Multilevel/Longitudinal Models Using Stata

Chuck Huber, PhD
StataCorp
chuber@stata.com

New York University
April 11, 2018

Outline

- The simulated data
- Single level models
- Two level models
- Longitudinal models
- Three level models
- Fixed vs random effects
- Multilevel models for binary data
- Multilevel models for survival data
- Multilevel structural equation models
- Bayesian multilevel models

The Simulated Dataset

- This dataset contains repeated measurements of total cholesterol (chol) and systolic blood pressure (sbp) over 7-8 years.
- The patients are nested within three different physician practices.
- The final measurement occurred when the patients survived a severe myocardial infarction.
- Time-to-death survival data were then collected on these patients.

The Simulated Dataset

```
. describe

Contains data from SecondMI.dta
  obs:         2,100
 vars:            13                          5 Jun 2016 20:33
 size:       109,200
--------------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
--------------------------------------------------------------------------------------
physician       float   %9.0g                 Physician ID
patient         float   %9.0g                 Patient ID
age_mi          float   %9.0g                 Age of first MI
age             float   %9.0g                 Age (years)
cage            float   %9.0g                 Centered Age
mage            float   %9.0g                 Age at Center of cage
sex             float   %9.0g      sex        Sex
sbp             float   %9.0g                 Systolic Blood Pressure (mm/Hg)
chol            float   %9.0g                 Total Cholesterol (mg/dL)
HiBP            float   %9.0g                 Hypertension (sbp>150)
time            float   %9.0g                 Days until death
dead            float   %9.0g      dead       Dead
zero            float   %9.0g                 Column of zeros for graphs
--------------------------------------------------------------------------------------
Sorted by: physician  patient  age
```

The Simulated Dataset

```
. list physician patient age_mi age cage mage sex sbp chol HiBP time dead if patient<=2, noobs sep(7) ab(9)
```

physician	patient	age_mi	age	cage	mage	sex	sbp	chol	HiBP	time	dead
1	1	59	53	-3	56	Male	144	214	0	14	censored
1	1	59	54	-2	56	Male	142	231	0	14	censored
1	1	59	55	-1	56	Male	150	230	0	14	censored
1	1	59	56	0	56	Male	147	239	0	14	censored
1	1	59	57	1	56	Male	146	255	0	14	censored
1	1	59	58	2	56	Male	154	268	1	14	censored
1	1	59	59	3	56	Male	155	286	1	14	censored
2	2	50	44	-3	47	Male	151	233	1	4	censored
2	2	50	45	-2	47	Male	139	238	0	4	censored
2	2	50	46	-1	47	Male	155	240	1	4	censored
2	2	50	47	0	47	Male	146	257	0	4	censored
2	2	50	48	1	47	Male	161	264	1	4	censored
2	2	50	49	2	47	Male	160	263	1	4	censored
2	2	50	50	3	47	Male	150	277	0	4	censored

The Simulated Dataset

```
. summ physician patient age_mi age cage mage sex sbp chol HiBP time dead
```

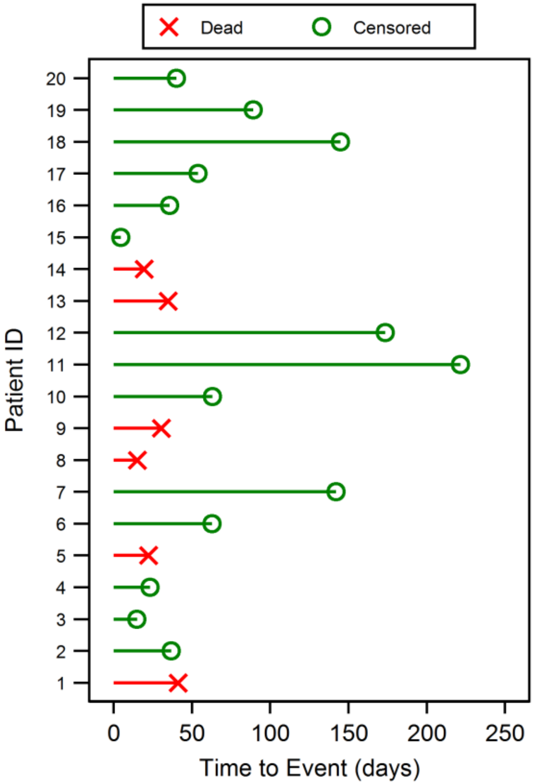
Variable	Obs	Mean	Std. Dev.	Min	Max
physician	2,100	2	.8166911	1	3
patient	2,100	150.5	86.62269	1	300
age_mi	2,100	54.78	3.273638	50	60
age	2,100	51.78	3.836484	44	60
cage	2,100	0	2.000476	-3	3
mage	2,100	51.78	3.273638	47	57
sex	2,100	.5166667	.4998412	0	1
sbp	2,100	148.4367	8.427155	121	180
chol	2,100	250.0533	18.55276	196	297
HiBP	2,100	.3995238	.4899172	0	1
time	2,100	26.9	50.15663	1	451
dead	2,100	.5133333	.4999412	0	1

The Simulated Dataset

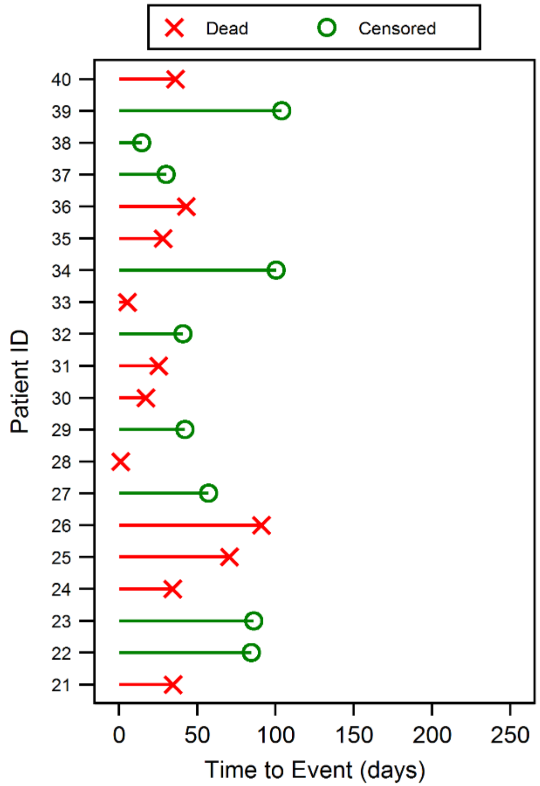
The Simulated Dataset

The Simulated Dataset

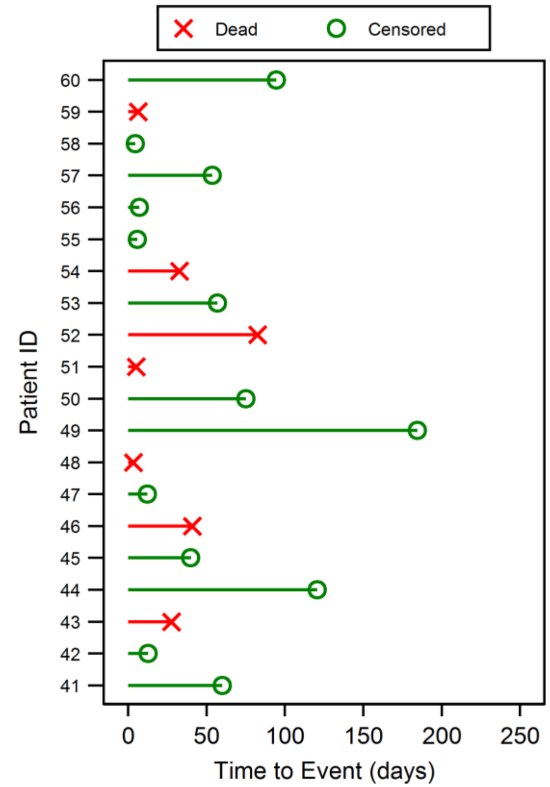
Physician 1



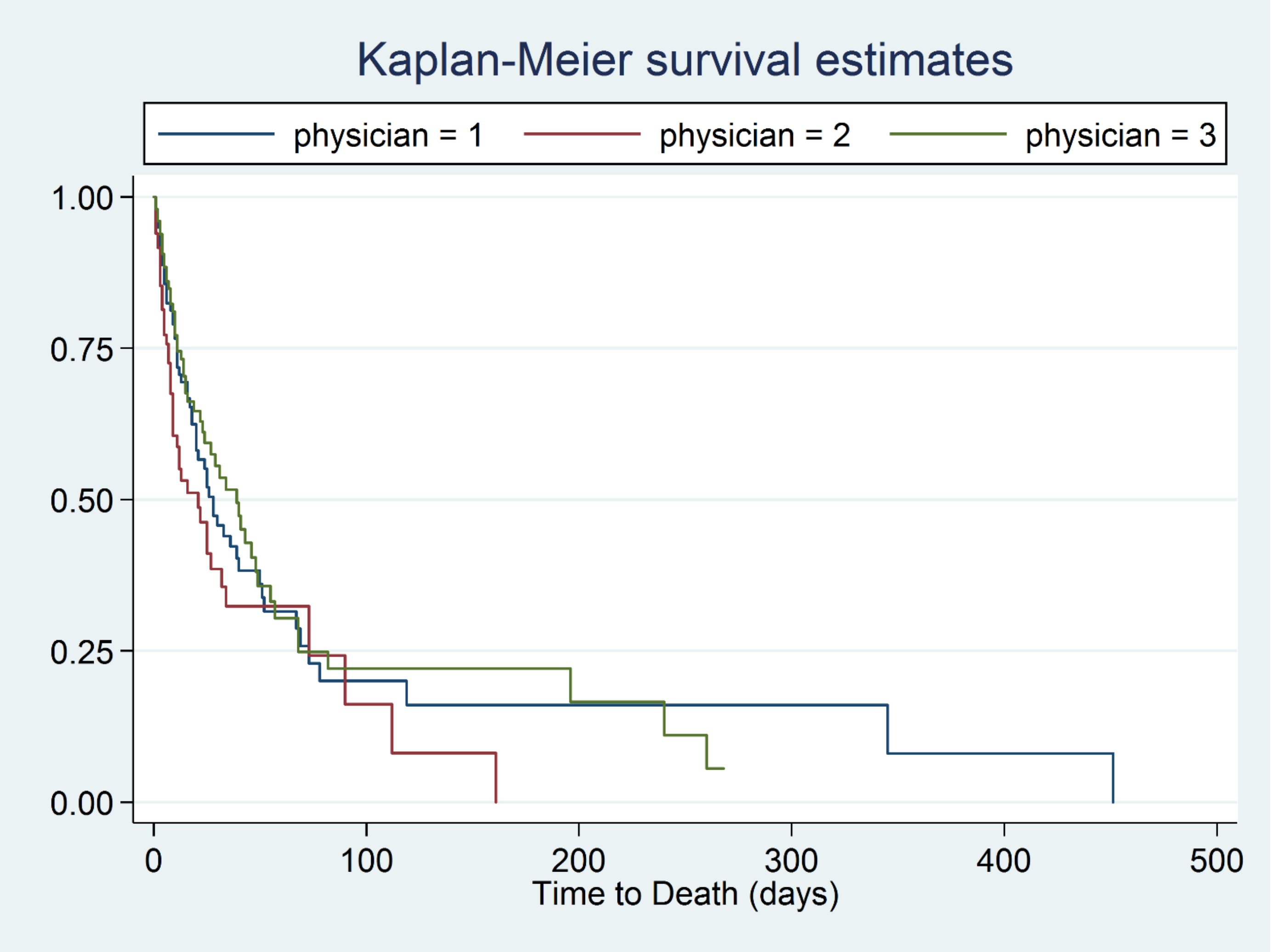
Physician 2



Physician 3



The Simulated Dataset



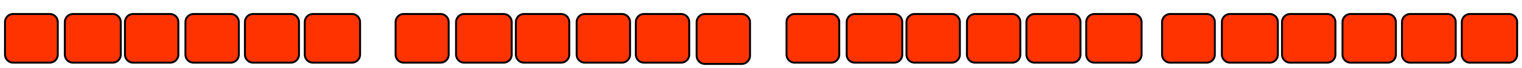
Outline

- ✓ The simulated data
- **Single level models**
- Two level models
- Longitudinal models
- Three level models
- Fixed vs random effects
- Multilevel models for binary data
- Multilevel models for survival data
- Multilevel structural equation models
- Bayesian multilevel models

Single Level Models

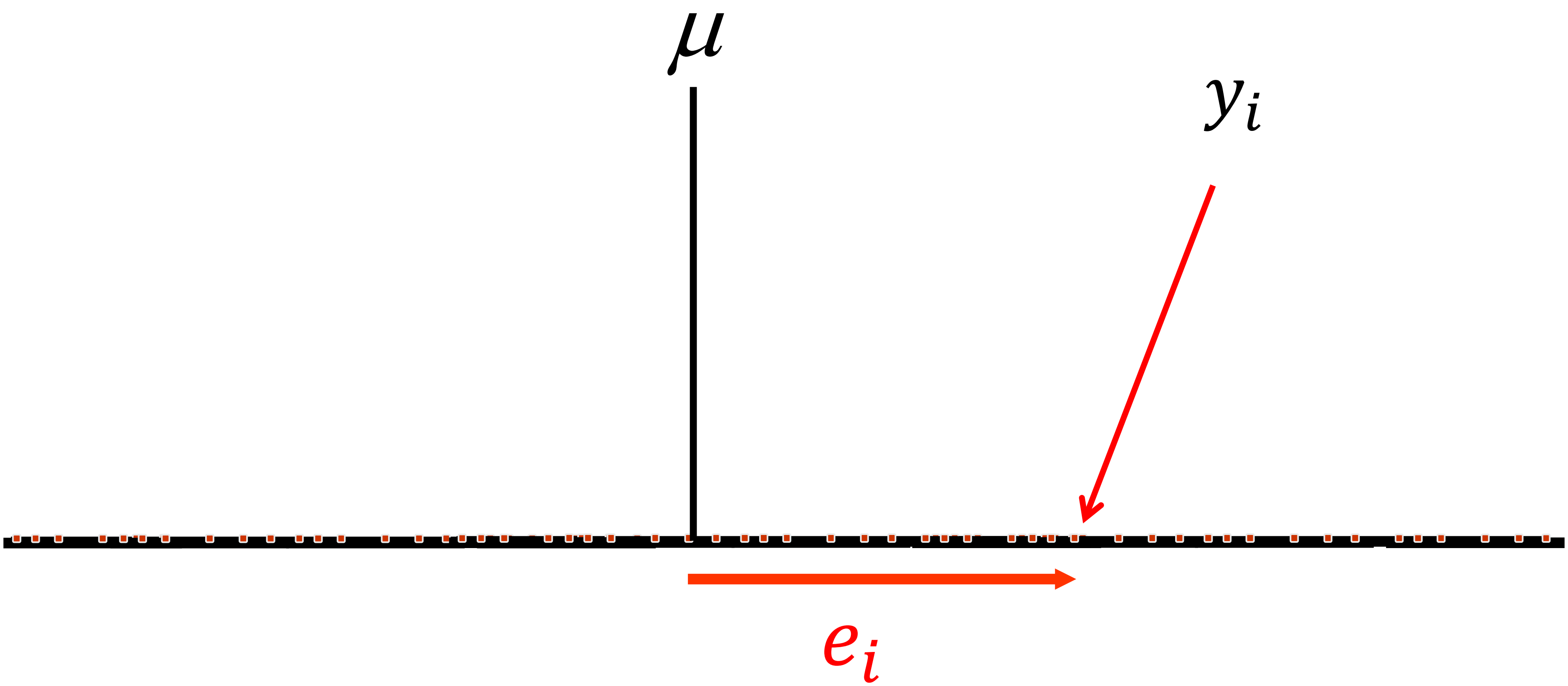
$$y_i = y_{patient}$$

Patients



We assume that observations of all patients are independent of each other

Single Level Models

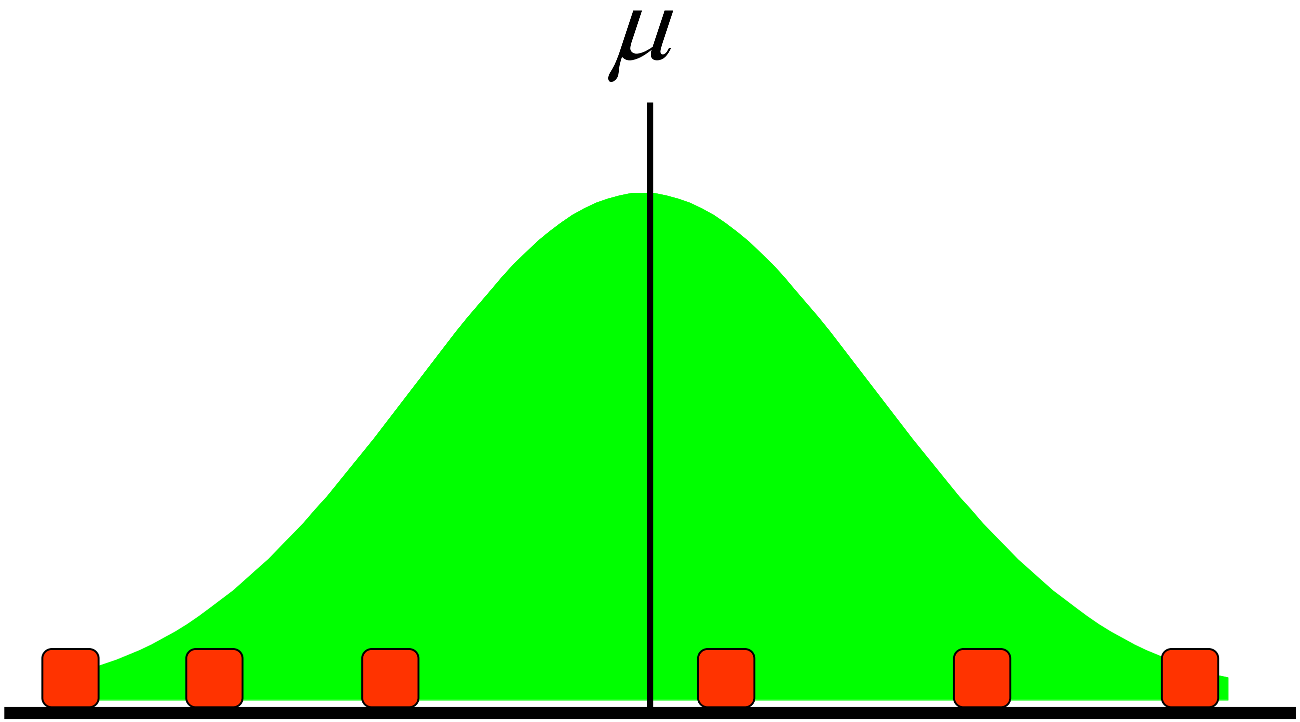


$$y_i = \mu + e_i$$

Single Level Models

$$\underbrace{y_i}_{\textbf{Observed}} = \underbrace{\mu}_{\textbf{Fixed}} + \underbrace{e_i}_{\textbf{Random}}$$

$$y_i = \mu + e_i$$



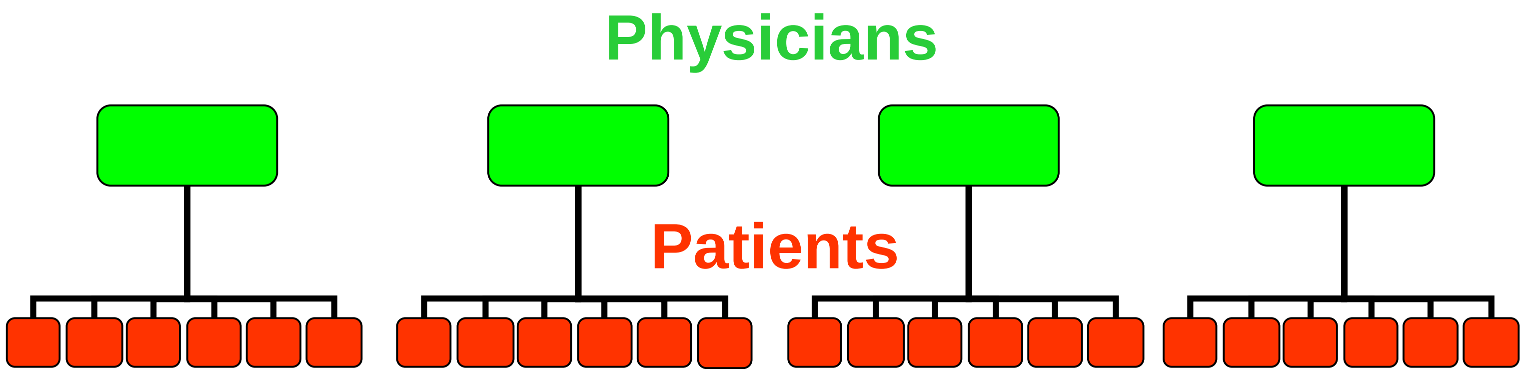
$$e_i \sim N(0, \sigma^2)$$

Outline

- ✓ The simulated data
- ✓ Single level models
- **Two level models**
- Longitudinal models
- Three level models
- Fixed vs random effects
- Multilevel models for binary data
- Multilevel models for survival data
- Multilevel structural equation models
- Bayesian multilevel models

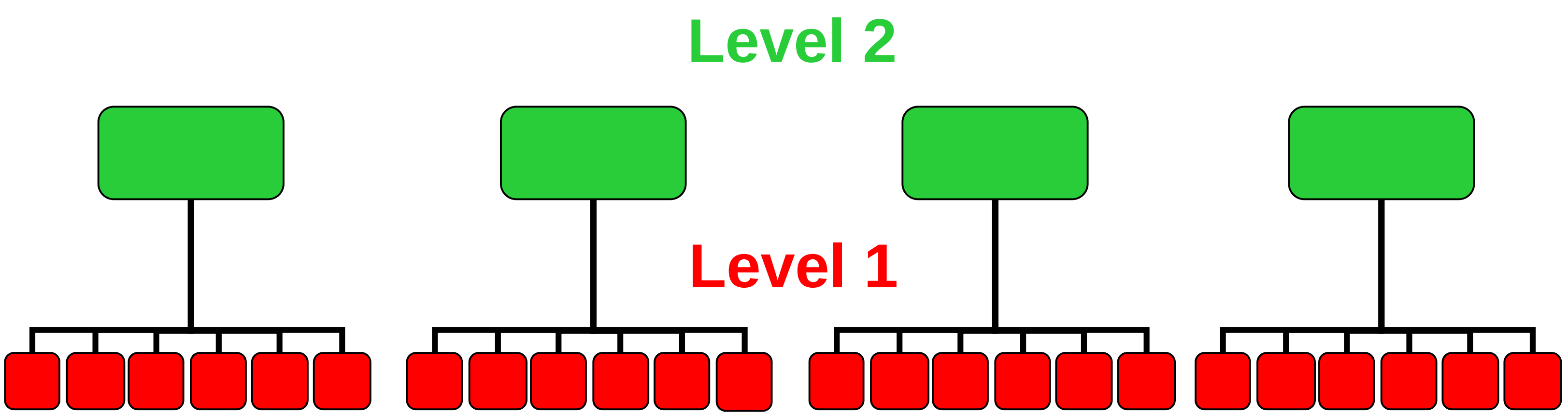
Data Analysis and Statistical Software

Two Level Models



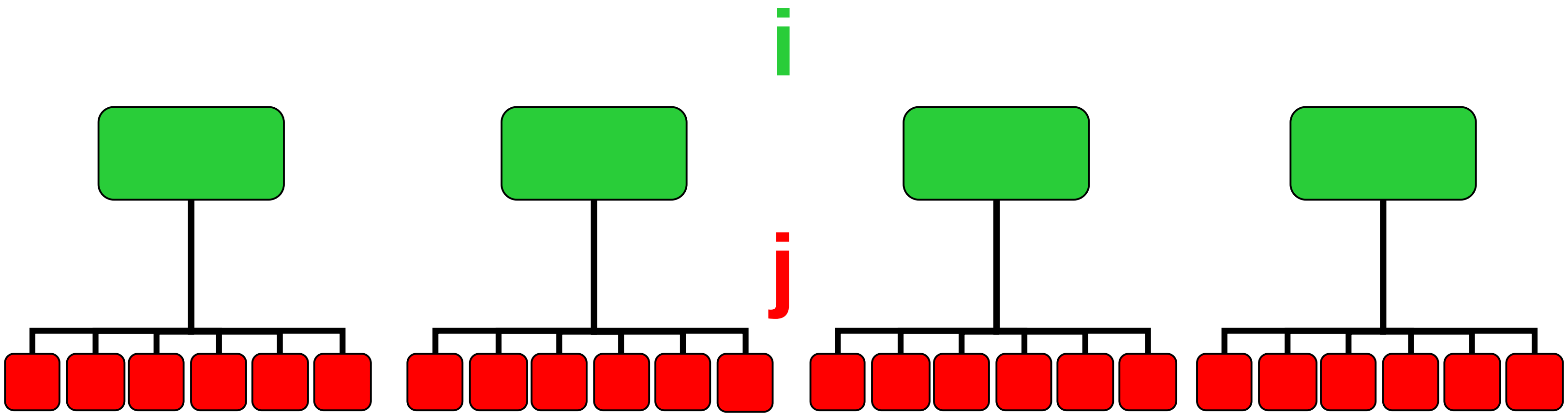
Patient observations within physician practice may be correlated

Two Level Models



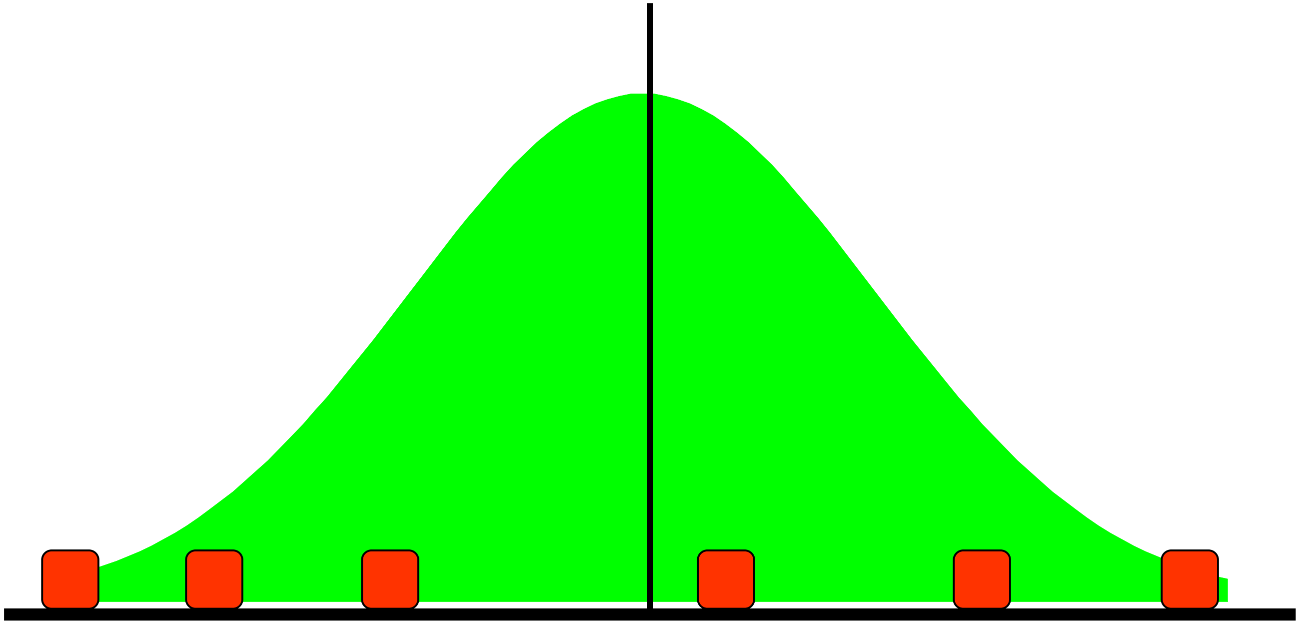
We can refer to the nesting structure in terms of “Levels”…

Two Level Models



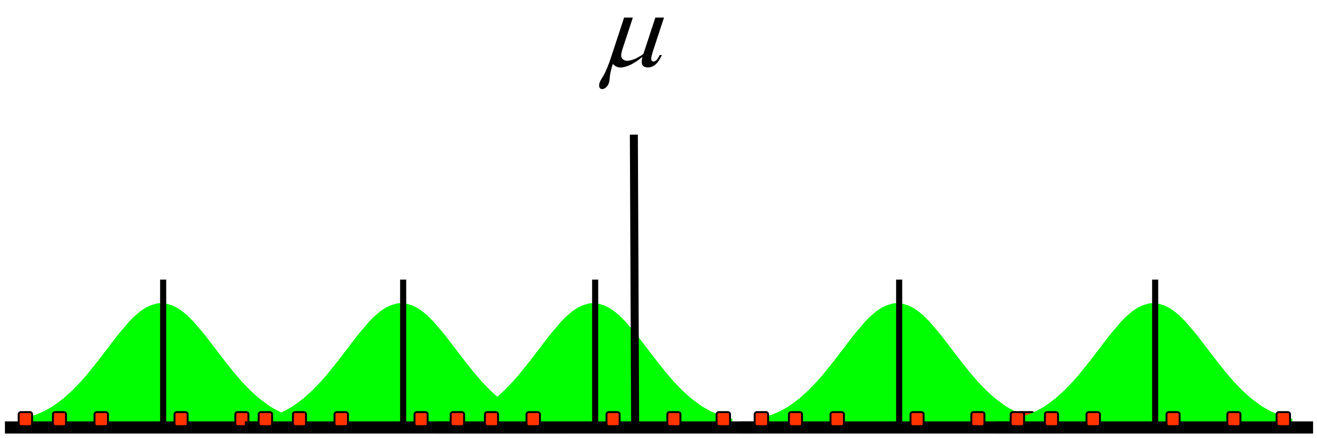
$$y_{ij} = y_{physician,patient}$$

Two Level Models

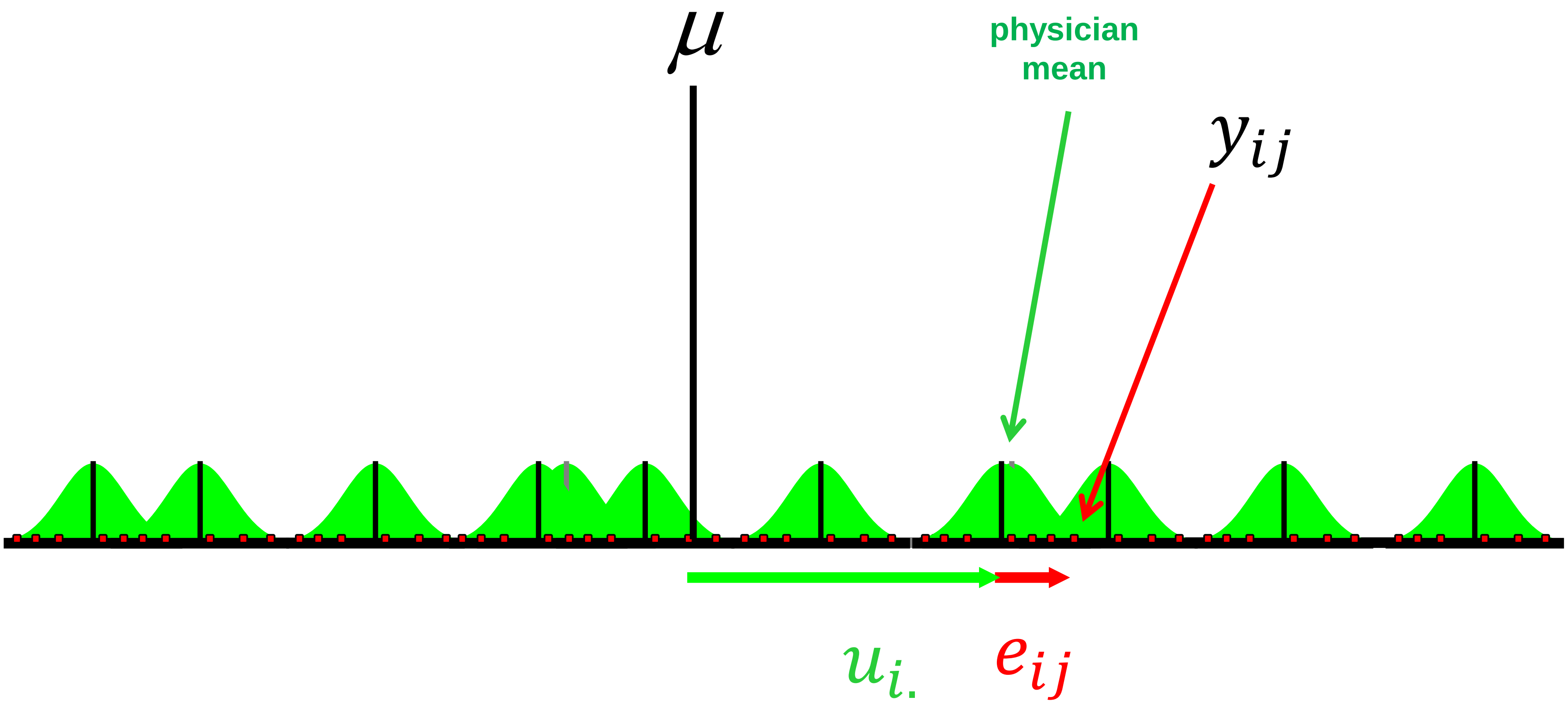


Patient observations within physician practice will vary about each physician's mean.

Two Level Models



Physicians' means will vary about the grand mean.



$$y_{ij} = \mu + u_{i.} + e_{ij}$$

Two Level Models

$$\underbrace{y_{ij}}_{\text{Observed}} = \underbrace{\mu}_{\text{Fixed}} + \underbrace{u_{i.} + e_{ij}}_{\text{Random}}$$

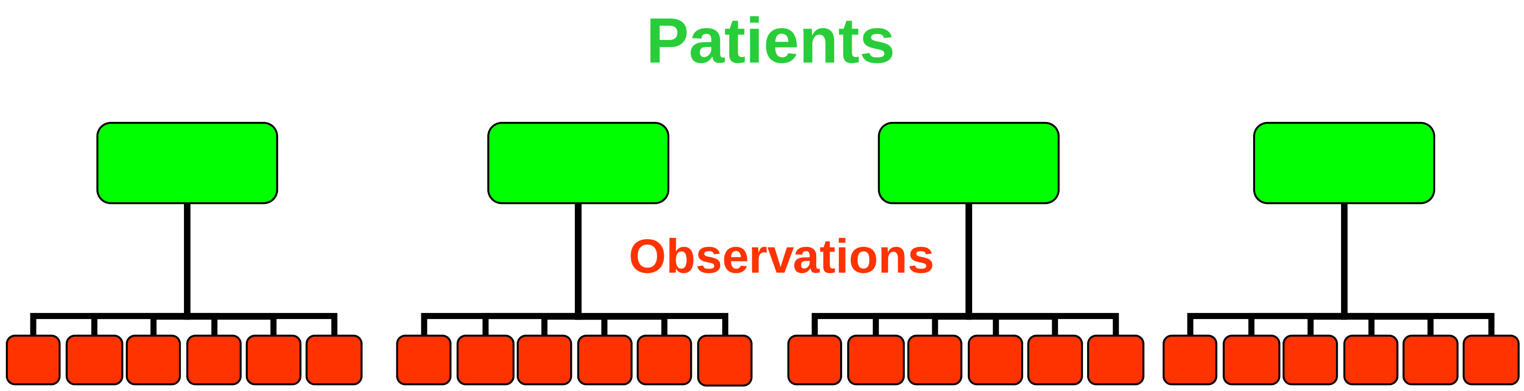
$$u_{i.} \sim N(0, \tau^2)$$

$$e_{ij} \sim N(0, \sigma^2)$$

Outline

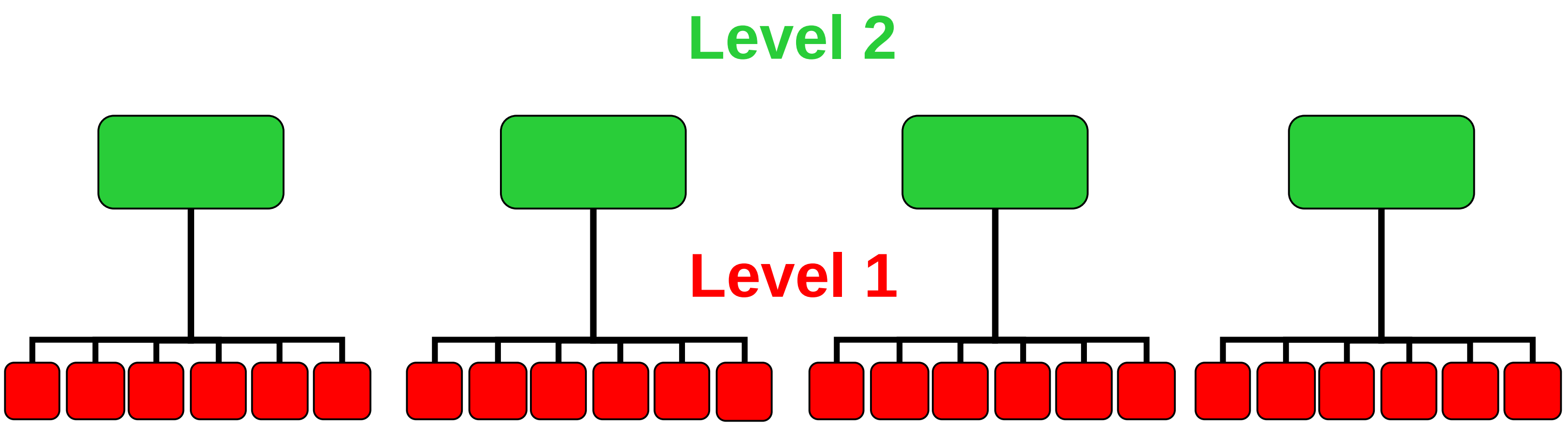
- ✓ The simulated data
- ✓ Single level models
- ✓ Two level models
- **Longitudinal models**
- Three level models
- Fixed vs random effects
- Multilevel models for binary data
- Multilevel models for survival data
- Multilevel structural equation models
- Bayesian multilevel models

Longitudinal Models



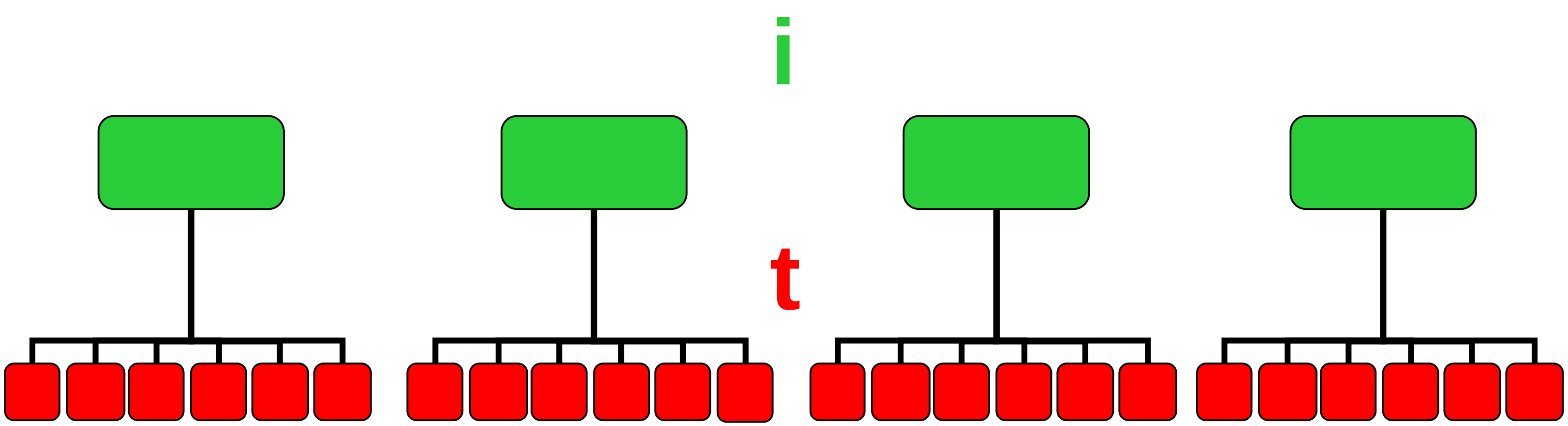
Repeated observations within patient are almost certainly correlated

Longitudinal Models



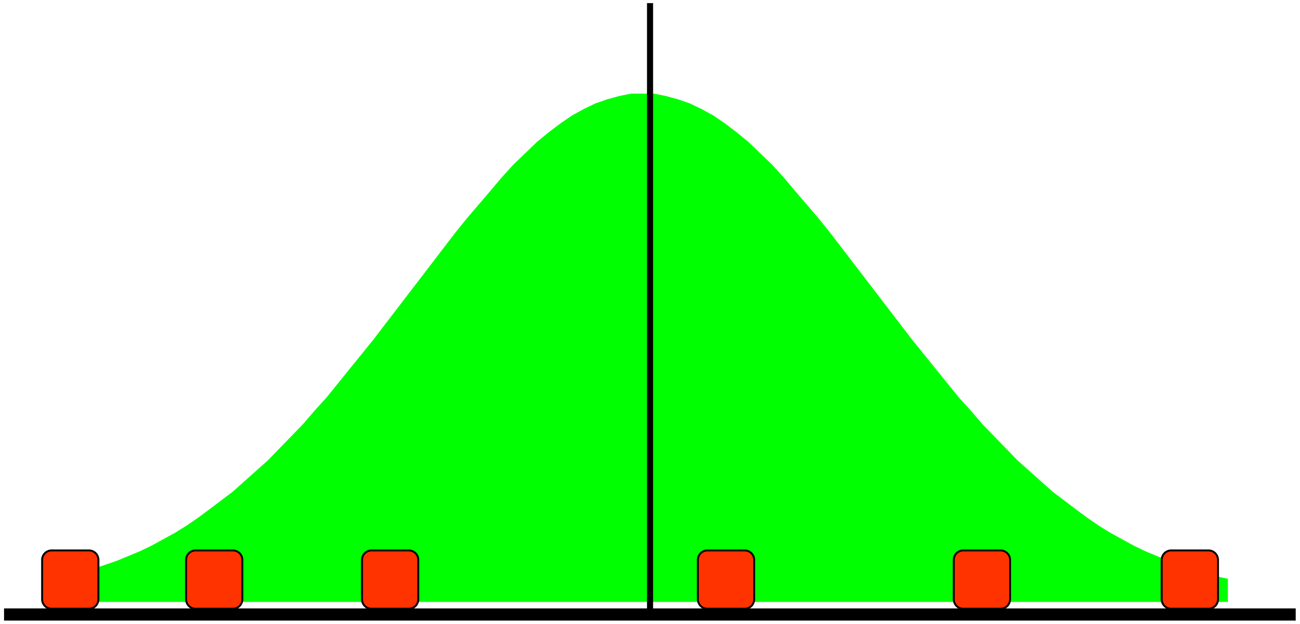
We can refer to the nesting structure in terms of "Levels"…

Longitudinal Models



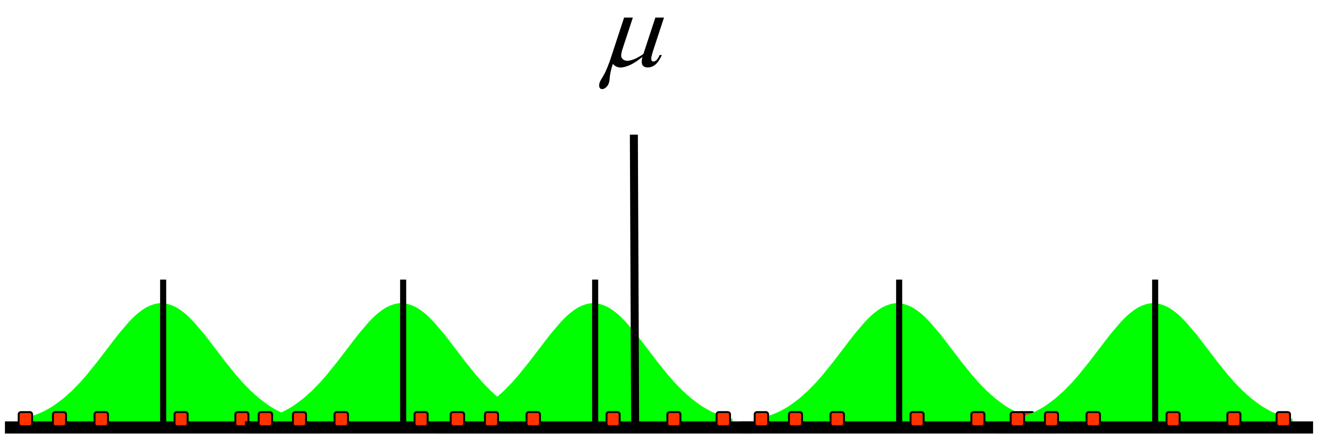
$$y_{it} = y_{patient,observation}$$

Longitudinal Models



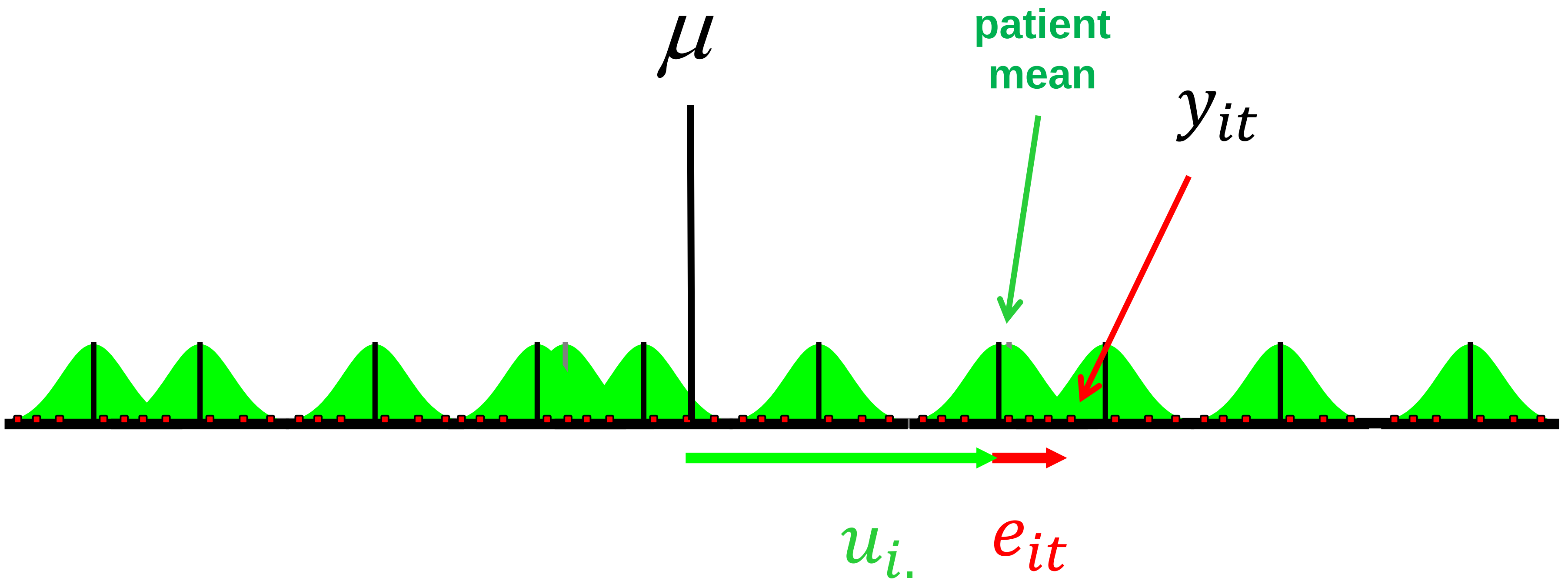
Repeated observations within patients will vary about each patient's mean.

Longitudinal Models



Patients' means will vary about the grand mean.

Longitudinal Models



$$y_{it} = \mu + u_{i.} + e_{it}$$

Longitudinal Models

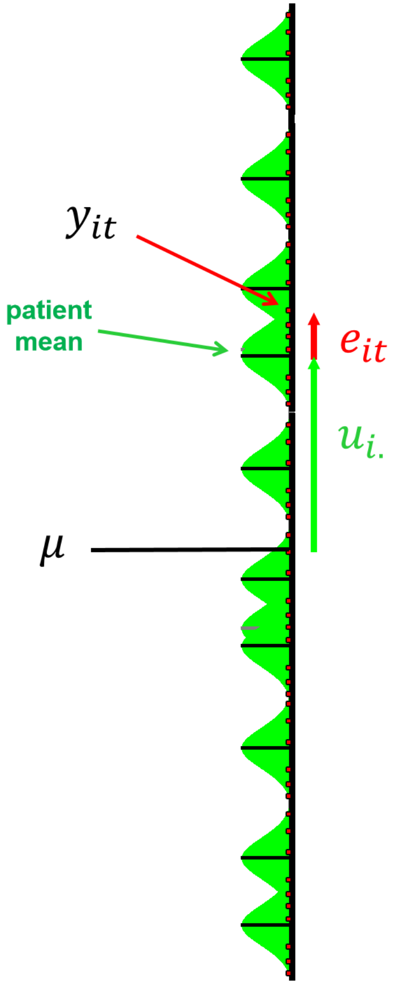
$$\underbrace{y_{it}}_{\text{Observed}} = \underbrace{\mu}_{\text{Fixed}} + \underbrace{u_{i.} + e_{it}}_{\text{Random}}$$

Observed **Fixed** **Random**

$$u_{i.} \sim N(0, \tau^2)$$

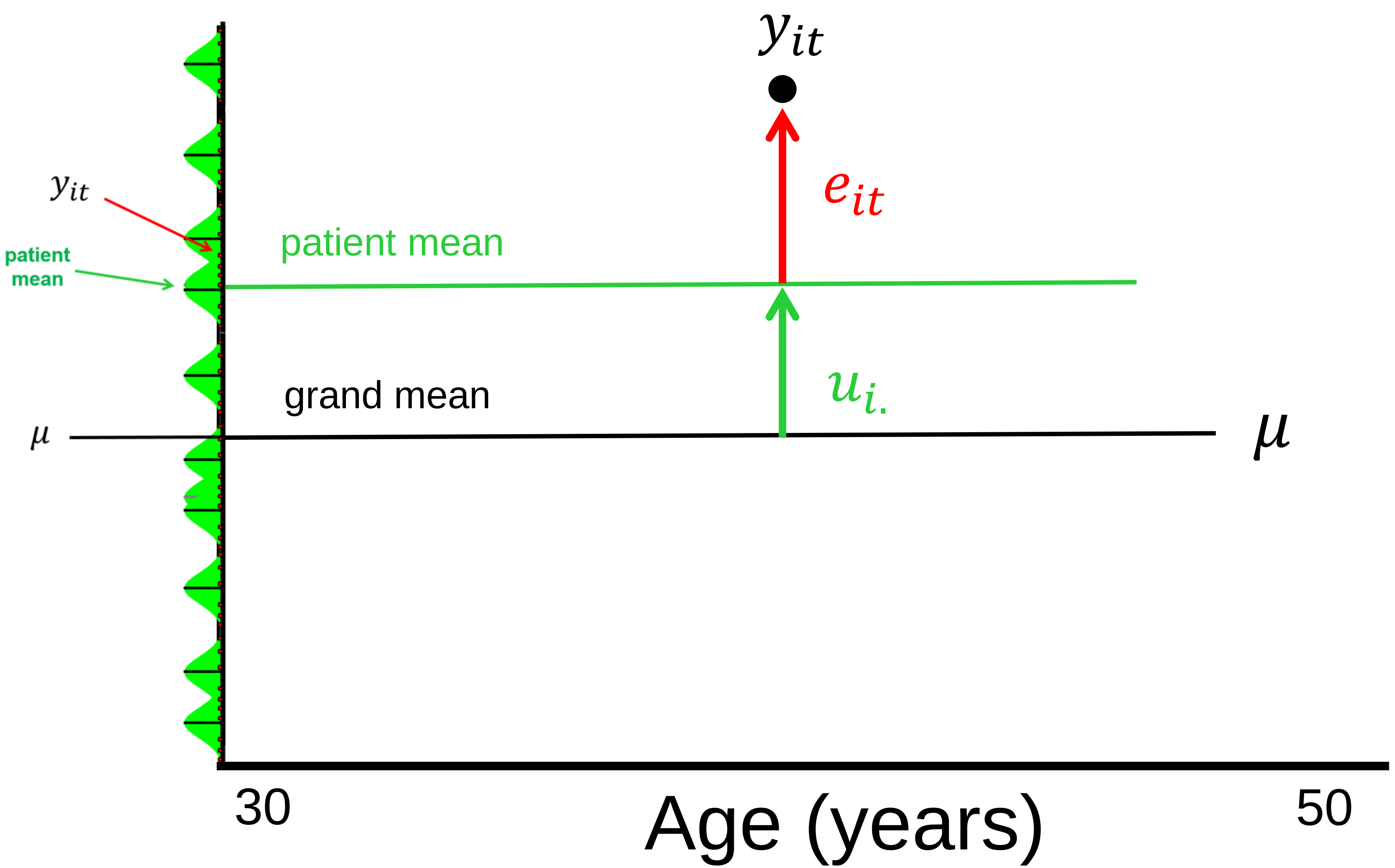
$$e_{it} \sim N(0, \sigma^2)$$

Longitudinal Models



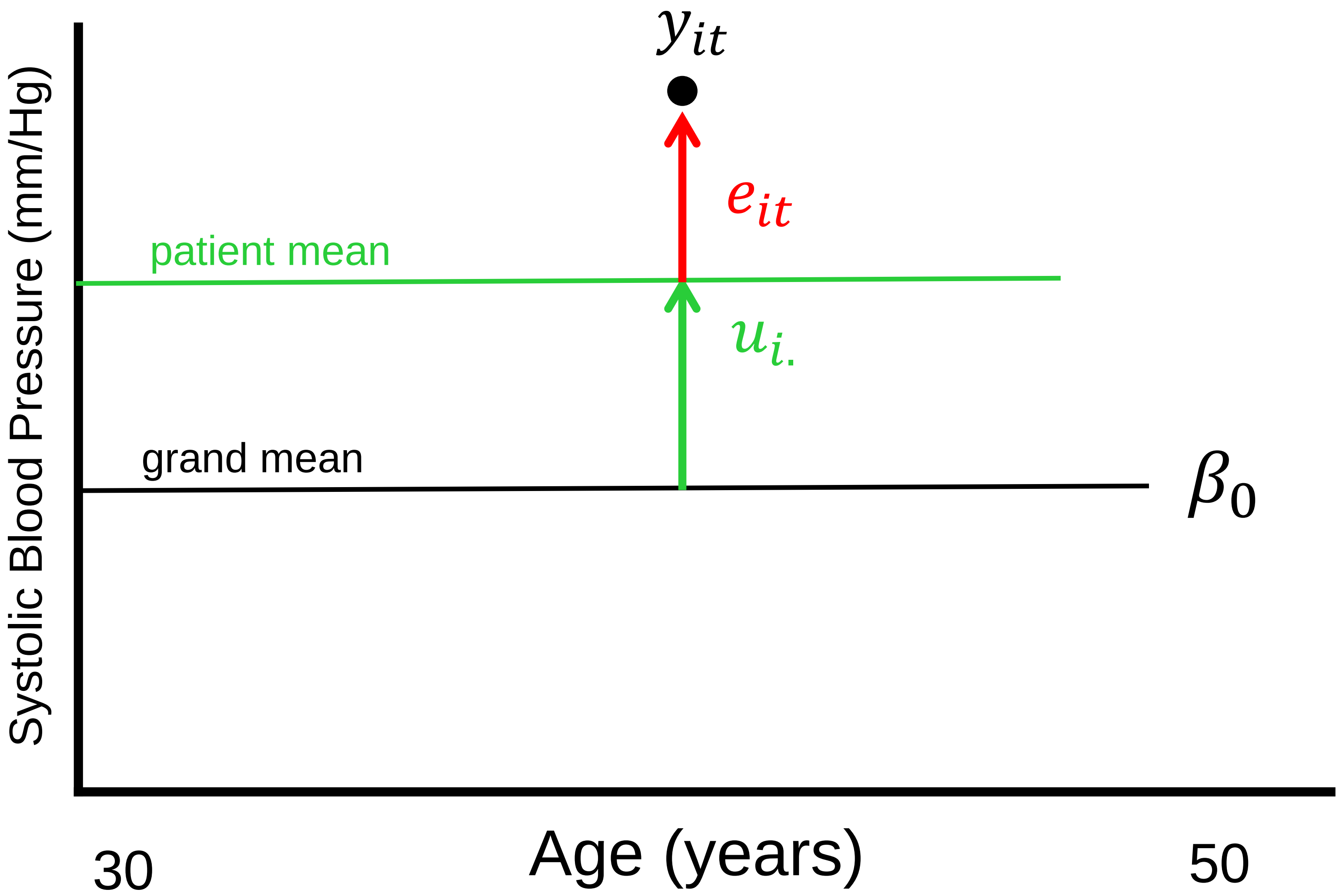
$$y_{it} = \mu + u_{i.} + e_{it}$$

$$y_{it} = \mu + u_{i.} + e_{it}$$

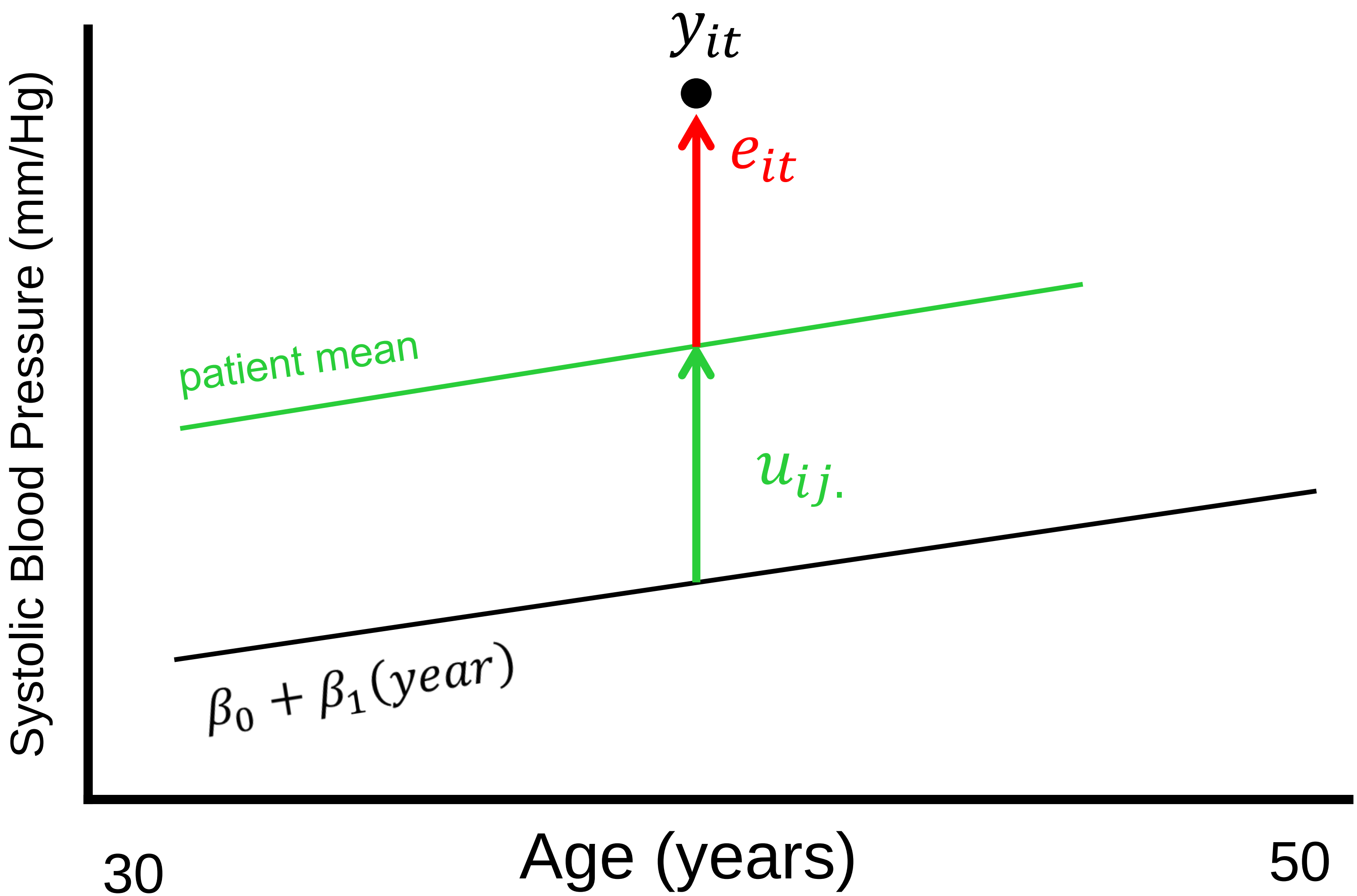


y_{it}
e_{it}
y_{it}
patient mean
patient mean
$u_{i.}$
grand mean
μ
μ
30
Age (years)
50

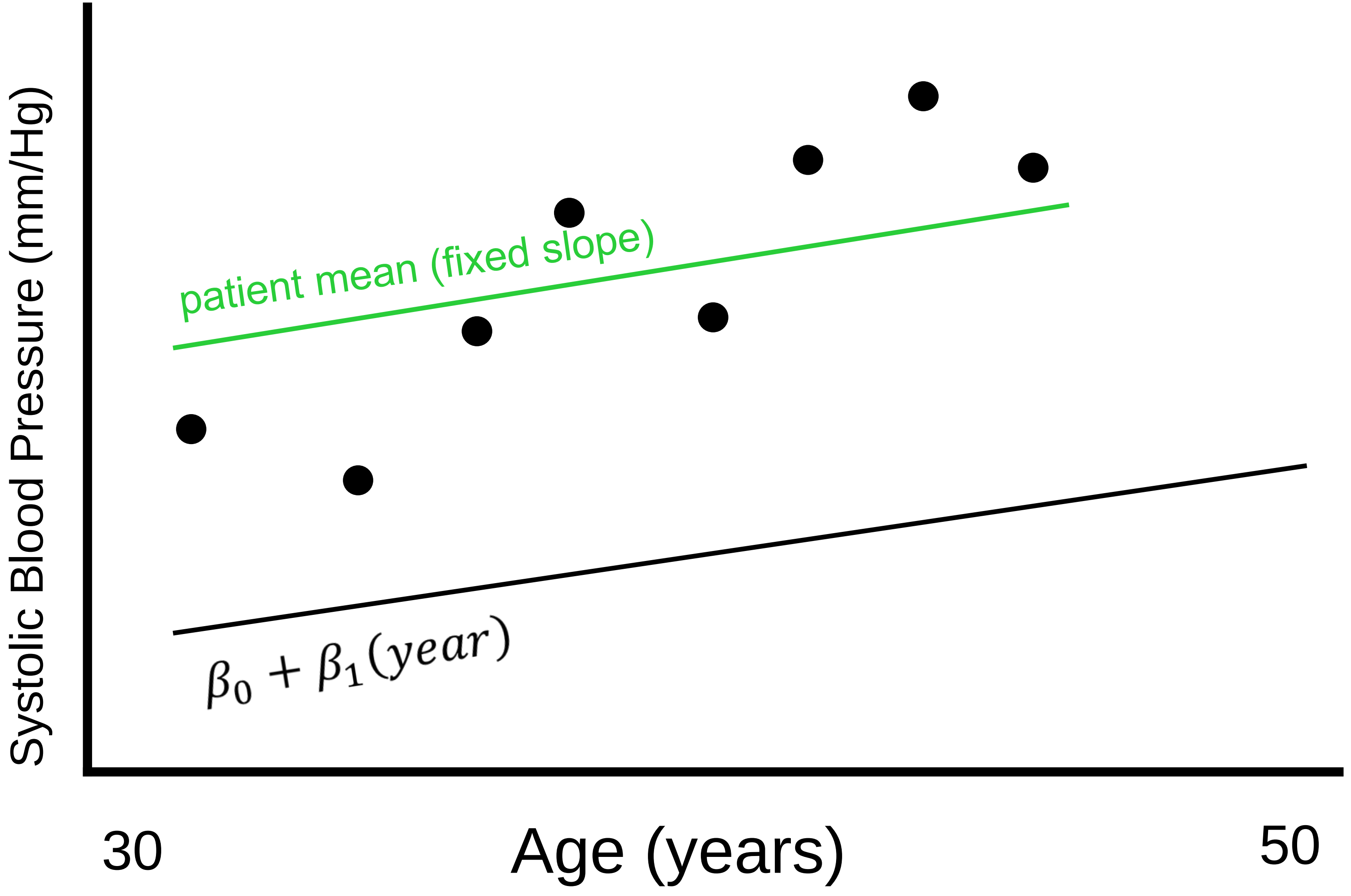
$$y_{it} = \beta_0 + u_{i.} + e_{it}$$



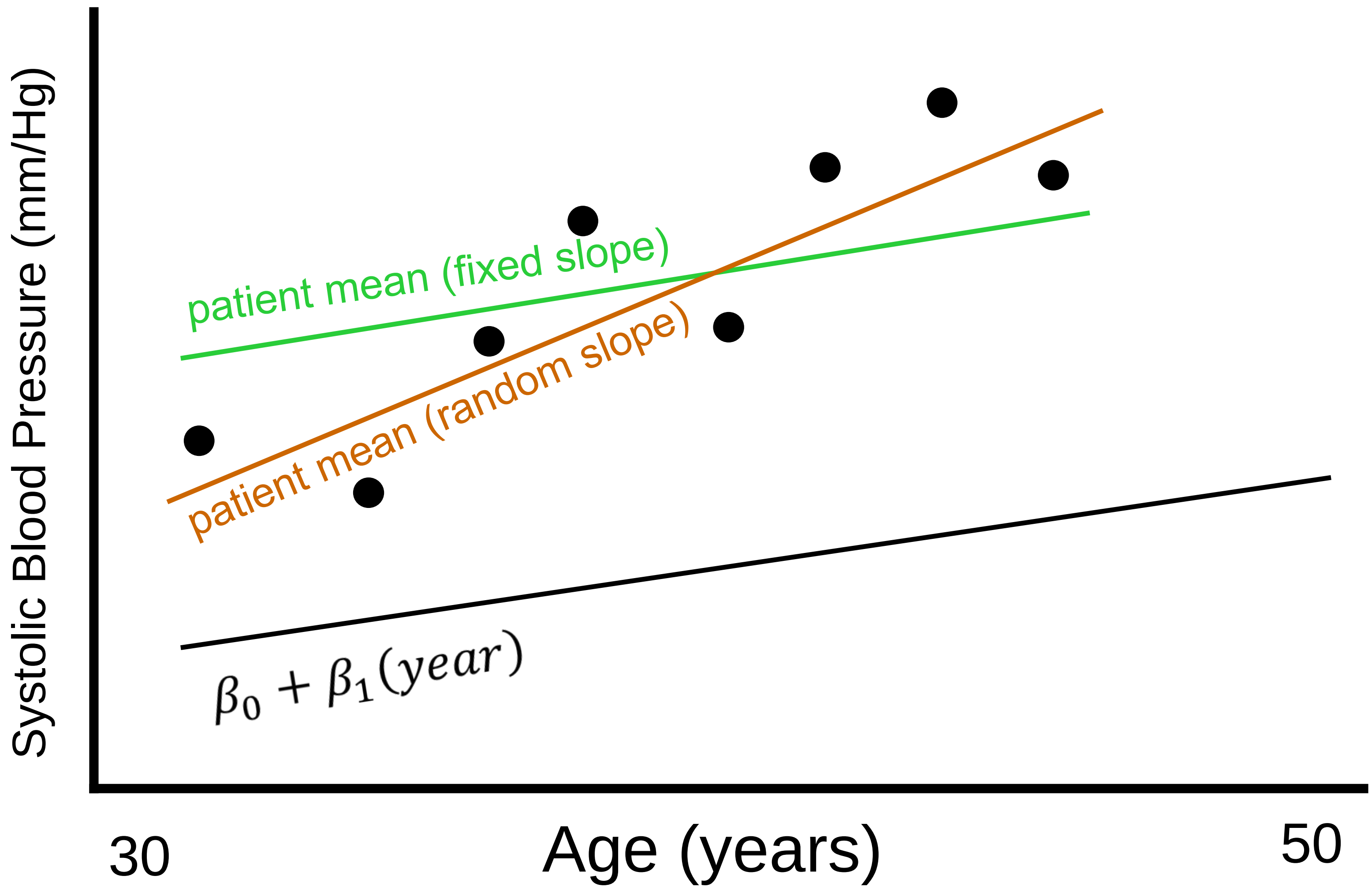
$$y_{it} = \beta_0 + \beta_1(age) + u_{i.} + e_{it}$$



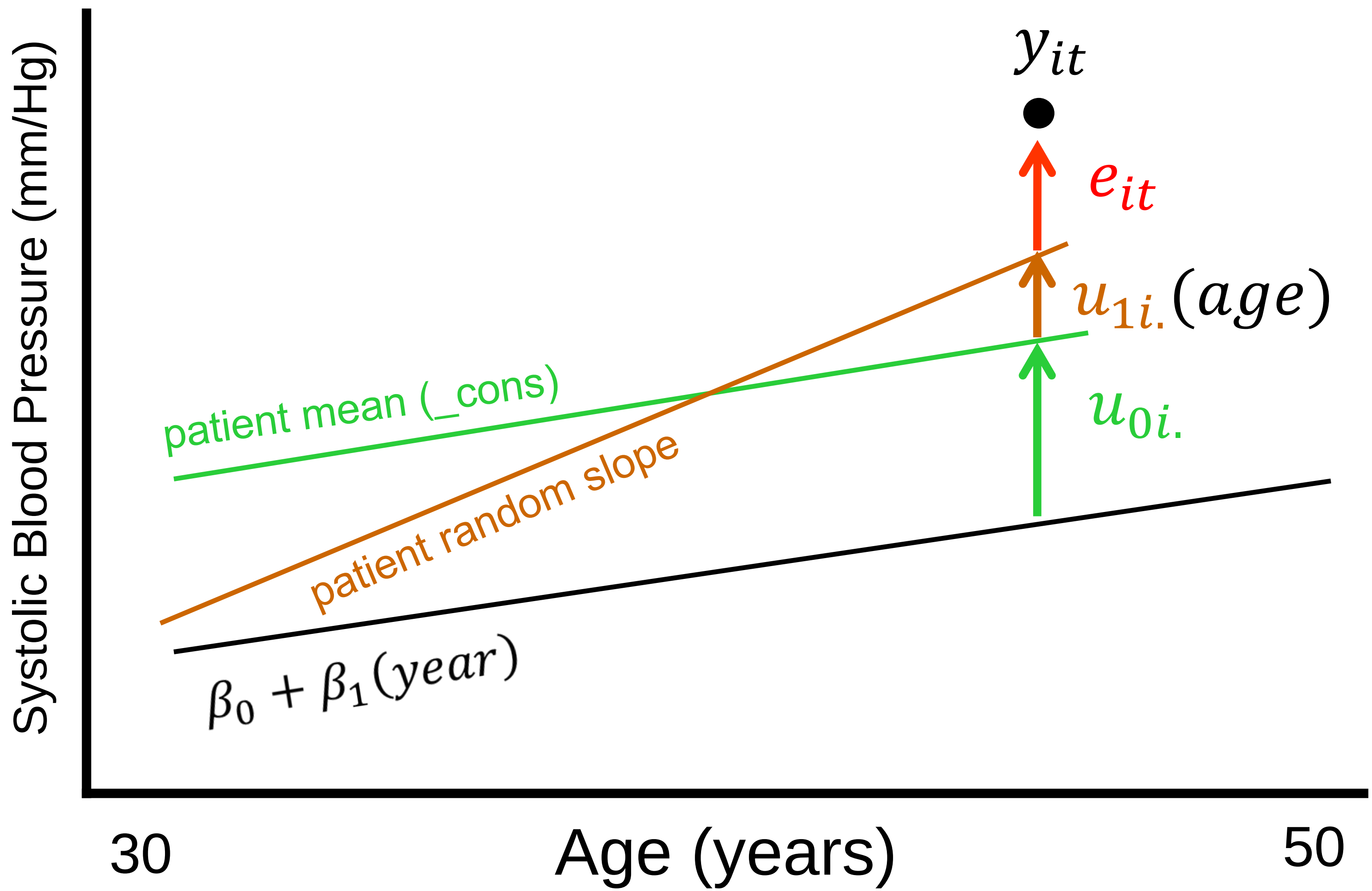
$$y_{it} = \beta_0 + \beta_1(age) + u_{i.} + e_{it}$$



$$y_{it} = \beta_0 + \beta_1(age) + u_{0i.} + u_{1i.}(age) + e_{it}$$



$$y_{ijk} = \beta_0 + \beta_1(age) + u_{0i.} + u_{1i.}(age) + e_{it}$$



Longitudinal Models

$$y_{it} = \beta_0 + \beta_1(age) + u_{0i.} + u_{1i.}(age) + e_{it}$$

$$u_{0i.} \sim N(0, \tau_0^2)$$

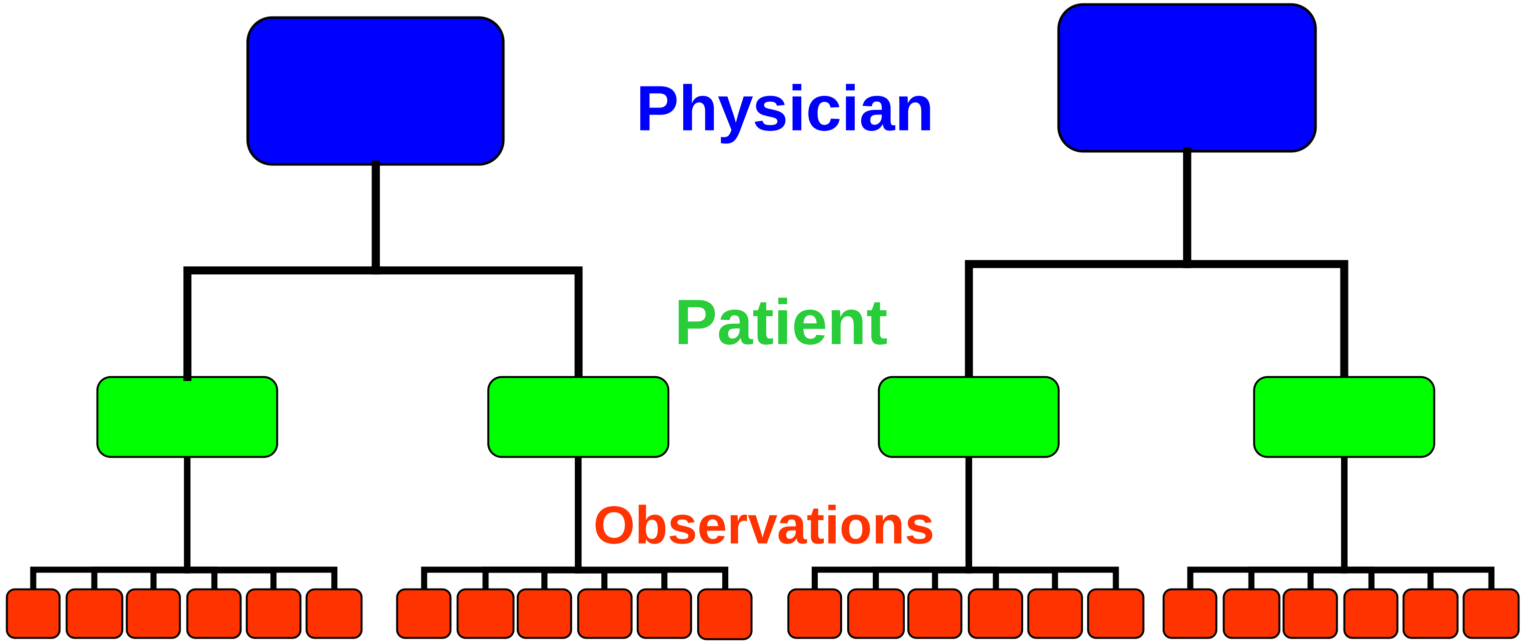
$$u_{1i.} \sim N(0, \tau_1^2)$$

$$e_{it} \sim N(0, \sigma^2)$$

Outline

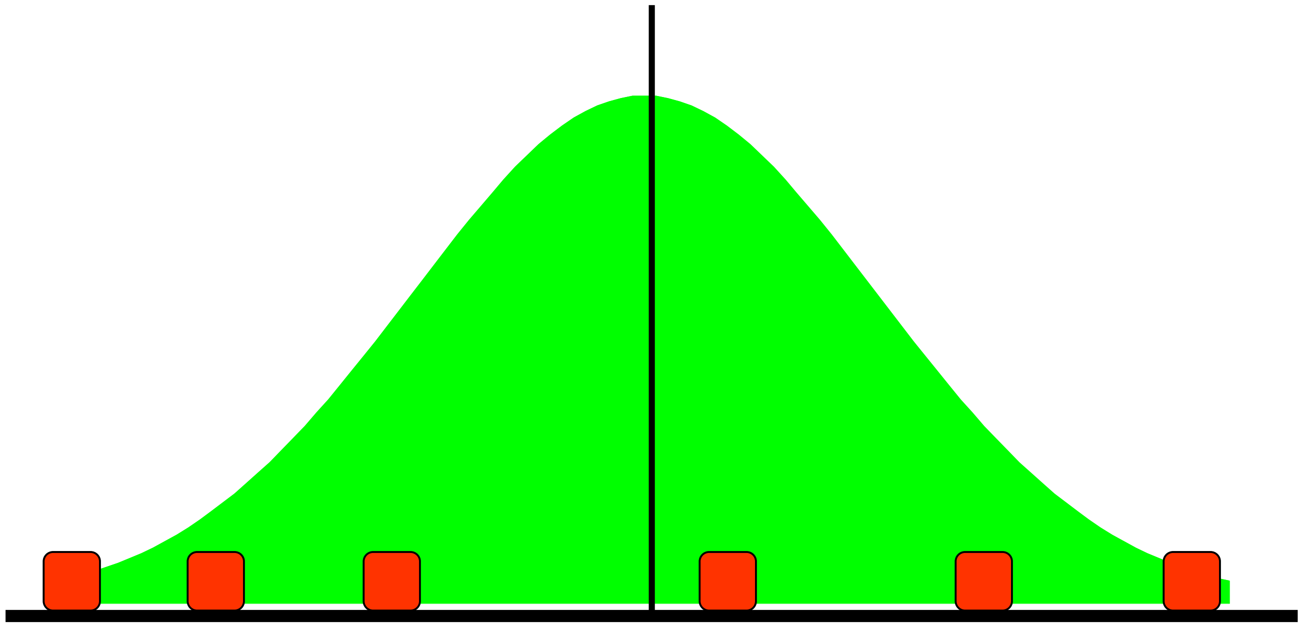
- ✓ The simulated data
- ✓ Single level models
- ✓ Two level models
- ✓ Longitudinal models
- **Three level models**
- Fixed vs random effects
- Multilevel models for binary data
- Multilevel models for survival data
- Multilevel structural equation models
- Bayesian multilevel models

Three Level Models



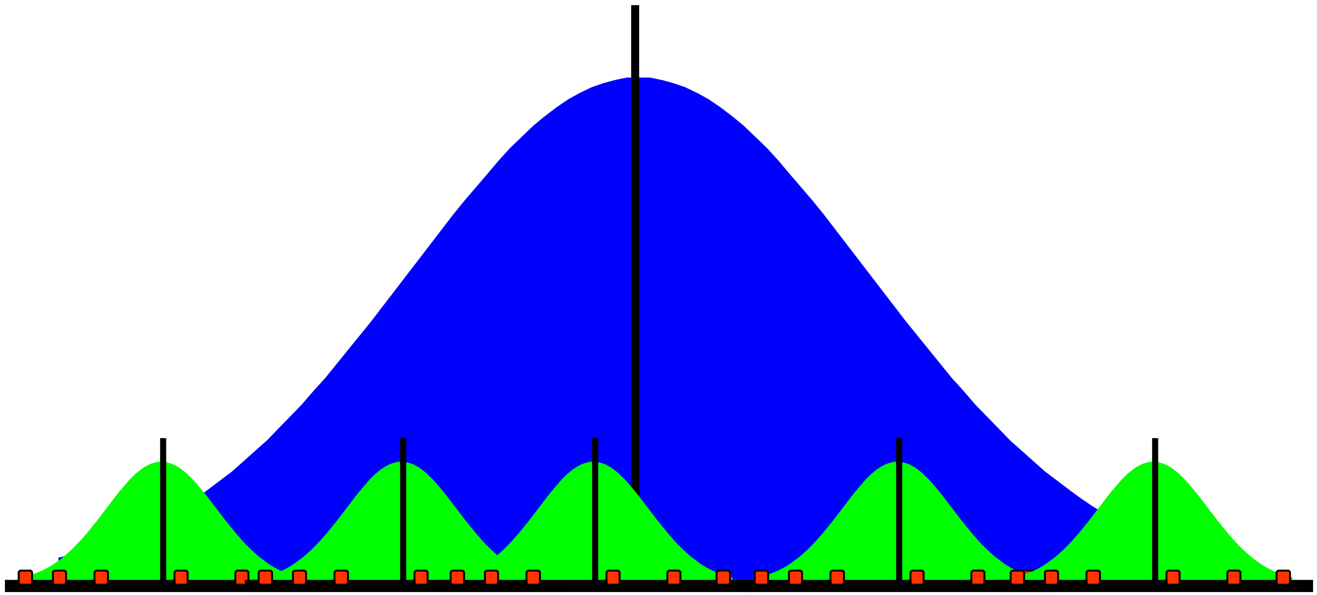
Repeated observations within patients treated by the same physician may be correlated

Three Level Models



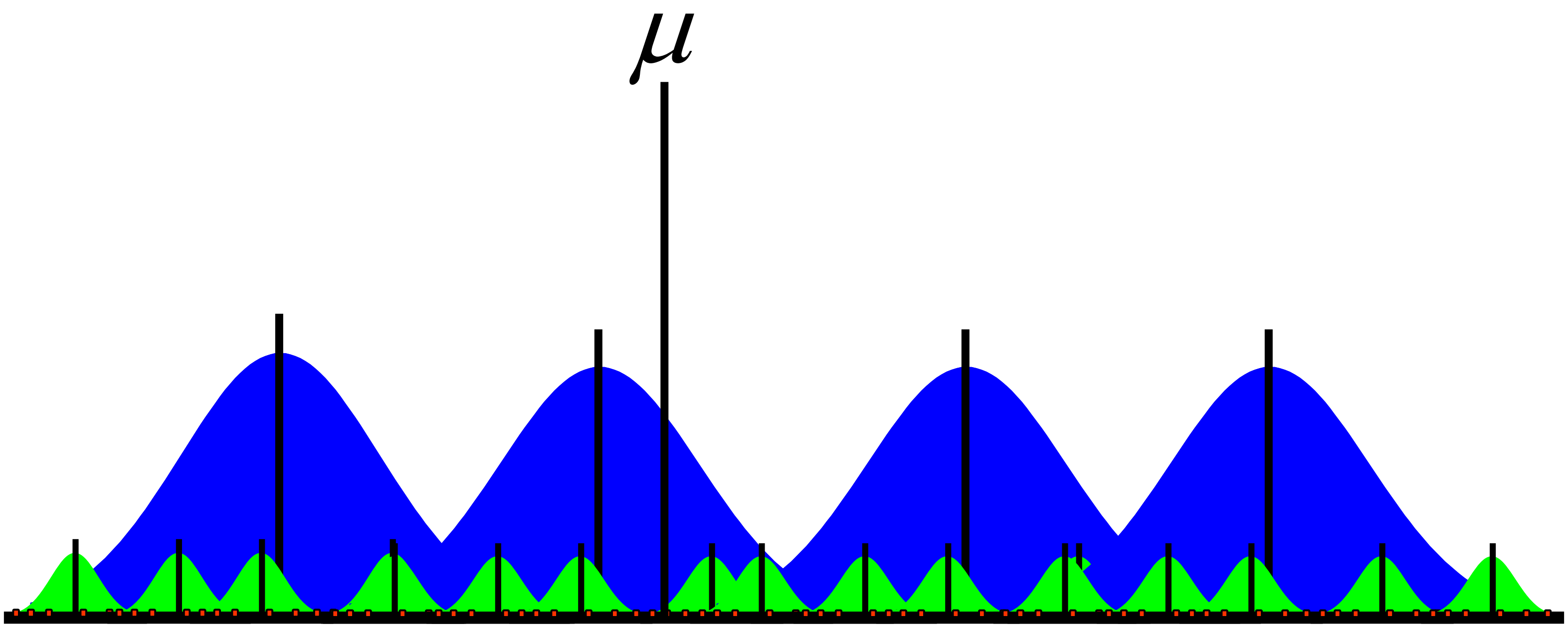
Repeated observations within patient will vary about each patient's mean.

Three Level Models



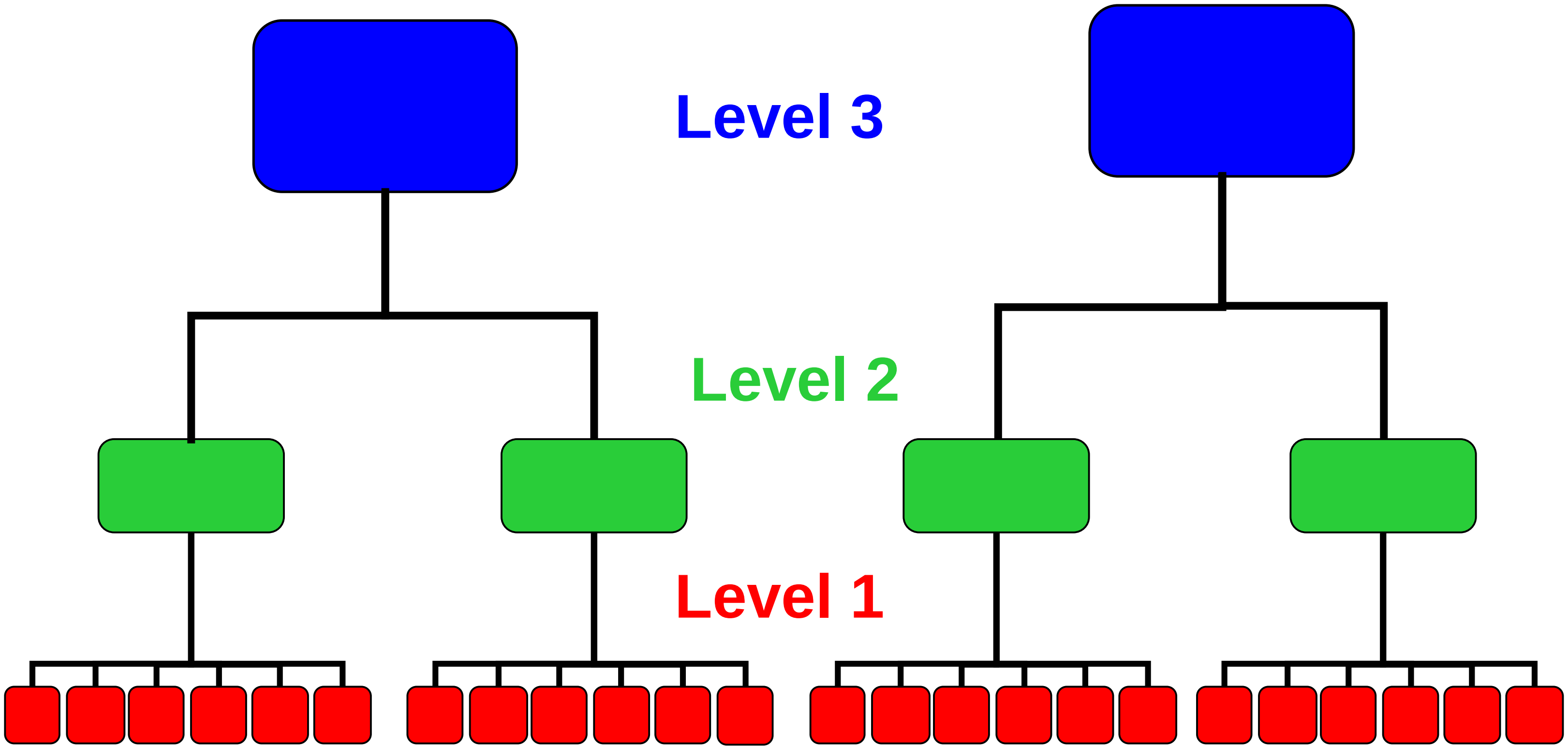
Patient's means treated by the same physician will vary about their physician mean.

Three Level Models

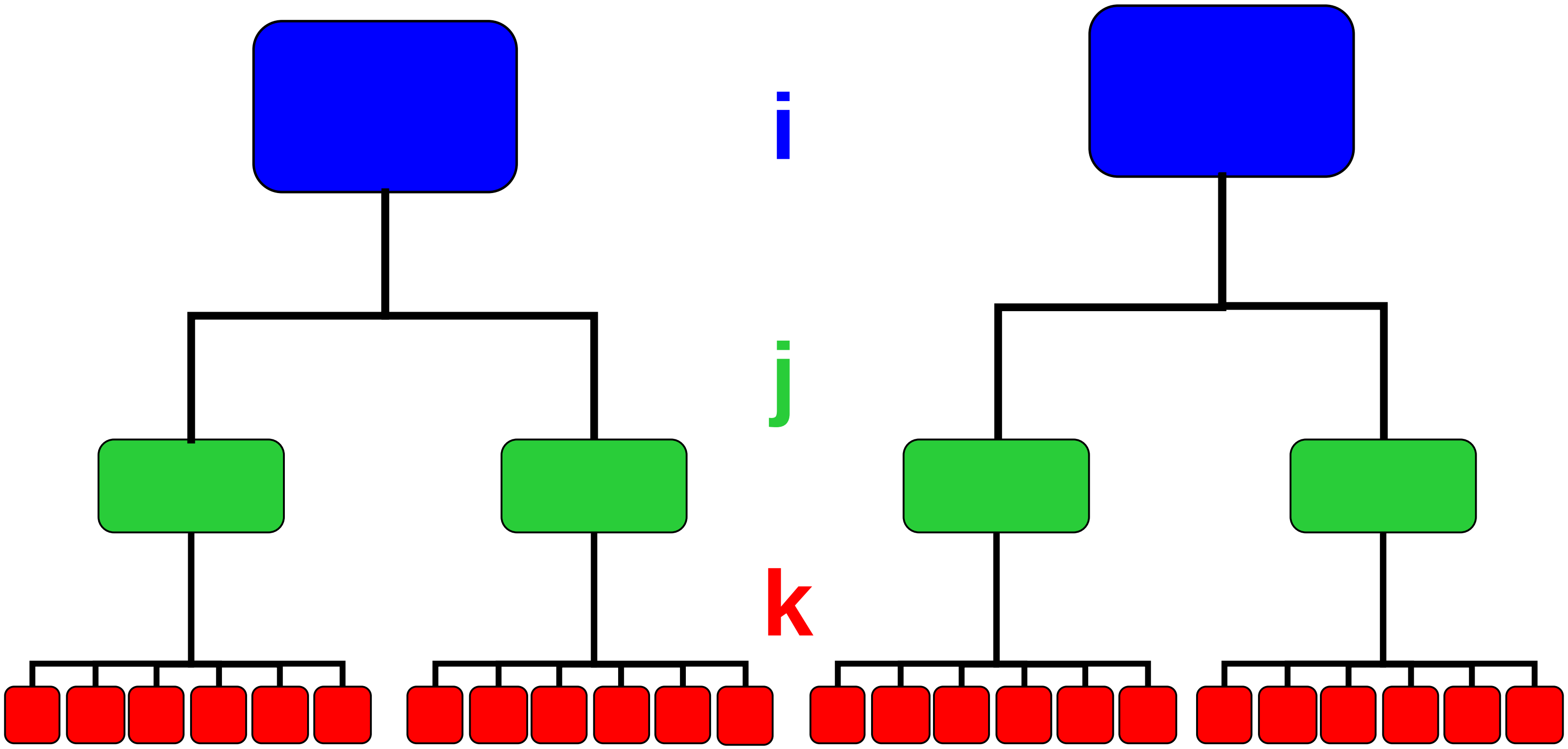


Physician means will vary about the grand mean.

Three Level Models



We can refer to the nesting structure in terms of “Levels”…



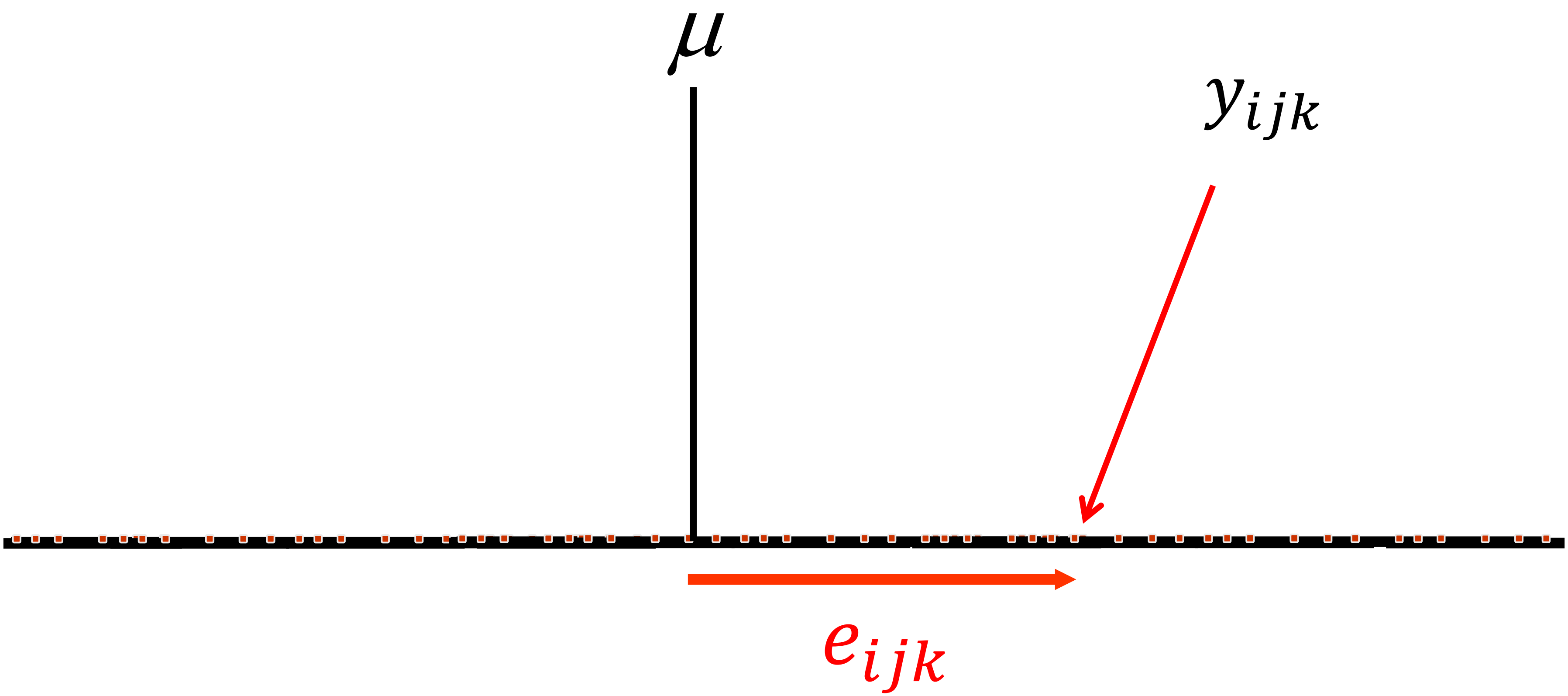
$$y_{ijk} = y_{physician, patient, observation}$$

Three Level Models

$$\underbrace{y_{ijk}}_{\textbf{Observed}} = \underbrace{\mu}_{\textbf{Fixed}} + \underbrace{e_{ijk}}_{\textbf{Random}}$$

Stata Data Analysis and Statistical Software

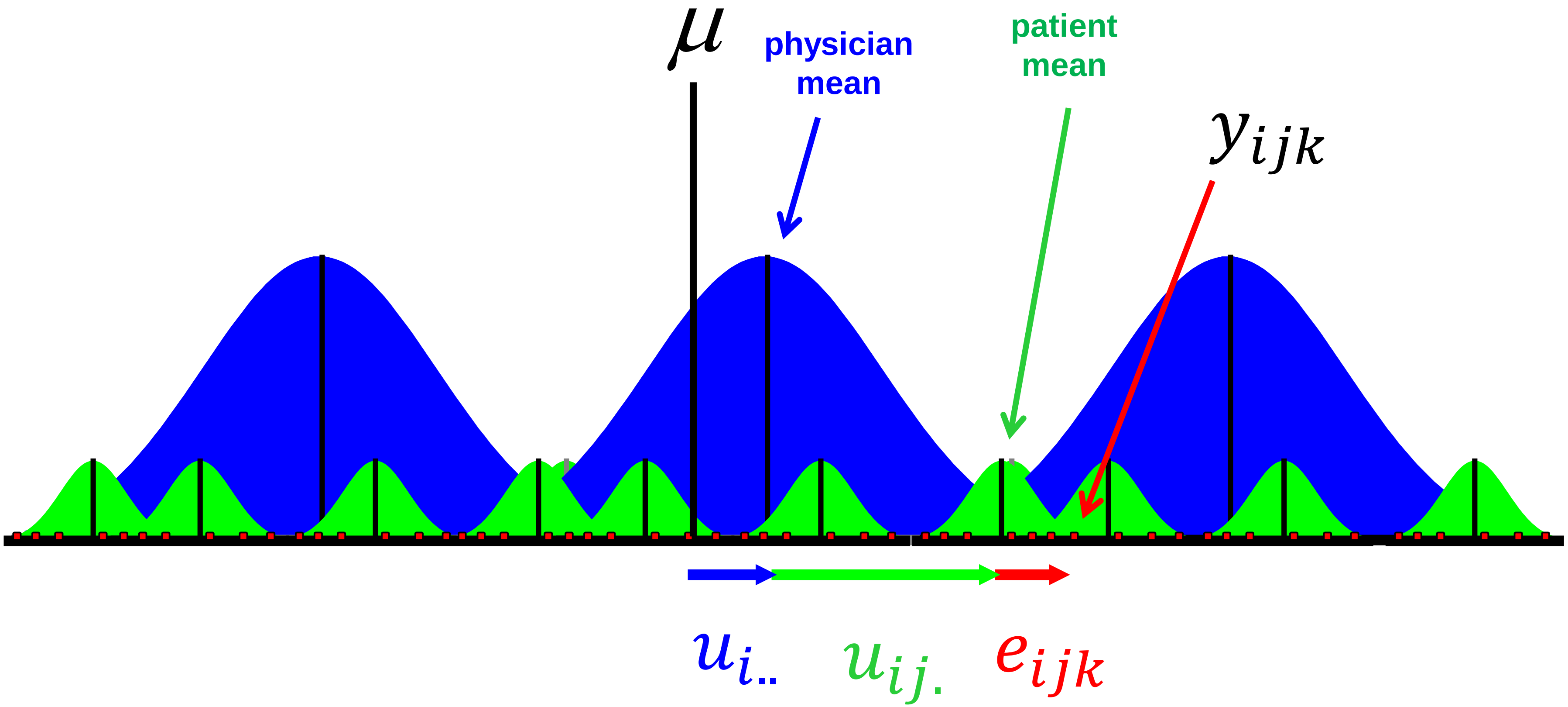
Three Level Models



$$y_{ijk} = \mu + e_{ijk}$$

Three Level Models

$$\underbrace{y_{ijk}}_{\textbf{Observed}} = \underbrace{\mu}_{\textbf{Fixed}} + \underbrace{u_{i..} + u_{jk.} + e_{ijk}}_{\textbf{Random}}$$



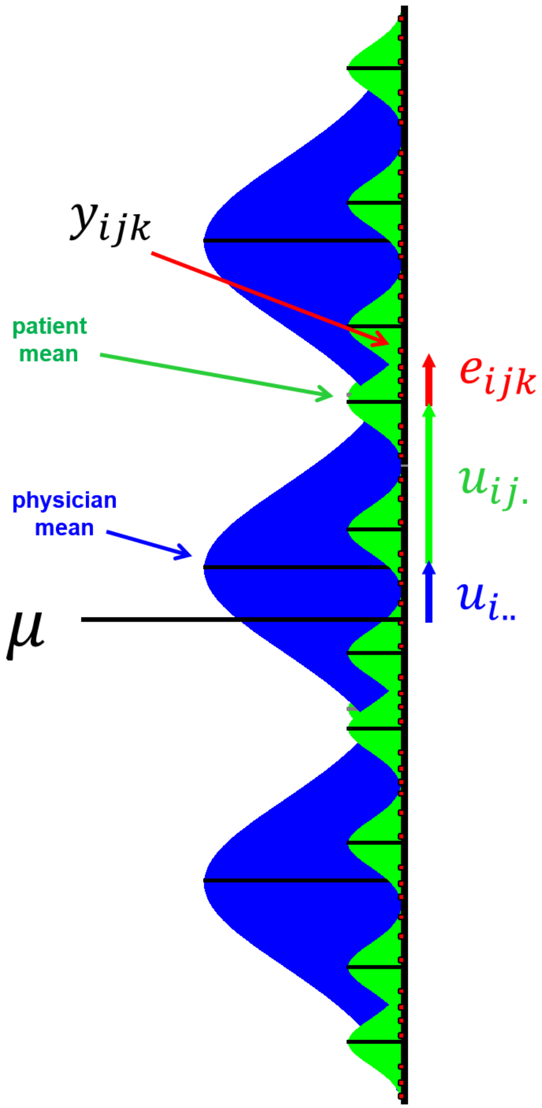
$$y_{ijk} = \mu + u_{i..} + u_{ij.} + e_{ijk}$$

$$y_{ijk} = \mu + u_{i..} + u_{ij.} + e_{ijk}$$

$$u_{i..} \sim N(0, \gamma^2)$$

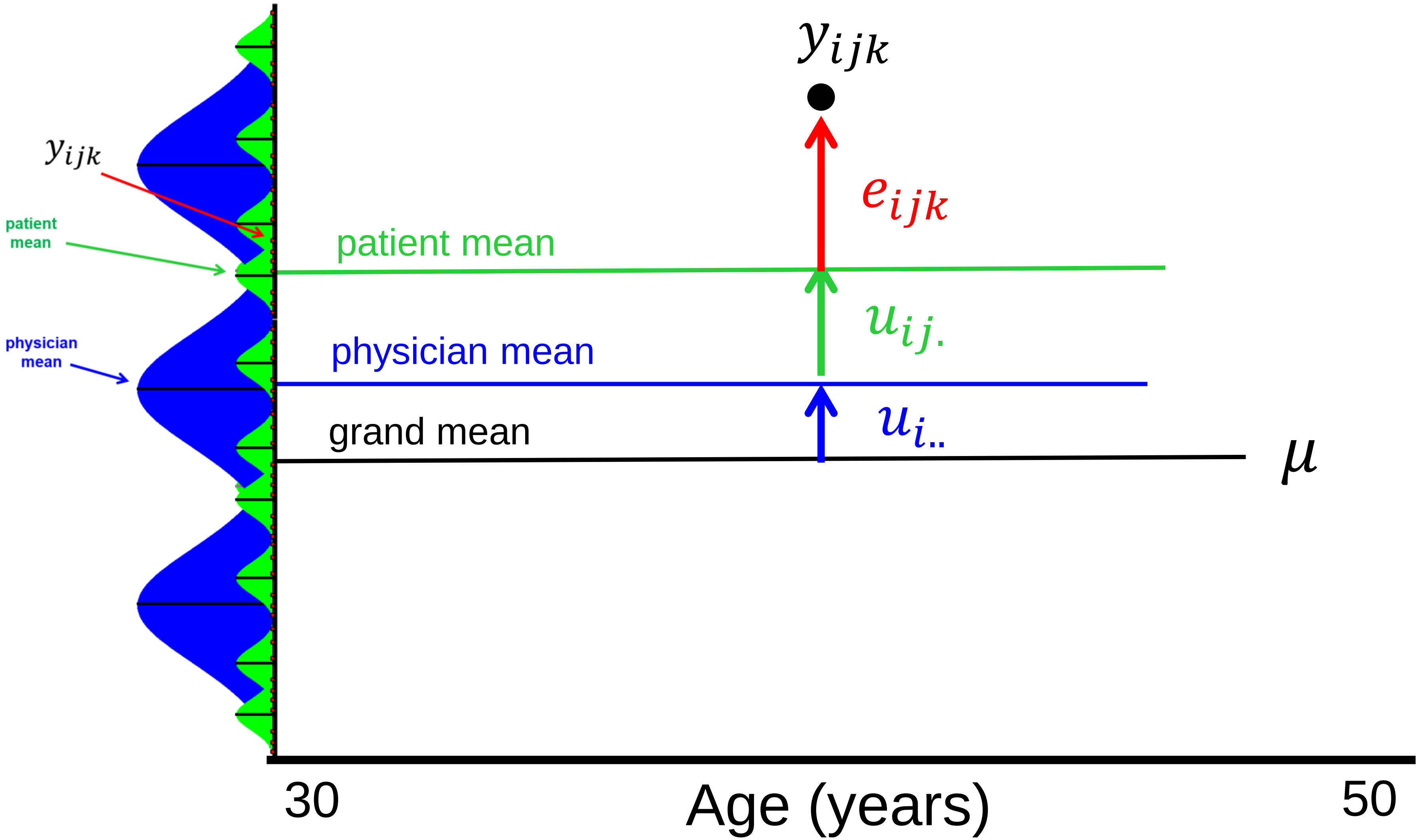
$$u_{ij.} \sim N(0, \tau^2)$$

$$e_{ijk} \sim N(0, \sigma^2)$$

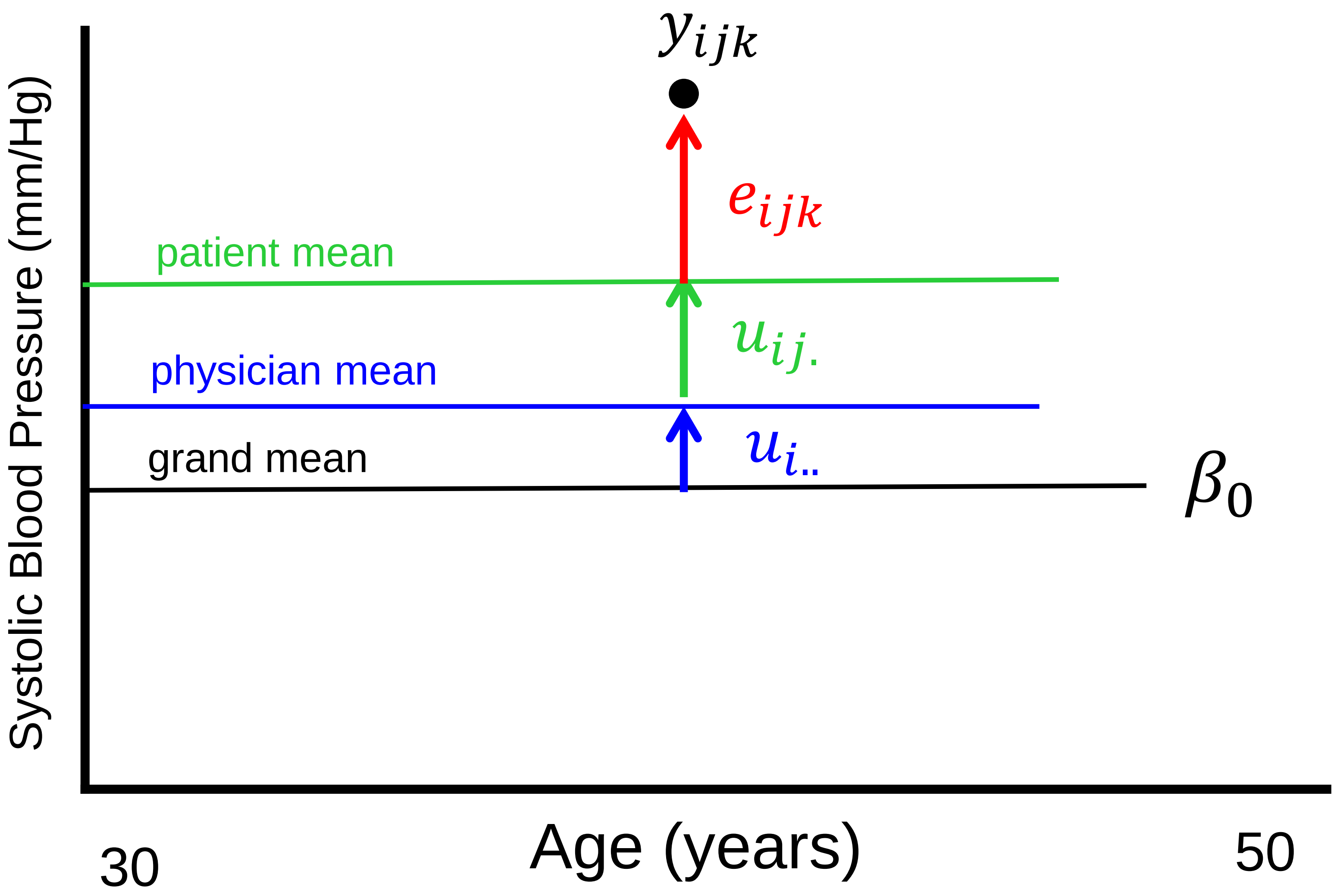


$$y_{ijk} = \mu + u_{i..} + u_{ij.} + e_{ijk}$$

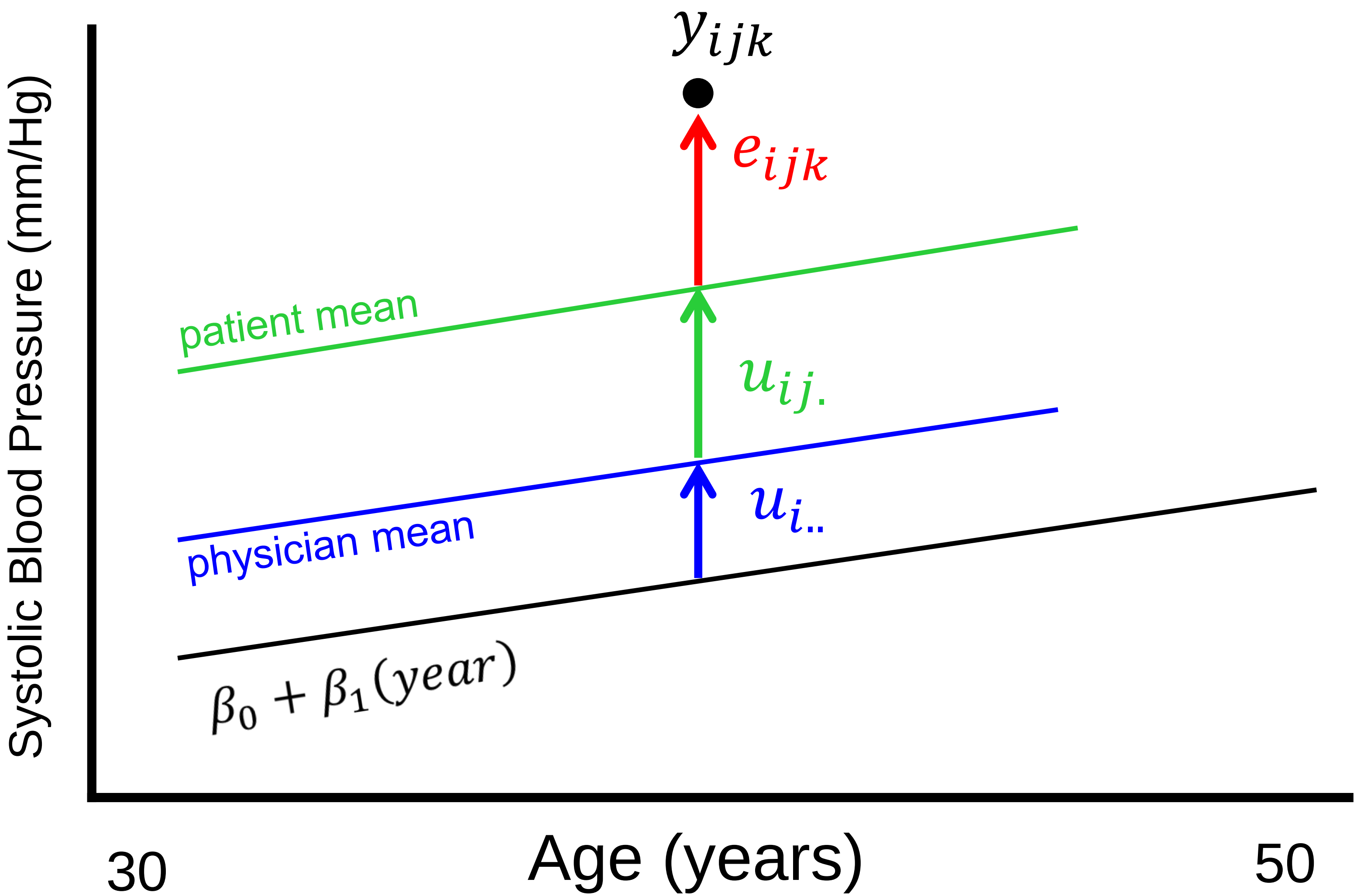
$$y_{ijk} = \mu + u_{i..} + u_{ij.} + e_{ijk}$$



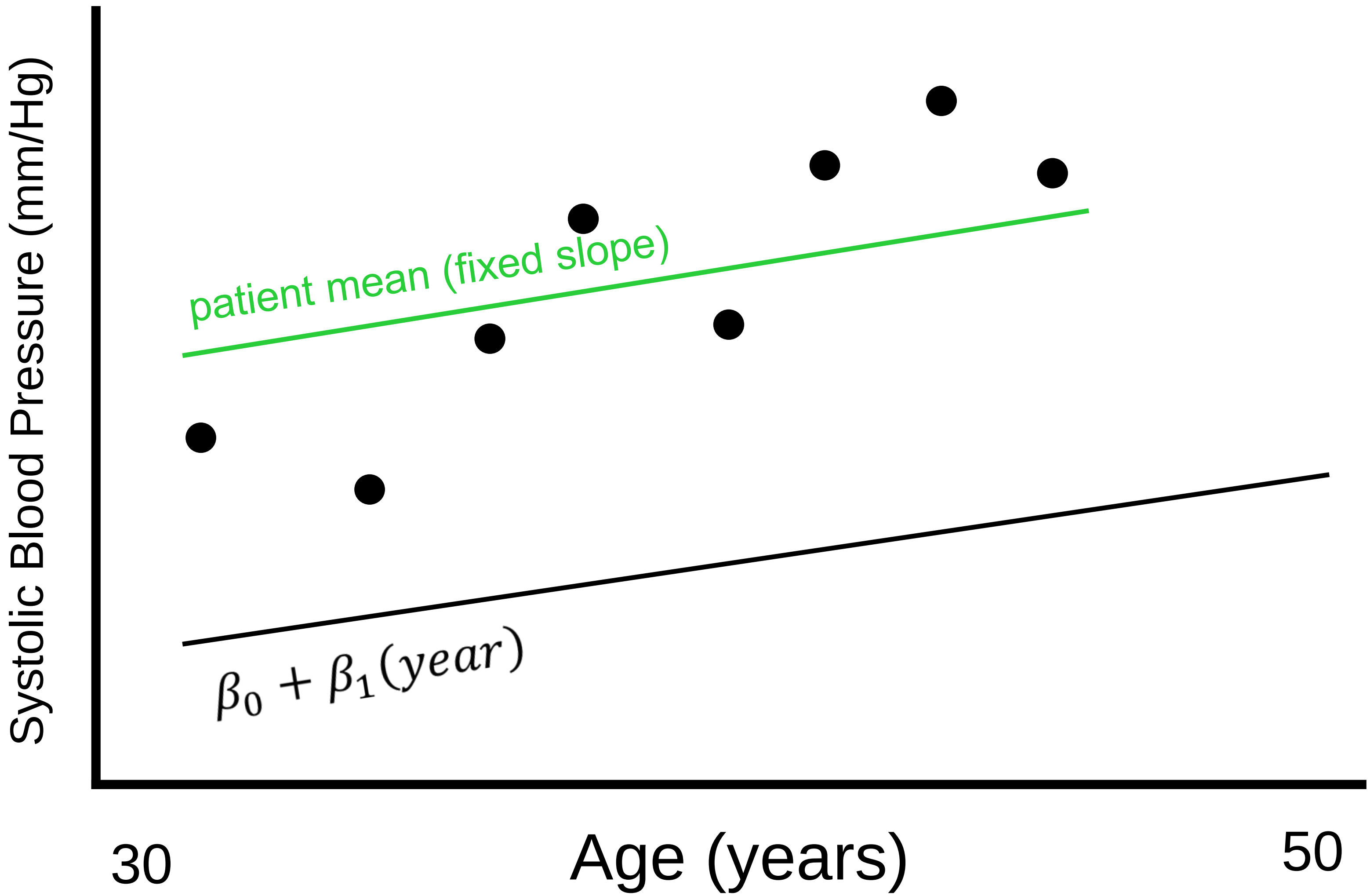
$$y_{ijk} = \beta_0 + u_{i..} + u_{ij.} + e_{ijk}$$



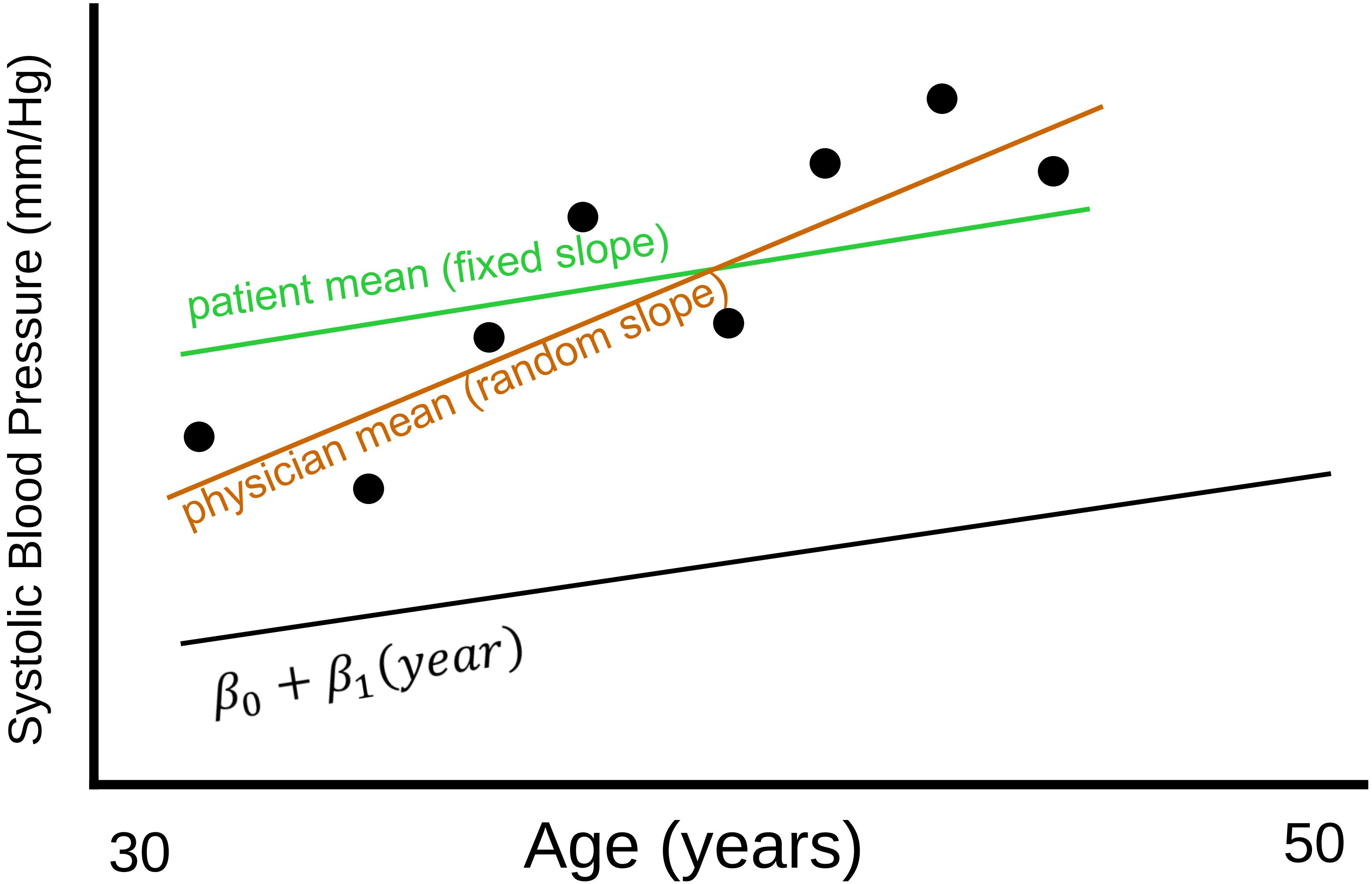
$$y_{ijk} = \beta_0 + \beta_1(age) + u_{i..} + u_{ij.} + e_{ijk}$$



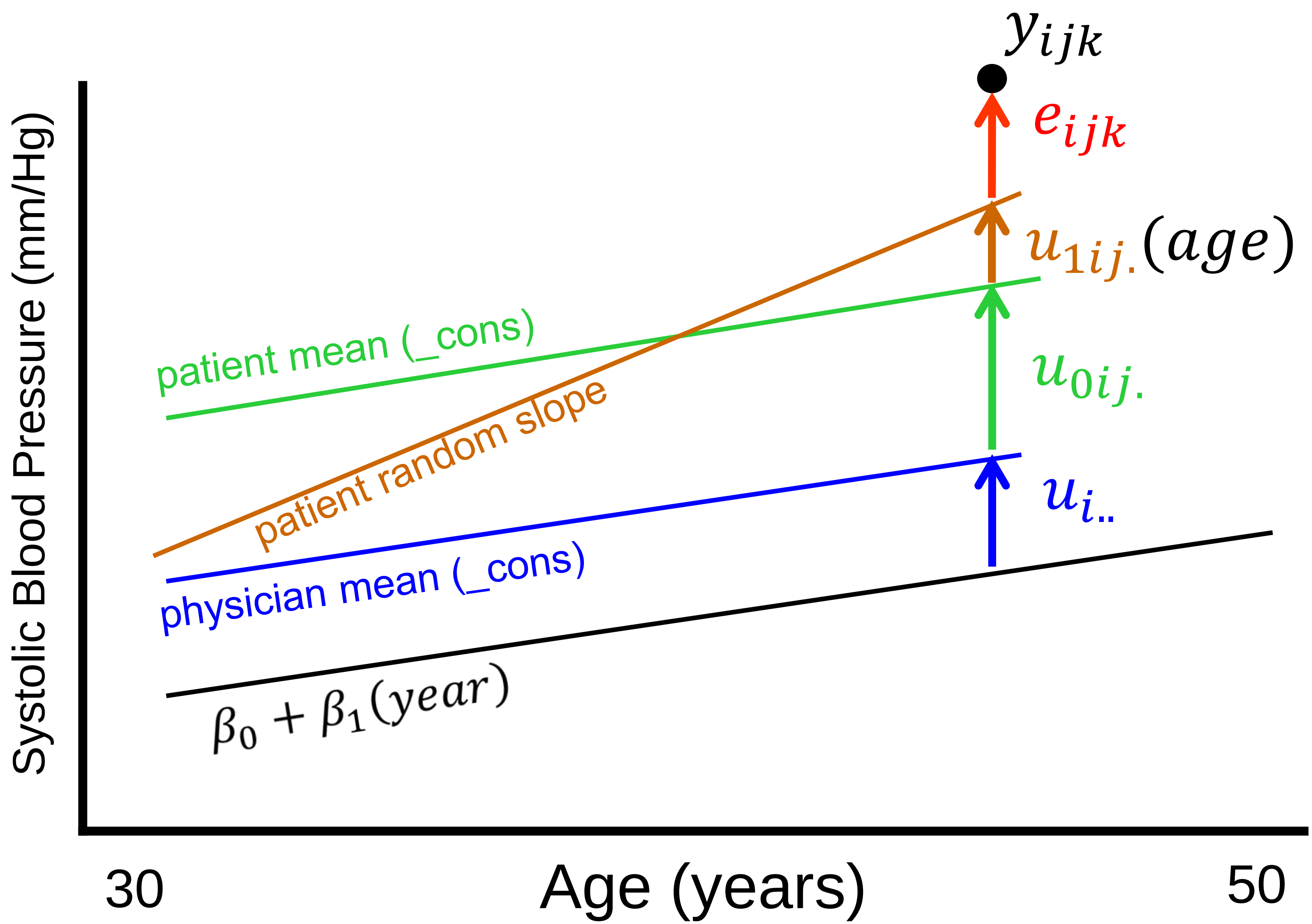
$$y_{ijk} = \beta_0 + \beta_1(age) + u_{i..} + u_{ij.} + e_{ijk}$$



$$y_{ijk} = \beta_0 + \beta_1(age) + u_{i..} + u_{0ij.} + u_{1ij.}(age) + e_{ijk}$$



$$y_{ijk} = \beta_0 + \beta_1(age) + u_{i..} + u_{0ij.} + u_{1ij.}(age) + e_{ijk}$$



$$y_{ijk} = \beta_0 + \beta_1(age) + u_{i..} + u_{0ij.} + u_{1ij.}(age) + e_{ijk}$$

$$u_{i..} \sim N(0, \gamma^2)$$

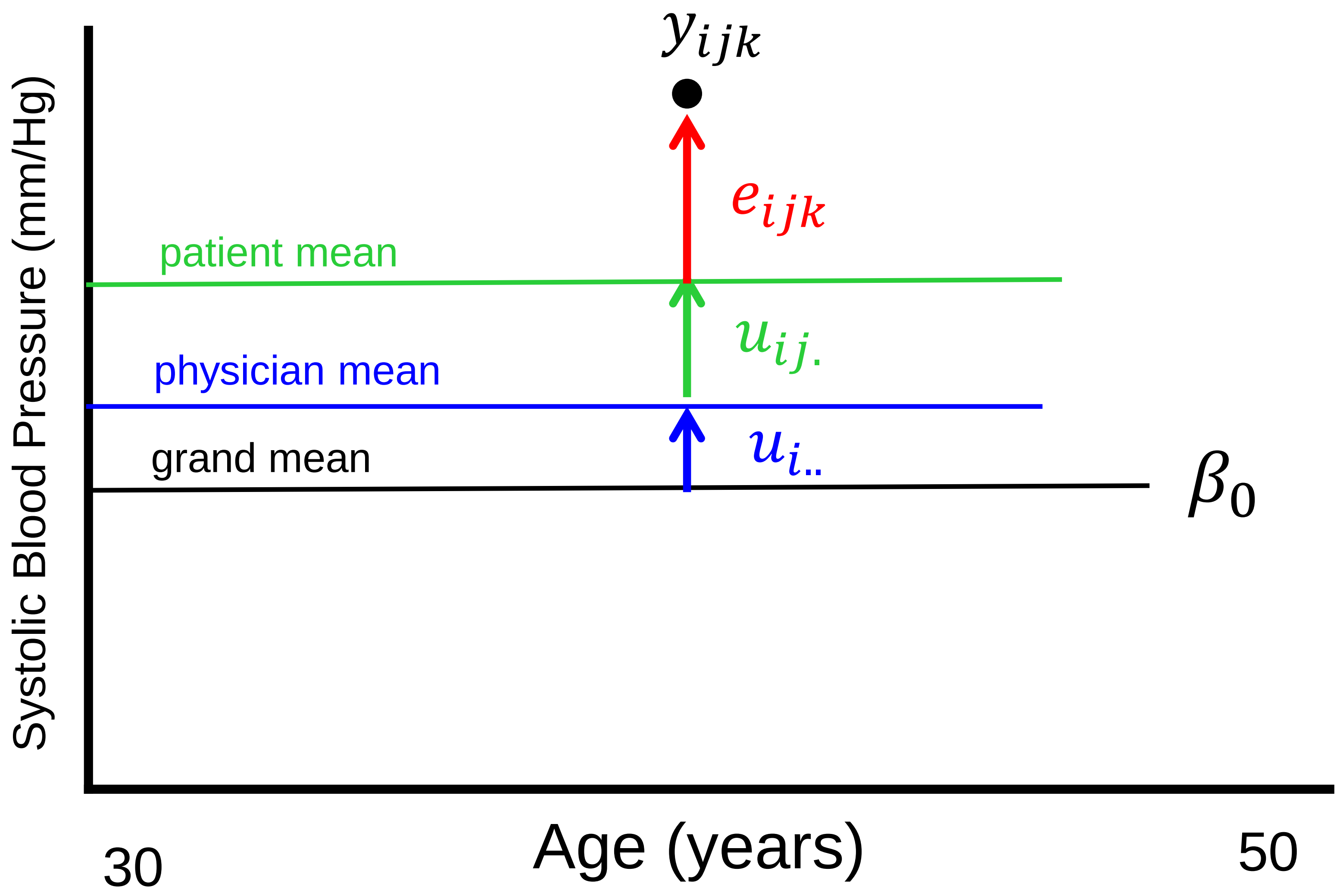
$$u_{0ij.} \sim N(0, \tau_0^2)$$

$$u_{1ij.} \sim N(0, \tau_1^2)$$

$$e_{ijk} \sim N(0, \sigma^2)$$

How do we do this in Stata?

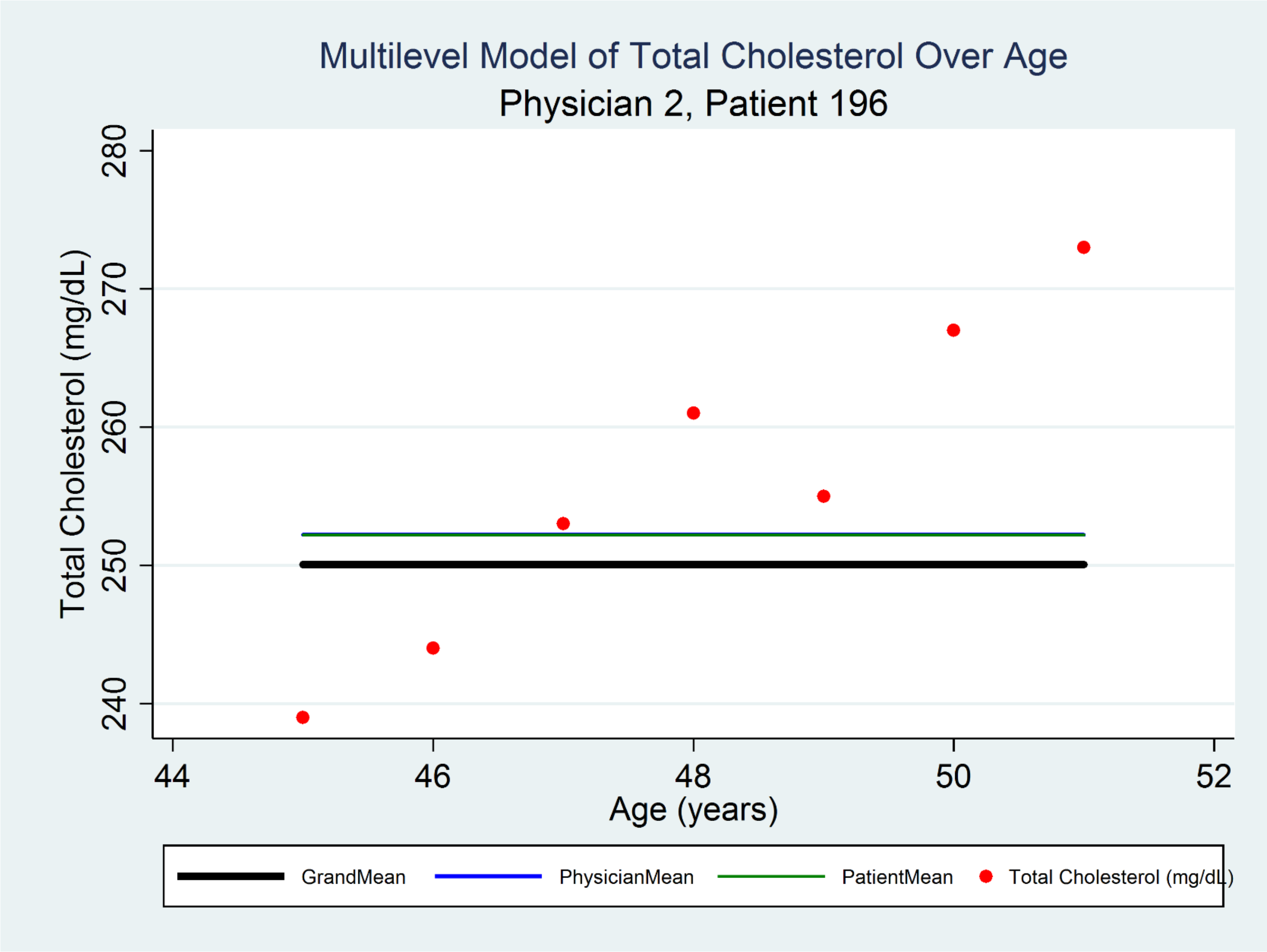
$$y_{ijk} = \beta_0 + u_{i..} + u_{ij.} + e_{ijk}$$



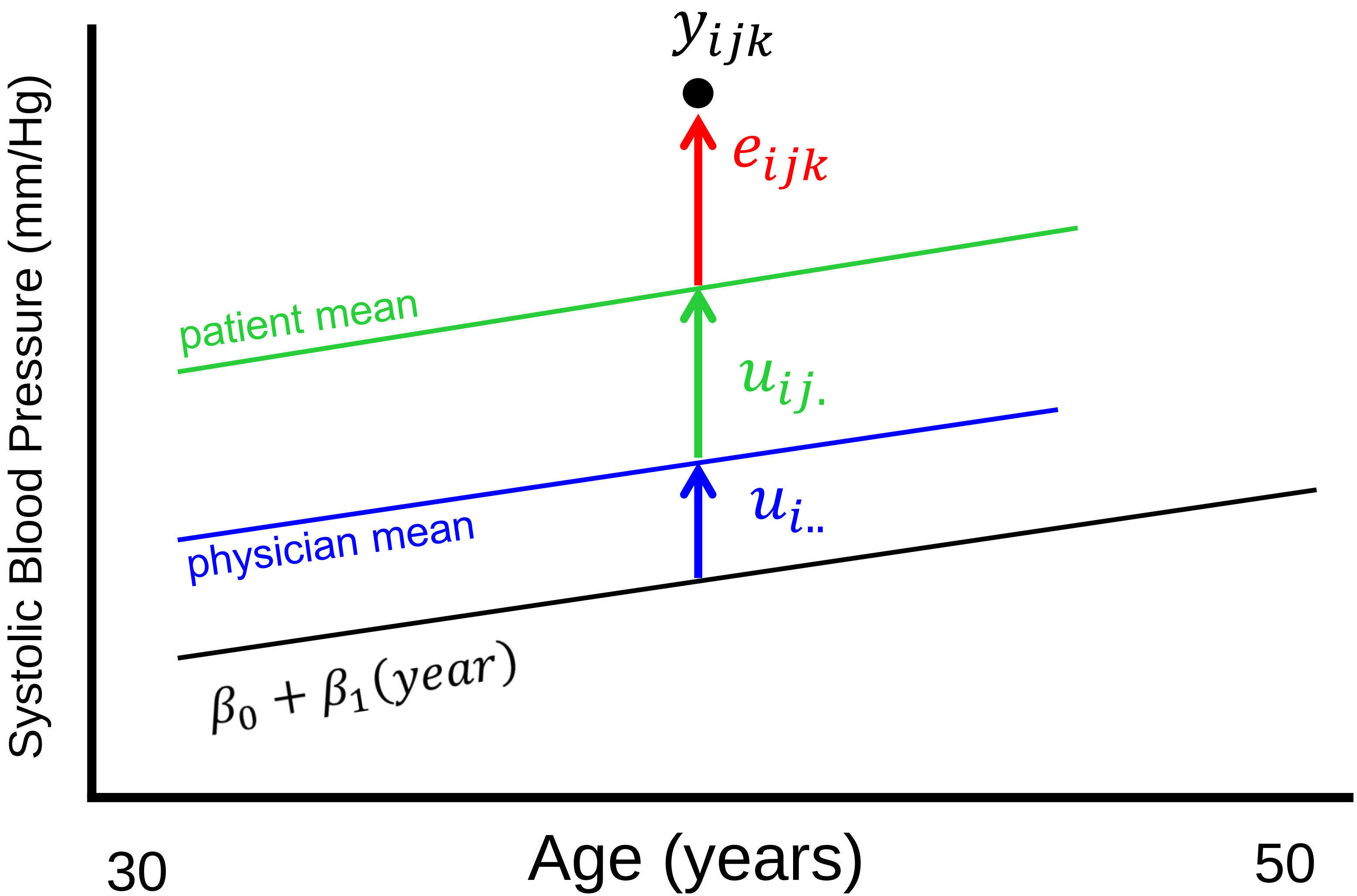
$$y_{ijk} = \beta_0 + u_{i..} + u_{ij.} + e_{ijk}$$

```
. mixed chol, || physician: || patient:

------------------------------------------------------------------------------
        chol |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   250.0533   1.253489   199.49   0.000     247.5965    252.5101
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
physician: Identity          |
                  var(_cons) |   4.228256   3.849052      .7100652    25.17818
-----------------------------+------------------------------------------------
patient: Identity            |
                  var(_cons) |   1.98e-19   5.10e-19      1.29e-21    3.06e-17
-----------------------------+------------------------------------------------
               var(Residual) |   339.8134   10.49483      319.8541    361.0182
------------------------------------------------------------------------------
LR test vs. linear model: chi2(2) = 19.14                 Prob > chi2 = 0.0001
```

$$y_{ijk} = \mu + u_{i..} + u_{ij.} + e_{ijk}$$



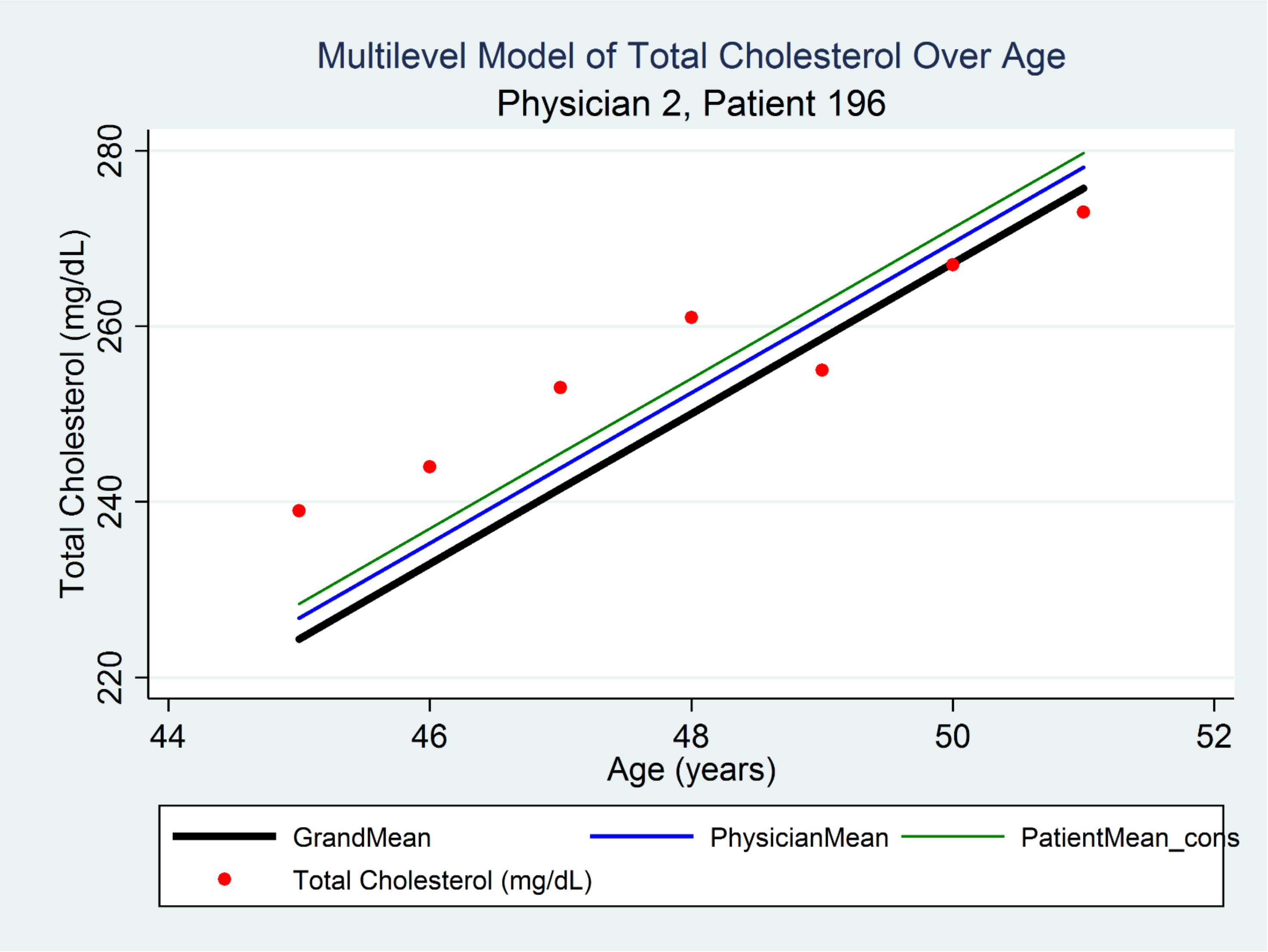
$$y_{ijk} = \beta_0 + \beta_1(age) + u_{i..} + u_{ij.} + e_{ijk}$$



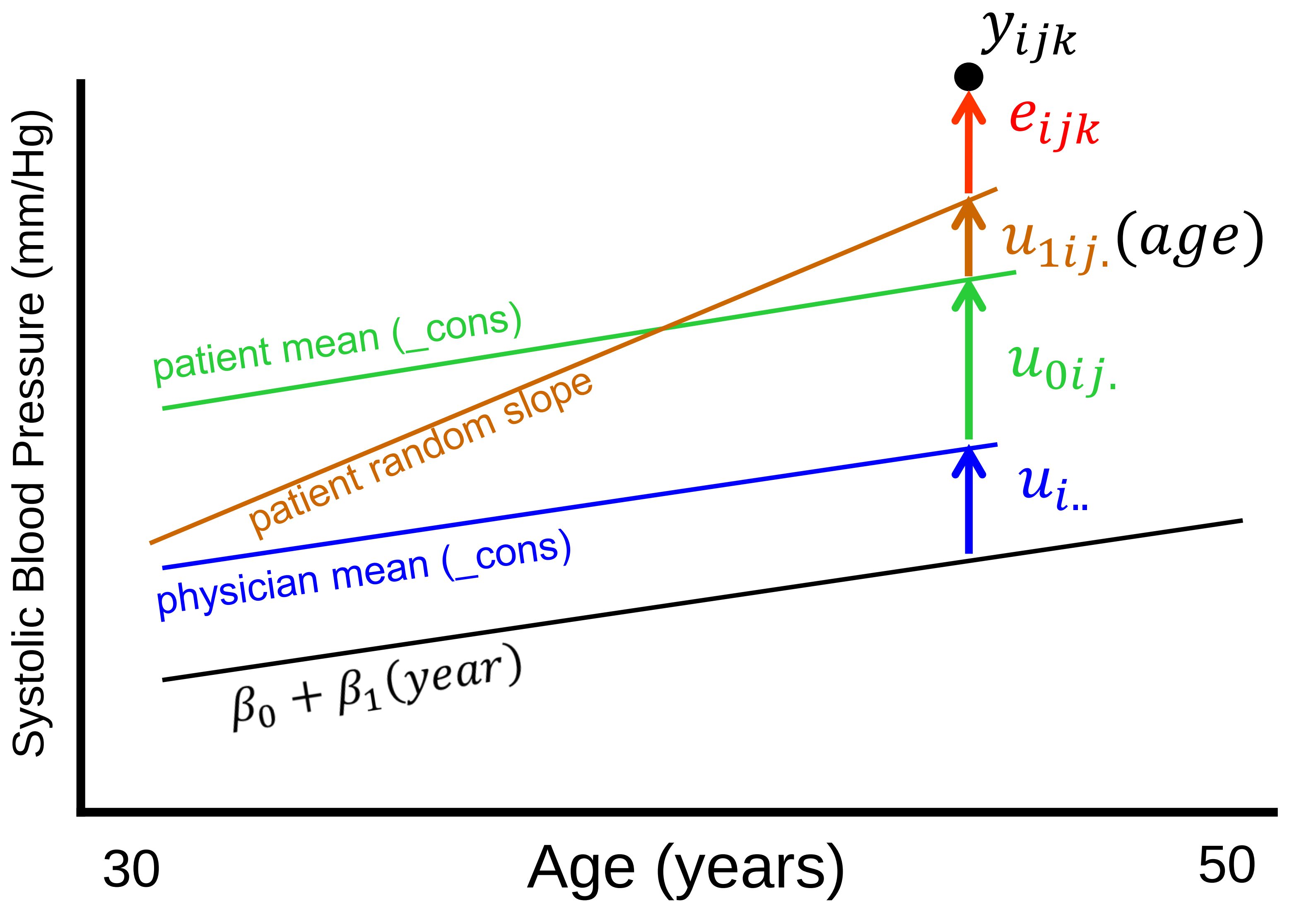
$$y_{ijk} = \beta_0 + \beta_1(cage) + u_{i..} + u_{ij.} + e_{ijk}$$

```
. mixed chol cage, || physician: || patient:

------------------------------------------------------------------------------
        chol |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cage |   8.552381   .0705304   121.26   0.000     8.414144    8.690618
       _cons |   250.0533    1.25345   199.49   0.000     247.5966    252.5101
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
physician: Identity          |
                  var(_cons) |   4.602921   3.848604      .8939747    23.69965
-----------------------------+------------------------------------------------
patient: Identity            |
                  var(_cons) |   5.079615   .9282736      3.550407    7.267475
-----------------------------+------------------------------------------------
               var(Residual) |   41.78608   1.392869      39.14337     44.6072
------------------------------------------------------------------------------
LR test vs. linear model: chi2(2) = 241.67                Prob > chi2 = 0.0000
```

$$y_{ijk} = \beta_0 + \beta_1(age) + u_{i..} + u_{ij.} + e_{ijk}$$



$$y_{ijk} = \beta_0 + \beta_1(age) + u_{i..} + u_{0ij.} + u_{1ij.}(age) + e_{ijk}$$

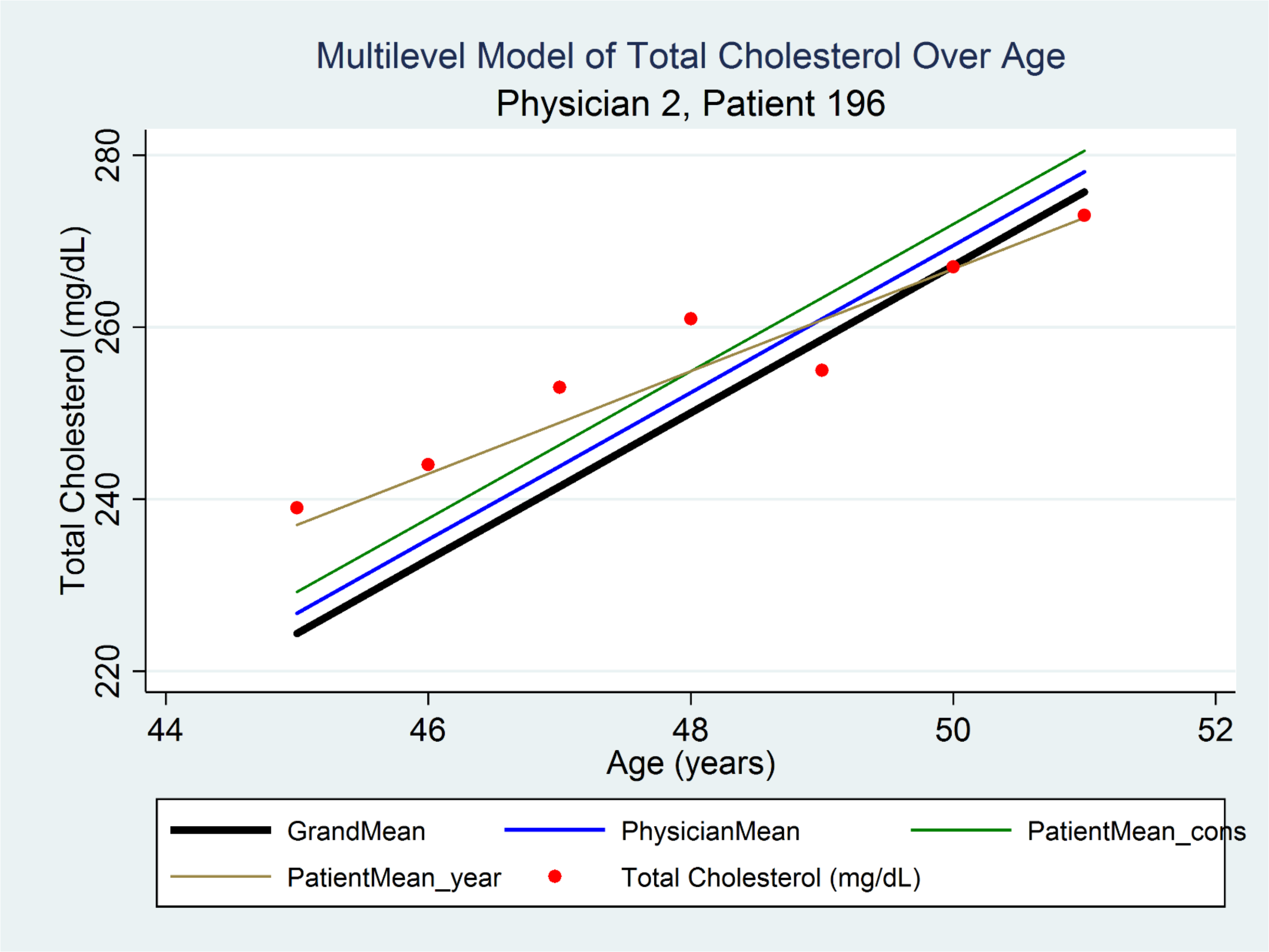


$$y_{ijk} = \beta_0 + \beta_1(age) + u_{i..} + u_{0ij.} + u_{1ij.}(age) + e_{ijk}$$

```
. mixed chol cage, || physician: || patient: cage, cov(indep)

------------------------------------------------------------------------------
        chol |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cage |   8.552381   .1242629    68.82   0.000      8.30883    8.795932
       _cons |   250.0533   1.253464   199.49   0.000     247.5966    252.5101
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
physician: Identity          |
                  var(_cons) |   4.603026   3.848763      .8939679    23.70091
-----------------------------+------------------------------------------------
patient: Independent         |
                   var(cage) |   3.768022   .3795468      3.092952    4.590432
                  var(_cons) |    7.59163   .9154443      5.993659    9.615636
-----------------------------+------------------------------------------------
               var(Residual) |   24.20199   .8837314      22.53043    25.99756
------------------------------------------------------------------------------
LR test vs. linear model: chi2(3) = 721.05                Prob > chi2 = 0.0000
```

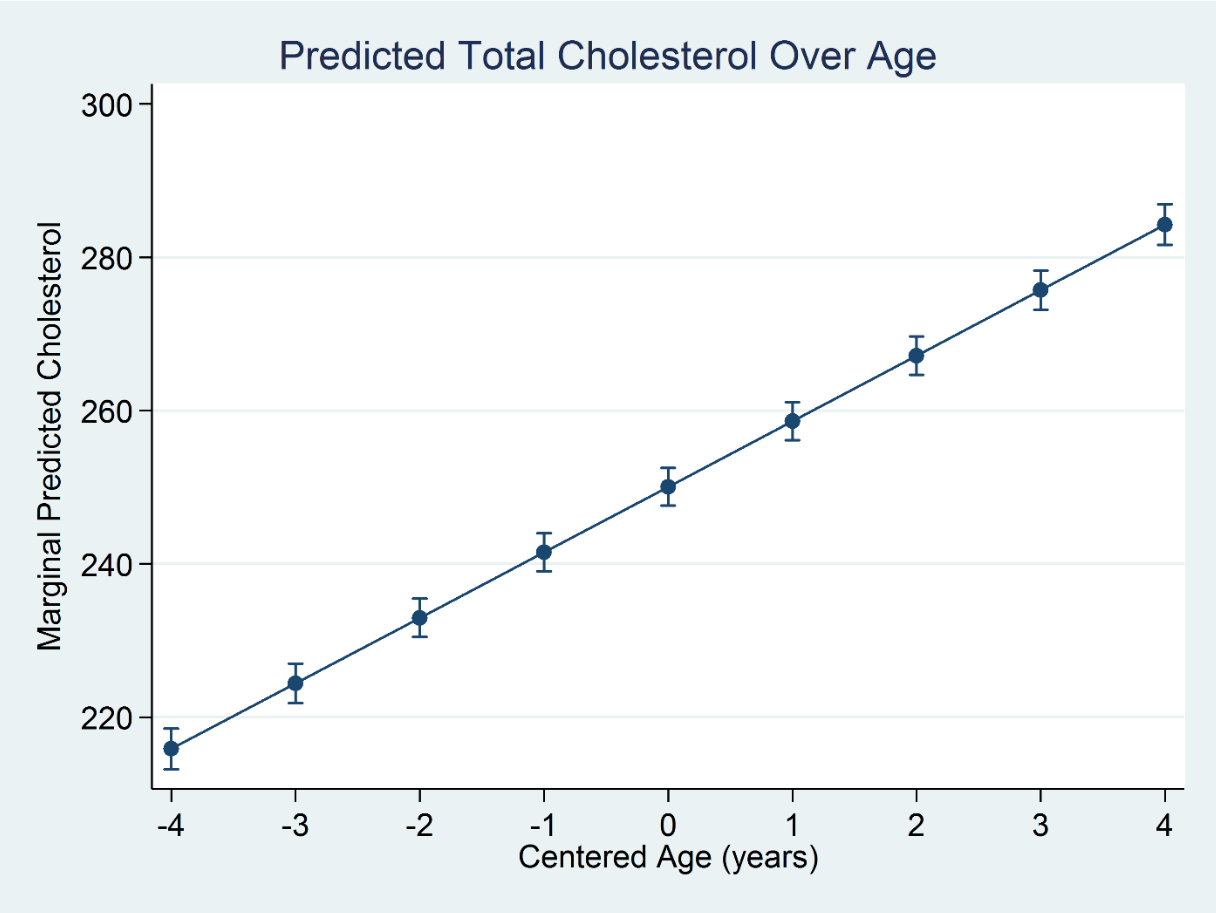
Data Analysis and Statistical Software

$$y_{ijk} = \beta_0 + \beta_1(age) + u_{i..} + u_{0ij.} + u_{1ij.}(age) + e_{ijk}$$



Data Analysis and Statistical Software

$$y_{ijk} = \beta_0 + \beta_1(age) + u_{i..} + u_{0ij.} + u_{1ij.}(age) + e_{ijk}$$




```
margins, at(cage=(-4(1)4))
mariginsplot
```

Mean and Centered Age

```
. list physician patient age_mi age cage mage sex sbp chol HiBP time dead if patient<=2, noobs sep(7) ab(9)
```

physician	patient	age_mi	age	cage	mage	sex	sbp	chol	HiBP	time	dead
1	1	59	53	-3	56	Male	144	214	0	14	censored
1	1	59	54	-2	56	Male	142	231	0	14	censored
1	1	59	55	-1	56	Male	150	230	0	14	censored
1	1	59	56	0	56	Male	147	239	0	14	censored
1	1	59	57	1	56	Male	146	255	0	14	censored
1	1	59	58	2	56	Male	154	268	1	14	censored
1	1	59	59	3	56	Male	155	286	1	14	censored
2	2	50	44	-3	47	Male	151	233	1	4	censored
2	2	50	45	-2	47	Male	139	238	0	4	censored
2	2	50	46	-1	47	Male	155	240	1	4	censored
2	2	50	47	0	47	Male	146	257	0	4	censored
2	2	50	48	1	47	Male	161	264	1	4	censored
2	2	50	49	2	47	Male	160	263	1	4	censored
2	2	50	50	3	47	Male	150	277	0	4	censored

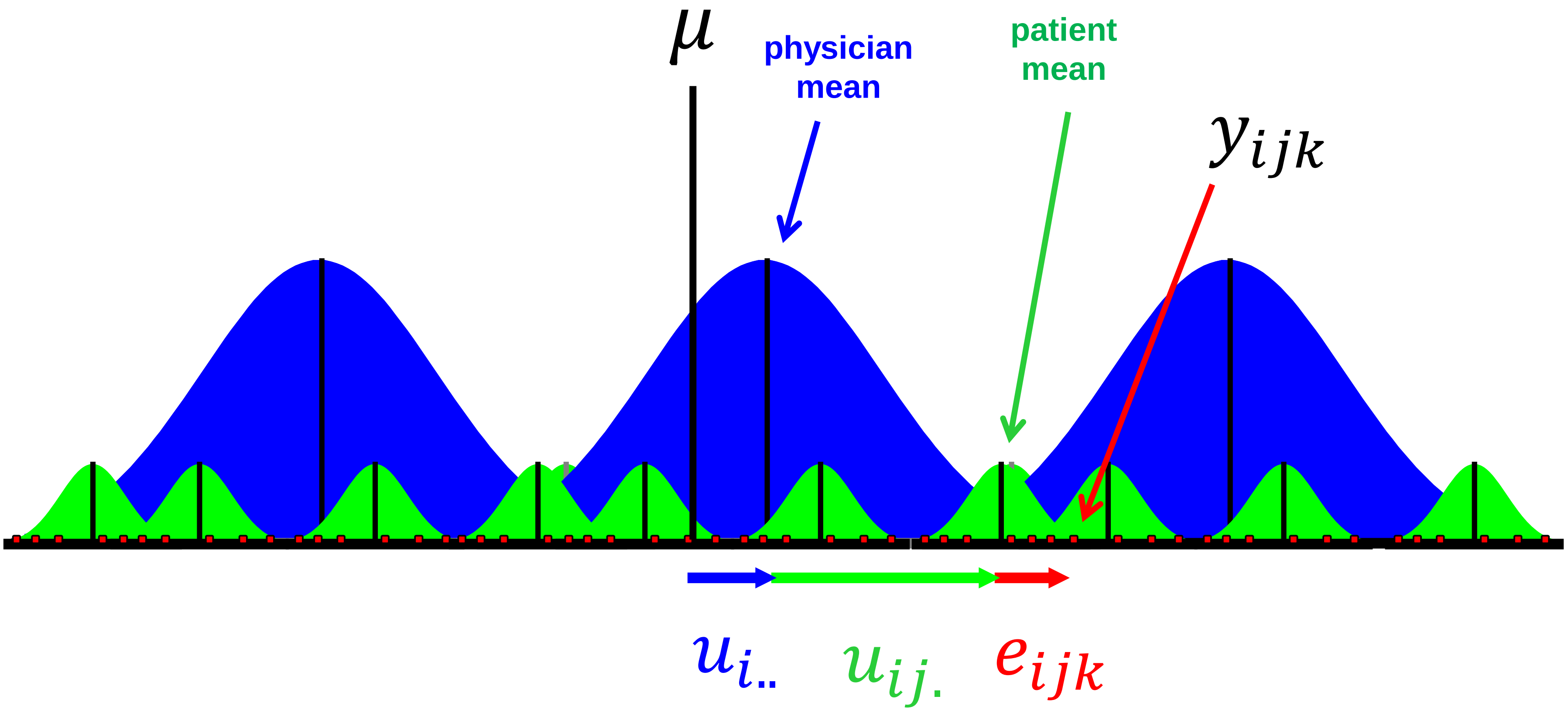
$$y_{ijk} = \beta_0 + \boldsymbol{\beta_1}(\boldsymbol{mage}) + \beta_2(age) + u_{i..} + u_{0ij.} + u_{1ij.}(age) + e_{ijk}$$

```
mixed chol mage cage, || physician: || patient: cage,cov(indep)
------------------------------------------------------------------------------
        chol |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        mage |   .1350396   .0588787     2.29   0.022     .0196395    .2504398
        cage |   8.552381   .1242629    68.82   0.000      8.30883    8.795932
       _cons |    243.061   3.311854    73.39   0.000     236.5699    249.5521
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
physician: Identity          |
                  var(_cons) |   4.912183   4.101832       .956079    25.23802
-----------------------------+------------------------------------------------
patient: Independent         |
                   var(cage) |   3.768022   .3795468      3.092952    4.590432
                  var(_cons) |   7.392755   .8992881      5.824549    9.383186
-----------------------------+------------------------------------------------
               var(Residual) |   24.20199   .8837314      22.53043    25.99756
------------------------------------------------------------------------------
LR test vs. linear model: chi2(3) = 725.88                Prob > chi2 = 0.0000
```

Outline

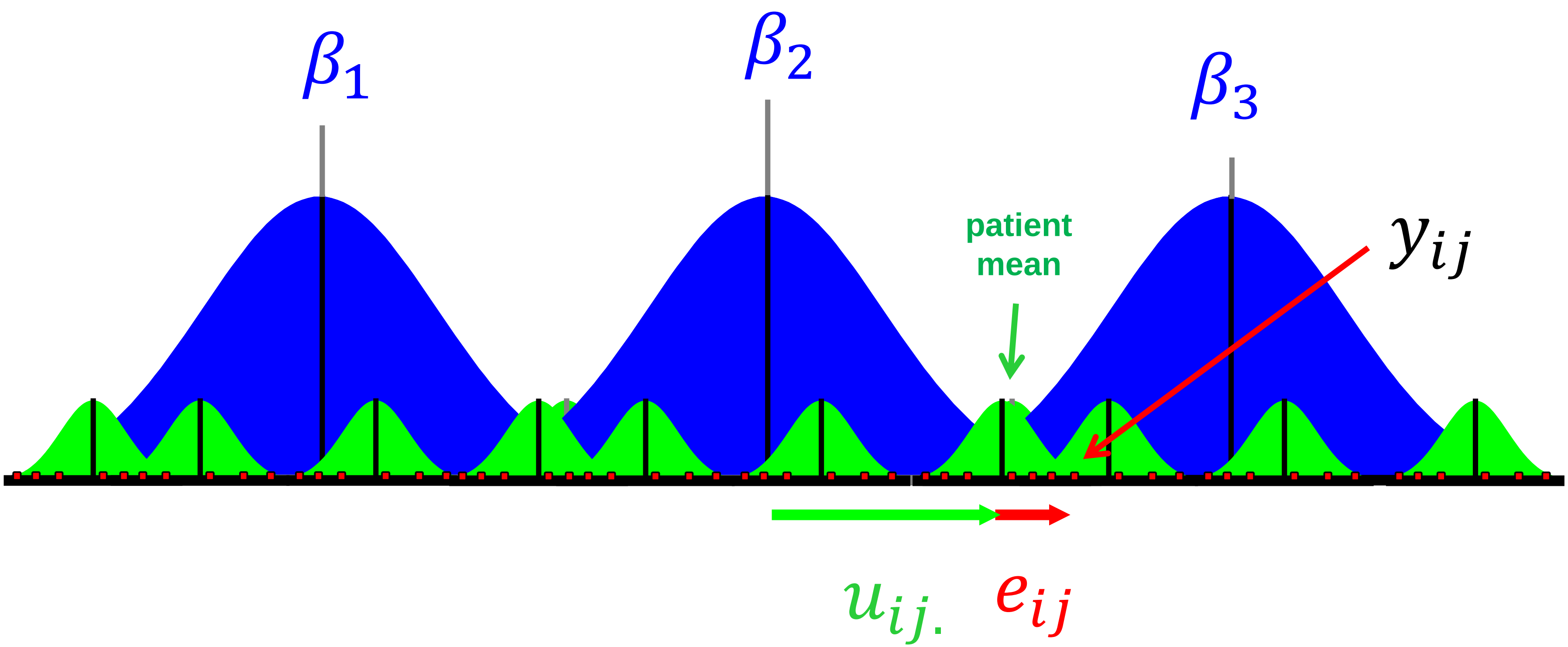
- ✓ The simulated data
- ✓ Single level models
- ✓ Two level models
- ✓ Longitudinal models
- ✓ Three level models
- **Fixed vs random effects**
- Multilevel models for binary data
- Multilevel models for survival data
- Multilevel structural equation models
- Bayesian multilevel models

Fixed versus Random Effects



$$y_{ijk} = \mu + u_{i..} + u_{ij.} + e_{ijk}$$

Data Analysis and Statistical Software



$$y_{ij} = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u_{i.} + e_{ij}$$

$$y_{ij} = \beta_1(0) + \beta_2(1) + \beta_3(0) + u_{i.} + e_{ij}$$

Fixed versus Random Effects

Physician treated as a fixed effect

$$\underbrace{y_{ij}}_{\text{Observed}} = \underbrace{\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3}_{\text{Fixed}} + \underbrace{u_{i.} + e_{ij}}_{\text{Random}}$$

Physician treated as a random effect

$$\underbrace{y_{ijk}}_{\text{Observed}} = \underbrace{\mu}_{\text{Fixed}} + \underbrace{u_{i..} + u_{ij.} + e_{ijk}}_{\text{Random}}$$

Physician Treated as a Random Effect

```
. mixed chol cage || physician: || patient: cage ,  ///
        stddev nolog noheader

------------------------------------------------------------------------------
        chol |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cage |   8.552381   .1242629    68.82   0.000      8.30883    8.795932
       _cons |   250.0533   1.253464   199.49   0.000     247.5966    252.5101
------------------------------------------------------------------------------


------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
physician: Identity          |
                  sd(_cons)  |   2.145466   .8969524      .9454987    4.868358
-----------------------------+------------------------------------------------
patient: Independent         |
                   sd(cage)  |   1.941139   .0977639      1.758679    2.142529
                  sd(_cons)  |   2.755291   .1661248      2.448195    3.100909
-----------------------------+------------------------------------------------
              sd(Residual)   |   4.919552   .0898183      4.746623     5.09878
------------------------------------------------------------------------------
LR test vs. linear model: chi2(3) = 721.05                 Prob > chi2 = 0.0000
```

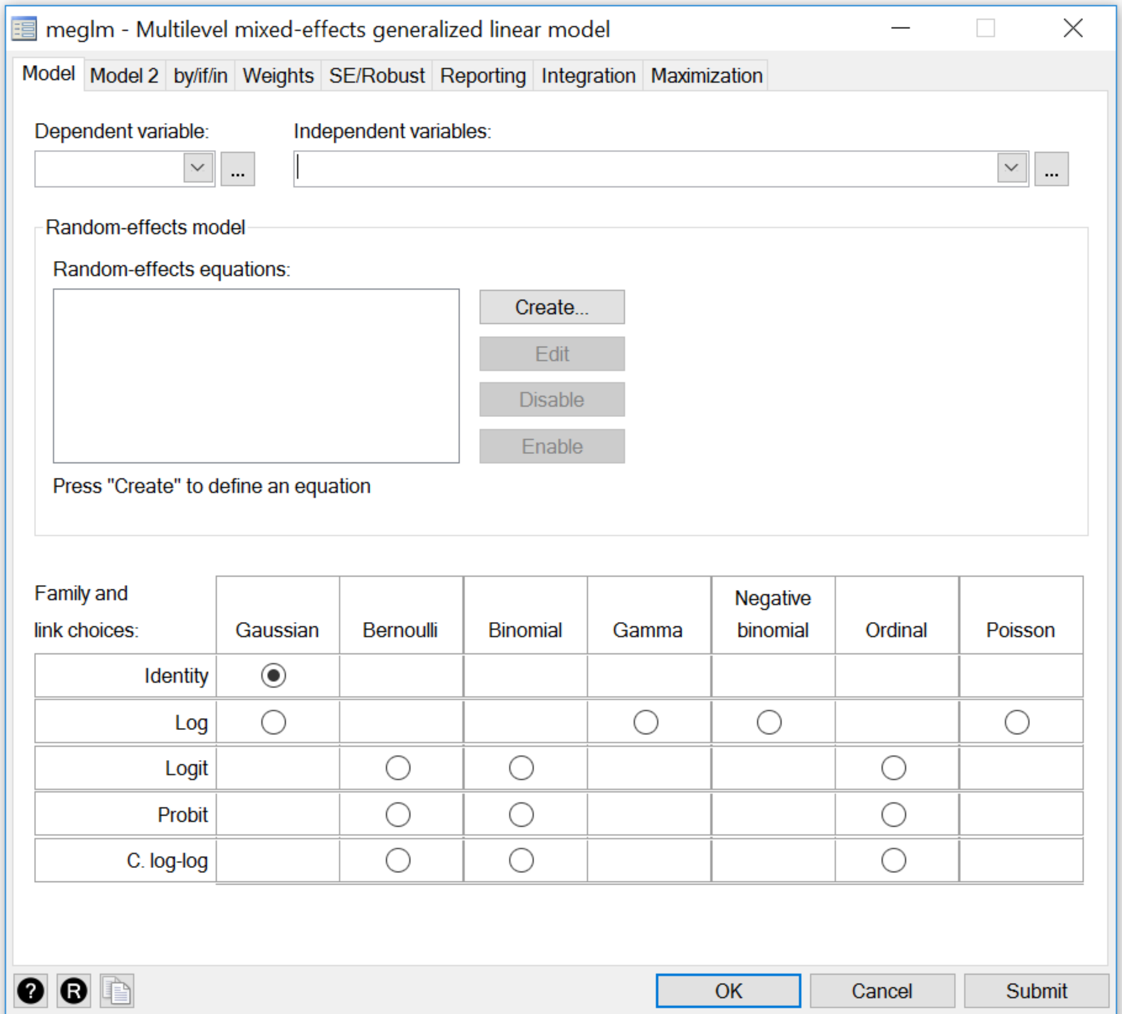
Physician Treated as a Fixed Effect

```
. mixed chol cage ibn.physician, nocons || patient: cage ,    ///
        stddev nolog noheader

------------------------------------------------------------------------------
        chol |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cage |   8.552381   .1242629    68.82   0.000      8.30883    8.795932
             |
   physician |
          1  |   250.4871    .330735   757.37   0.000     249.8389    251.1354
          2  |   252.4686    .330735   763.36   0.000     251.8203    253.1168
          3  |   247.2043    .330735   747.44   0.000     246.5561    247.8525
------------------------------------------------------------------------------

------------------------------------------------------------------------------
  Random-effects Parameters  |   Estimate   Std. Err.     [95% Conf. Interval]
-----------------------------+------------------------------------------------
patient: Independent         |
                    sd(cage) |   1.941139   .0977639      1.758679    2.142529
                   sd(_cons) |   2.735167   .1648911      2.430349    3.078216
-----------------------------+------------------------------------------------
               sd(Residual) |   4.919552   .0898183      4.746624    5.098781
------------------------------------------------------------------------------
LR test vs. linear model: chi2(2) = 533.65                Prob > chi2 = 0.0000
```

Outline

- ✓ The simulated data
- ✓ Single level models
- ✓ Two level models
- ✓ Longitudinal models
- ✓ Three level models
- ✓ Fixed vs random effects
- **Multilevel models for binary data**
- Multilevel models for survival data
- Multilevel structural equation models
- Bayesian multilevel models

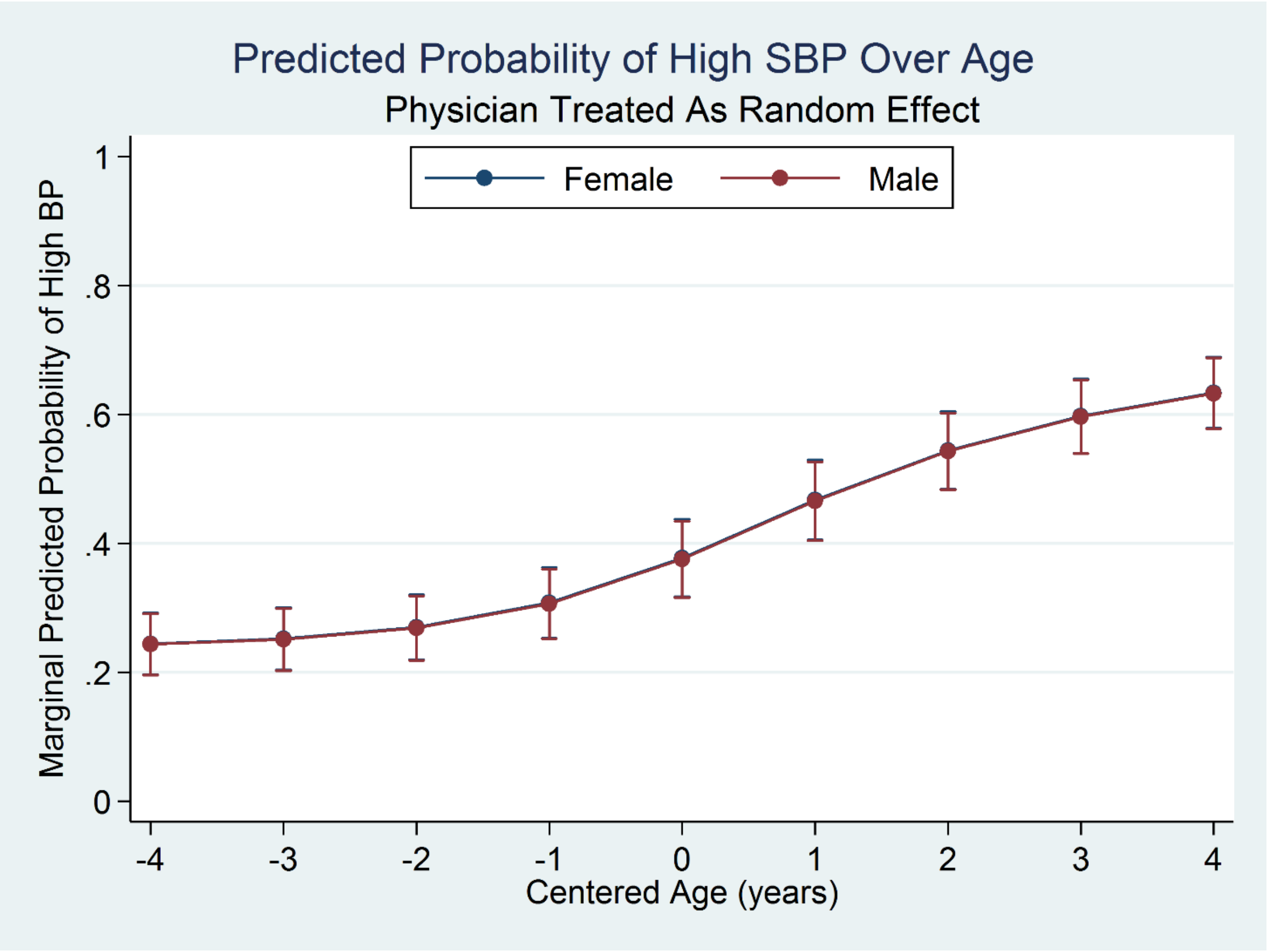
Multilevel Models for GLMs



Multilevel Logistic Regression

```
. melogit HiBP cage i.sex || physician: || patient: cage, or nolog

Mixed-effects logistic regression               Number of obs     =      2,100

Integration method: mvaghermite                 Integration pts.  =          7
                                                Wald chi2(2)      =      66.75
Log likelihood = -1274.1098                     Prob > chi2       =     0.0000
------------------------------------------------------------------------------------
            HiBP | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----------------+------------------------------------------------------------------
            cage |   1.498393   .0741657     8.17   0.000     1.359859     1.65104
                 |
             sex |
           Male  |   .9906147   .1350889    -0.07   0.945     .7582763    1.294142
           _cons |   .5702334   .0833962    -3.84   0.000     .4281195    .7595219
-----------------+------------------------------------------------------------------
physician        |
      var(_cons) |   .0347844   .0397053                      .0037133    .3258408
-----------------+------------------------------------------------------------------
physician>patient |
       var(cage) |   .3894626   .0728996                      .2698596    .5620742
      var(_cons) |   .5086553   .1356728                      .3015671    .8579525
------------------------------------------------------------------------------------
```

Multilevel Logistic Regression

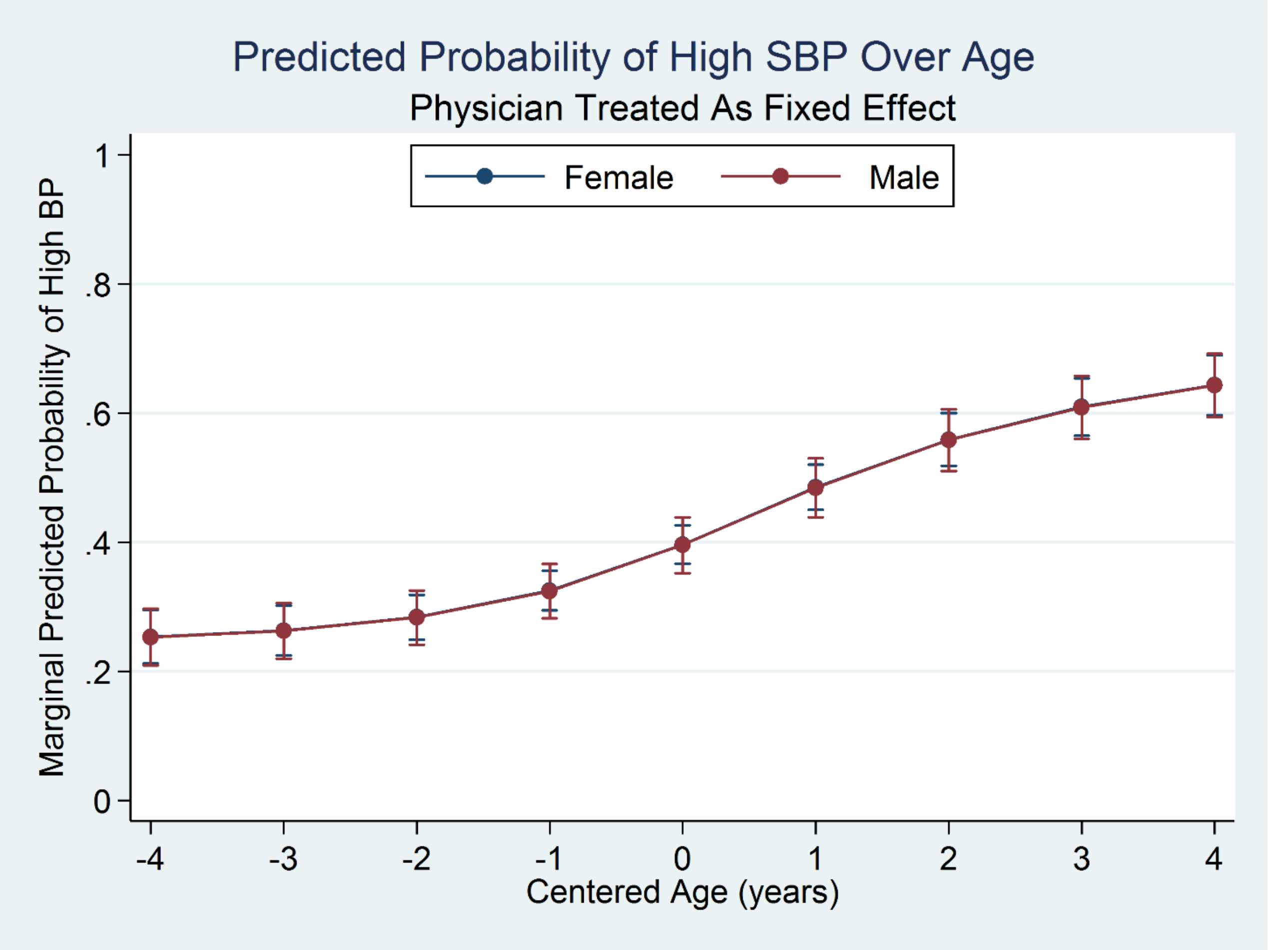



```
margins sex, at(cage=(-4(1)4))
marginsplot
```

Multilevel Logistic Regression

```
. melogit HiBP cage i.sex ibn.physician, nocons || patient: cage, or nolog

Integration method: mvaghermite                  Integration pts.   =          7
                                                 Wald chi2(5)       =     127.54
Log likelihood = -1270.7301                      Prob > chi2        =     0.0000
------------------------------------------------------------------------------
        HiBP | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        cage |   1.498201   .0741319     8.17   0.000     1.359728    1.650775
             |
         sex |
       Male  |   .9945755   .1349149    -0.04   0.968     .7623803    1.297489
             |
   physician |
          1  |   .7676037   .1029827    -1.97   0.049     .5901179    .9984708
          2  |   .4559589   .0636761    -5.62   0.000      .346779     .599513
          3  |   .5271314   .0731874    -4.61   0.000     .4015478    .6919912
             |
       _cons |          1  (omitted)
-------------+----------------------------------------------------------------
patient      |
   var(cage)|   .3891293    .072832                      .2696357    .5615786
  var(_cons)|   .4948104   .1335798                      .2915056    .8399058
------------------------------------------------------------------------------
```

Multilevel Logistic Regression



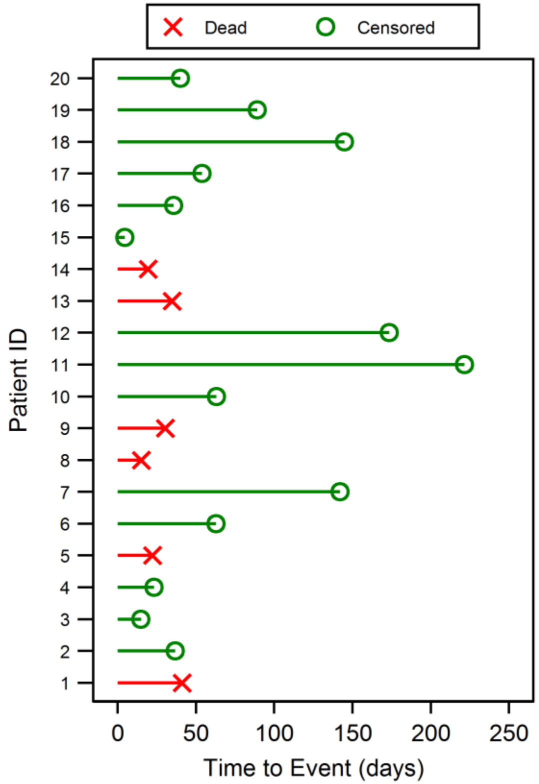

```
margins sex, at(cage=(-4(1)4))
marginsplot
```

Outline

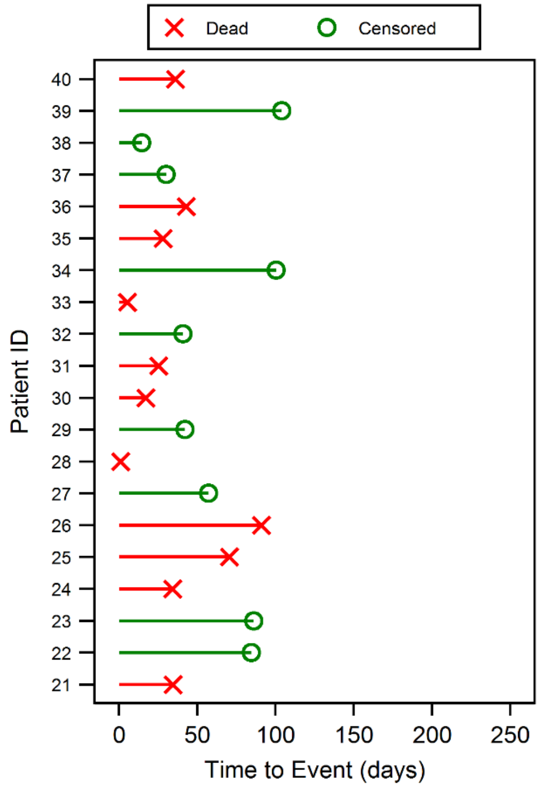
- ✓ The simulated data
- ✓ Single level models
- ✓ Two level models
- ✓ Longitudinal models
- ✓ Three level models
- ✓ Fixed vs random effects
- ✓ Multilevel models for binary data
- **Multilevel models for survival data**
- Multilevel structural equation models
- Bayesian multilevel models

Multilevel Survival Data

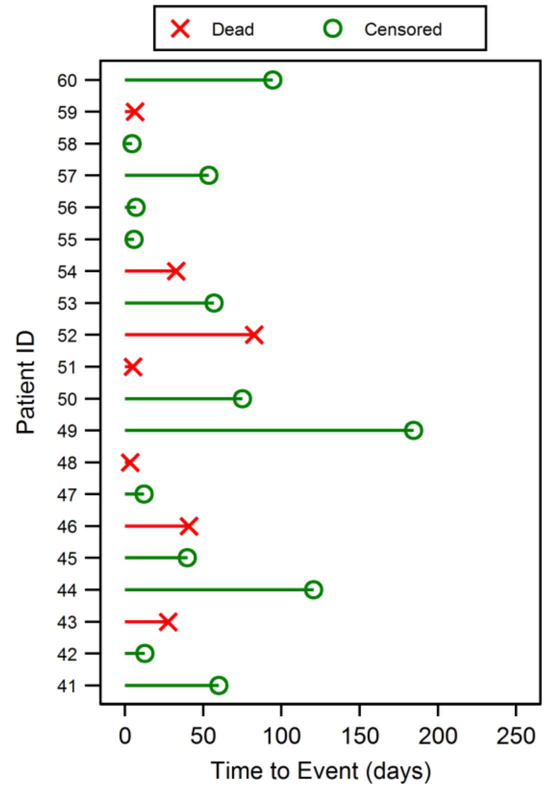
Physician 1



Physician 2



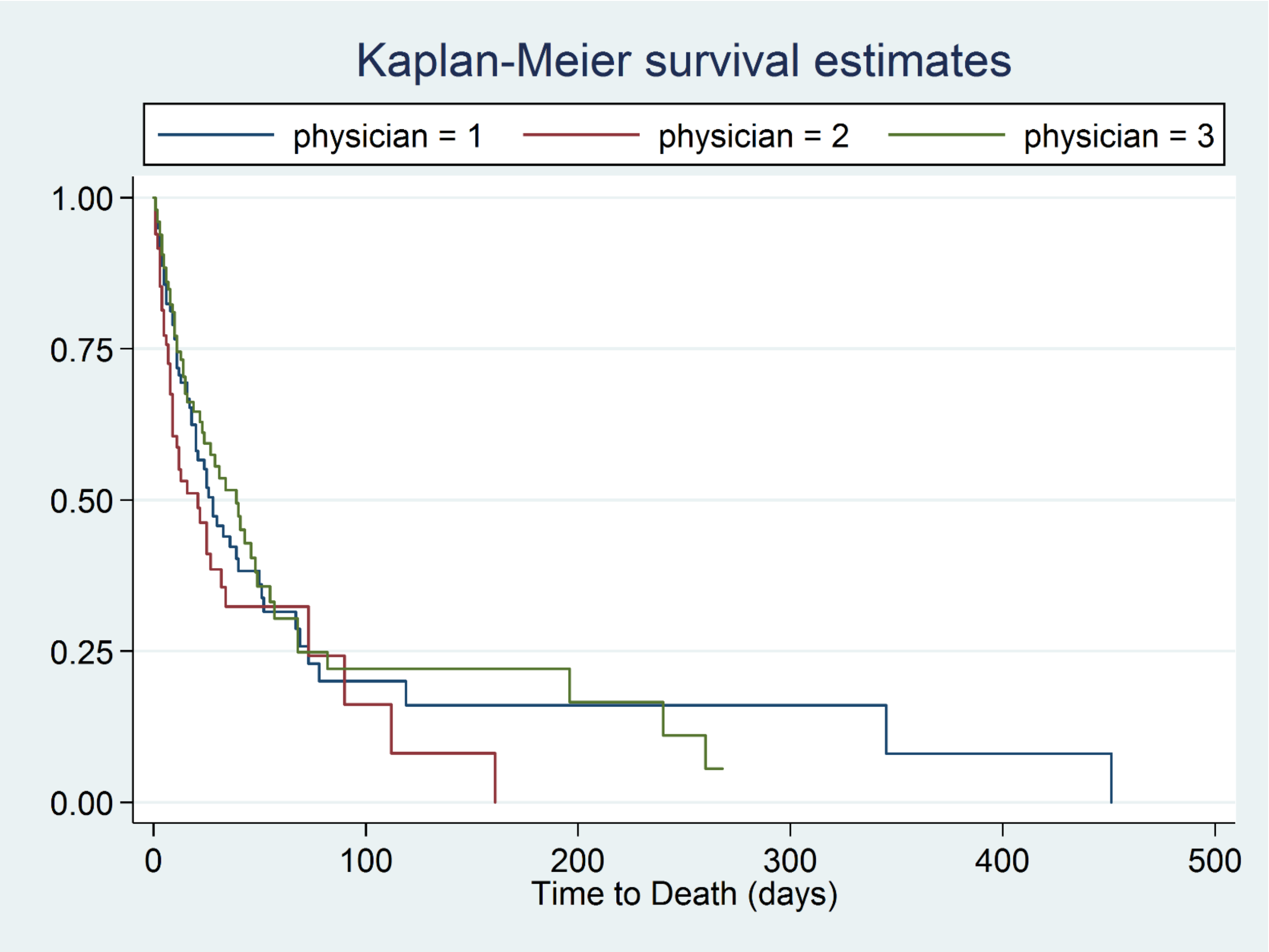
Physician 3



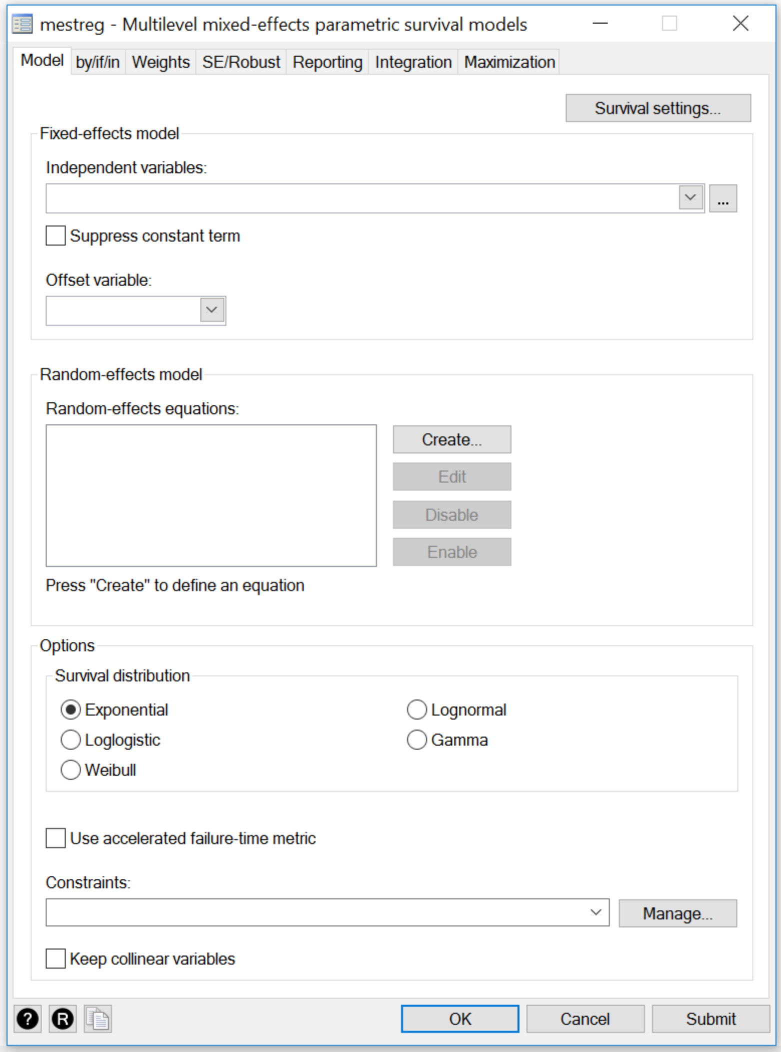
Multilevel Survival Data

```
. use SecondMI.dta, clear

. keep if age_mi==age
(1,800 observations deleted)

. stset time, failure(dead)

     failure event:  dead != 0 & dead < .
obs. time interval:  (0, time]
 exit on or before:  failure

------------------------------------------------------------------------------
        300  total observations
          0  exclusions
------------------------------------------------------------------------------
        300  observations remaining, representing
        154  failures in single-record/single-failure data
       8070  total analysis time at risk and under observation
                                                at risk from t =         0
                                     earliest observed entry t =         0
                                          last observed exit t =       451
```

Multilevel Survival Data



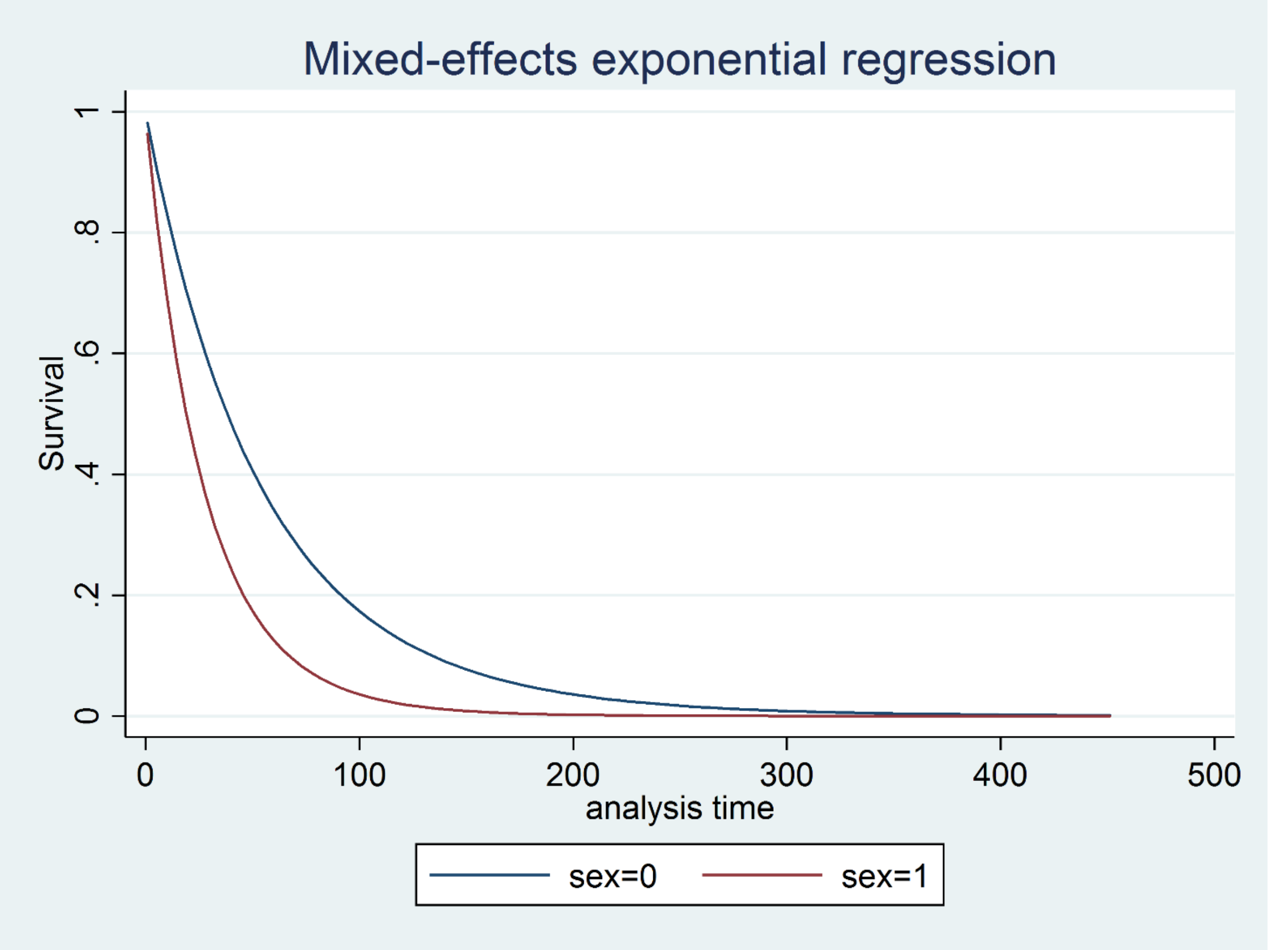
Multilevel Survival Data



Multilevel Survival Data

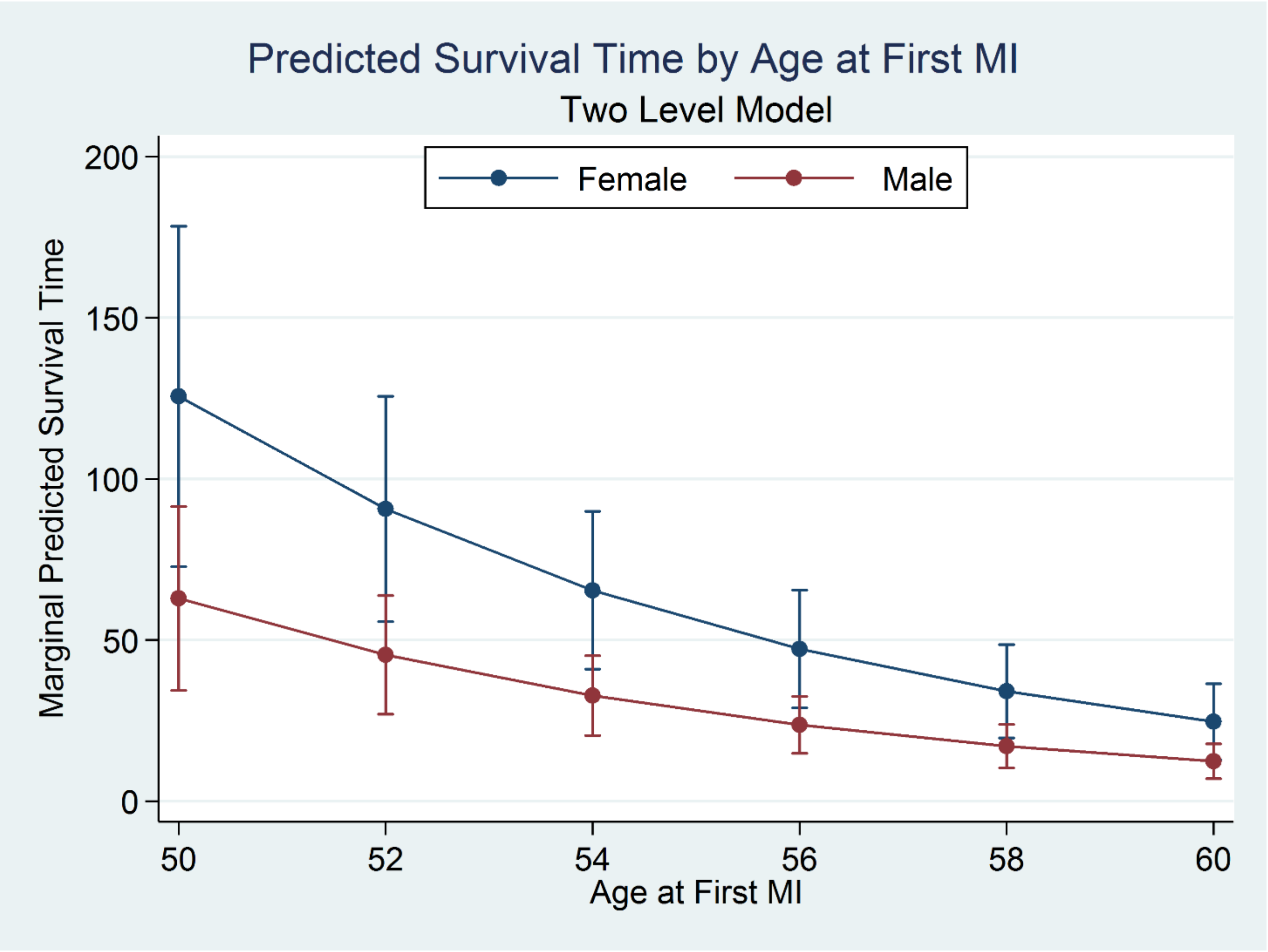
```
. mestreg age_mi i.sex || physician:, dist(exponential) nolog

         failure _d:  dead
   analysis time _t:  time

Mixed-effects exponential regression            Number of obs     =        300
Group variable:       physician                 Number of groups  =          3

                                                Obs per group:
                                                              min =        100
                                                              avg =      100.0
                                                              max =        100

Integration method: mvaghermite                 Integration pts.  =          7

                                                Wald chi2(2)      =      74.37
Log likelihood = -727.51036                     Prob > chi2       =     0.0000
------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      age_mi |   1.177106   .0296204     6.48   0.000      1.12046    1.236616
             |
         sex |
        Male |   1.997485   .3299564     4.19   0.000      1.44503     2.76115
       _cons |   2.37e-06   3.25e-06    -9.45   0.000     1.62e-07    .0000347
-------------+----------------------------------------------------------------
physician    |
   var(_cons)|   .0664489   .0725349                      .0078222    .5644816
------------------------------------------------------------------------------
LR test vs. exponential model: chibar2(01) = 4.30     Prob >= chibar2 = 0.0190
```

Multilevel Survival Data




```
stcurve, survival marginal at1(sex=0) at2(sex=1)
```

Multilevel Survival Data

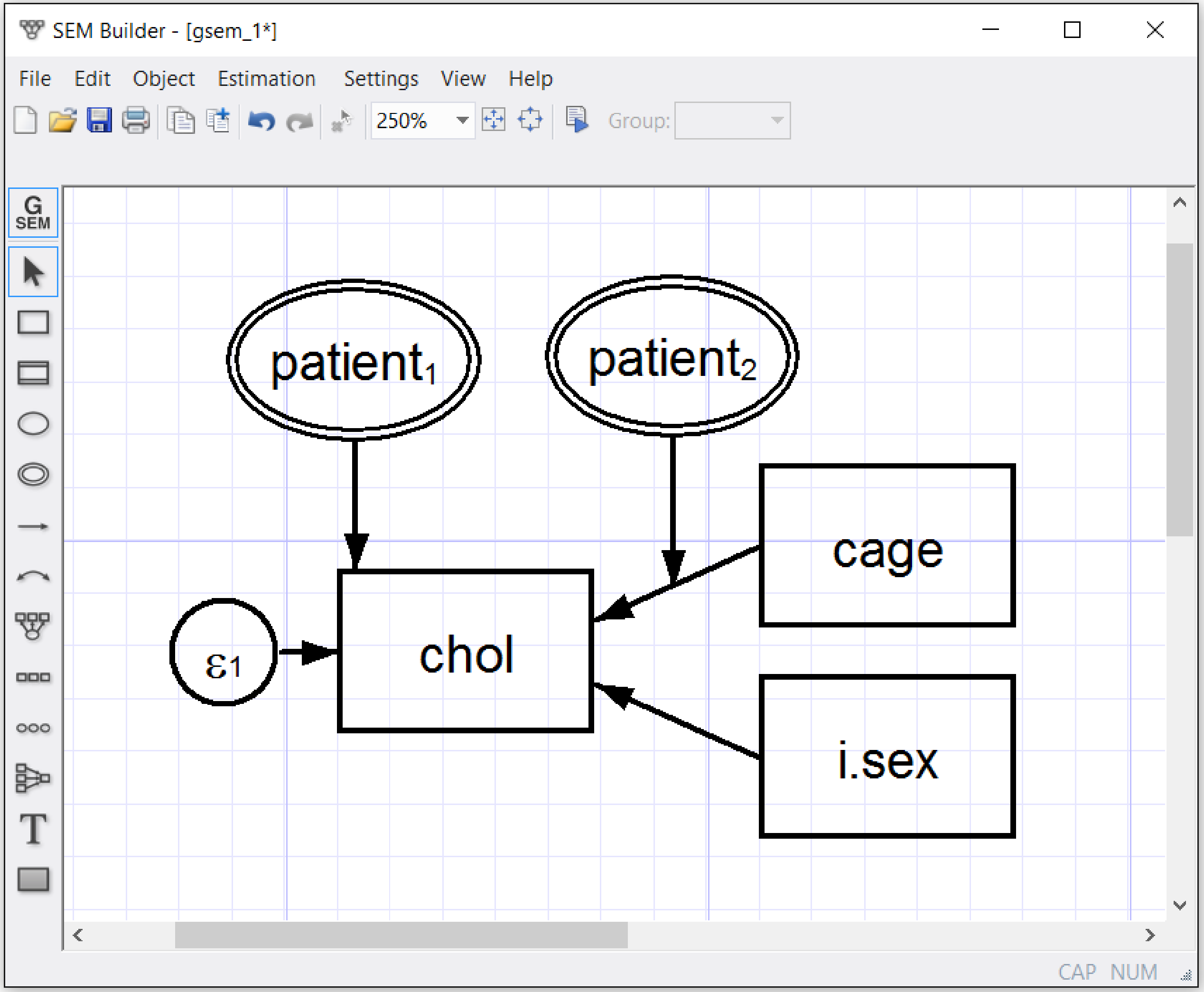



```
margins sex, at(age_mi=(50(2)60))
marginsplot
```

Outline

- The simulated data
- Single level models
- Two level models
- Longitudinal models
- Three level models
- Fixed vs random effects
- Multilevel models for binary data
- Multilevel models for survival data
- **Multilevel structural equation models**
- Bayesian multilevel models

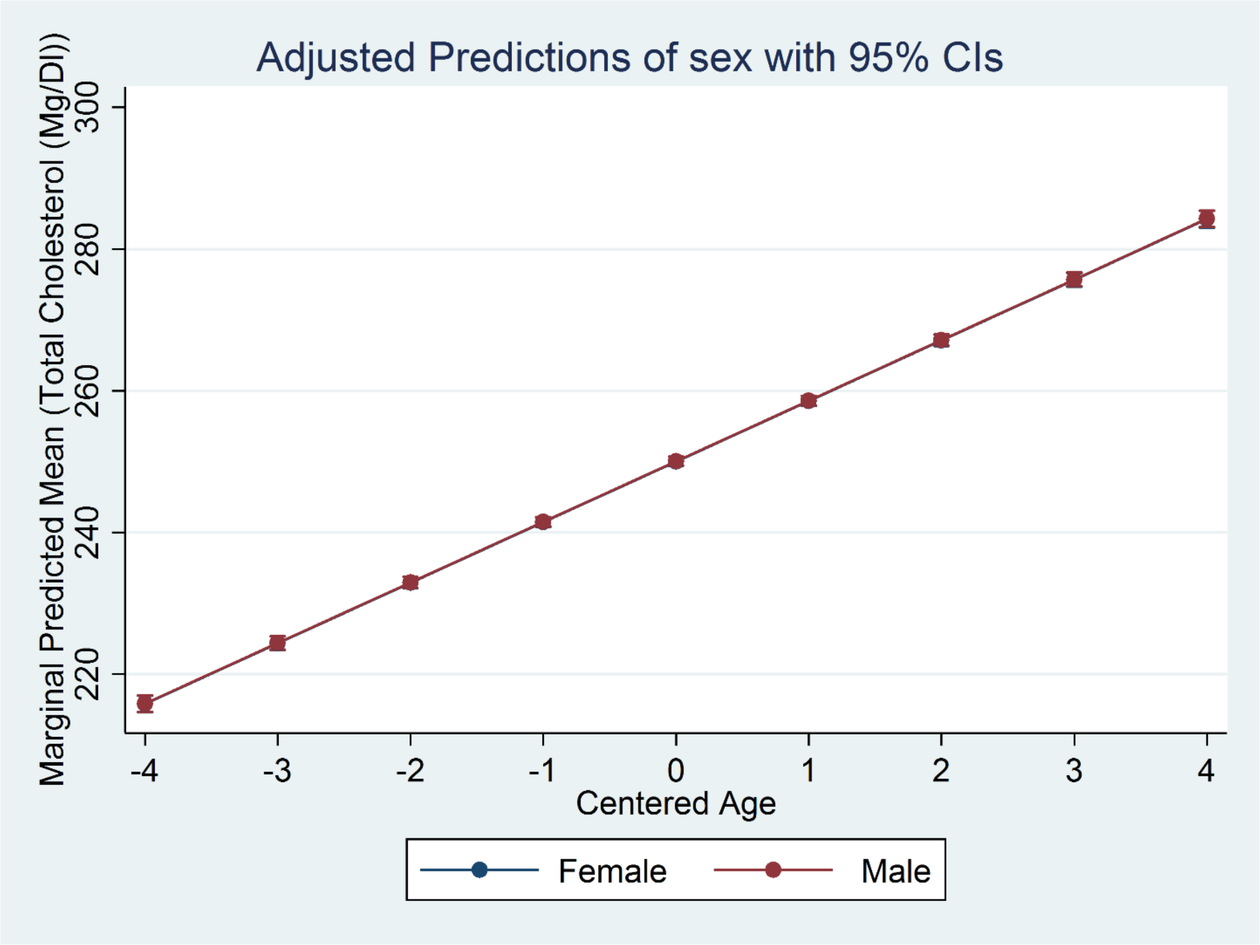
Multilevel Structural Equation Models



Multilevel Structural Equation Models

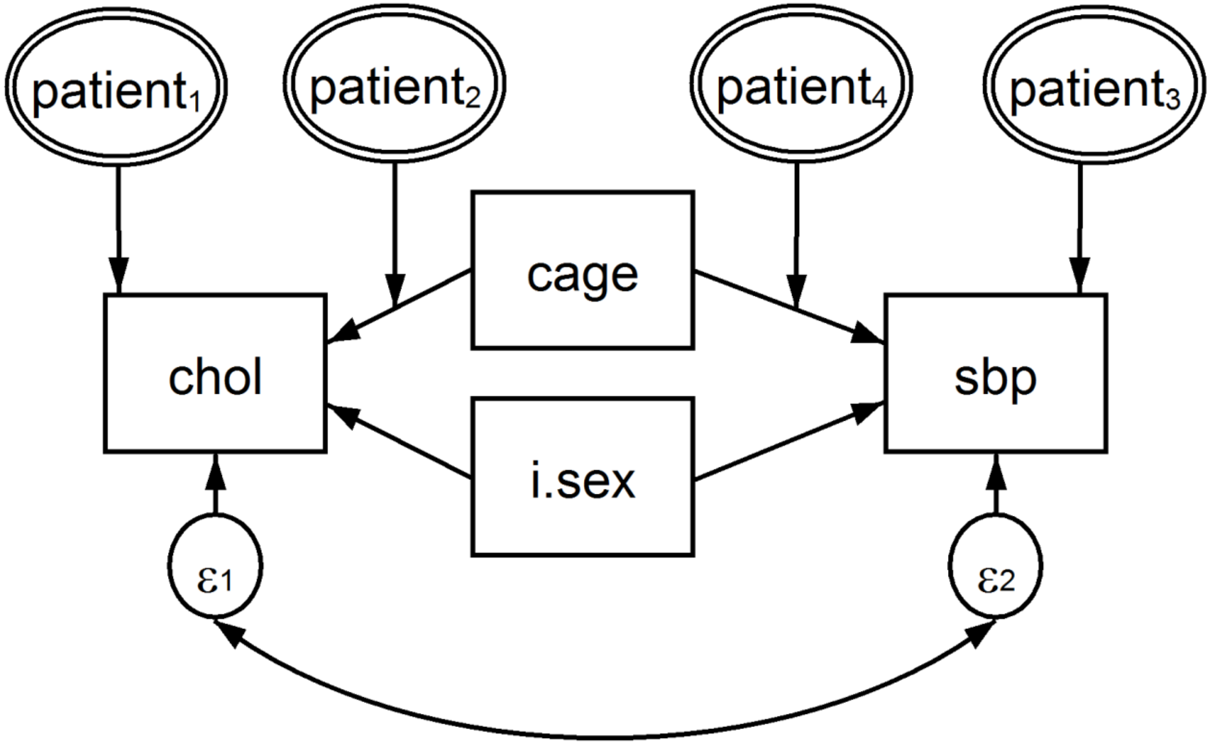
```
. gsem (chol <- cage i.sex M1[patient] M2[patient]#c.cage),   ///
>       covstruct(_lexogenous, diagonal)                       ///
>       latent(M1 M2) nolog

Generalized structural equation model          Number of obs     =      2,100
Response         : chol
Family           : Gaussian
Link             : identity
Log likelihood = -6803.8521

 ( 1)  [chol]M1[patient] = 1
 ( 2)  [chol]c.cage#M2[patient] = 1

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
chol <-      |
        cage |   8.552381   .1242628    68.82   0.000      8.30883    8.795932
             |
         sex |
       Male  |   .0555538   .4570642     0.12   0.903    -.8402757    .9513833
             |
 M1[patient] |          1  (constrained)
             |
     c.cage# |
 M2[patient] |          1  (constrained)
             |
       _cons |   250.0246   .3285356   761.03   0.000     249.3807    250.6685
-------------+----------------------------------------------------------------
var(M1[pati~t])| 12.19324   1.284093                      9.919221    14.98859
var(M2[pati~t])| 3.768018   .3795464                      3.092949    4.590428
-------------+----------------------------------------------------------------
var(e.chol)  |     24.202   .8837321                      22.53045    25.99757
------------------------------------------------------------------------------
```

Multilevel Structural Equation Models



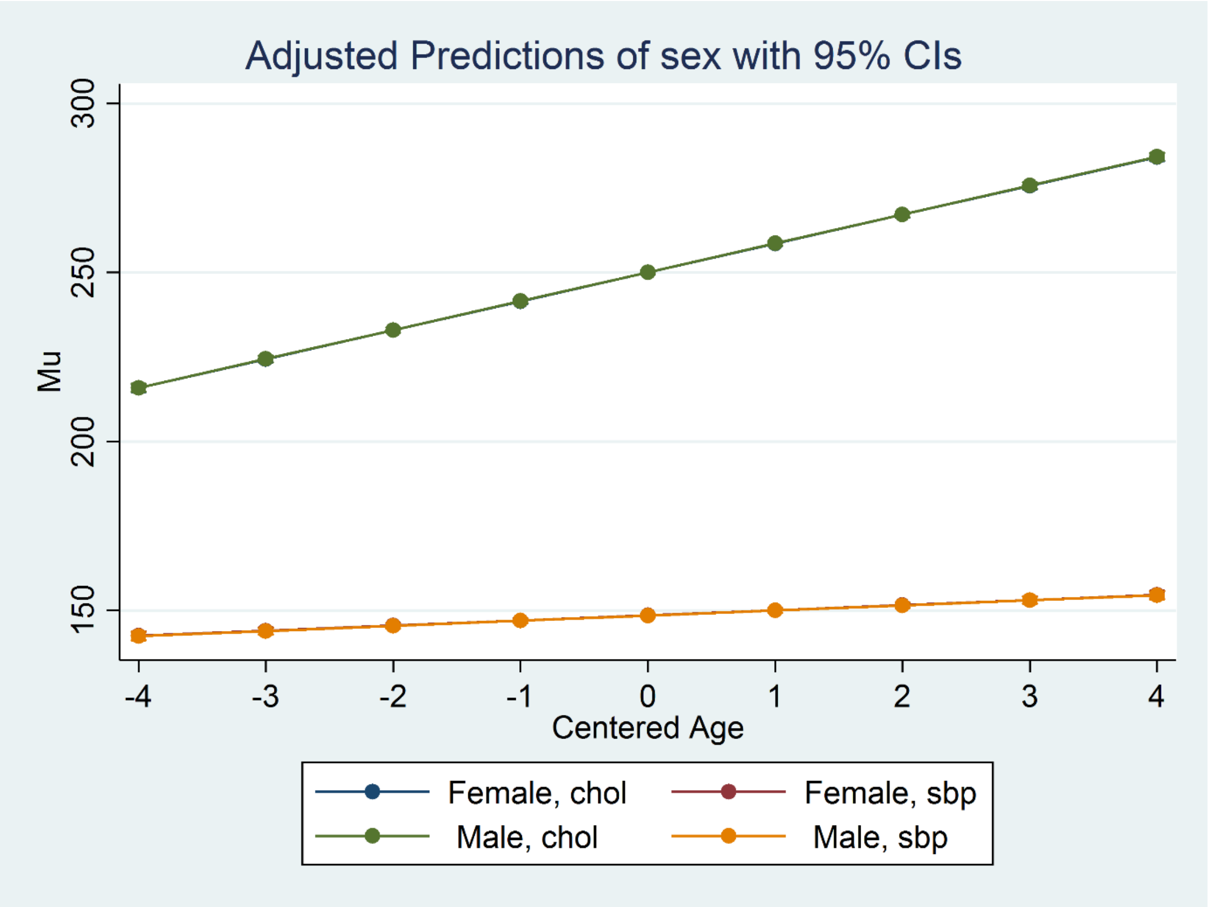

```
margins sex, at(age_mi=(50(2)60))
marginsplot
```

Multilevel Structural Equation Models




```
gsem    (chol <- cage i.sex M1[patient] M2[patient]#c.cage)         ///
        (sbp  <- cage i.sex M3[patient] M4[patient]#c.cage),        ///
        cov(e.sbp*e.chol)                                           ///
        covstruct(_lexogenous, diagonal)                            ///
        latent(M1 M2 M3 M4)
```

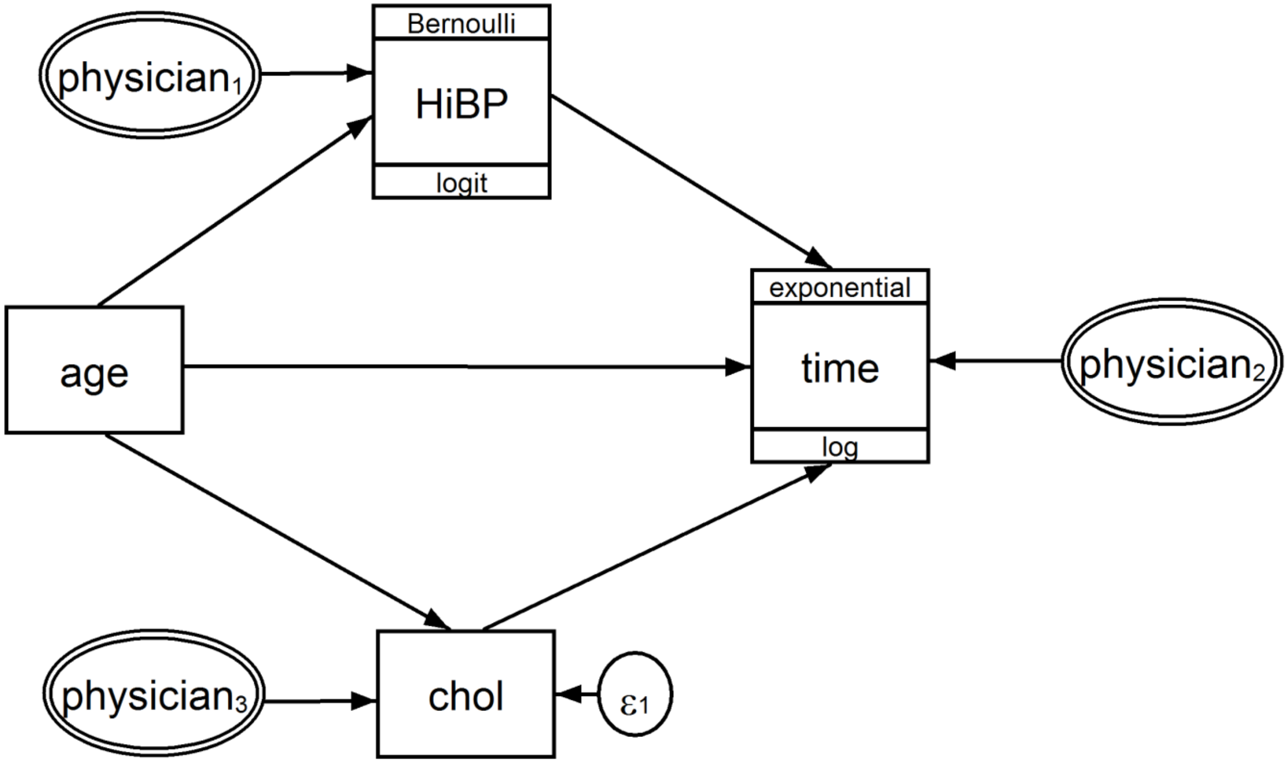
Multilevel Structural Equation Models




```
margins sex, at(age_mi=(50(2)60))
marginsplot
```

Data Analysis and Statistical Software

Multilevel Structural Equation Models




```
gsem (time <- age chol HiBP M2[physician], family(exponential, failure(dead) ph) link(log)) ///
     (HiBP <- age M1[physician], family(bernoulli) link(logit))                            ///
     (chol <- age M3[physician])                                                           ///
     , covstruct(_lexogenous, diagonal)                                                    ///
     latent(M1 M2 M3) nolog
```

Stata Data Analysis and Statistical Software

Multilevel Structural Equation Models

```
Generalized structural equation model          Number of obs     =        300

Response       : time                          No. of failures   =        154
Family         : exponential                   Time at risk      =       8070
Form           : proportional hazards
Link           : log

Response       : HiBP
Family         : Bernoulli
Link           : logit

Response       : chol
Family         : Gaussian
Link           : identity

Log likelihood = -2006.1937

 ( 1)  [time]M2[physician] = 1
 ( 2)  [HiBP]M1[physician] = 1
 ( 3)  [chol]M3[physician] = 1
------------------------------------------------------------------------------------
                   |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------------+----------------------------------------------------------------
time <-            |
               age |   .1847911   .0254149     7.27   0.000     .1349787    .2346034
              chol |   .0041927   .0102822     0.41   0.683      -.01596    .0243455
              HiBP |   .2141293   .1718165     1.25   0.213    -.1226248    .5508834
                   |
    M2[physician]  |          1  (constrained)
                   |
             _cons |  -15.11604   2.963422    -5.10   0.000    -20.92424   -9.307841
-------------------+----------------------------------------------------------------
HiBP <-            |
               age |   .0371905   .0359236     1.04   0.301    -.0332184    .1075993
                   |
    M1[physician]  |          1  (constrained)
                   |
             _cons |  -1.794614   1.970548    -0.91   0.362    -5.656816    2.067589
-------------------+----------------------------------------------------------------
chol <-            |
               age |   .3163586   .1506689     2.10   0.036     .0210531    .6116641
                   |
    M3[physician]  |          1  (constrained)
                   |
             _cons |   258.3299   8.300959    31.12   0.000     242.0603    274.5995
-------------------+----------------------------------------------------------------
var(M2[physician]) |   .0865751   .0902101                      .0112319    .6673191
var(M1[physician]) |   .0095812   .0414325                      2.00e-06    45.95349
var(M3[physician]) |   1.645804   1.935408                      .1642083    16.49534
-------------------+----------------------------------------------------------------
        var(e.chol)|   70.42731   5.779812                      59.96321    82.71748
------------------------------------------------------------------------------------
```

Multilevel Structural Equation Models

```
. estat eform time HiBP
```

	exp(b)	Std. Err.	z	P>|z|	[95% Conf.	Interval]
time						
age	1.202967	.0305733	7.27	0.000	1.144512	1.264407
chol	1.004202	.0103254	0.41	0.683	.9841666	1.024644
HiBP	1.238783	.2128433	1.25	0.213	.8845955	1.734785
M2[physician]	2.718282	.	.	.	.	.
_cons	2.72e-07	8.07e-07	-5.10	0.000	8.18e-10	.0000907
HiBP						
age	1.037891	.0372847	1.04	0.301	.9673273	1.113601
M1[physician]	2.718282	.	.	.	.	.
_cons	.1661917	.3274886	-0.91	0.362	.0034936	7.905739

Outline

- ✓ The simulated data
- ✓ Single level models
- ✓ Two level models
- ✓ Longitudinal models
- ✓ Three level models
- ✓ Fixed vs random effects
- ✓ Multilevel models for binary data
- ✓ Multilevel models for survival data
- ✓ Multilevel structural equation models
- **Bayesian multilevel models**

Bayesian Multilevel Models

```
bayesmh chol i.patient i.patient#c.cage,
        likelihood(normal({var_0})) noconstant
        prior({chol:i.patient}, normal({chol:_cons},{var_patient}))
        prior({chol:i.patient#c.cage}, normal({chol:cage},{var_cage}))
        prior({chol:_cons}, normal(0, 100))
        prior({chol:cage}, normal(0, 100))
        prior({var_0}, igamma(0.001, 0.001))
        prior({var_patient}, igamma(0.001, 0.001))
        prior({var_cage}, igamma(0.001, 0.001))
        block({var_0}, gibbs)
        block({var_patient}, gibbs)
        block({var_cage}, gibbs)
        block({chol:i.patient}, gibbs)
        block({chol:i.patient#c.cage}, gibbs)
        block({chol:cage}, gibbs)
        block({chol:_cons}, gibbs)
        burnin(5000) mcmcsize(10000) thinning(1) rseed(14)
        dots notable
```

Bayesian Multilevel Models

Using the new Bayes prefix in Stata 15:

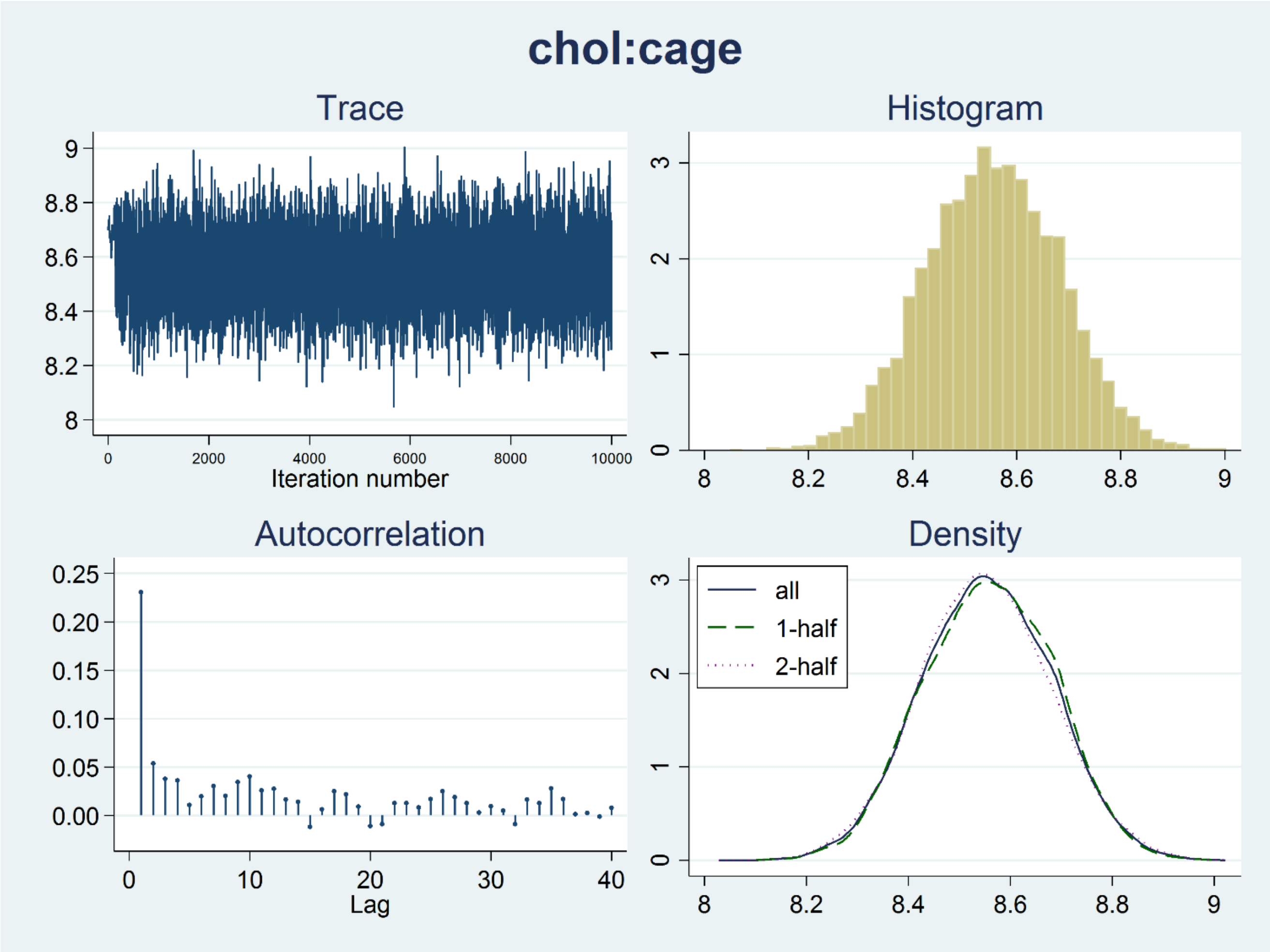
```
bayes: mixed chol cage || physician: || patient: age
```

Bayesian Multilevel Models

```
. bayesstats summary {chol:cage _cons} {var_0} {var_patient} {var_cage}

Posterior summary statistics                      MCMC sample size =     10,000
```

	Mean	Std. Dev.	MCSE	Median	Equal-tailed [95% Cred.	Interval]
chol						
cage	8.554648	.1270552	.001871	8.554884	8.309157	8.800845
_cons	249.4838	4.138322	.439259	249.9238	249.3812	250.3708
var_0	41.98637	186.7146	17.6519	24.24116	22.58511	26.41207
var_patient	12.16671	1.780897	.148846	12.23375	9.609332	15.09232
var_cage	3.74957	.5837284	.053908	3.777186	3.01292	4.610533

Bayesian Multilevel Models



Outline

- ✓ The simulated data
- ✓ Single level models
- ✓ Two level models
- ✓ Longitudinal models
- ✓ Three level models
- ✓ Fixed vs random effects
- ✓ Multilevel models for binary data
- ✓ Multilevel models for survival data
- ✓ Multilevel structural equation models
- ✓ Bayesian multilevel models

For more information

- Videos
  - Introduction to Multilevel Models Part 1
  - Introduction to Multilevel Models Part 2
- Blogs
  - Introduction to Multilevel Models Part 1
  - Introduction to Multilevel Models Part 2
  - How to simulate multilevel/longitudinal data

Questions?

chuber@stata.com