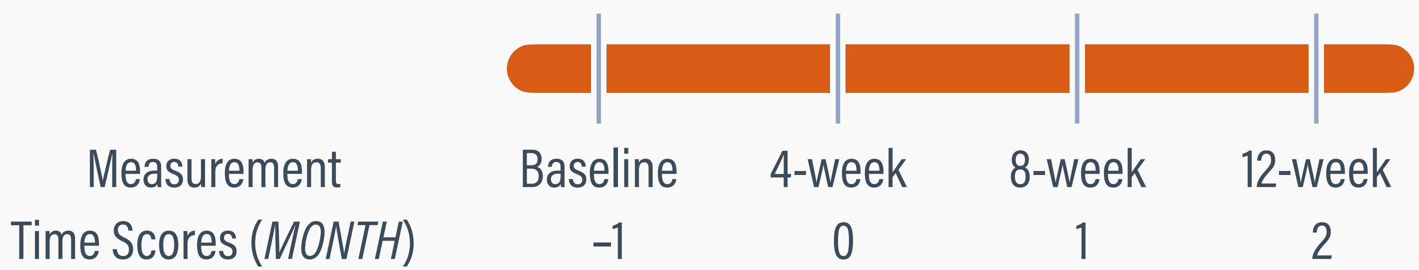
Establishing a meaningful zero value is important!



ANALYSIS MODEL

- Growth model with a quadratic fixed effect but no quadratic random effect (common curvature across persons)

$$BREATHCO_{ti} = \beta_{0i} + \beta_{1i}(MONTH_{ti}) + \beta_2(MONTH_{ti}^2) + \varepsilon_{ti}$$



BLIMP SCRIPT

```
DATA: clinicaltrial.dat;
VARIABLES: person wave month breathCO drinks cigs condition gender nicdep quit16 dropout;
CLUSTERID: person;
MISSING: 999;
FIXED: month;
MODEL:
breathCO ~ month (month^2) | month;          « Quadratic fixed effect
BURN: 5000;
ITERATIONS: 10000;
SEED: 90291;
```

ANALYSIS MODEL

- Growth model with a quadratic fixed effect and a quadratic random effect (person-specific curvature)

$$BREATHCO_{ti} = \beta_{0i} + \beta_{1i}(MONTH_{ti}) + \beta_{2i}(MONTH_{ti}^{2}) + \varepsilon_{ti}$$

Measurement	Baseline	4-week	8-week	12-week
Time Scores (*MONTH*)	–1	0	1	2

BLIMP SCRIPT

```
DATA: clinicaltrial.dat;
VARIABLES: person wave month breathCO drinks cigs condition gender nicdep quit16 dropout;
CLUSTERID: person;
MISSING: 999;
FIXED: month;
MODEL:
breathCO ~ month (month^2) | month (month^2);          « Quadratic fixed and random effects
BURN: 5000;
ITERATIONS: 10000;
SEED: 90291;
```

COEFFICIENT OUTPUT

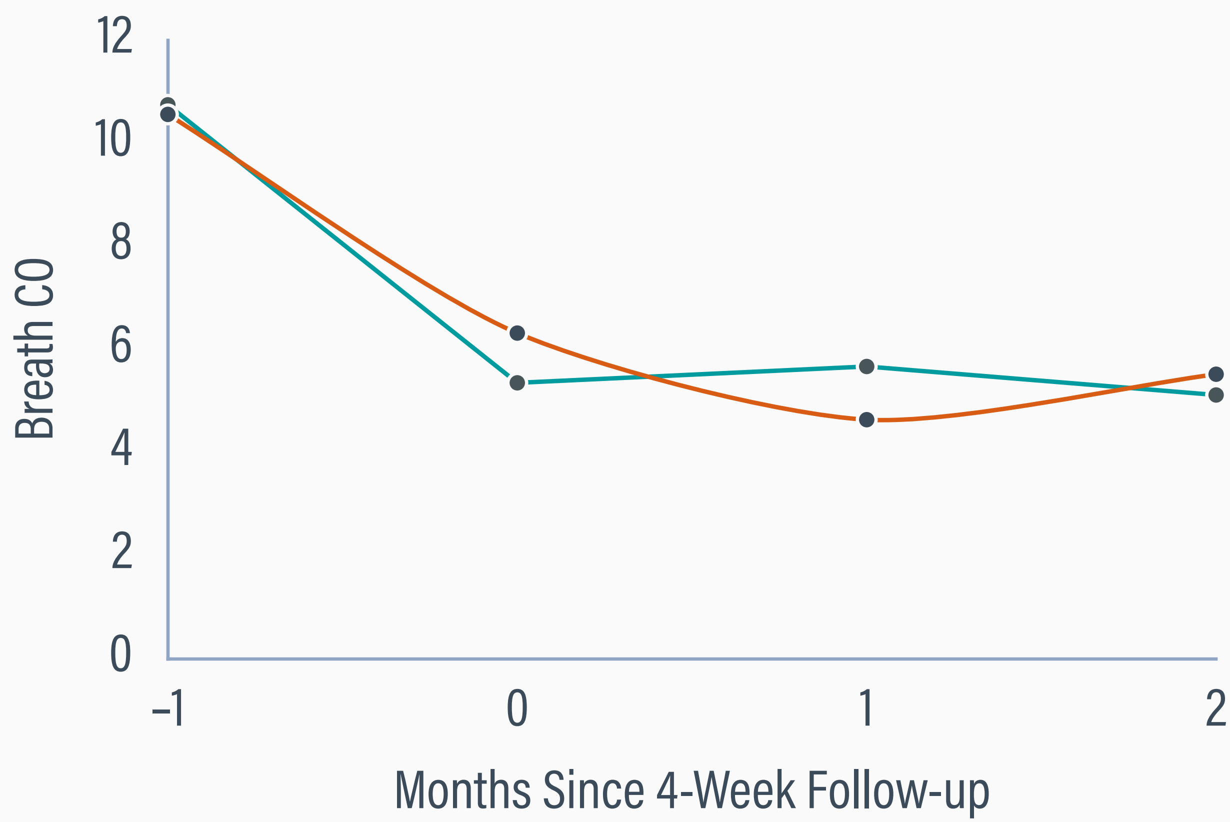
```
OUTCOME MODEL ESTIMATES:

  Summaries based on 10000 iterations using 2 chains.

Outcome Variable:  breathCO

Parameters                           Median     StdDev       2.5%      97.5%        PSR      N_Eff
                                 -------------------------------------------------------------------
Variances:
  ...

Coefficients:
  Intercept                           6.308      0.448      5.433      7.200      1.008    613.994
  month                              -2.961      0.342     -3.620     -2.276      1.003   1324.276
  (month^2)                           1.281      0.200      0.886      1.673      1.002   2075.754

Standardized Coefficients:
  ...
                                 -------------------------------------------------------------------
```

Curvature (change in linear trend over time)

Instantaneous linear change at MONTH = 0

Model-predicted mean at MONTH = 0

OBSERVED VS. PREDICTED MEANS



VARIANCE COMPONENT OUTPUT

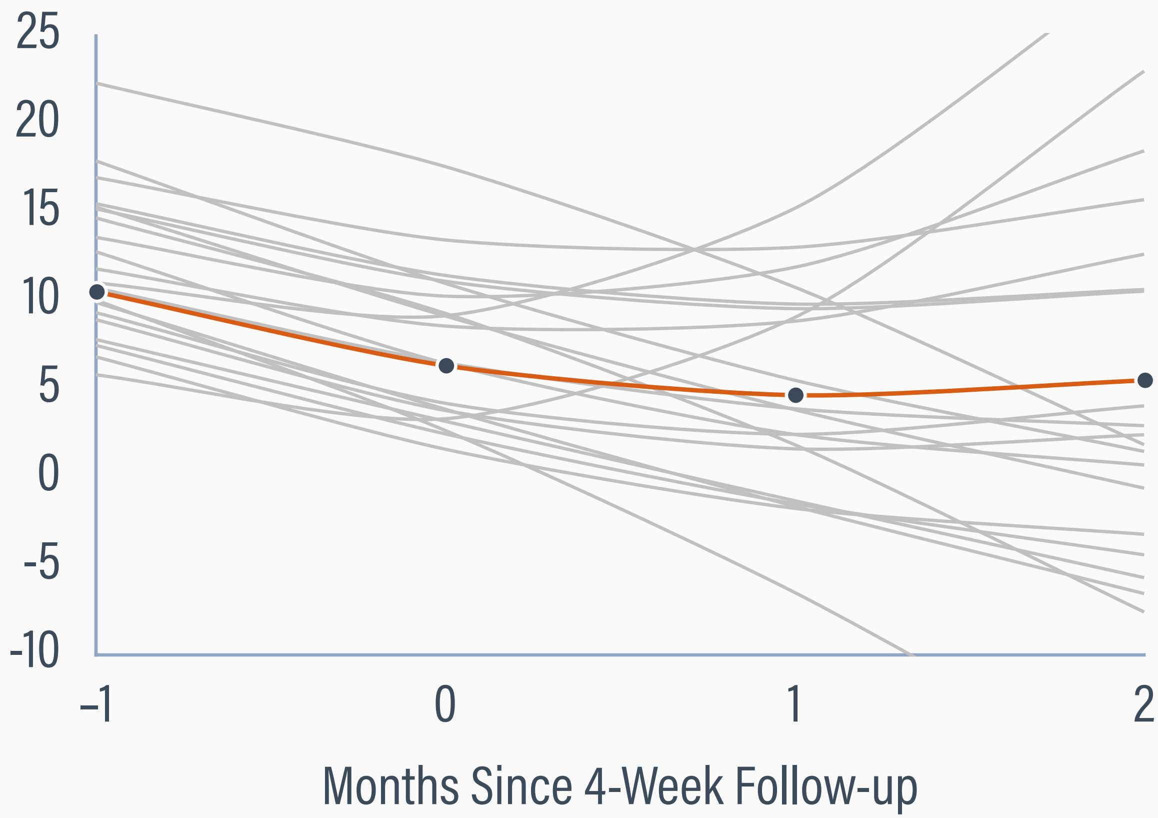
```
OUTCOME MODEL ESTIMATES:

  Summaries based on 10000 iterations using 2 chains.

Outcome Variable:  breathCO

Parameters                          Median     StdDev       2.5%      97.5%        PSR      N_Eff

-------------------------------------------------------------------
Variances:
  L2 : Var(Intercept)               20.849      3.718     14.824     29.308      1.005    536.477
  L2 : Cov(month,Intercept)          2.370      1.984     -1.178      6.574      1.002    855.422
  L2 : Var(month)                   11.426      2.184      7.727     16.181      1.005   1119.236
  L2 : Cov(month^2,Intercept)       -1.244      1.169     -3.843      0.733      1.004    286.116
  L2 : Cov(month^2,month)           -4.975      1.165     -7.661     -3.055      1.011    408.096
  L2 : Var(month^2)                  2.387      0.707      1.280      4.055      1.017    237.353
  Residual Var.                     10.823      1.183      8.779     13.436      1.005    965.422

Coefficients:
  ...
                               -------------------------------------------------------------------
```

Curvature variance

Variance of instantaneous linear change rates at MONTH = 0

Variance of intercepts at MONTH = 0

SPAGHETTI PLOT



EFFECT SIZE OUTPUT

```
OUTCOME MODEL ESTIMATES:

  Summaries based on 10000 iterations using 2 chains.

Outcome Variable:  breathCO

Parameters                           Median     StdDev       2.5%      97.5%        PSR      N_Eff

-------------------------------------------------------------------
Variances:
  ...

Proportion Variance Explained
  by Coefficients                     0.119      0.025      0.072      0.171      1.003   1345.826
  by Level-2 Random Intercepts        0.463      0.043      0.382      0.548      1.001   1702.503
  by Level-2 Random Slopes            0.166      0.029      0.114      0.227      1.005    823.026
  by Level-1 Residual Variation       0.248      0.031      0.193      0.315      1.004   1007.734

                               -------------------------------------------------------------------
```

BAYESIAN WALD TEST

- Asparouhov and Muthén (2021) proposed a Bayesian Wald test that mimics familiar likelihood-based Wald tests

$$T = (\theta - \theta_0)' \Sigma_\theta^{-1} (\theta - \theta_0)$$

- *T* is the sum of squared standardized differences (chi-square metric) between the posterior means and null hypothesis

BLIMP SCRIPT

```
DATA: clinicaltrial.dat;
VARIABLES: person wave month breathCO drinks cigs condition gender nicdep quit16 dropout;
CLUSTERID: person;
MISSING: 999;
FIXED: month;
MODEL:
breathCO ~ month@beta1 (month^2)@beta2 | month (month^2);
TEST: beta2 = 0;                                  « Single-parameter test
BURN: 5000;
ITERATIONS: 10000;
SEED: 90291;
```

MODEL FIT OUTPUT

```
MODEL FIT:

  INFORMATION CRITERIA

    Marginal Likelihood
      DIC2                    4076.173
      WAIC                    4138.998

...

  WALD TESTS (Asparouhov & Muthén, 2021)

  Test #1

    Full:
      [1]  breathCO ~ Intercept month@beta1 (month^2)@beta2 | Intercept month

    Restricted:
      [1]  breathCO ~ Intercept month@beta1 (month^2)@beta2 | Intercept month

    Constraints in Restricted:
      [1]  beta2 = 0

    Wald Statistic (Chi-Square)                 41.007
    Number of Parameters Tested (df)                 1
    Probability                                  0.000
```

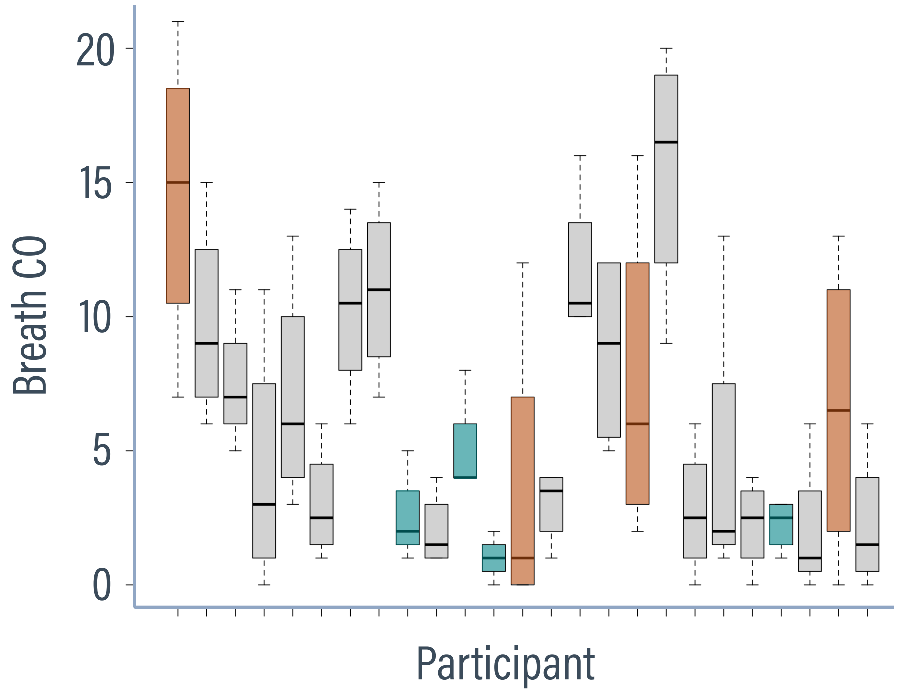
MODEL COMPARISON

Criterion	Linear	Quadratic Fixed	Quadratic Random
Marginal DIC	4196.55	4138.49	4076.17
Marginal WAIC	4258.86	4201.08	4139.00
R^2 Fixed effects	.09	.12	.12
R^2 Intercepts	.43	.45	.46
R^2 Slopes	.06	.08	.17
R^2 Residual	.41	.35	.24

MODELING HETEROGENEOUS VARIATION IN BLIMP

VARIANCE HETEROGENEITY

- Participants exhibit substantial differences in their within-person variation of breath CO
- The standard growth model assumes constant within-cluster variation



MODELING WITHIN-CLUSTER VARIATION

- Cluster-specific residual variances that do not depend on other variables (Kasim & Raudenbush, 1998)
- Location and scale models where within-cluster variation is modeled as a function of level-1 and level-2 predictors
- Cluster-specific residual variation in location and scale models can be an outcome and a level-2 predictor

HETEROGENOUS VARIANCE SCRIPT

```
DATA: clinicaltrial.dat;
VARIABLES: person wave month breathCO drinks cigs condition gender nicdep quit16 dropout;
CLUSTERID: person;
MISSING: 999;
FIXED: month;
MODEL:
breathCO ~ month (month^2) | month (month^2);
BURN: 5000;
ITERATIONS: 10000;
SEED: 90291;
OPTIONS: hev;
```

» Kasim & Raudenbush (1998)

ANALYSIS MODEL

- Location and scale models introduce a regression with the logarithm of the within-cluster variation as the outcome

$$BREATHCO_{ti} = \beta_{0i} + \beta_{1i}(MONTH_{ti}) + \beta_{2i}(MONTH_{ti}^2) + \varepsilon_{ti}$$
$$\sigma^2_{\varepsilon_{ti}} = \exp(\gamma_0 + \epsilon_i)$$

- Level-1 and level-2 variables can predict within-cluster variation, and variation can predict distal outcomes

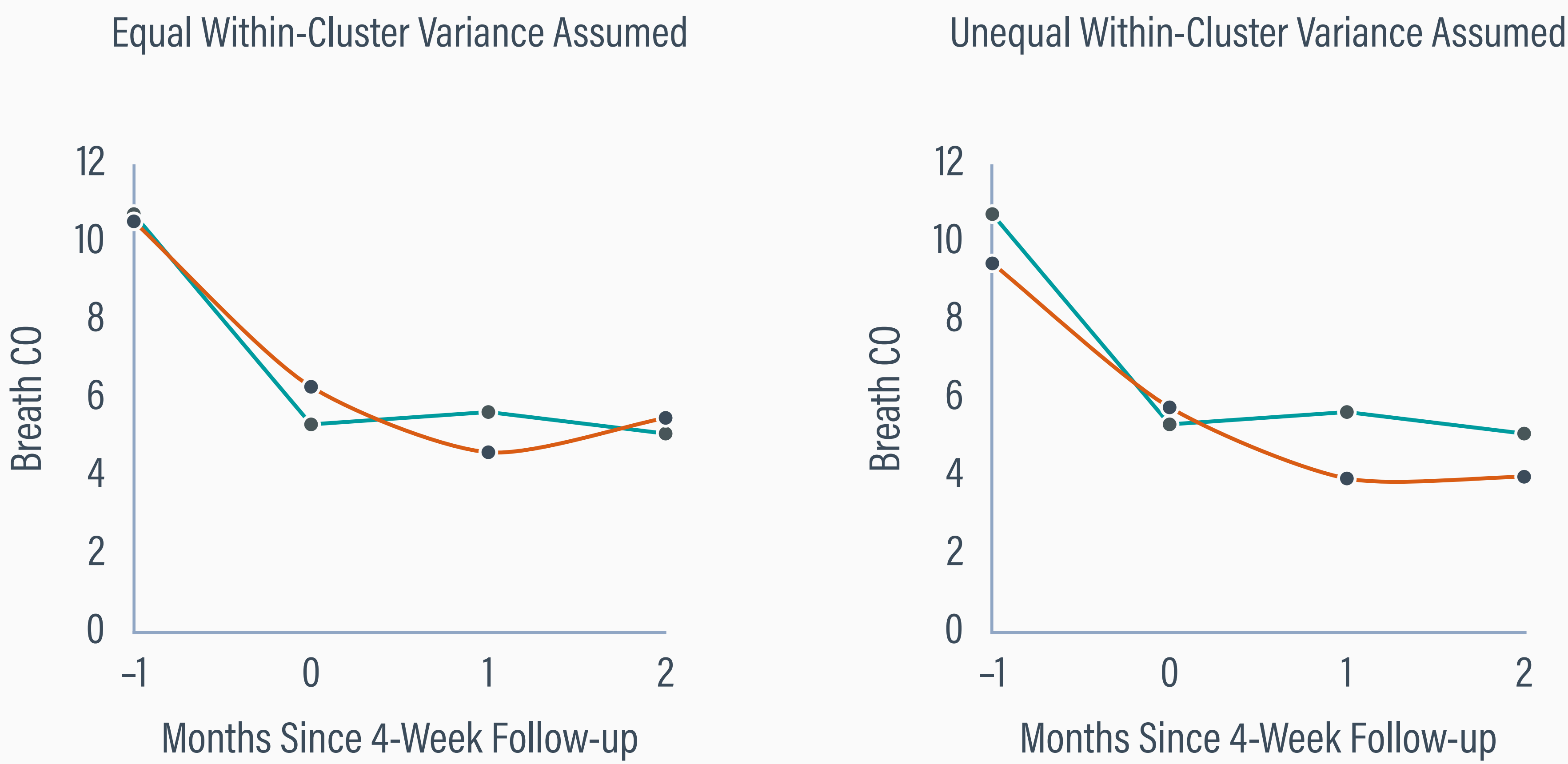
LOCATION AND SCALE MODEL SCRIPT

```
DATA: clinicaltrial.dat;
VARIABLES: person wave month breathCO drinks cigs female condition nicdep quit16 dropout;
CLUSTERID: person;
LATENT: person = logvar;            « Define level-2 latent variable called logvar
MISSING: 999;
FIXED: month;
MODEL:
breathCO ~ month (month^2) | month (month^2);
var(breathCO) ~ 1@logvar;           « Log of within-cluster variation equal to the latent variable
BURN: 10000;
ITERATIONS: 20000;
SEED: 90291;
```

COEFFICIENT OUTPUT

```
OUTCOME MODEL ESTIMATES:

  Summaries based on 20000 iterations using 2 chains.

Outcome Variable:  breathCO

Parameters                              Median     StdDev       2.5%      97.5%        PSR      N_Eff
                                ----------------------------------------------------------------------
Variances:
  ...

Coefficients:
  Intercept                              5.780      0.453      4.944      6.729      1.002     98.688
  month                                 -2.758      0.286     -3.309     -2.168      1.009    240.943
  (month^2)                              0.934      0.191      0.535      1.309      1.003    185.840

 ...
                                ----------------------------------------------------------------------
```

OBSERVED VS. PREDICTED MEANS



VARIANCE COMPONENT OUTPUT

```
OUTCOME MODEL ESTIMATES:

  Summaries based on 10000 iterations using 2 chains.

Outcome Variable:  breathCO

Parameters                          Median    StdDev      2.5%     97.5%       PSR     N_Eff
                                  -----------------------------------------------------------
Variances:
  L2 : Var(Intercept)               16.105     3.550    10.065    23.972     1.008   250.216
  L2 : Cov(month,Intercept)          1.205     1.431    -1.362     4.226     1.003   509.765
  L2 : Var(month)                    5.148     1.527     2.767     8.746     1.011   209.675
  L2 : Cov(month^2,Intercept)       -2.449     1.033    -4.653    -0.554     1.007   173.358
  L2 : Cov(month^2,month)           -2.625     0.861    -4.639    -1.251     1.023   168.084
  L2 : Var(month^2)                  1.612     0.557     0.673     2.875     1.038    97.502
  Mean Residual Var.                 6.739     1.013     4.854     8.880     1.021   133.344
  Q25% Residual Var.                 2.195     0.509     1.410     3.389     1.011   135.047
  Q50% Residual Var.                 6.702     1.185     4.676     9.308     1.014   189.349
  Q75% Residual Var.                20.614     3.840    14.321    29.243     1.012   236.991

 ...
```

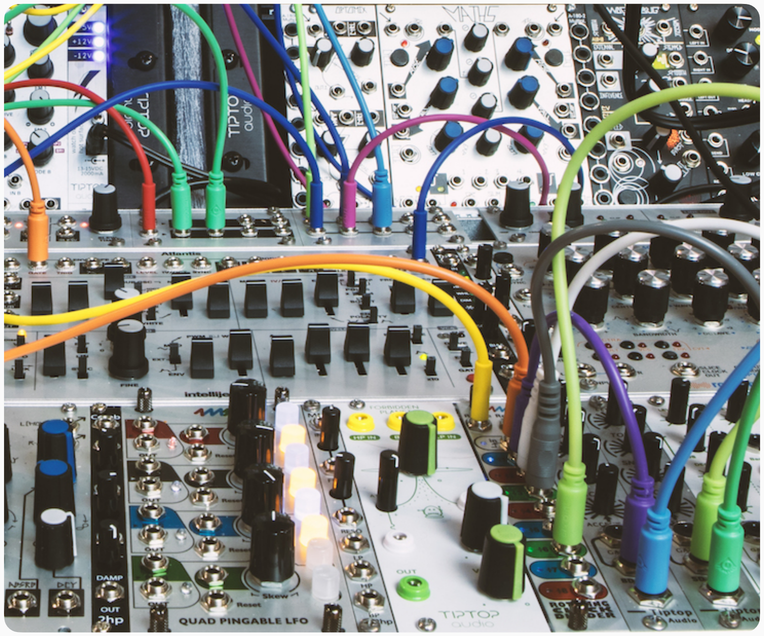
= exp(quantiles of within-cluster variance on log metric)

= exp(mean within-cluster variance on log metric)

INCOMPLETE PREDICTOR VARIABLES IN BLIMP

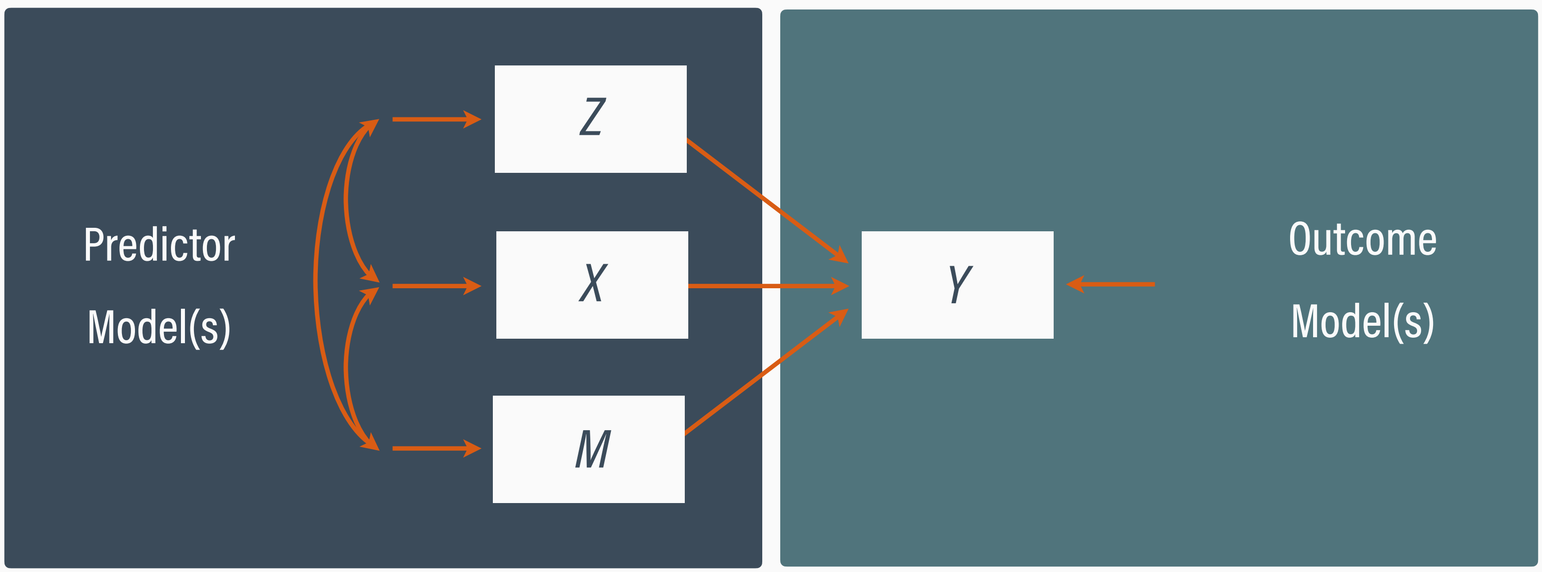
FACTORED REGRESSION MODELING

- Incomplete predictors require a model
- Factored regression uses a modular specification where a sequence of models replaces a multivariate function



FACTORED REGRESSION OVERVIEW

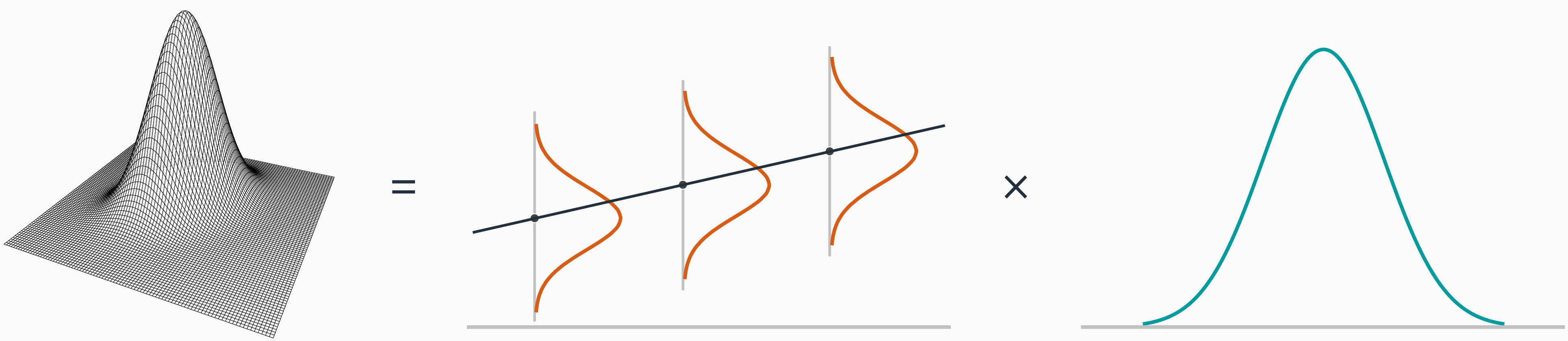
- Factored regression specifications invoke separate models for incomplete predictor variables and outcomes



BIVARIATE FACTORED REGRESSION

Bivariate Distribution = Univariate Outcome Model × Univariate Predictor Model

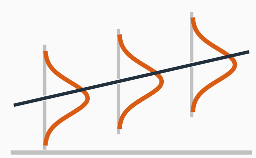
$$f(Y, X) = f(Y|X) \times f(X)$$



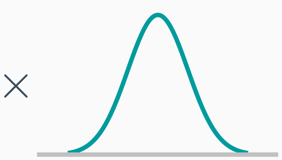
MISSING PREDICTOR VARIABLES

- The factored regression specification dictates the distribution of missing predictor scores

$$f(X|Y) \propto f(Y|X) \times f(X) =$$



$\times$



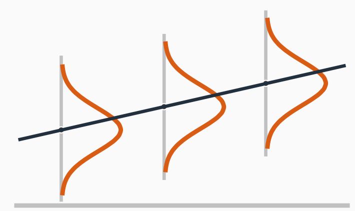
- The distribution of X given Y is the product of the two model-implied distributions (normal curves)

SIMPLE REGRESSION ILLUSTRATION

- The factored regression invokes two linear regressions

Factorization

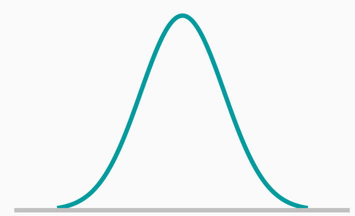
$$f(Y, X) = f(Y\,|X) \times f(X)$$



Fitted models

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

$$X = \gamma_0 + \epsilon$$



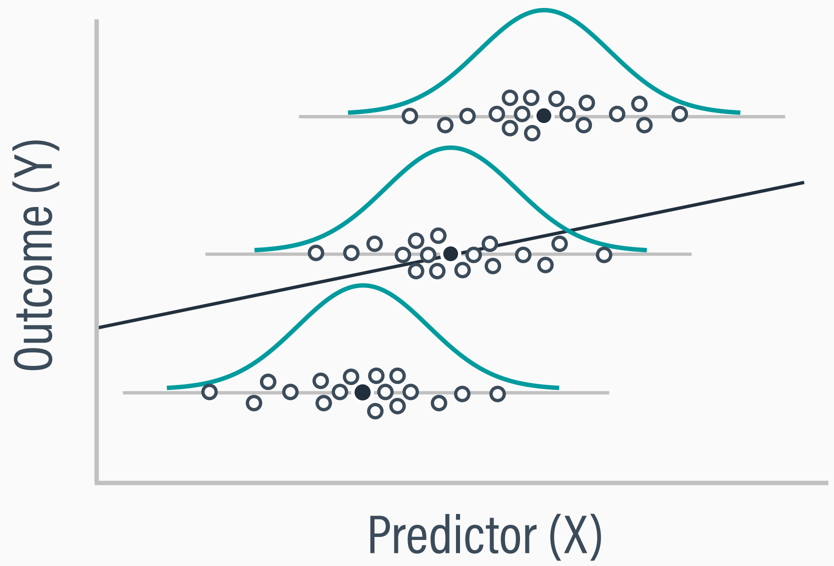
PREDICTOR IMPUTATIONS

- Two sets of model parameters define the center and spread of the imputations

$$X_{mis} \sim N\left(\widehat{X}, \sigma^2_{X|Y}\right)$$

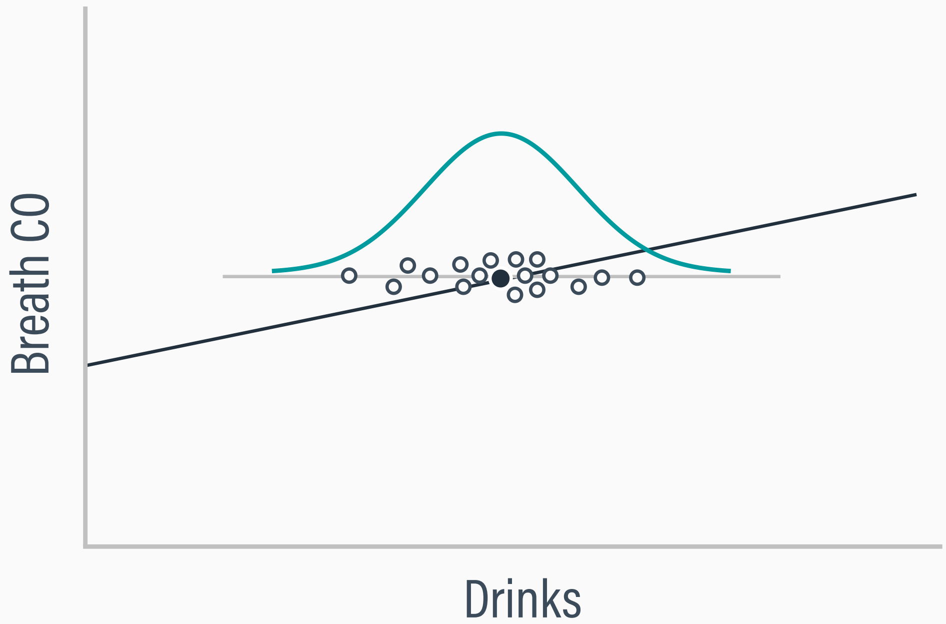
$$\sigma^2_{X|Y} = \left(\frac{1}{\sigma^2_{\epsilon}} + \frac{\beta_1^2}{\sigma^2_{\varepsilon}}\right)^{-1}$$

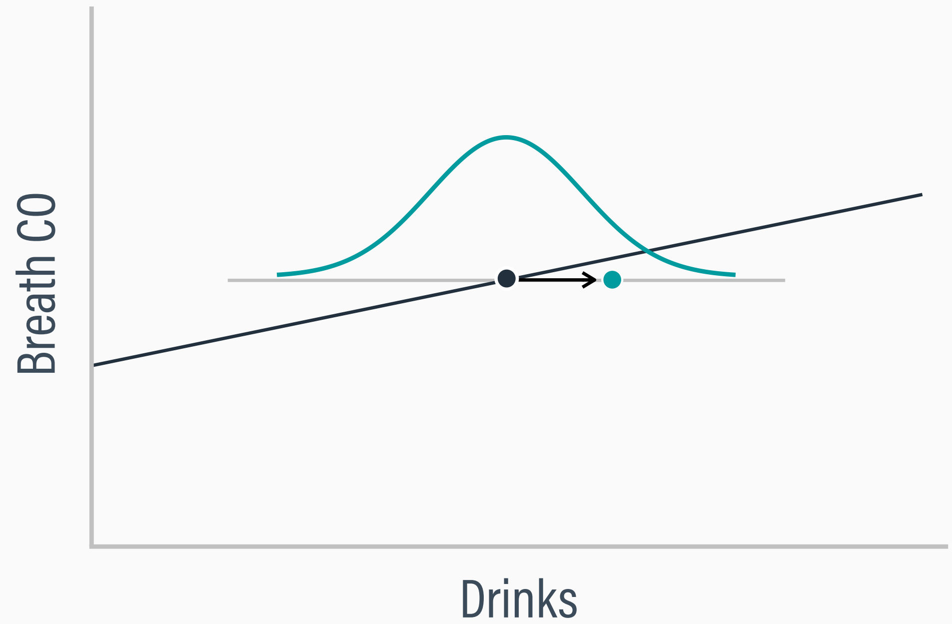
$$\widehat{X} = \sigma^2_{X|Y} \times \left(\frac{\gamma_0}{\sigma^2_{\epsilon}} + \frac{\beta_1(Y - \beta_0)}{\sigma^2_{\varepsilon}}\right)$$



IMPUTATION EXAMPLE

Imputation = predicted value + random normal noise





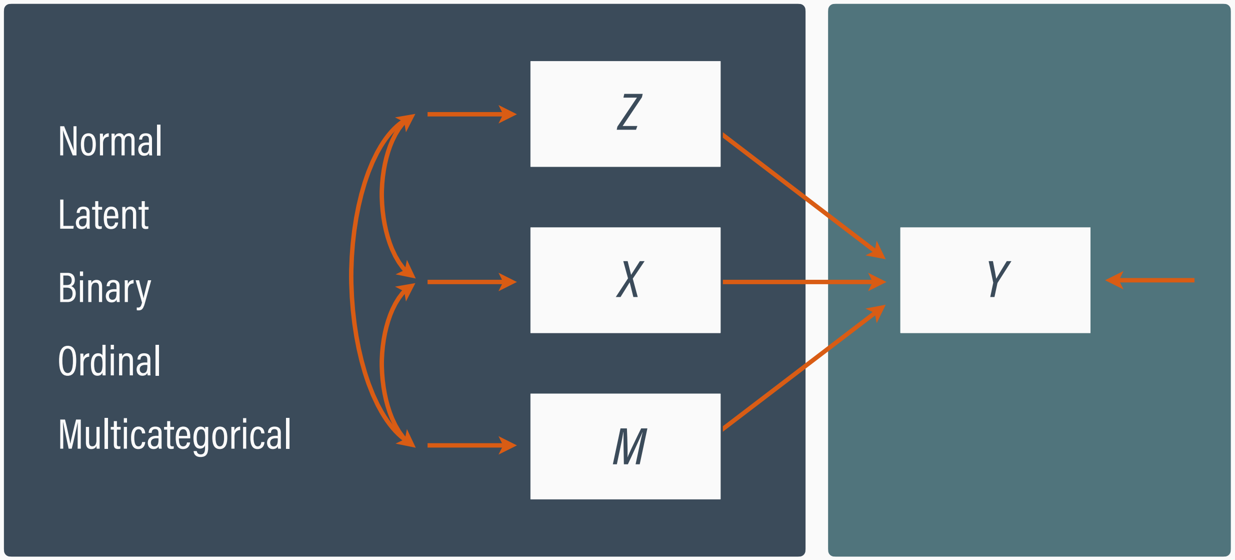
BLIMP DEFAULT SPECIFICATION

Multivariate Distribution = Univariate Outcome Model × Multivariate Predictor Model

$$f(Y,M,X,Z) = f(Y|M,X,Z) \times f(M,X,Z)$$

MODEL:

y ~ z x m;



CLINICAL TRIAL DATA

Predictors

Outcome

Variable	Definition	Missing %	Scale
PERSON	Person-level (level-2) identifier	0	Integer index
WAVE	Data collection wave	0	Integer index (1 to 4)
MONTH	Months relative to 4-week follow-up	0	Integer index (–1 to 2)
BREATHCO	Breath CO reading	23.94	Numeric (0 to 46)
DRINKS	Number of drinks per drinking day	15.00	Numeric (0 to 34)
CIGS	Number of cigarettes per smoking day	14.85	Numeric (0 to 50)
FEMALE	Gender dummy code	0	0 = Male, 1 = Female
CONDITION	Treatment condition	0	0 = Single-medication, 1 = Dual-medication
NICDEP	Nicotine dependence dummy code	6.67	0 = Very Low, to Low 1 = Moderate to Very High
QUIT16	Quit smoking at 4-month follow-up	33.94	0 = Did not quit, 1 = Quit smoking
DROPOUT	Dropout indicator	13.33	0 = In study 1 = Dropped out

ILLUSTRATIVE LINEAR MODEL

- Two incomplete predictors: DRINKS is a time-varying random covariate, and NICDEP is a binary level-2 predictor

$$BREATHCO_{ti} = \beta_{0i} + \beta_{1i}(MONTH_{ti}) + \beta_{2i}\left(DRINKS_{ti} - \mu_{i(D)}\right) + \beta_3(NICDEP_i) + \varepsilon_{ti}$$

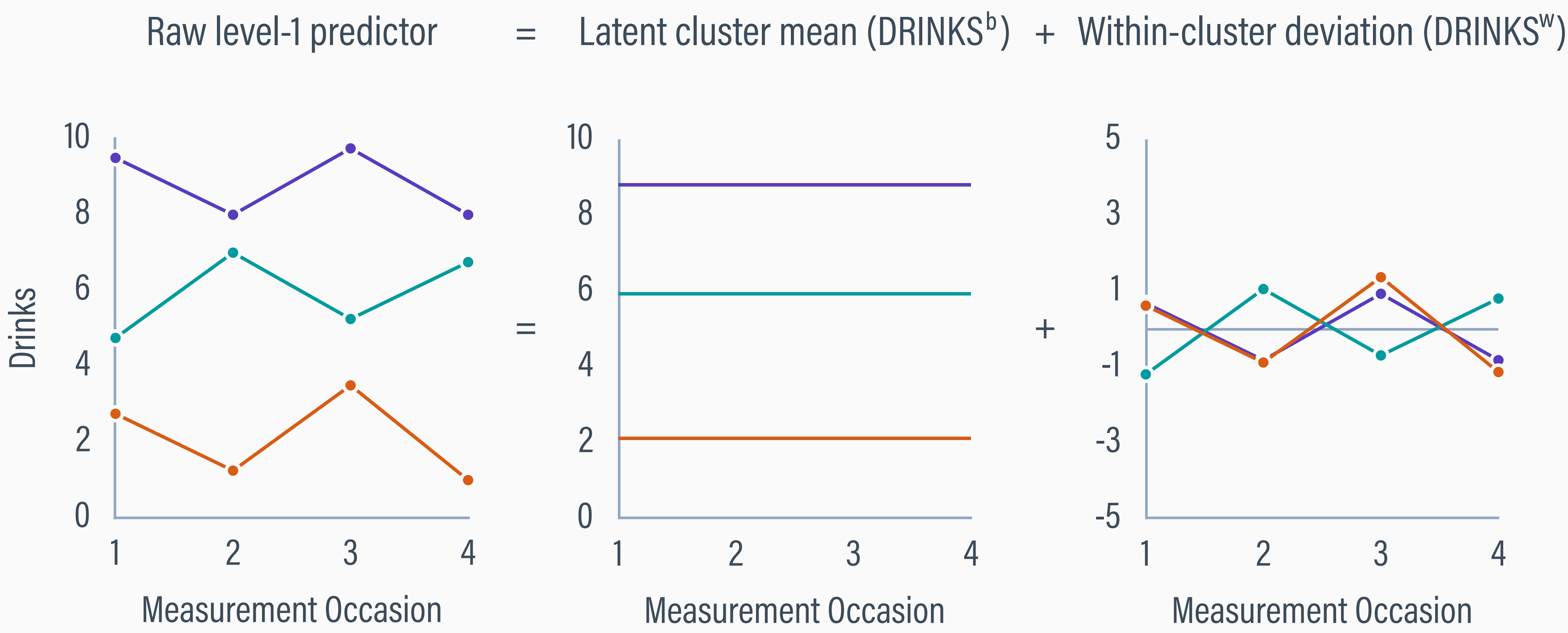
- DRINKS is centered at its latent group means

DISAGGREGATING LEVEL-1 PREDICTORS

- Blimp disaggregates predictors with nonzero ICCs into within-cluster and between-cluster components

$$DRINKS_{ti} = \overbrace{\left(DRINKS_{ti} - \mu_{i(D)}\right)}^{\text{Within-cluster deviation of a score around the latent group mean}} + \underbrace{\left(\mu_{i(D)} - \mu_{(D)}\right)}_{\text{Between-cluster deviation of a latent group mean around the grand mean}} = DRINKS_{ti}^{w} + DRINKS_{i}^{b}$$

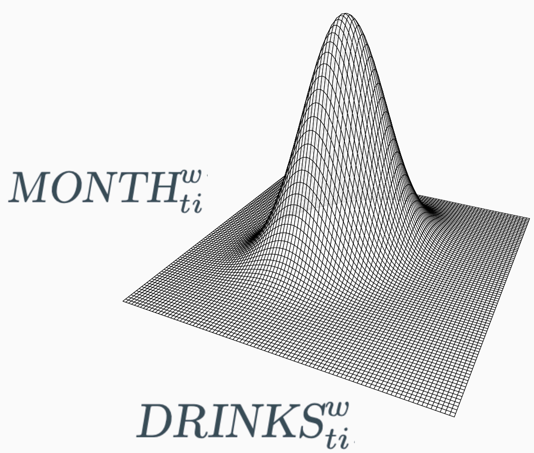
PARTITION FOR THREE PARTICIPANTS



PREDICTOR DISTRIBUTIONS

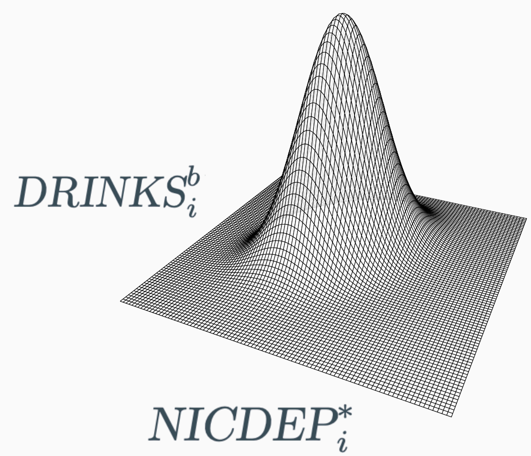
Level-1

$$\begin{pmatrix} MONTH^{w}_{ti} \\ DRINKS^{w}_{ti} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{(M)} \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^{2w}_{M} & \sigma^{w}_{MD} \\ \sigma^{w}_{DM} & \sigma^{2w}_{D} \end{pmatrix}\right)$$



Level-2

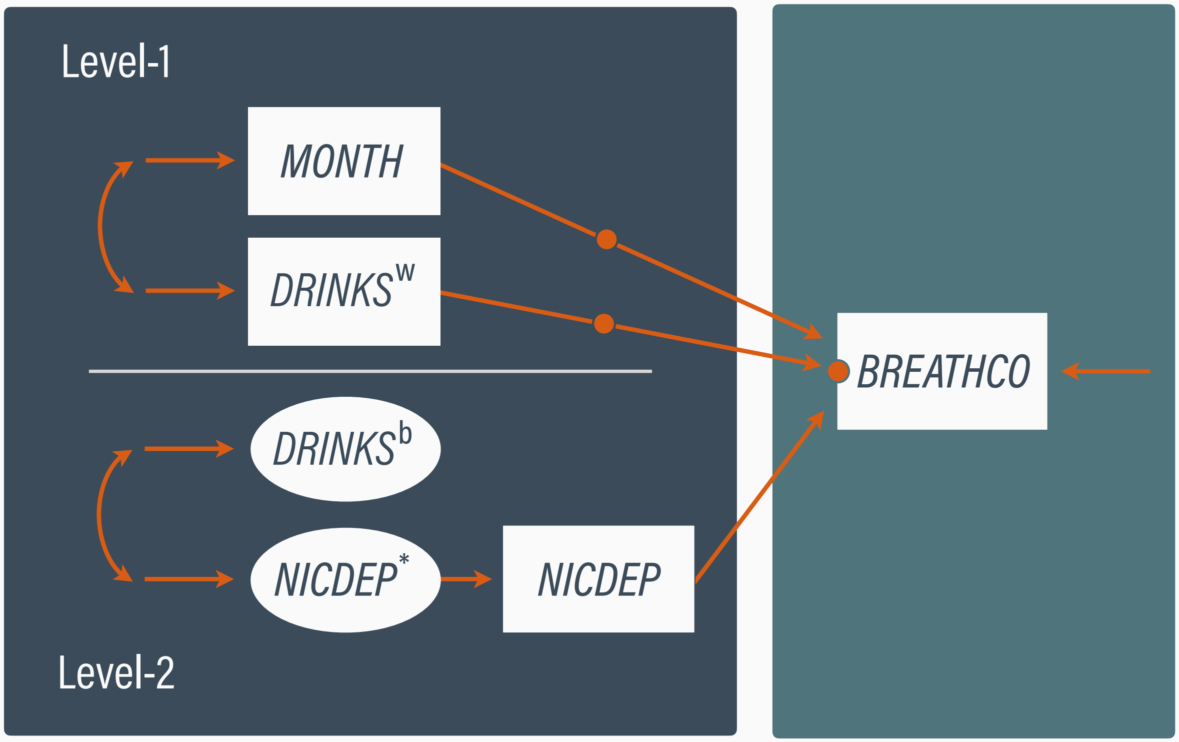
$$\begin{pmatrix} DRINKS^{b}_{i} \\ NICDEP^{*}_{i} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{(D)} \\ \mu_{(N^{*})} \end{pmatrix}, \begin{pmatrix} \sigma^{2b}_{D} & \sigma^{b}_{DN^{*}} \\ \sigma^{b}_{N^{*}D} & 1 \end{pmatrix}\right)$$



PATH DIAGRAM

Joint Distribution = Univariate Outcome Model × Multivariate Predictor Model

ORDINAL: nicdep;
CENTER:
groupmean = drinks;
MODEL:
breathCO ~ month drinks nicdep
  | month drinks;



DISTRIBUTION OF IMPUTATIONS

- The distribution of the DRINKS imputations depends on the focal model and a secondary model linking it to the predictors

$$\text{var}\left(DRINKS_{ti(mis)}|BREATHCO_{ti}, MONTH_{ti}, NICDEP_{i}\right) = \left(\frac{1}{\sigma^2_\epsilon} + \frac{\beta^2_{2i}}{\sigma^2_\varepsilon}\right)^{-1}$$

- Random coefficients induce heteroscedastic variation that depends on the magnitude of a person's random slope

LATENT CONTEXTUAL EFFECTS

- Linear growth model with an incomplete predictor at each level and latent group means as an additional level-2 covariate

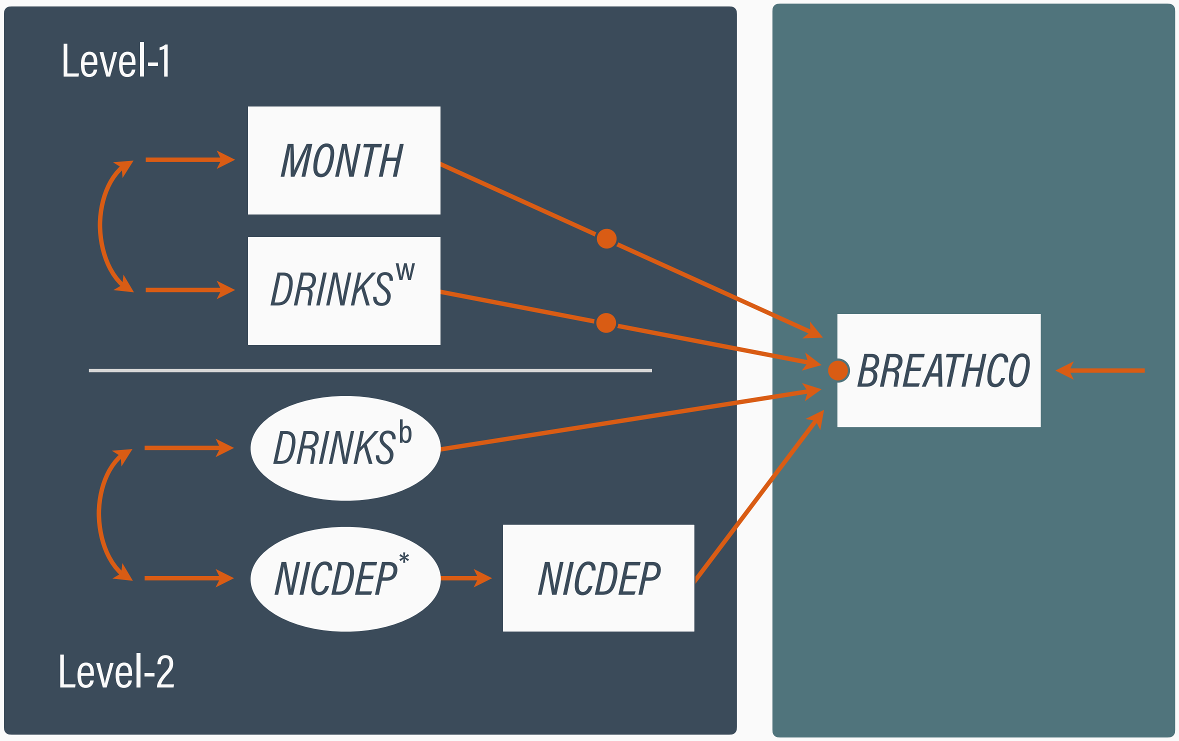
$$BREATHCO_{ti} = \beta_{0i} + \beta_{1i}(MONTH_{ti}) + \beta_{2i}(DRINKS_i^{\,w}) + \beta_3(DRINKS_i^{\,b}) + \beta_4(NICDEP_i) + \varepsilon_{ti}$$

- The influence of drinking is partitioned into a within- and between-person regression slopes

PATH DIAGRAM

Joint Distribution = Univariate Outcome Model × Multivariate Predictor Model

ORDINAL: nicdep;
CENTER:
groupmean = drinks;
grandmean = drinks.mean;
MODEL:
breathCO ~ month drinks nicdep
drinks.mean | month drinks;



CLINICAL TRIAL DATA

Predictors
Outcome

Variable	Definition	Missing %	Scale
PERSON	Person-level (level-2) identifier	0	Integer index
WAVE	Data collection wave	0	Integer index (1 to 4)
MONTH	Months relative to 4-week follow-up	0	Integer index (–1 to 2)
BREATHCO	Breath CO reading	23.94	Numeric (0 to 46)
DRINKS	Number of drinks per drinking day	15.00	Numeric (0 to 34)
CIGS	Number of cigarettes per smoking day	14.85	Numeric (0 to 50)
FEMALE	Gender dummy code	0	0 = Male, 1 = Female
CONDITION	Treatment condition	0	0 = Single-medication, 1 = Dual-medication
NICDEP	Nicotine dependence dummy code	6.67	0 = Very Low, to Low 1 = Moderate to Very High
QUIT16	Quit smoking at 4-month follow-up	33.94	0 = Did not quit, 1 = Quit smoking
DROPOUT	Dropout indicator	13.33	0 = In study 1 = Dropped out

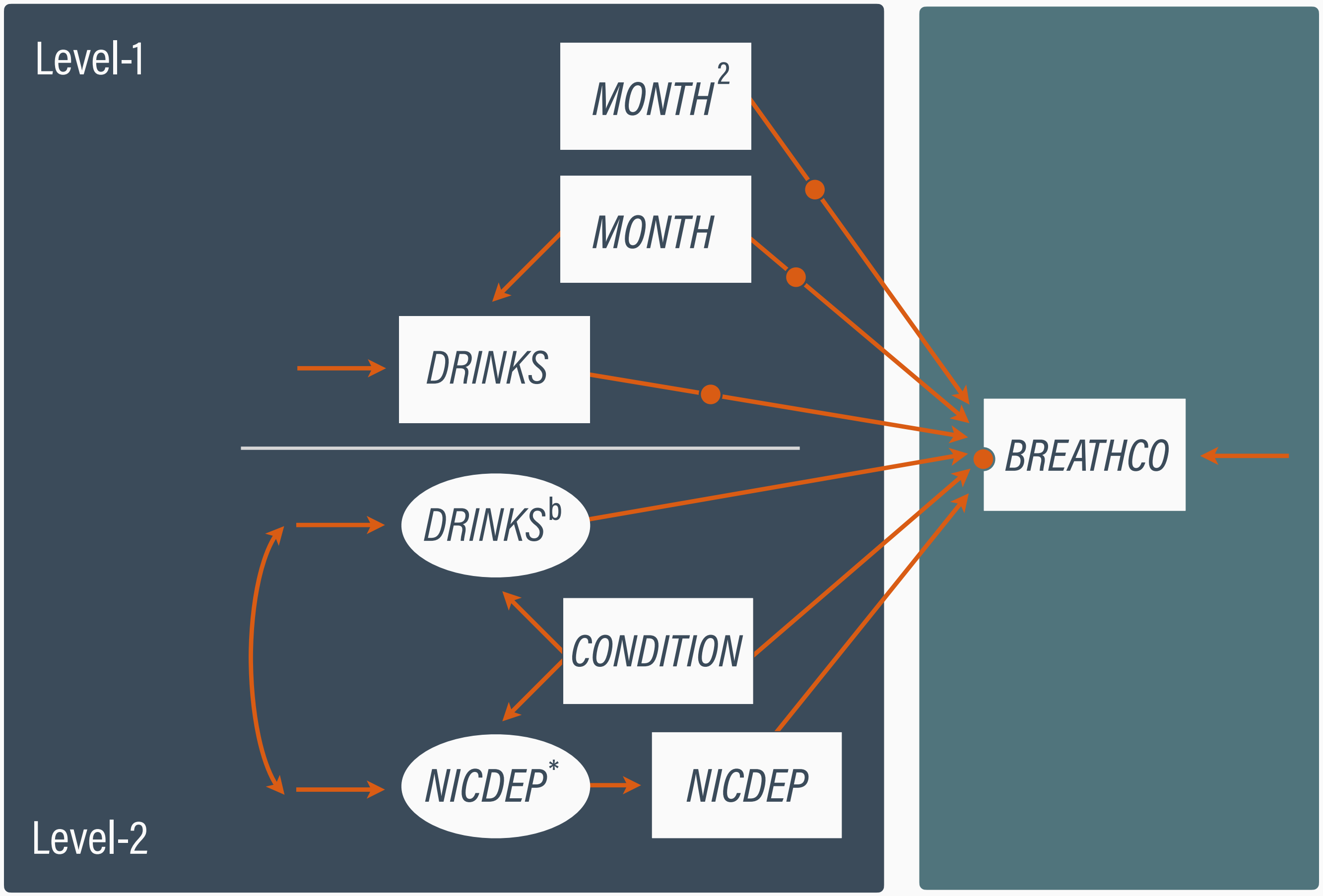
ANALYSIS MODEL

- Quadratic growth model with a random time-varying covariate, latent group means, and and time-invariant predictors

$$BREATHCO_{ti} = \beta_{0i} + \beta_{1i}(MONTH_{ti}) + \beta_{2i}(MONTH_{ti}^{2}) + \beta_{3i}(DRINKS_{ti}^{w})$$
$$+ \beta_{4}\left(DRINKS_{i}^{b} - \mu_{(DRINKS)}\right) + \beta_{5}(CONDITION_{i}) + \beta_{6}\left(NICDEP_{i} - \mu_{(NICDEP)}\right) + \varepsilon_{ti}$$

Measurement	Baseline	4-week	8-week	12-week
Time Scores (MONTH)	–1	0	1	2

PATH DIAGRAM



BLIMP SCRIPT

```
DATA: clinicaltrial.dat;
VARIABLES: person wave month breathCO drinks cigs female condition nicdep quit16 dropout;
CLUSTERID: person;
ORDINAL: condition nicdep;
MISSING: 999;
FIXED: month condition;
CENTER: groupmean = drinks, grandmean = drinks.mean nicdep;
MODEL:
breathCO ~ month (month^2) drinks drinks.mean condition nicdep | month (month^2) drinks;
BURN: 10000;
ITERATIONS: 20000;
SEED: 90291;
```

PSR DIAGNOSTIC OUTPUT

```
BURN-IN POTENTIAL SCALE REDUCTION (PSR) OUTPUT:

  NOTE: Split chain PSR is being used. This splits each chain's
        iterations to create twice as many chains.

  Comparing iterations across 2 chains     Highest PSR   Parameter #
                            251 to 500           1.299             7
                            501 to 1000          1.103            34
                            751 to 1500          1.107            15
                           1001 to 2000          1.038            27
                            ...     ...           ...            ...
                           4001 to 8000          1.103             4
                           4251 to 8500          1.055            10
                           4501 to 9000          1.038            10
                           4751 to 9500          1.025            10
                           5001 to 10000         1.016            10
```

» Worst PSR < 1.05

COEFFICIENT OUTPUT

```
OUTCOME MODEL ESTIMATES:

  Summaries based on 20000 iterations using 2 chains.

Outcome Variable:  breathCO

Parameters                       Median    StdDev      2.5%     97.5%       PSR     N_Eff
                                ------------------------------------------------------------
Variances:
  ...

Coefficients:
  Intercept                       5.682     0.591     4.504     6.810     1.003  1125.433
  month                          -2.456     0.372    -3.191    -1.736     1.002  1598.430
  drinks                          0.312     0.133     0.065     0.582     1.009   554.416
  drinks.mean[person]             0.319     0.278    -0.210     0.887     1.013   446.532
  condition                       1.345     0.740    -0.066     2.815     1.002  1100.831
  nicdep                          2.741     0.767     1.207     4.220     1.001  1095.901
  (month^2)                       1.076     0.209     0.663     1.486     1.001  2296.637

Standardized Coefficients:
  ...
                                ------------------------------------------------------------
```

VARIANCE COMPONENT OUTPUT

```
OUTCOME MODEL ESTIMATES:

  Summaries based on 20000 iterations using 2 chains.

Outcome Variable:  breathCO

Parameters                            Median    StdDev      2.5%     97.5%       PSR     N_Eff
                                ------------------------------------------------------------------
Variances:
  L2 : Var(Intercept)                 18.483     3.575    12.495    26.628     1.001  1203.014
  L2 : Cov(month,Intercept)            5.756     2.161     2.058    10.481     1.002  1234.223
  L2 : Var(month)                     10.041     2.251     6.337    15.134     1.000  1197.078
  L2 : Cov(drinks,Intercept)           1.410     0.697     0.182     2.965     1.005   452.609
  L2 : Cov(drinks,month)              -0.066     0.532    -1.004     1.098     1.013   273.380
  L2 : Var(drinks)                     0.441     0.248     0.108     1.073     1.030   162.376
  L2 : Cov(month^2,Intercept)         -2.763     1.214    -5.548    -0.762     1.000   801.204
  L2 : Cov(month^2,month)             -4.263     1.124    -6.811    -2.435     1.002   709.560
  L2 : Cov(month^2,drinks)             0.013     0.263    -0.574     0.477     1.014   251.194
  L2 : Var(month^2)                    2.058     0.657     1.035     3.597     1.003   445.892
  Residual Var.                       10.196     1.178     8.134    12.723     1.004  1370.317

Coefficients:
  ...
                                ------------------------------------------------------------------
```

EFFECT SIZE OUTPUT

```
OUTCOME MODEL ESTIMATES:

  Summaries based on 20000 iterations using 2 chains.

Outcome Variable:   breathCO

Parameters                           Median     StdDev       2.5%      97.5%        PSR      N_Eff

-------------------------------------------------------------------
Variances:
  ...

Proportion Variance Explained
  by Coefficients                     0.192      0.036      0.126      0.266      1.001   1456.191
  by Level-2 Random Intercepts        0.352      0.048      0.260      0.446      1.010    447.559
  by Level-2 Random Slopes            0.235      0.045      0.162      0.337      1.013    319.565
  by Level-1 Residual Variation       0.215      0.029      0.163      0.278      1.003   1453.998

                               -------------------------------------------------------------------
```

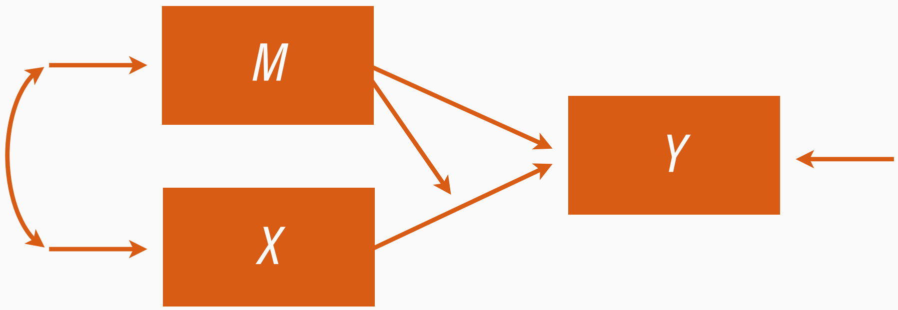
INTERACTION EFFECTS IN BLIMP

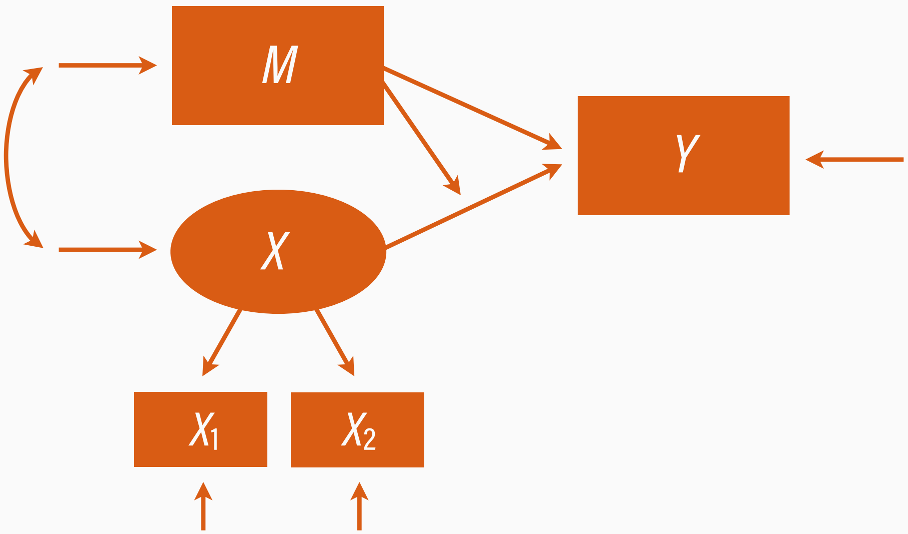
INCOMPLETE PRODUCT TERMS

- Factored regression specifications readily accommodate incomplete interactive and other nonlinear effects
- Product terms appear as deterministic functions of lower-order terms in the focal regression (growth) model
- The models and distributions of the predictors are unchanged

INTERACTION EFFECTS

Multivariate Distribution = Univariate Outcome Model × Predictor Models





PREDICTOR IMPUTATIONS

- The product term induces heteroscedastic variation, such that the spread of the X imputations depends on the value of M

$$X_{mis} \sim \left(\widehat{X}, \sigma^2_{X|Y,M}\right)$$

$$\sigma^2_{X|Y,M} = \left(\frac{1}{\sigma^2_{\epsilon}} + \frac{(\beta_1 + \beta_3 M)^2}{\sigma^2_{\varepsilon}}\right)^{-1}$$

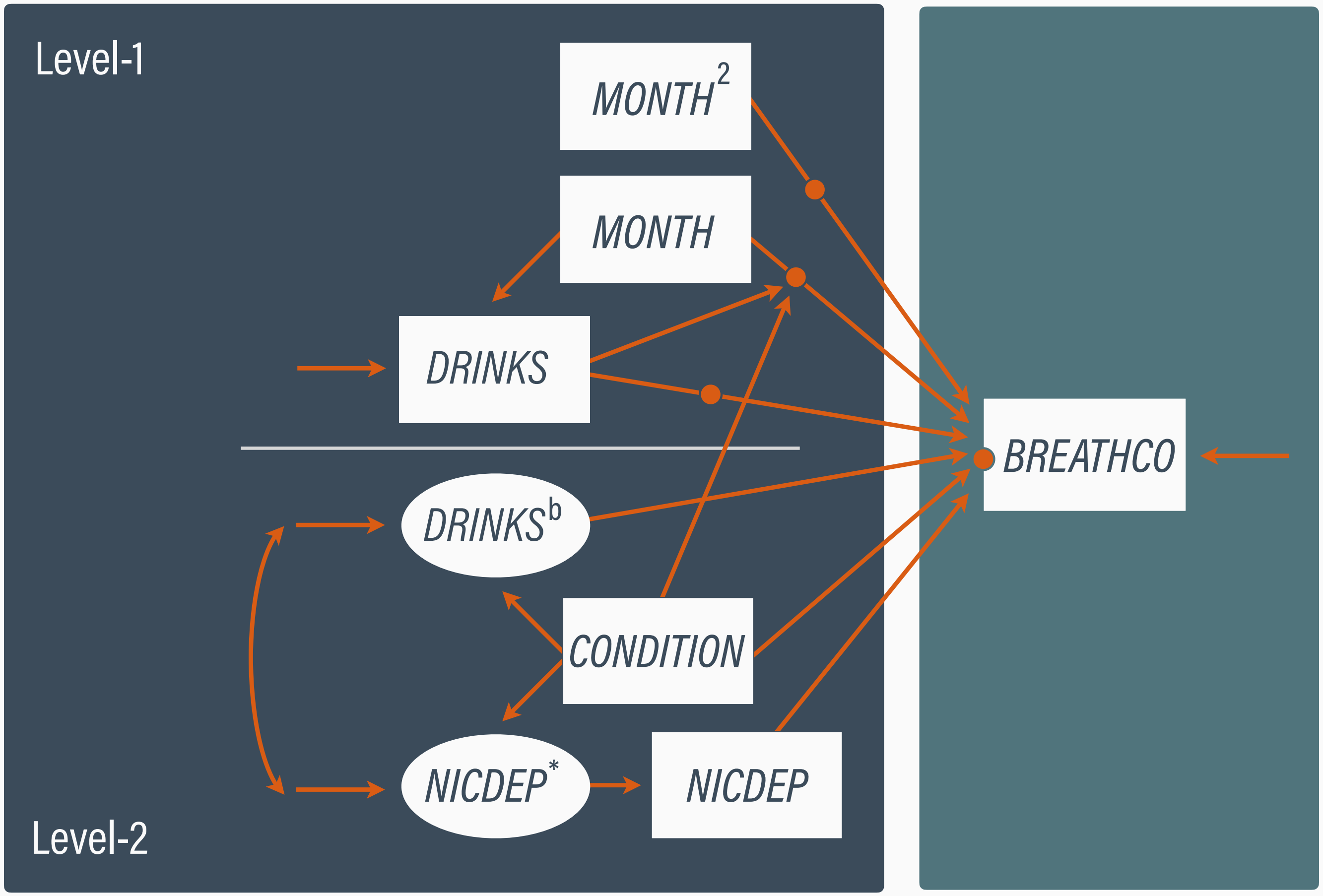
$$\widehat{X} = \sigma^2_{X|Y,M} \times \left(\frac{\gamma_0 + \gamma_1 M}{\sigma^2_{\epsilon}} + \frac{(\beta_1 + \beta_3 M)(Y - \beta_0 - \beta_2 M)}{\sigma^2_{\varepsilon}}\right)$$

ANALYSIS MODEL

- Quadratic growth model with a random time-varying covariate, time-invariant predictors, and within-cluster and cross-level interaction effects

$$
\begin{aligned}
BREATHCO_{ti} &= \beta_{0i} + \beta_{1i}(MONTH_{ti}) + \beta_{2i}(MONTH_{ti}^{2}) + \beta_{3i}(DRINKS_{ti}^{w}) \\
&+ \beta_{4}\left(DRINKS_{i}^{b} - \mu_{(DRINKS)}\right) + \beta_{5}(CONDITION_{i}) + \beta_{6}\left(NICDEP_{i} - \mu_{(NICDEP)}\right) \\
&+ \underbrace{\beta_{7}(MONTH_{ti})(DRINKS_{ti}^{w})}_{\text{Within-person interaction}} + \underbrace{\beta_{8}(MONTH_{ti})(CONDITION_{i})}_{\text{Cross-level interaction}} + \varepsilon_{ti}
\end{aligned}
$$

PATH DIAGRAM



BLIMP SCRIPT

```
DATA: clinicaltrial.dat;
VARIABLES: person wave month breathCO drinks cigs female condition nicdep quit16 dropout;
CLUSTERID: person;
ORDINAL: condition nicdep;
MISSING: 999;
FIXED: month condition;
CENTER: groupmean = drinks, grandmean = drinks.mean nicdep;
MODEL:
breathCO ~ month (month^2) drinks drinks.mean condition nicdep
   month*drinks month*condition | month (month^2) drinks;
BURN: 10000;
ITERATIONS: 20000;
SEED: 90291;
```

« Product terms

PSR DIAGNOSTIC OUTPUT

```
BURN-IN POTENTIAL SCALE REDUCTION (PSR) OUTPUT:

  NOTE: Split chain PSR is being used. This splits each chain's
        iterations to create twice as many chains.

  Comparing iterations across 2 chains     Highest PSR   Parameter #
                         251 to 500              1.518             9
                         501 to 1000             1.203             6
                         751 to 1500             1.118            12
                        1001 to 2000             1.074             4
                         ...     ...               ...           ...
                        4001 to 8000             1.047             4
                        4251 to 8500             1.041             5
                        4501 to 9000             1.044            15
                        4751 to 9500             1.043             4
                        5001 to 10000            1.034             9
```

» Worst PSR < 1.05

COEFFICIENT OUTPUT

```
OUTCOME MODEL ESTIMATES:

  Summaries based on 20000 iterations using 2 chains.

Outcome Variable:   breathCO

Parameters                         Median     StdDev       2.5%      97.5%        PSR      N_Eff
                                -----------------------------------------------------------------
Variances:
  ...

Coefficients:
  Intercept                         5.545      0.558      4.436      6.619      1.001   1134.035
  month                            -2.210      0.419     -3.033     -1.382      1.000   1638.356
  drinks                            0.368      0.131      0.118      0.623      1.000    722.465
  drinks.mean[person]               0.222      0.252     -0.257      0.749      1.008    508.331
  condition                         1.309      0.729     -0.113      2.754      1.004   1147.243
  nicdep                            2.682      0.763      1.193      4.213      1.002   1087.752
  (month^2)                         0.886      0.216      0.465      1.314      1.001   2295.107
  month*drinks                     -0.263      0.099     -0.453     -0.065      1.001   1020.609
  month*condition                  -0.129      0.416     -0.937      0.704      1.001   1861.082

Standardized Coefficients:
  ...
                                -----------------------------------------------------------------
```

Cross-level interaction

Within-person interaction

BLIMP SCRIPT: CONDITIONAL EFFECTS

```
MODEL:
breathCO ~ month (month^2) drinks drinks.mean condition nicdep
    month*drinks month*condition | month (month^2) drinks;
SIMPLE: month | drinks;
SIMPLE: month | condition;
```

« Monthly change by levels of the moderators

CONDITIONAL EFFECT OUTPUT

```
Conditional Effects                    Median     StdDev       2.5%      97.5%        PSR      N_Eff
                              -------------------------------------------------------------------
...

  month | drinks @ +1 SD
    Intercept                           6.695      0.746      5.209      8.132      1.000    869.797
    Slope                              -2.141      0.430     -2.973     -1.280      1.000   1402.021

  month | drinks @ 0
    Intercept                           5.545      0.558      4.436      6.619      1.001   1134.035
    Slope                              -1.322      0.298     -1.907     -0.733      1.000   1904.117

  month | drinks @ -1 SD
    Intercept                           4.387      0.635      3.158      5.641      1.001   1077.089
    Slope                              -0.511      0.428     -1.332      0.338      1.001   1190.200

  ...
                              -----------------------------------------------------------------
```

CONDITIONAL EFFECT OUTPUT

```
Conditional Effects              Median     StdDev       2.5%      97.5%        PSR      N_Eff
                                 ---------------------------------------------------------------
...
  month | condition @ 0
    Intercept                     5.545      0.558      4.436      6.619      1.001   1134.035
    Slope                        -1.322      0.298     -1.907     -0.733      1.000   1904.117

  month | condition @ 1
    Intercept                     6.859      0.625      5.649      8.102      1.002    834.610
    Slope                        -1.448      0.333     -2.087     -0.782      1.001   1482.718
                                 ________________________________
```

STRUCTURAL MODELS IN BLIMP

STRUCTURAL MODELS

- Level-1 variables can predictor other level-1 variables
- Level-2 variables and the random effects (or latent group means) of level-1 variables can predict level-2 variables
- Outcome variables can be any combination of normal, skewed continuous, latent, binary, ordinal, multicategorical, and count

CLINICAL TRIAL DATA

Predictors

Outcomes

Variable	Definition	Missing %	Scale
PERSON	Person-level (level-2) identifier	0	Integer index
WAVE	Data collection wave	0	Integer index (1 to 4)
MONTH	Months relative to 4-week follow-up	0	Integer index (–1 to 2)
BREATHCO	Breath CO reading	23.94	Numeric (0 to 46)
DRINKS	Number of drinks per drinking day	15.00	Numeric (0 to 34)
CIGS	Number of cigarettes per smoking day	14.85	Numeric (0 to 50)
FEMALE	Gender dummy code	0	0 = Male, 1 = Female
CONDITION	Treatment condition	0	0 = Single-medication, 1 = Dual-medication
NICDEP	Nicotine dependence dummy code	6.67	0 = Very Low, to Low 1 = Moderate to Very High
QUIT16	Quit smoking at 4-month follow-up	33.94	0 = Did not quit, 1 = Quit smoking
DROPOUT	Dropout indicator	13.33	0 = In study 1 = Dropped out

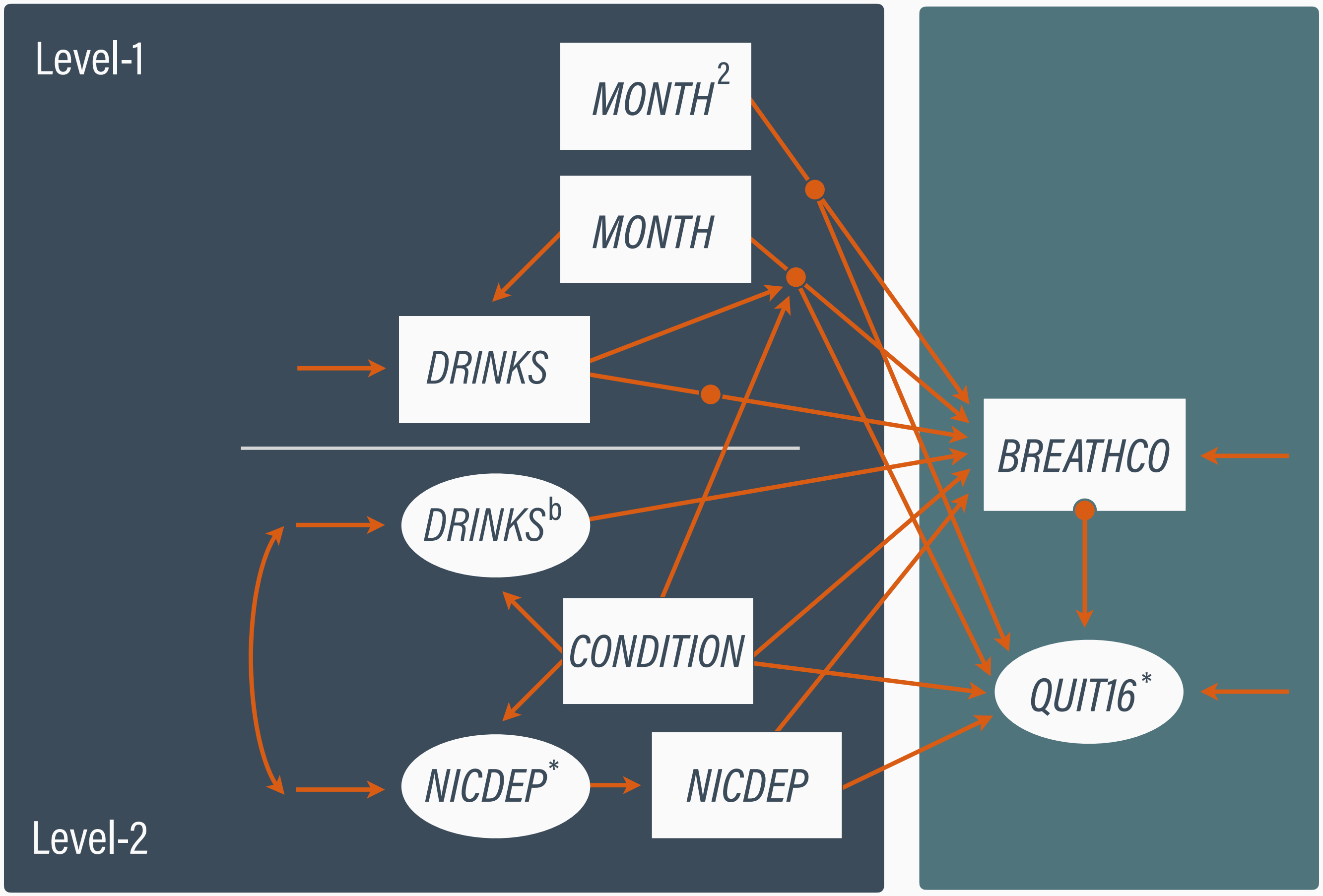
ANALYSIS MODEL

- Quadratic growth model with trajectories and time-invariant covariates predicting quit status at the 4-month follow-up

$$
\begin{aligned}
BREATHCO_{ti} &= \beta_{0i} + \beta_{1i}(MONTH_{ti}) + \beta_{2i}(MONTH_{ti}^{2}) + \beta_{3i}(DRINKS_{ti}^{w}) \\
&+ \beta_{4}\left(DRINKS_{i}^{b} - \mu_{(DRINKS)}\right) + \beta_{5}(CONDITION_{i}) + \beta_{6}\left(NICDEP_{i} - \mu_{(NICDEP)}\right) \\
&+ \beta_{7}(MONTH_{ti})(DRINKS_{ti}^{w}) + \beta_{8}(MONTH_{ti})(CONDITION_{i}) + \varepsilon_{ti}
\end{aligned}
$$

$$
\operatorname{logit}(QUIT16_{i}) = \gamma_{0} + \gamma_{1}(\beta_{0i}) + \gamma_{2}(\beta_{1i}) + \gamma_{3}(\beta_{2i}) + \gamma_{4}(CONDITION_{i}) + \gamma_{5}(NICDEP_{i})
$$

PATH DIAGRAM



BLIMP SCRIPT

```
DATA: clinicaltrial.dat;
VARIABLES: person wave month breathCO drinks cigs female condition nicdep quit16 dropout;
CLUSTERID: person;
ORDINAL: condition nicdep quit16;
MISSING: 999;
RANDOMEFFECT:
ranicepts = breathCO | 1 [person];
ranslopes = breathCO | month [person];
rancurves = breathCO | month^2 [person];
```

« Define random effects as latent variables

BLIMP SCRIPT, CONTINUED

```
FIXED: month condition;
CENTER: groupmean = drinks, grandmean = drinks.mean nicdep;
MODEL:
breathCO ~ month (month^2) drinks drinks.mean condition nicdep
    month*drinks month*condition | month (month^2) drinks;
logit(quit16) ~ ranicepts ranslopes rancurves condition nicdep;     « Logistic outcome model
BURN: 60000;
ITERATIONS: 150000;
SEED: 90291;
```

PSR DIAGNOSTIC OUTPUT

```
BURN-IN POTENTIAL SCALE REDUCTION (PSR) OUTPUT:

  NOTE: Split chain PSR is being used. This splits each chain's
        iterations to create twice as many chains.

  Comparing iterations across 2 chains     Highest PSR   Parameter #
                              1501 to 3000       1.362          10
                              3001 to 6000       1.158           6
                              4501 to 9000       1.145          30
                              6001 to 12000      1.169          34
                               ...     ...        ...          ...
                             24001 to 48000      1.076           6
                             25501 to 51000      1.038          36
                             27001 to 54000      1.046          36
                             28501 to 57000      1.038          36
                             30001 to 60000      1.045           6
```

» Worst PSR < 1.05

COEFFICIENT OUTPUT

OUTCOME MODEL ESTIMATES:

```
  Summaries based on 150000 iterations using 2 chains.

...

Outcome Variable:  logit(quit16)

Parameters                          Median      StdDev        2.5%       97.5%         PSR       N_Eff
                                ------------------------------------------------------------------
Coefficients:
  Intercept                          0.001       0.764      -1.608       1.441       1.000    1952.803
  ranicepts                         -0.921       0.789      -3.541      -0.421       1.005     109.498
  ranslopes                         -1.519       0.995      -3.152       0.898       1.013     148.055
  rancurves                         -3.079       2.298      -6.462       2.756       1.013      98.556
  condition                         -0.241       1.084      -2.409       1.913       1.000    2414.594
  nicdep                            -2.077       1.145      -4.677      -0.152       1.001    2022.065

Odds Ratios:
  Intercept                          1.001       1.509       0.200       4.223       1.000    2774.290
  ranicepts                          0.398       0.162       0.029       0.657       1.004     121.467
  ranslopes                          0.219       0.998       0.043       2.456       1.007     279.311
  rancurves                          0.046      33.830       0.002      15.733       1.001    4426.390
  condition                          0.785       3.300       0.090       6.774       1.001    3848.116
  nicdep                             0.125       0.430       0.009       0.859       1.000    8129.363

Proportion Variance Explained
  ...
                                ------------------------------------------------------------------
```

MISSING NOT AT RANDOM MODELS IN BLIMP

MISSING NOT AT RANDOM MODELS

- A missing not at random (MNAR) process occurs when the unseen score values carry information about missingness
- Selection and pattern mixture modeling are the two primary frameworks, both are available for longitudinal analyses
- Selection models feature an additional regression with a binary missing data or dropout indicator as the dependent variable

LONGITUDINAL MNAR PROCESSES

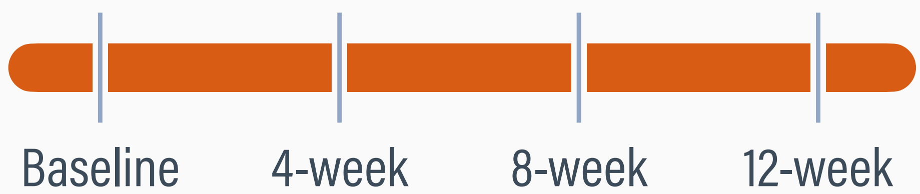
- Random coefficient-dependent missingness occurs when one's underlying growth trajectory is responsible for missing data
- Outcome-dependent missingness occurs when the unseen value of the dependent variable at a particular occasion predicts nonresponse

MISSING DATA INDICATORS

- A missing data indicator codes each instance of the outcome variable as 0 = observed and 1 = missing
- If most participants permanently attrit, then the binary dropout indicator is coded as 0 = observed at occasion t, 1 = dropped out at occasion t, and NA = dropped out prior to occasion t

DROPOUT INDICATOR CODING

Completer pattern	Obs 0	Obs 0	Obs 0	Obs 0
Dropout pattern 1	Obs 0	Obs 0	Obs 0	Mis 1
Dropout pattern 2	Obs 0	Obs 0	Mis 1	Mis NA
Dropout pattern 3	Obs 0	Mis 1	Mis NA	Mis NA



CLINICAL TRIAL DATA

Predictors

Outcomes

Variable	Definition	Missing %	Scale
PERSON	Person-level (level-2) identifier	0	Integer index
WAVE	Data collection wave	0	Integer index (1 to 4)
MONTH	Months relative to 4-week follow-up	0	Integer index (–1 to 2)
BREATHCO	Breath CO reading	23.94	Numeric (0 to 46)
DRINKS	Number of drinks per drinking day	15.00	Numeric (0 to 34)
CIGS	Number of cigarettes per smoking day	14.85	Numeric (0 to 50)
FEMALE	Gender dummy code	0	0 = Male, 1 = Female
CONDITION	Treatment condition	0	0 = Single-medication, 1 = Dual-medication
NICDEP	Nicotine dependence dummy code	6.67	0 = Very Low, to Low 1 = Moderate to Very High
QUIT16	Quit smoking at 4-month follow-up	33.94	0 = Did not quit, 1 = Quit smoking
DROPOUT	Dropout indicator	13.33	0 = In study 1 = Dropped out

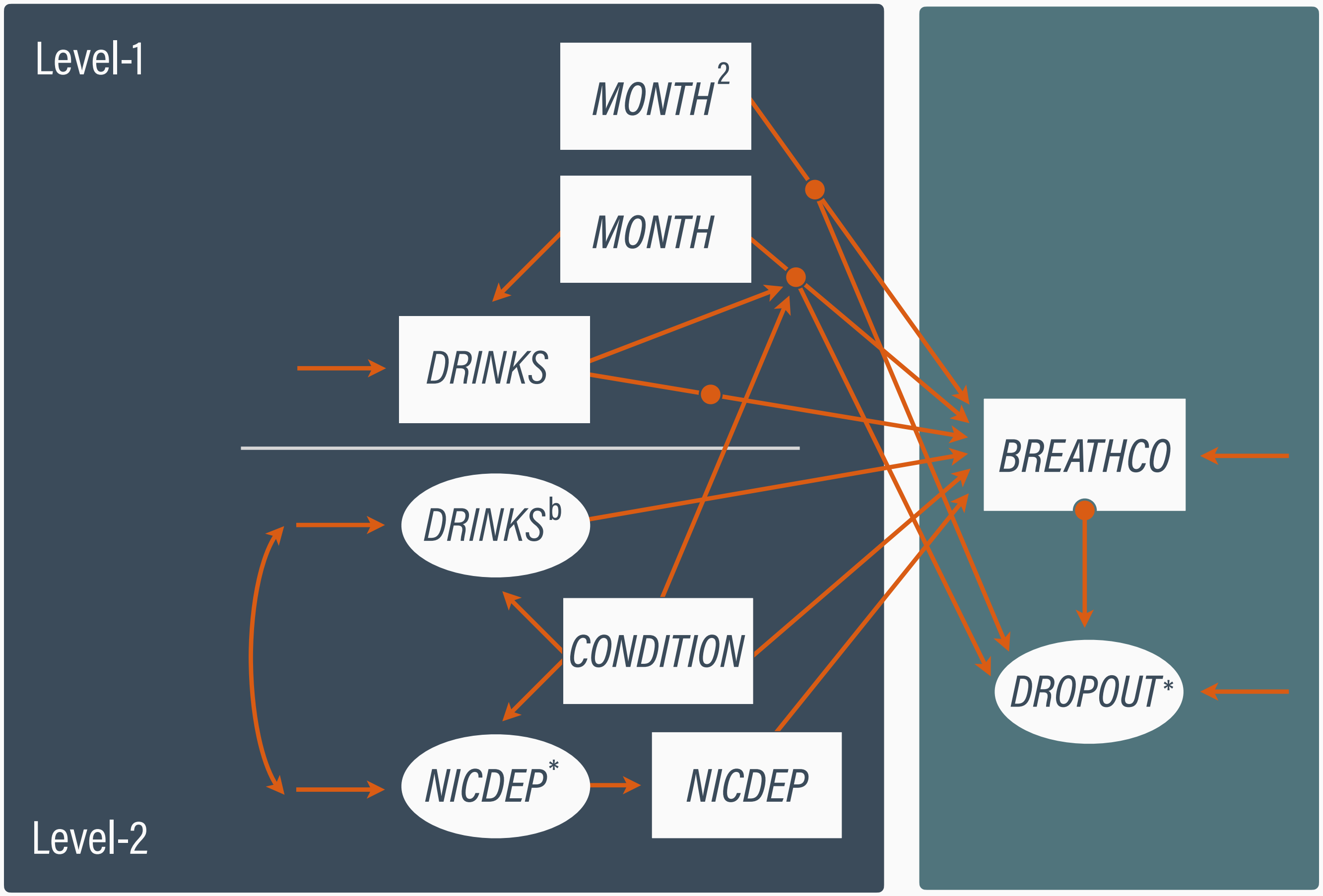
WU-CARROLL MODEL

- Growth model with occasion-specific dummy codes and random coefficients predicting the dropout indicator

$$
\begin{aligned}
BREATHCO_{ti} &= \beta_{0i} + \beta_{1i}(MONTH_{ti}) + \beta_{2i}(MONTH^{2}_{ti}) + \beta_{3i}(DRINKS^{w}_{ti}) \\
&+ \beta_4\left(DRINKS^{b}_{i} - \mu_{(DRINKS)}\right) + \beta_5(CONDITION_i) + \beta_6\left(NICDEP_i - \mu_{(NICDEP)}\right) \\
&+ \beta_7(MONTH_{ti})(DRINKS^{w}_{ti}) + \beta_8(MONTH_{ti})(CONDITION_i) + \varepsilon_{ti}
\end{aligned}
$$

$$
DROPOUT^{*}_{ti} = \gamma_0 + \gamma_1(T_{2i}) + \gamma_2(T_{3i}) + \gamma_3(T_{4i}) + \gamma_4(\beta_{0i}) + \gamma_5(\beta_{1i}) + \gamma_6(\beta_{2i}) + \epsilon_{ti}
$$

PATH DIAGRAM



BLIMP SCRIPT

```
DATA: clinicaltrial.dat;
VARIABLES: person wave month breathCO drinks cigs female condition nicdep quit16 dropout;
CLUSTERID: person;
ORDINAL: condition nicdep dropout;
MISSING: 999;
RANDOMEFFECT:
ranicepts = breathCO | 1 [person];
ranslopes = breathCO | month [person];
rancurves = breathCO | month^2 [person]
```

BLIMP SCRIPT, CONTINUED

```
MODEL:
breathCO ~ month (month^2) drinks drinks.mean condition nicdep
    month*drinks month*condition | month (month^2) drinks;
dropout ~ 1@-3 (month == 0) (month == 1) (month == 2) ranicepts ranslopes rancurves | 1@0;
BURN: 50000;
ITERATIONS: 100000;
SEED: 90291;
```

PSR DIAGNOSTIC OUTPUT

```
BURN-IN POTENTIAL SCALE REDUCTION (PSR) OUTPUT:

  NOTE: Split chain PSR is being used. This splits each chain's
        iterations to create twice as many chains.

  Comparing iterations across 2 chains     Highest PSR   Parameter #
                         1251 to 2500            1.477            35
                         2501 to 5000            1.514            37
                         3751 to 7500            1.290            43
                         5001 to 10000           1.357            36
                          ...     ...             ...            ...
                        20001 to 40000           1.063            42
                        21251 to 42500           1.023            32
                        22501 to 45000           1.023             6
                        23751 to 47500           1.031             6
                        25001 to 50000           1.045             6
```

» Worst PSR < 1.05

COEFFICIENT OUTPUT

```
OUTCOME MODEL ESTIMATES:

  Summaries based on 100000 iterations using 2 chains.

Outcome Variable:  breathCO

Parameters                          Median    StdDev      2.5%     97.5%       PSR     N_Eff
                                ------------------------------------------------------------------
Variances:
  ...

Coefficients:
  Intercept                          5.089     0.667     3.778     6.385     1.002   305.714
  month                             -2.456     0.496    -3.448    -1.500     1.001   295.011
  drinks                             0.360     0.143     0.091     0.650     1.003   393.015
  drinks.mean[person]                0.260     0.247    -0.205     0.765     1.001   443.548
  condition                          1.541     0.712     0.205     2.995     1.001   772.555
  nicdep                             2.919     0.775     1.378     4.395     1.002   714.591
  month^2                            0.959     0.255     0.461     1.466     1.003   306.829
  month*drinks                      -0.280     0.098    -0.473    -0.089     1.001   768.577
  month*condition                   -0.142     0.410    -0.943     0.660     1.002  2068.035

Standardized Coefficients:
  ...

                                ------------------------------------------------------------------
```

COEFFICIENT OUTPUT

```
OUTCOME MODEL ESTIMATES:

  Summaries based on 100000 iterations using 2 chains.

...

Outcome Variable:   dropout

Parameters                            Median     StdDev       2.5%      97.5%        PSR      N_Eff
                                -------------------------------------------------------------------
Variances:
  Residual Var.                        1.000      0.000      1.000      1.000        nan        nan

Coefficients:
  Intercept                           -3.000      0.000     -3.000     -3.000        nan        nan
  ranicepts                           -0.062      0.068     -0.200      0.075      1.007    313.659
  ranslopes                           -0.112      0.610     -1.196      1.193      1.013    170.750
  rancurves                           -0.281      1.446     -2.831      2.895      1.017    173.331
  (month == 0)                         2.191      0.149      1.858      2.445      1.003    727.572
  (month == 1)                         1.378      0.207      0.938      1.748      1.001   3130.294
  (month == 2)                         1.396      0.208      0.966      1.782      1.000   5385.965

Thresholds:
  Tau 1                                0.000      0.000      0.000      0.000        nan        nan

  ...
                                -------------------------------------------------------------------
```

DIGGLE-KENWARD MODEL

- Growth model with occasion-specific dummy codes and concurrent and lagged outcome scores predicting dropout

$$BREATHCO_{ti} = \beta_{0i} + \beta_{1i}(MONTH_{ti}) + \beta_{2i}(MONTH^2_{ti}) + \beta_{3i}(DRINKS^w_{ti})$$
$$+ \beta_4\left(DRINKS^b_i - \mu_{(DRINKS)}\right) + \beta_5(CONDITION_i) + \beta_6\left(NICDEP_i - \mu_{(NICDEP)}\right)$$
$$+ \beta_7(MONTH_{ti})(DRINKS^w_{ti}) + \beta_8(MONTH_{ti})(CONDITION_i) + \varepsilon_{ti}$$

$$DROPOUT^*_{ti} = \gamma_0 + \gamma_1(T_{2i}) + \gamma_2(T_{3i}) + \gamma_3(T_{4i})$$
$$+ \gamma_4(BREATHCO_{ti}) + \gamma_5(BRTHCOLAG_{ti}) + \epsilon_{ti}$$

BLIMP SCRIPT

```
DATA: clinicaltrial.dat;
VARIABLES: person wave month breathCO drinks cigs female condition nicdep quit16 dropout;
CLUSTERID: person;
ORDINAL: condition nicdep dropout;
MISSING: 999;
TRANSFORM:
brthCOlag = lag1(breathCO, month, person);              « Compute lagged outcome
```

BLIMP SCRIPT, CONTINUED

```
FIXED: month condition;
CENTER: groupmean = drinks, grandmean = drinks.mean nicdep;
MODEL:
breathCO ~ month (month^2) drinks drinks.mean condition nicdep
    month*drinks month*condition | month (month^2) drinks;
brthCOlag ~ breathCO month  drinks condition nicdep | 1@0;
dropout ~ 1@-3 (month == 0) (month == 1) (month == 2) breathCO brthCOlag | 1@0;
BURN: 20000;
ITERATIONS: 10000;
SEED: 90291;
```

PSR DIAGNOSTIC OUTPUT

```
BURN-IN POTENTIAL SCALE REDUCTION (PSR) OUTPUT:

  NOTE: Split chain PSR is being used. This splits each chain's
        iterations to create twice as many chains.

  Comparing iterations across 2 chains     Highest PSR   Parameter #
                          501 to 1000          1.507              9
                         1001 to 2000          1.113              6
                         1501 to 3000          1.149              6
                         2001 to 4000          1.108              6
                          ...     ...            ...            ...
                         8001 to 16000         1.035              6
                         8501 to 17000         1.009              6
                         9001 to 18000         1.013             41
                         9501 to 19000         1.016             41
» Worst PSR < 1.05      10001 to 20000         1.010              4
```

COEFFICIENT OUTPUT

```
OUTCOME MODEL ESTIMATES:

  Summaries based on 10000 iterations using 2 chains.

Outcome Variable:  breathCO

Parameters                             Median     StdDev       2.5%      97.5%        PSR      N_Eff
                                ------------------------------------------------------------------
Variances:
  ...

Coefficients:
  Intercept                             5.769      0.601      4.662      6.953      1.005    468.534
  month                                -2.159      0.425     -2.993     -1.314      1.004    525.492
  drinks                                0.263      0.129      0.018      0.518      1.006    300.078
  drinks.mean[person]                   0.408      0.293     -0.127      1.004      1.010    222.795
  condition                             1.548      0.756      0.061      2.980      1.004    493.066
  nicdep                                2.954      0.771      1.466      4.505      1.002    586.808
  (month^2)                             0.895      0.216      0.475      1.316      1.002    588.550
  month*drinks                         -0.184      0.097     -0.379      0.002      1.013    394.952
  month*condition                      -0.173      0.409     -1.001      0.602      1.001   1299.504

Standardized Coefficients:
  ...

                                ------------------------------------------------------------------
```

COEFFICIENT OUTPUT

```
OUTCOME MODEL ESTIMATES:

  Summaries based on 10000 iterations using 2 chains.

...

Outcome Variable:  dropout

Parameters                         Median     StdDev       2.5%      97.5%        PSR      N_Eff
                                ------------------------------------------------------------------
Variances:
  Residual Var.                     1.000      0.000      1.000      1.000        nan        nan

Coefficients:
  Intercept                        -3.000      0.000     -3.000     -3.000        nan        nan
  breathCO                          0.047      0.029     -0.014      0.100      1.008    178.917
  brthCOlag                        -0.029      0.023     -0.075      0.015      1.007    218.258
  (month == 0)                      2.292      0.165      1.967      2.618      1.001    962.839
  (month == 1)                      1.324      0.209      0.903      1.722      1.002    654.210
  (month == 2)                      1.294      0.224      0.834      1.715      1.004    441.393

Thresholds:
  Tau 1                             0.000      0.000      0.000      0.000        nan        nan

Standardized Coefficients:
  ...
                                ------------------------------------------------------------------
```

MODEL COMPARISON

Parameter	MAR		MNAR-WC		MNAR-DK	
	Mdn	*SD*	*Mdn*	*SD*	*Mdn*	*SD*
Intercept	5.55	0.56	5.09	0.67	5.77	0.60
MONTH	–2.21*	0.42	-2.46*	0.50	-2.16*	0.43
$MONTH^2$	0.89*	0.22	0.96*	0.26	0.90*	0.22
$DRINKS^w$	0.37*	0.13	0.36*	0.14	0.26*	0.13
$DRINKS^b$	0.22	0.25	0.26	0.25	0.41	0.29
CONDITION	1.31	0.73	1.54*	0.71	1.55*	0.76
NICDEP	2.68*	0.76	2.92*	0.78	2.95*	0.77
$TIME \times DRINKS^w$	–0.26*	0.10	-0.28*	0.10	-0.18	0.10
$TIME \times CONDITION$	–0.13	0.42	-0.14	0.41	-0.17	0.41
R^2 Fixed effects	.21	.04	.24	.04	.20	.04
R^2 Intercepts	.34	.05	.33	.05	.40	.05
R^2 Slopes	.25	.05	.24	.04	.20	.04

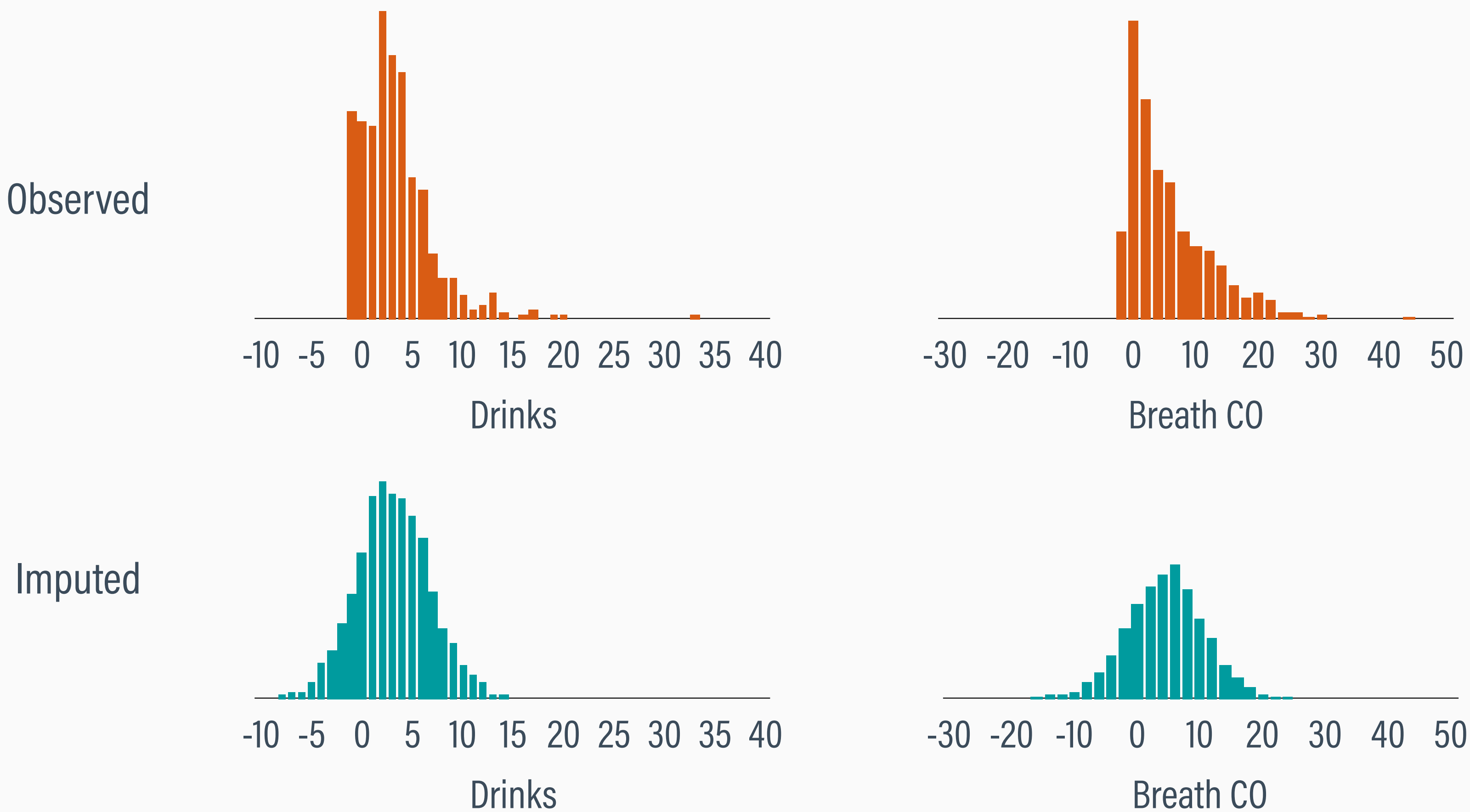
* = significant at $p < .05$

MODELING NONNORMAL DATA IN BLIMP

NONNORMAL DATA

- Generating normally distributed imputations for skewed variables usually produces implausible out-of-range values
- Means and coefficients are often unaffected, but other estimates can be distorted
- Conduct sensitivity analyses by generating imputations from nonnormal distributions and/or applying transformations

OBSERVED VS. IMPUTED DATA



YEO-JOHNSON TRANSFORMATION

- The Yeo–Johnson normal distribution defines a transformed (latent) variable X* with a particular mean and variance
- The distribution invokes a shape parameter λ that links the nonnormal X scores to a normalized (latent) variable
- Incomplete variables are modeled on the normalized metric, and the MCMC algorithm iteratively estimates λ

TRANSFORMATION FUNCTION

- The Yeo–Johnson power function subsumes a range of common transformations used in practice (e.g., logarithmic, inverse, Box–Cox)

$$X_i^* = \begin{cases} ((X_i + 1)^\lambda - 1)/\lambda & \text{if } X_i \geq 0 \text{ and } \lambda \neq 0 \\ \ln(X_i + 1) & \text{if } X_i \geq 0 \text{ and } \lambda = 0 \\ -((-X_i + 1)^{2-\lambda} - 1)/(2 - \lambda) & \text{if } X_i < 0 \text{ and } \lambda \neq 2 \\ -\ln(-X_i + 1) & \text{if } X_i < 0 \text{ and } \lambda = 2 \end{cases}$$

SEQUENTIAL SPECIFICATION

- Modeling skewed predictors requires a so-called sequential specification that factorizes the multivariate predictor distribution into a sequence of univariate regressions
- Conceptually, a path model is used to to link predictors
- Each incomplete predictor is an outcome in its own regression

NONNORMAL PREDICTORS

- The nonnormal predictor appears as a normalized (latent) variable in its own regression model
- The nonnormal version of the variable appears as a predictor in the focal regression (growth) model
- An inverse transformation converts the normalized latent imputes to nonnormal imputes on the original metric

PATH DIAGRAM

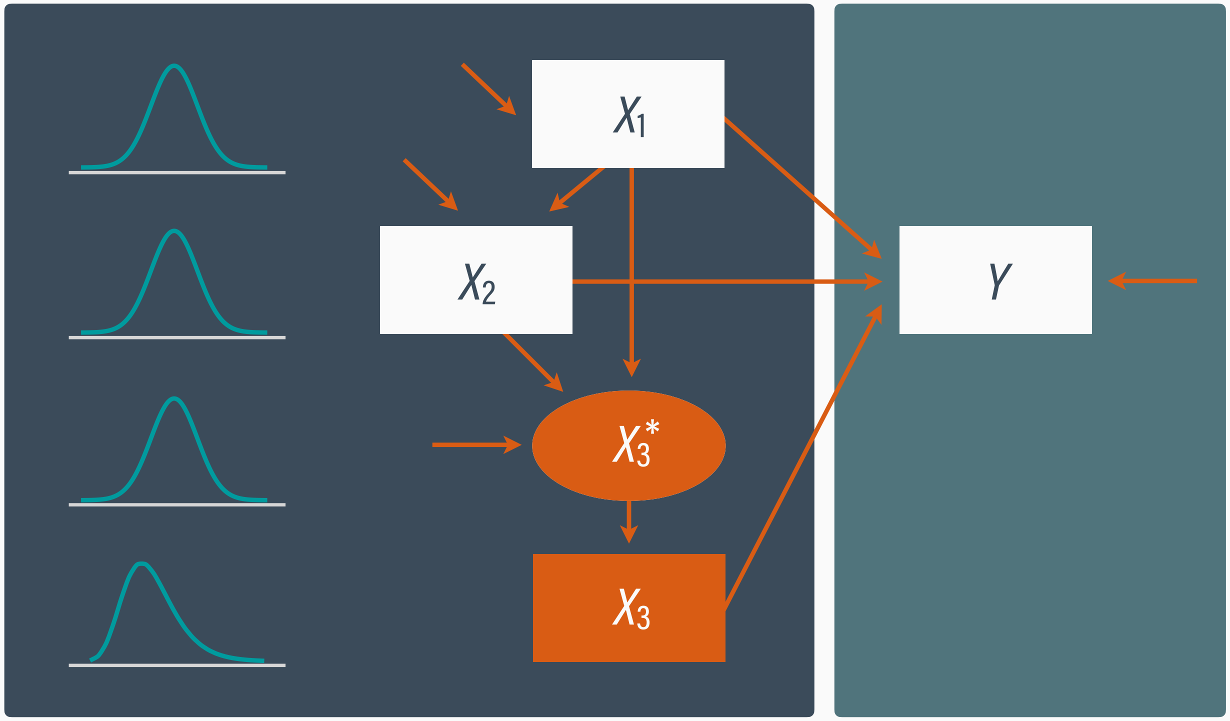
Joint Distribution = Univariate Outcome Model × Univariate Predictor Models

MODEL:

x2 ~ x1;

» yjt(x3) ~ x2 x1;

y ~ x1 x2 x3;



ANALYSIS MODEL

Model With Normalized Predictor as Outcome

$$DRINKS^{*}_{ti} = \gamma_{0i} + \gamma_1(MONTH_{ti}) + \gamma_2(NICDEP_i) + \gamma_3(CONDITION_i) + \epsilon_{ti}$$

Focal Model with Skewed Predictor

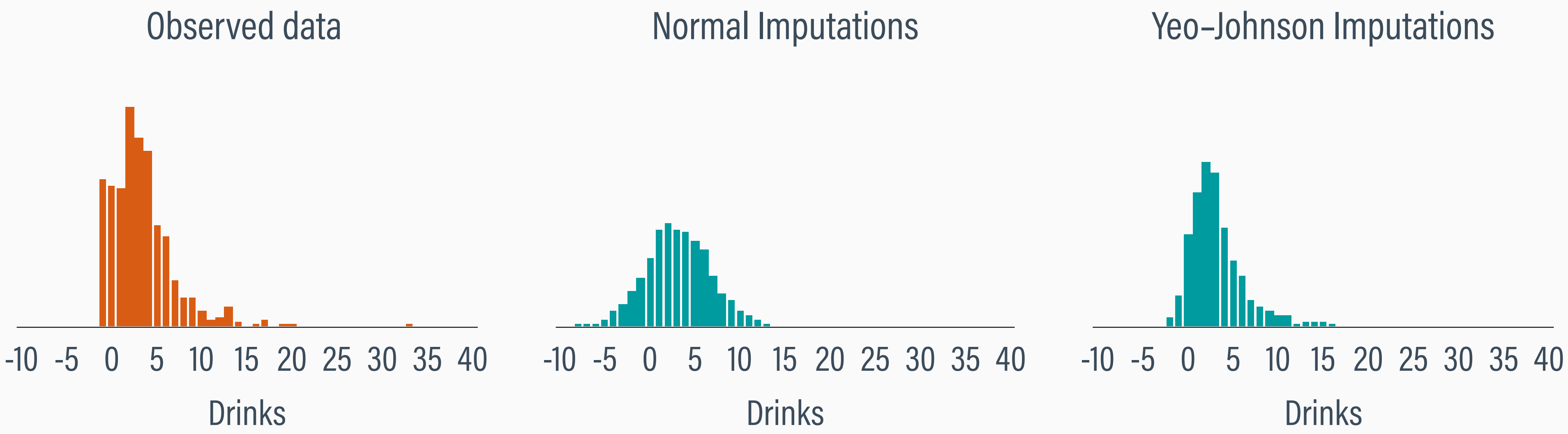
$$\begin{aligned} BREATHCO_{ti} &= \beta_{0i} + \beta_{1i}(MONTH_{ti}) + \beta_{2i}(MONTH^2_{ti}) + \beta_{3i}(DRINKS^{w}_{ti}) \\ &+ \beta_4\left(DRINKS^{b}_{i} - \mu_{(DRINKS)}\right) + \beta_5(CONDITION_i) + \beta_6\left(NICDEP_i - \mu_{(NICDEP)}\right) \\ &+ \beta_7(MONTH_{ti})(DRINKS^{w}_{ti}) + \beta_8(MONTH_{ti})(CONDITION_i) + \varepsilon_{ti} \end{aligned}$$

BLIMP MODEL SCRIPT

- Centering the skewed variable at its median value of 4 in the normalized model improves MCMC convergence

```
MODEL:
focal.model:
breathCO ~ month (month^2) drinks drinks.mean condition nicdep
    month*drinks month*condition | month (month^2) drinks;
predictor.model:
yjt(drinks – 4) ~ time nicdep condition;
```

OBSERVED VS. IMPUTED DATA



COEFFICIENT OUTPUT

```
OUTCOME MODEL ESTIMATES:

  Summaries based on 20000 iterations using 2 chains.

Outcome Variable:  breathCO

Parameters                          Median     StdDev       2.5%      97.5%        PSR      N_Eff
                                ------------------------------------------------------------------

Variances:
  ...

Coefficients:
  Intercept                          5.535      0.568      4.441      6.680      1.002    964.847
  month                             -2.236      0.418     -3.054     -1.403      1.001   1940.700
  drinks                             0.343      0.134      0.086      0.616      1.003    615.314
  drinks.mean[person]                0.248      0.318     -0.345      0.906      1.011    415.224
  condition                          1.403      0.718     -0.021      2.824      1.001   1198.627
  nicdep                             2.787      0.751      1.299      4.242      1.002   1124.305
  (month^2)                          0.917      0.212      0.503      1.331      1.001   2519.477
  month*drinks                      -0.240      0.095     -0.428     -0.054      1.001    838.642
  month*condition                   -0.154      0.408     -0.955      0.642      1.000   2773.145

Standardized Coefficients:
  ...

                                ------------------------------------------------------------------
```

COEFFICIENT OUTPUT

```
OUTCOME MODEL ESTIMATES:

  Summaries based on 20000 iterations using 2 chains.

...

Outcome Variable:  yjt(drinks - 4)

Parameters                         Median     StdDev       2.5%      97.5%        PSR      N_Eff
                                --------------------------------------------------------------------

Variances:
  L2 : Var(Intercept)               2.978      0.537      2.080      4.184      1.001   3091.170
  Residual Var.                     5.380      0.383      4.691      6.198      1.000   4958.543

Coefficients:
  Intercept                         0.044      0.246     -0.430      0.537      1.000   3103.794
  month                            -0.909      0.090     -1.084     -0.730      1.001   5346.106
  nicdep                           -0.372      0.354     -1.057      0.320      1.002   2400.282
  condition                        -0.454      0.338     -1.105      0.227      1.001   3199.748

Transformation:
  Yeo-Johnson (lambda)              0.628      0.025      0.577      0.678      1.003   1921.032

Standardized Coefficients:
  ...
                                --------------------------------------------------------------------
```

NONNORMAL OUTCOMES

- A nonnormal outcome appears as a normalized (latent) variable in the focal regression (growth) model
- Normalization changes the dependent variable's metric, so estimates will differ due to rescaling
- Blimp can save original and normalized imputes for inspection

NORMALIZED OUTCOME MODEL

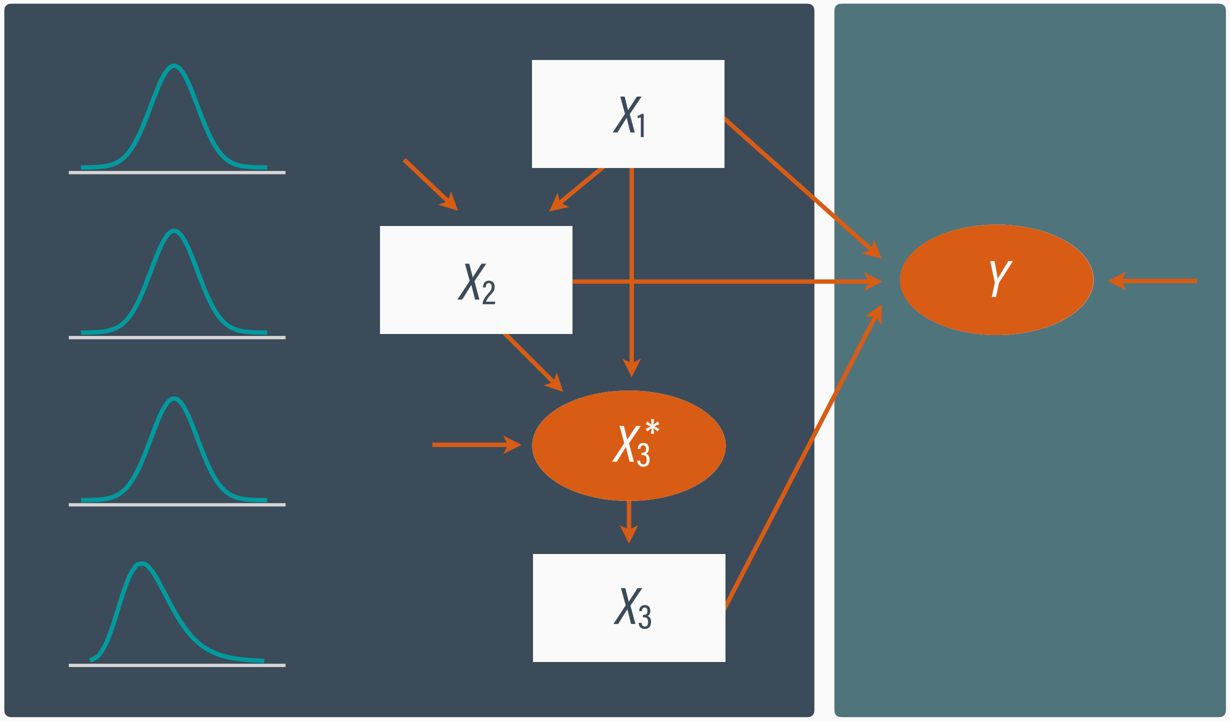
Joint Distribution = Univariate Outcome Model × Univariate Predictor Models

MODEL:

x2 ~ x1;

yjt(x3) ~ x2 x1;

» yjt(y) ~ x1 x2 x3;



ANALYSIS MODEL

Model With Normalized Predictor as Outcome

$$DRINKS^*_{ti} = \gamma_{0i} + \gamma_1(MONTH_{ti}) + \gamma_2(NICDEP_i) + \gamma_3(CONDITION_i) + \epsilon_{ti}$$

Focal Model with Skewed Predictor and Normalized Outcome

$$\begin{aligned} BREATHCO^*_{ti} &= \beta_{0i} + \beta_{1i}(\mathrm{MONTH}_{ti}) + \beta_{2i}(\mathrm{MONTH}^2_{ti}) + \beta_{3i}(\mathrm{DRINKS}^w_{ti}) \\ &+ \beta_4\left(DRINKS^b_i - \mu_{(DRINKS)}\right) + \beta_5(CONDITION_i) + \beta_6\left(NICDEP_i - \mu_{(NICDEP)}\right) \\ &+ \beta_7(MONTH_{ti})(DRINKS^w_{ti}) + \beta_8(MONTH_{ti})(CONDITION_i) + \varepsilon_{ti} \end{aligned}$$

BLIMP MODEL SCRIPT

- Centering the skewed variables at their median (4 and 8) in the normalized models improves MCMC convergence

```
MODEL:
focal.model:
yjt(breathCO – 8) ~ month (month^2) drinks drinks.mean condition nicdep
    month*drinks month*condition | month (month^2) drinks;
predictor.model:
yjt(drinks – 4) ~ time nicdep condition;
```

COEFFICIENT OUTPUT

```
OUTCOME MODEL ESTIMATES:

  Summaries based on 20000 iterations using 2 chains.

Outcome Variable:  yjt(breathCO - 8)

Parameters                         Median     StdDev       2.5%      97.5%        PSR      N_Eff
                                -------------------------------------------------------------------
Variances:
  ...

Coefficients:
  Intercept                        -5.274      0.670     -6.614     -3.985      1.003    710.678
  month                            -2.936      0.407     -3.735     -2.138      1.001   1161.525
  drinks                            0.159      0.119     -0.069      0.399      1.001    595.354
  drinks.mean[person]               0.456      0.310     -0.159      1.072      1.003    345.554
  condition                         1.499      0.691      0.141      2.860      1.003   1067.435
  nicdep                            2.847      0.715      1.414      4.244      1.001   1176.606
  (month^2)                         1.213      0.223      0.778      1.646      1.001   1433.258
  month*drinks                     -0.158      0.088     -0.335      0.011      1.004   1375.647
  month*condition                  -0.194      0.383     -0.938      0.568      1.001   1832.944

Standardized Coefficients:
  ...

                                -------------------------------------------------------------------
```

MODEL COMPARISON

Parameter	Normal Imputations *Mdn*	*SD*	Y–J Predictor *Mdn*	*SD*	Y–J Predictor & Outcome *Mdn*	*SD*
Intercept	5.55	0.56	5.54	0.57	−5.27	0.67
MONTH	−2.21*	0.42	−2.24*	0.42	−2.94*	0.41
$MONTH^2$	0.89*	0.22	0.92*	0.21	1.21*	0.22
$DRINKS^w$	0.37*	0.13	0.34*	0.13	0.16*	0.12
$DRINKS^b$	0.22	0.25	0.25	0.32	0.46	0.31
CONDITION	1.31	0.73	1.40	0.72	1.50*	0.69
NICDEP	2.68*	0.76	2.79*	0.75	2.85*	0.72
$TIME \times DRINKS^w$	−0.26*	0.10	−0.24*	0.10	−0.16*	0.09
$TIME \times CONDITION$	−0.13	0.42	−0.15	0.41	−0.19	0.38
R^2 Fixed effects	.21	.04	.21	.04	.23	.04
R^2 Intercepts	.34	.05	.34	.05	.39	.04
R^2 Slopes	.25	.05	.24	.05	.18	.04

* = significant at $p < .05$

For more information go to

WWW.APPLIEDMISSINGDATA.COM