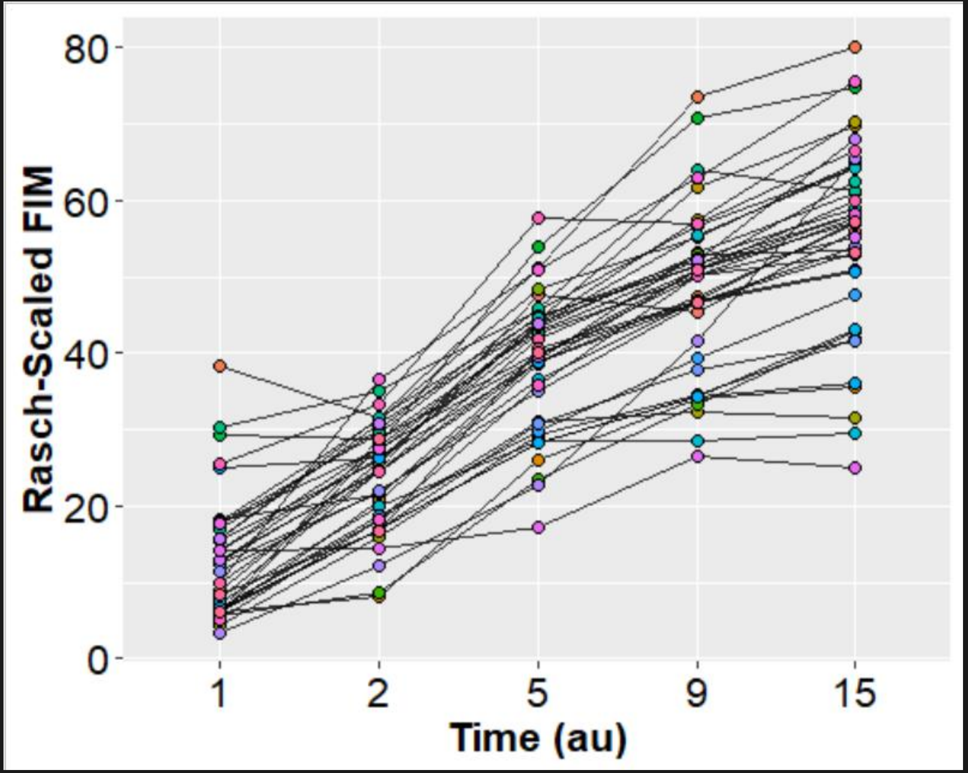
Mixed-Effects Models for Longitudinal Study Designs

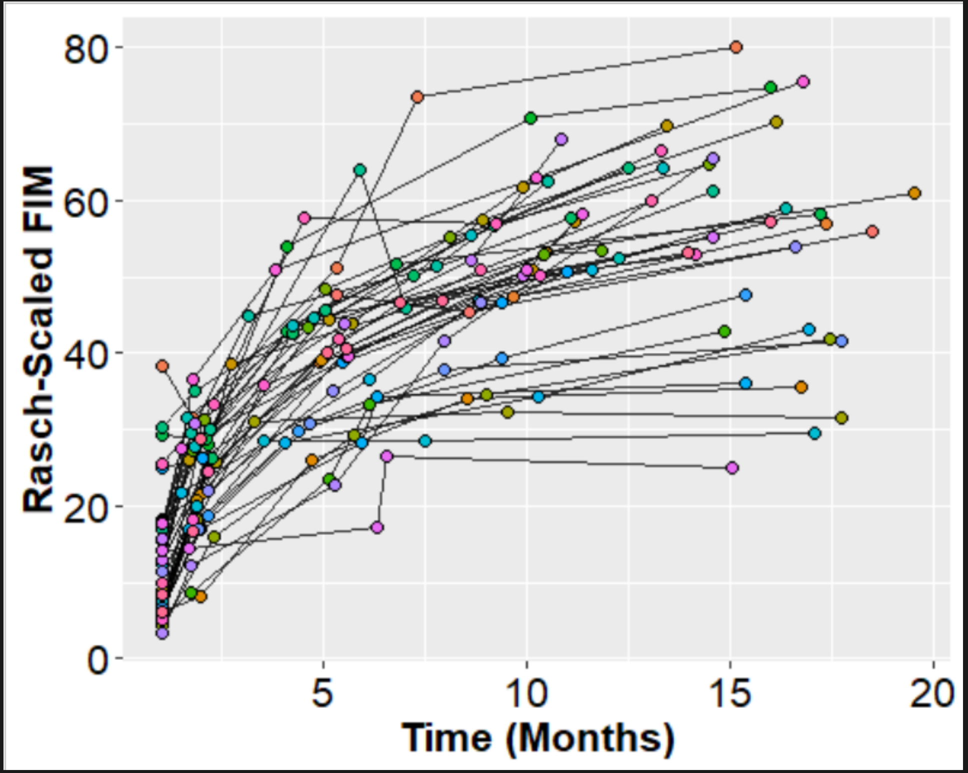
Keith Lohse | PhD | Pstat | (he/him) |
Associate Professor | Physical Therapy and Neurology |
Washington University School of Medicine in Saint Louis |
lohse@wustl.edu
keithlohse

Longitudinal Data

- RM ANOVA



- Mixed-Effect Regression



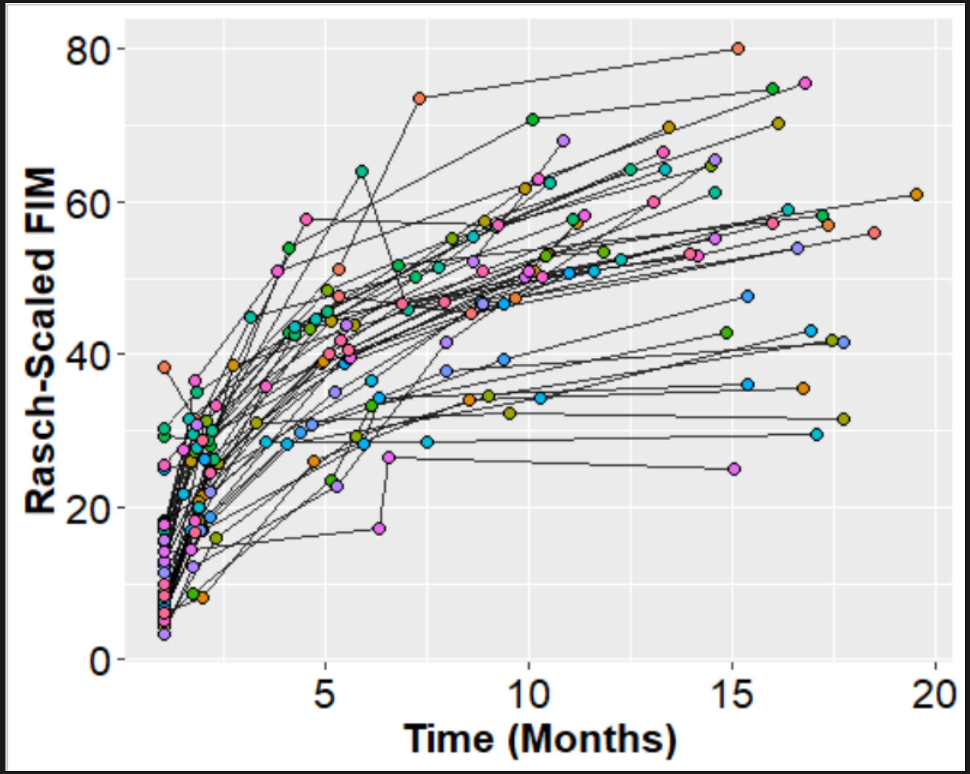
Longitudinal Data

- **RM ANOVA** assumes that there are categorical differences between the repeated measures, and the focus is on mean differences between conditions.

  - While this is often a weakness, it can be a strength when the proper “shape” of the time curve is unclear/complicated.

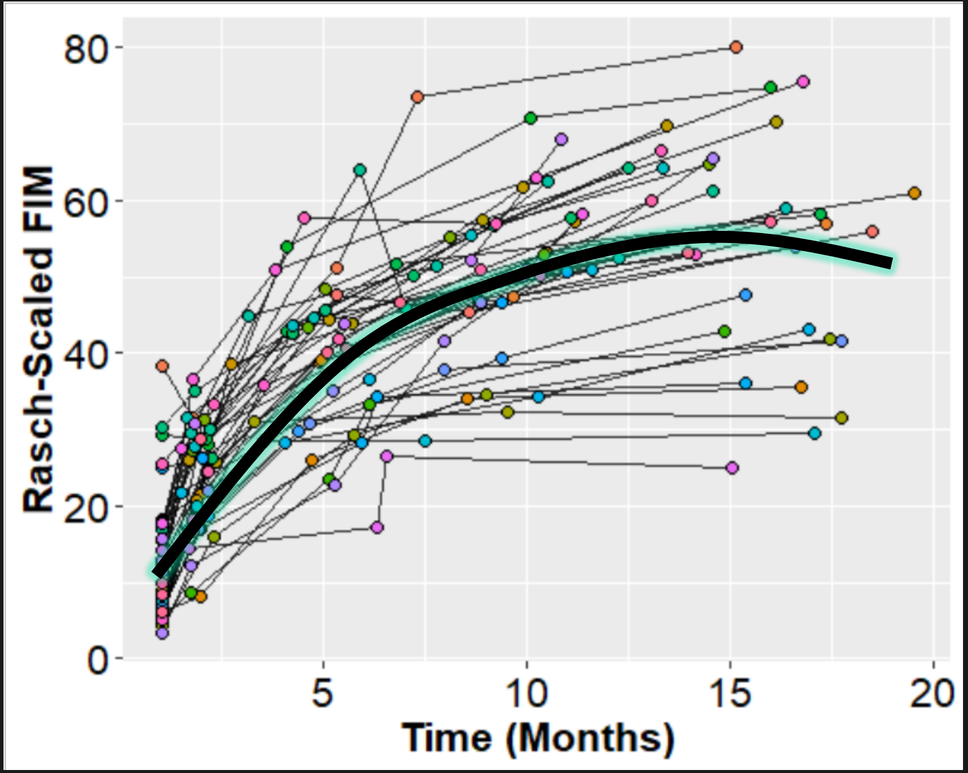
  - E.g., if you have 5 levels of time, a categorical Time factor would use 4 degrees of freedom, equivalent to a 4^{th} order polynomial.

Longitudinal Data



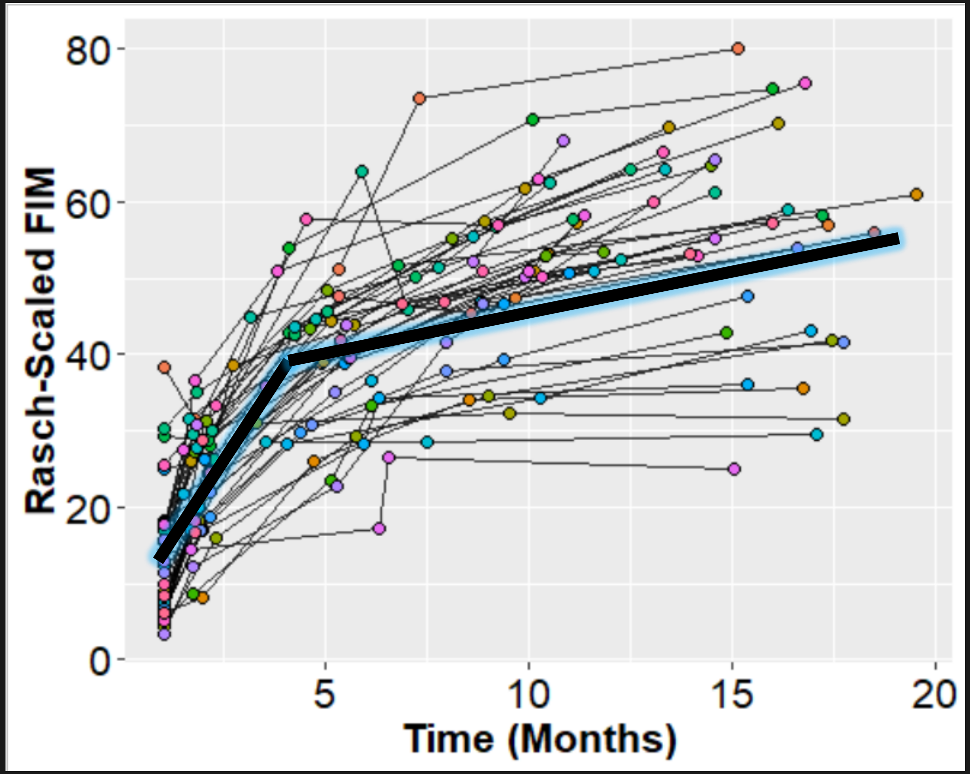
- In many cases, however, the **mixed-effects model** is going to be preferable, because there is variability in time and/or there is a more parsimonious shape to the time curve.

Longitudinal Data



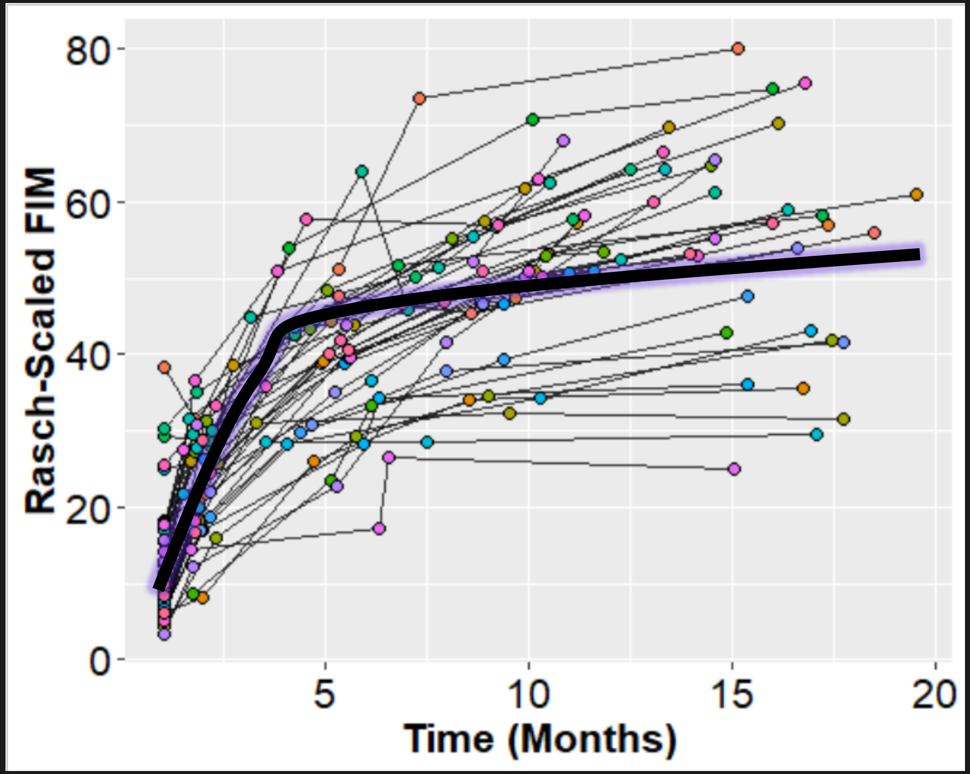
- In many cases, however, the **mixed-effects model** is going to be preferable, because there is variability in time and/or there is a more parsimonious shape to the time curve.
  - **Polynomials**

Longitudinal Data



- In many cases, however, the **mixed-effects model** is going to be preferable, because there is variability in time and/or there is a more parsimonious shape to the time curve.
    - **Polynomials**
    - **Splines**

Longitudinal Data

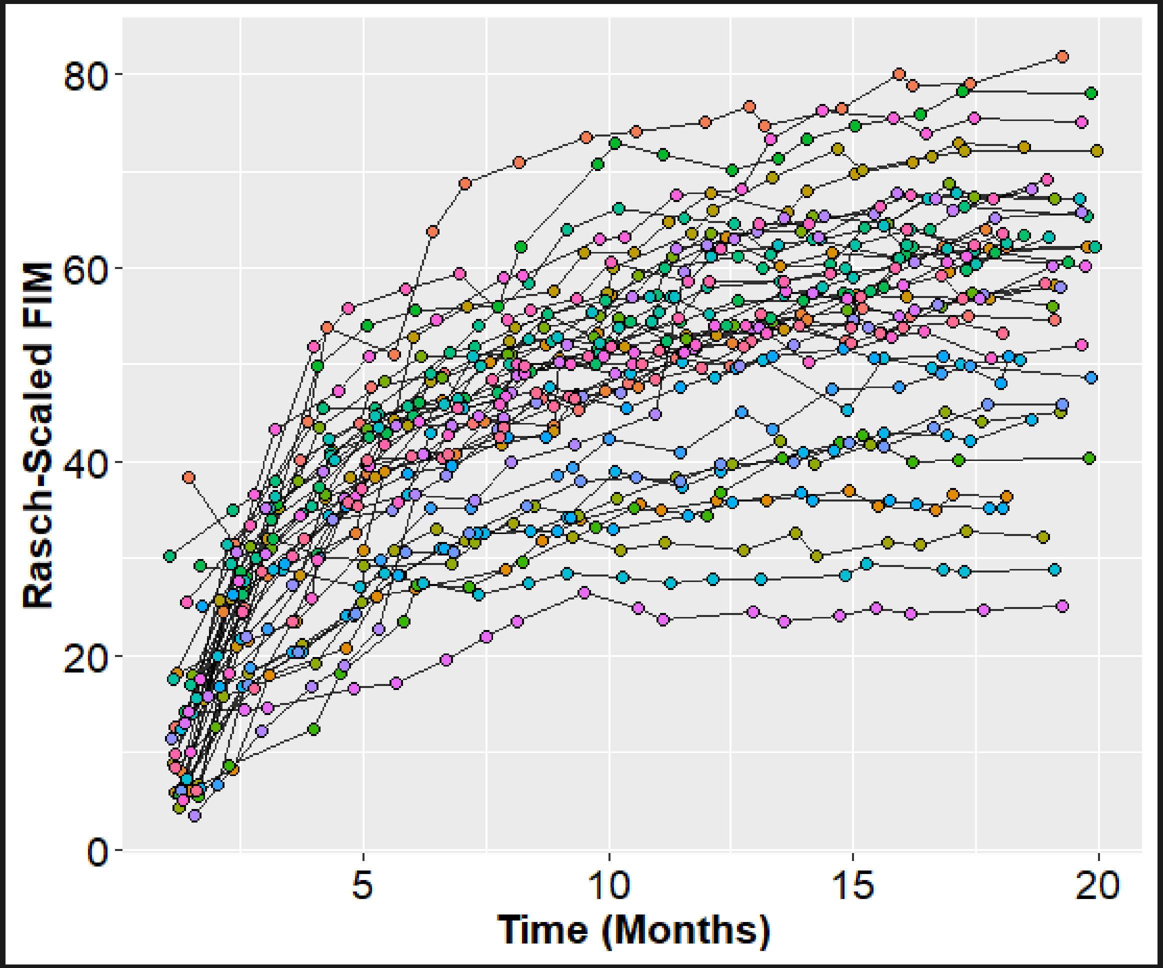


- In many cases, however, the **mixed-effects model** is going to be preferable, because there is variability in time and/or there is a more parsimonious shape to the time curve.
  - **Polynomials**
  - **Splines**
  - **Nonlinear functions**

Steps to Building Longitudinal Models

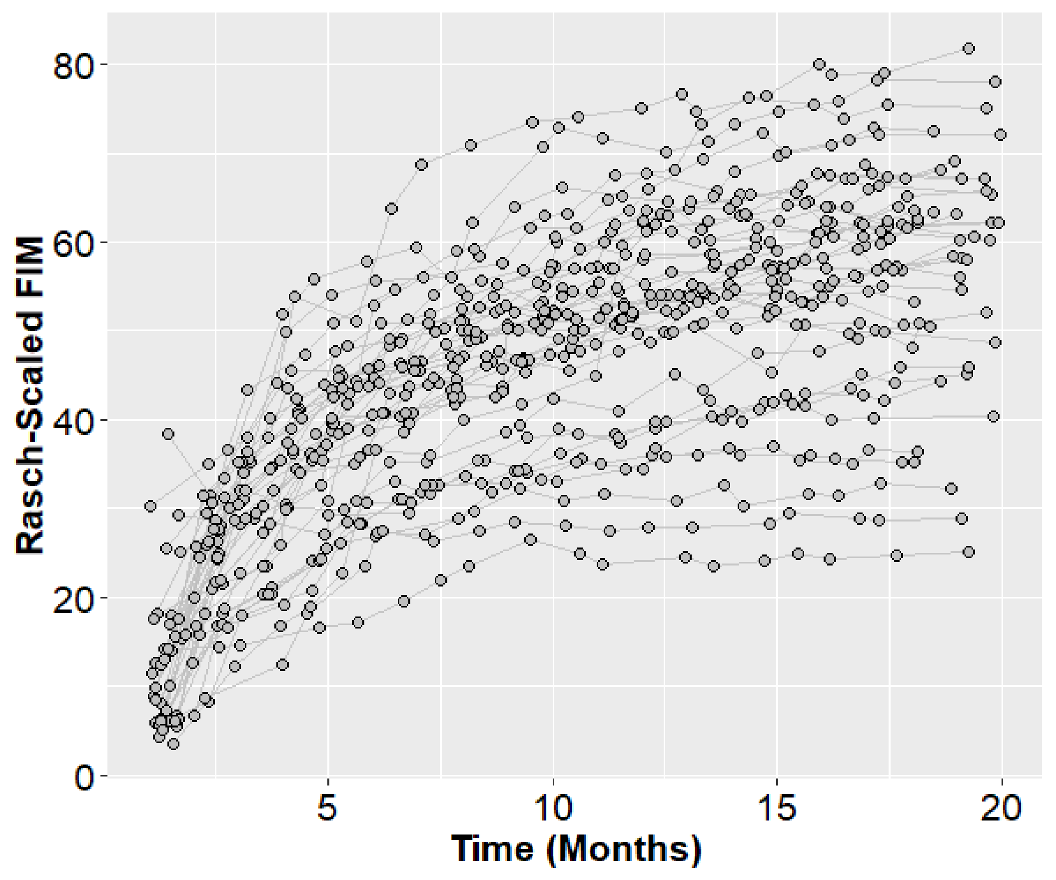
- **Unconditional Models:** we fit a series of models to determine how our dependent variable changes over time.
- **Model Comparison:** we can statistically evaluate our candidate models to determine a reasonable shape for change over time and how this differs from person to person.
- **Conditional Model:** adding fixed effects of interest, we can see which variables affect where people start and which variables affect how people change over time.
- **Model Comparison/Hypothesis Testing:** in an applied context, we are often interested in hypothesis tests of this conditional fixed-effects.
  - But note that you can test hypotheses about random-effects as well.
- **Regression Diagnostics and Assumption Checks:** As with OLS regression, we have a number of assumptions about our residuals (and random-effects).
  - We can check for outliers, influence and collinearity.
  - We will also often encounter “convergence” warnings in the ML estimation.

Unconditional Time Models



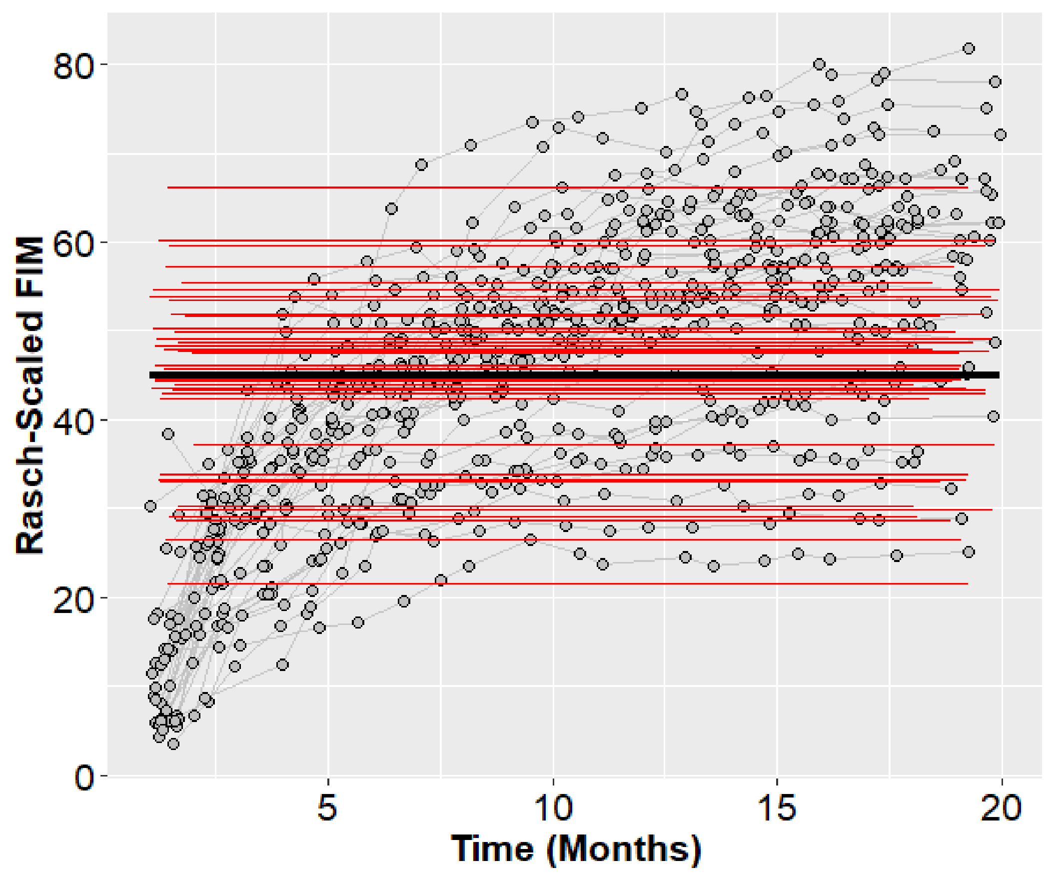
Unconditional Time Models

- **Adding Fixed and Random Effects**
  - Random Intercepts Model



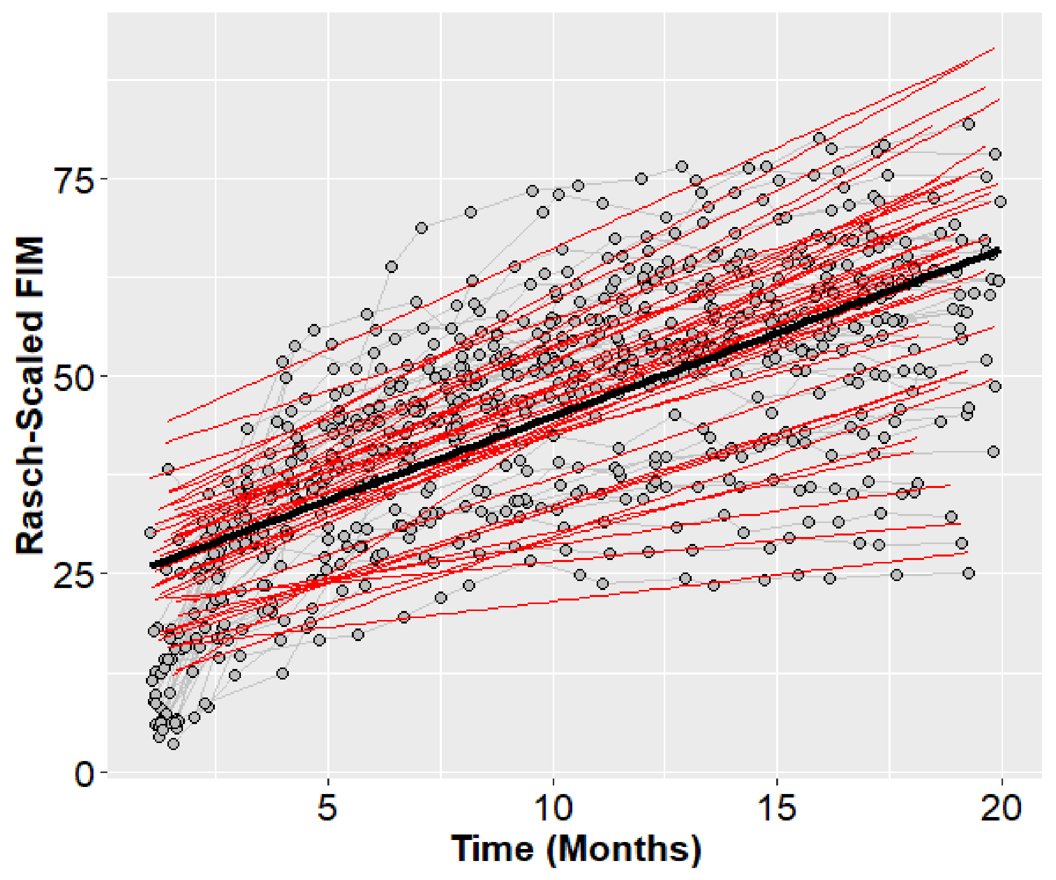
Unconditional Time Models

- **Adding Fixed and Random Effects**
  - Random Intercepts Model
    - $\beta_0 + U_{0j} + \epsilon_{ij}$



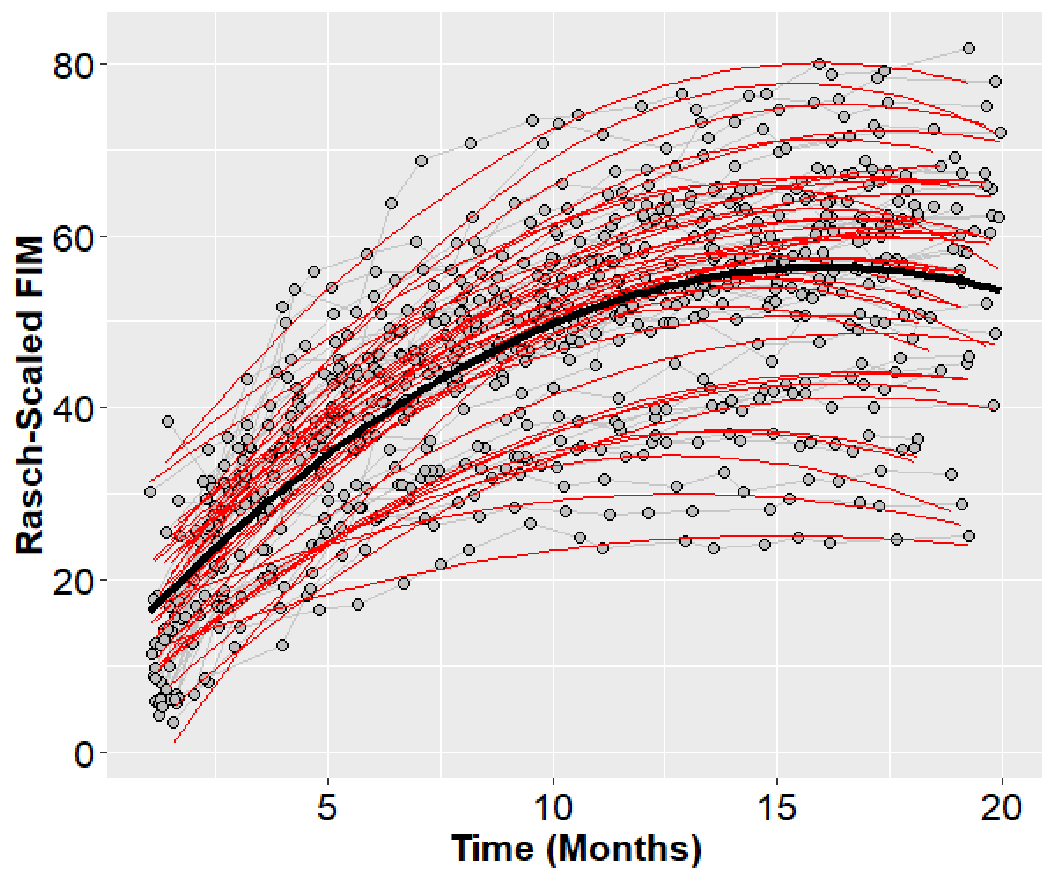
Unconditional Time Models

- **Adding Fixed and Random Effects**
  - Random Intercepts Model
    - $\beta_0 + U_{0j} + \epsilon_{ij}$
  - Random Slopes Model
    - $\left(\beta_0 + U_{0j}\right) + (\beta_1 + U_{1j})(X_{1_{ij}}) + \epsilon_{ij}$



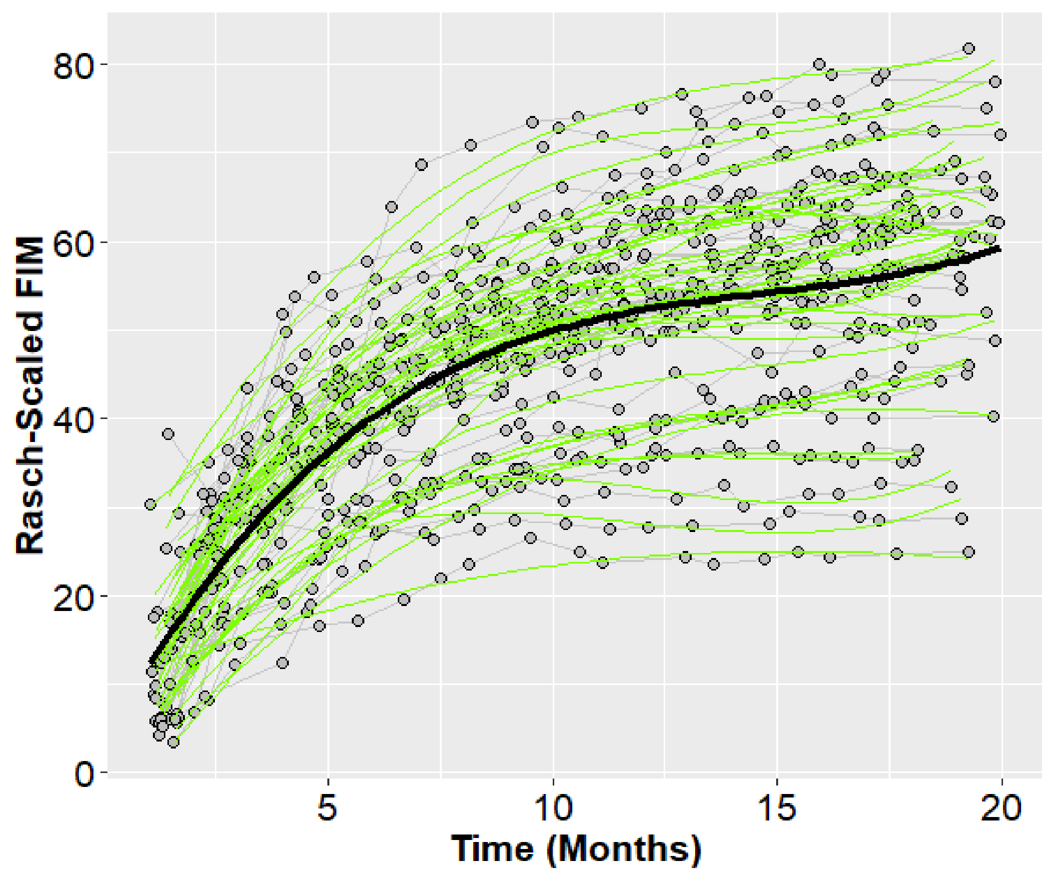
Unconditional Time Models

- **Adding Fixed and Random Effects**
  - Random Intercepts Model
    - $\beta_0 + U_{0j} + \epsilon_{ij}$
  - Random Slopes Model
    - $(\beta_0 + U_{0j}) + (\beta_1 + U_{1j})(X_{1ij}) + \epsilon_{ij}$
  - Quadratic Random Slopes Model
    - $(\beta_0 + U_{0j}) + (\beta_1 + U_{1j})(X_{1ij}) + (\beta_2 + U_{2j})(X_{1\,ij}^2) + \epsilon_{ij}$



Unconditional Time Models

- **Adding Fixed and Random Effects**
  - Random Intercepts Model
    - $\beta_0 + U_{0j} + \epsilon_{ij}$
  - Random Slopes Model
    - $(\beta_0 + U_{0j}) + (\beta_1 + U_{1j})(X_{1ij}) + \epsilon_{ij}$
  - Quadratic Random Slopes Model
    - $(\beta_0 + U_{0j}) + (\beta_1 + U_{1j})(X_{1ij}) + (\beta_2 + U_{2j})(X_{1\,ij}^2) + \epsilon_{ij}$
- And so on!
  - [Cubic random slopes model shown.]



Unconditional Time Models

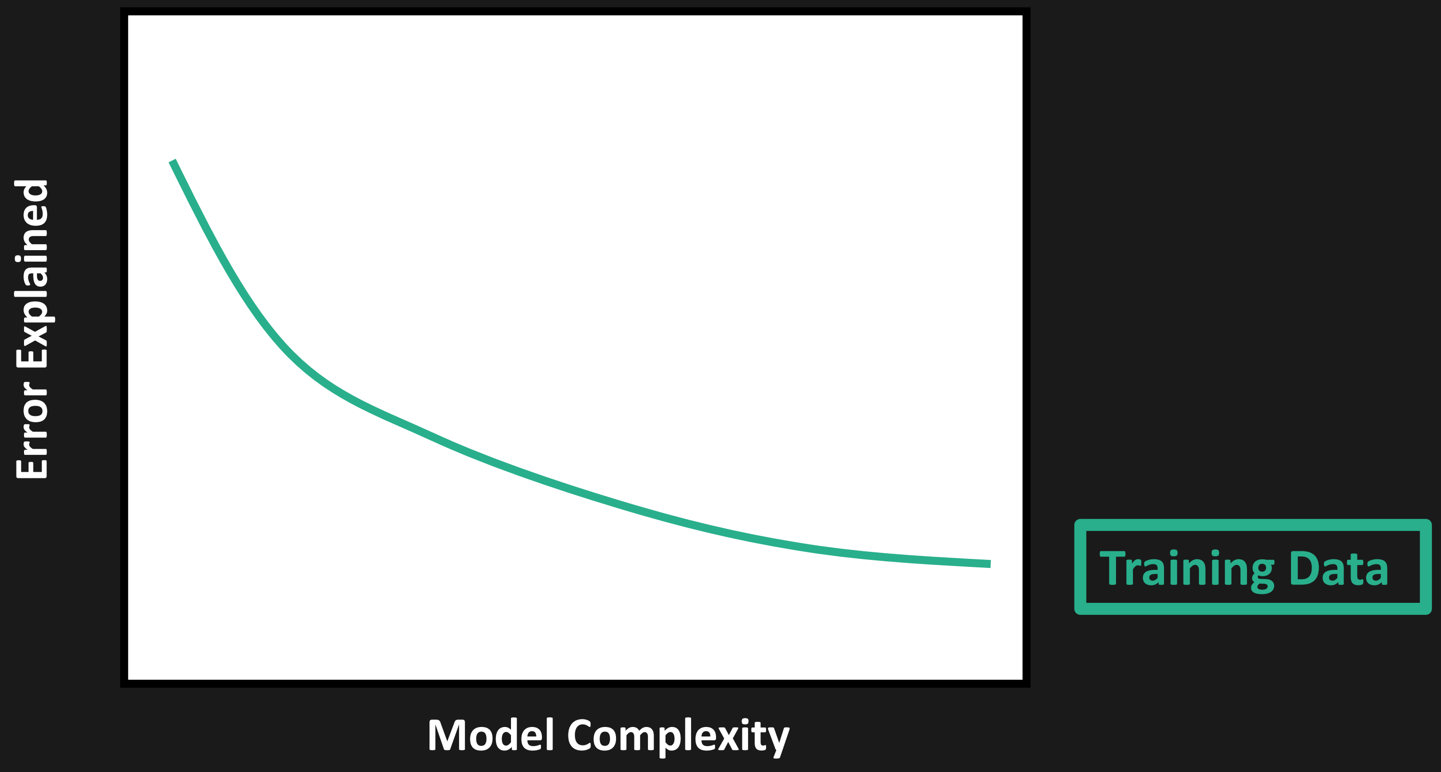
- Here we are adopting a successive polynomial approach.
  - Splines and truly nonlinear functions are a more advanced topic for another time.

- Note that you also don’t have to add fixed- and random-effects at the same time.
  - E.g, ***Fixed Slope Random Intercept*** model allows intercepts to vary but assumes the same slope for everyone: $(\beta_0 + U_{0j}) + (\beta_1)(X_{1_{ij}}) + \epsilon_{ij}$

- When do we stop?
  - Overfitting: There is a tradeoff between our models complexity and our models generalizability.

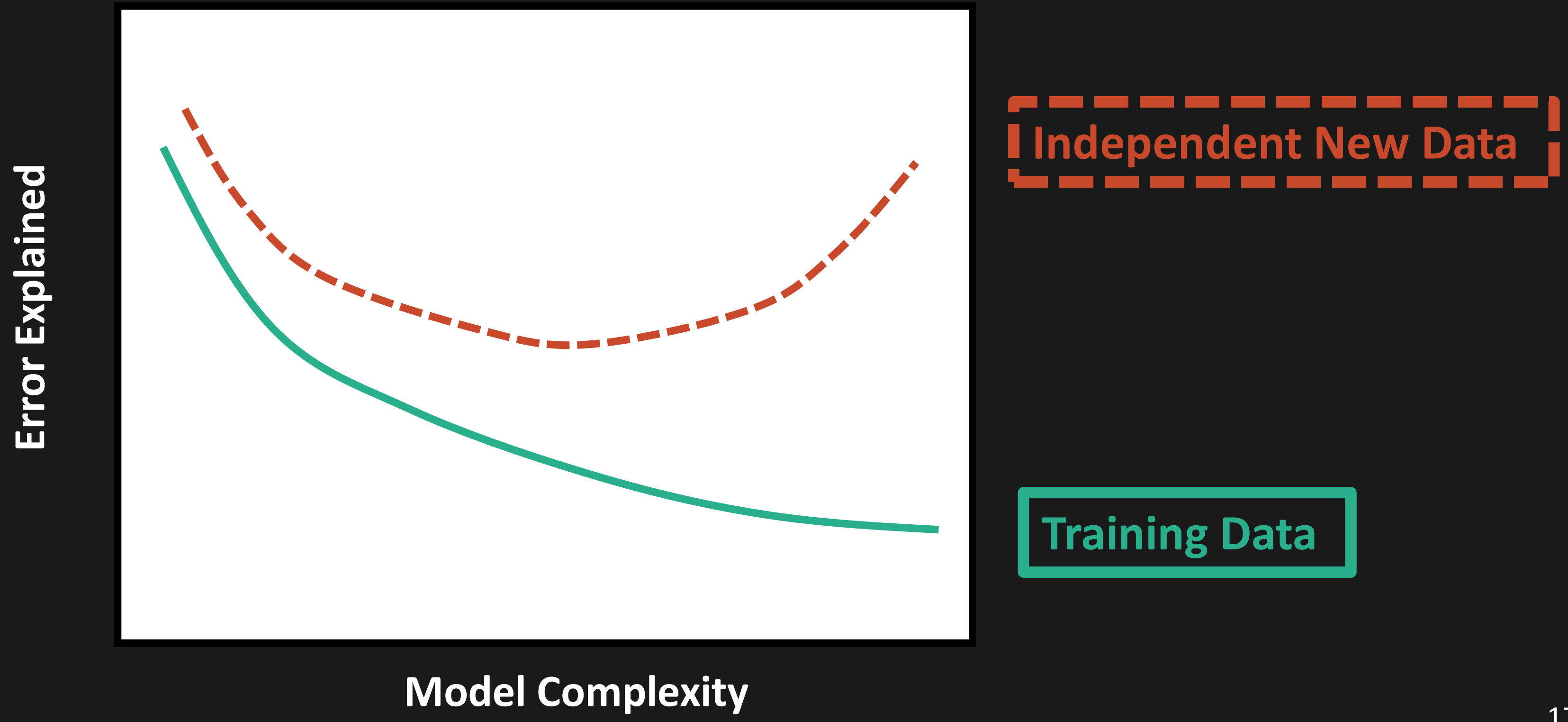
Overfitting

- A complex model will explain the current data very well, but it will not generalize to a new sample.



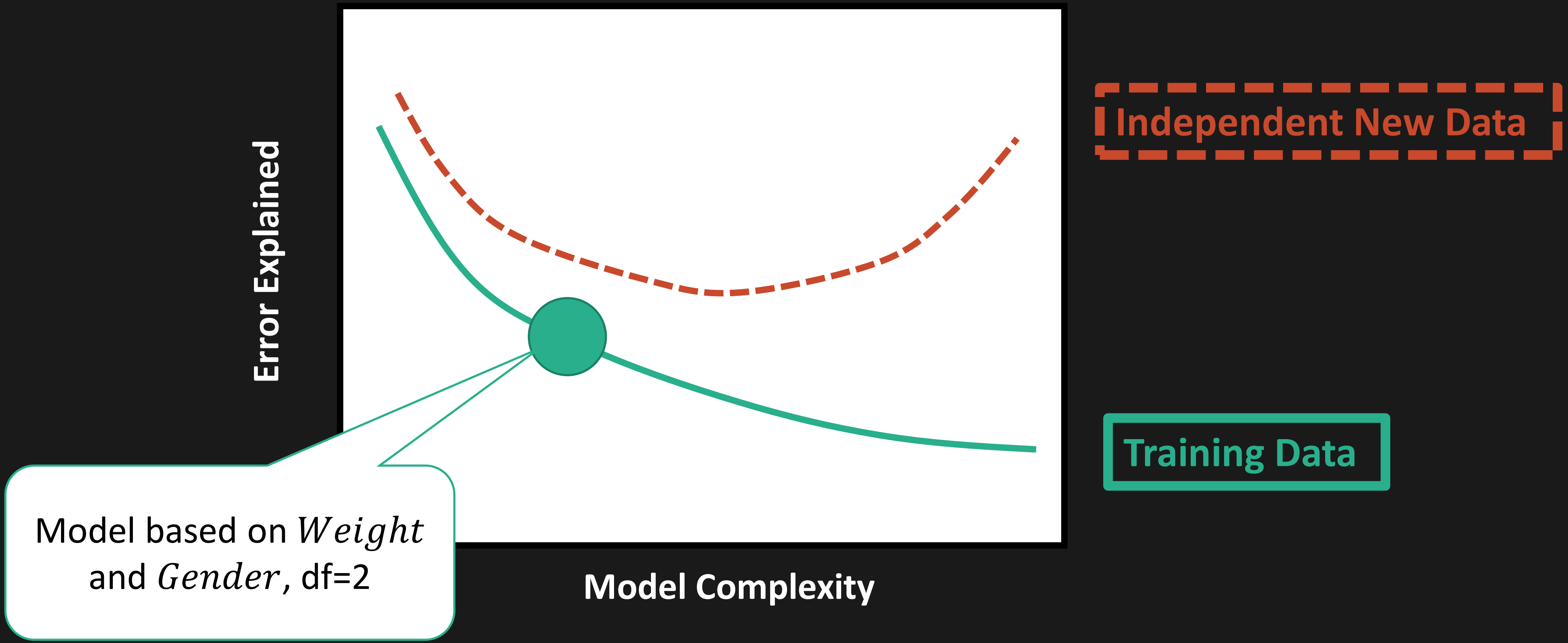
Overfitting

- A complex model will explain the current data very well, but it will not generalize to a new sample.



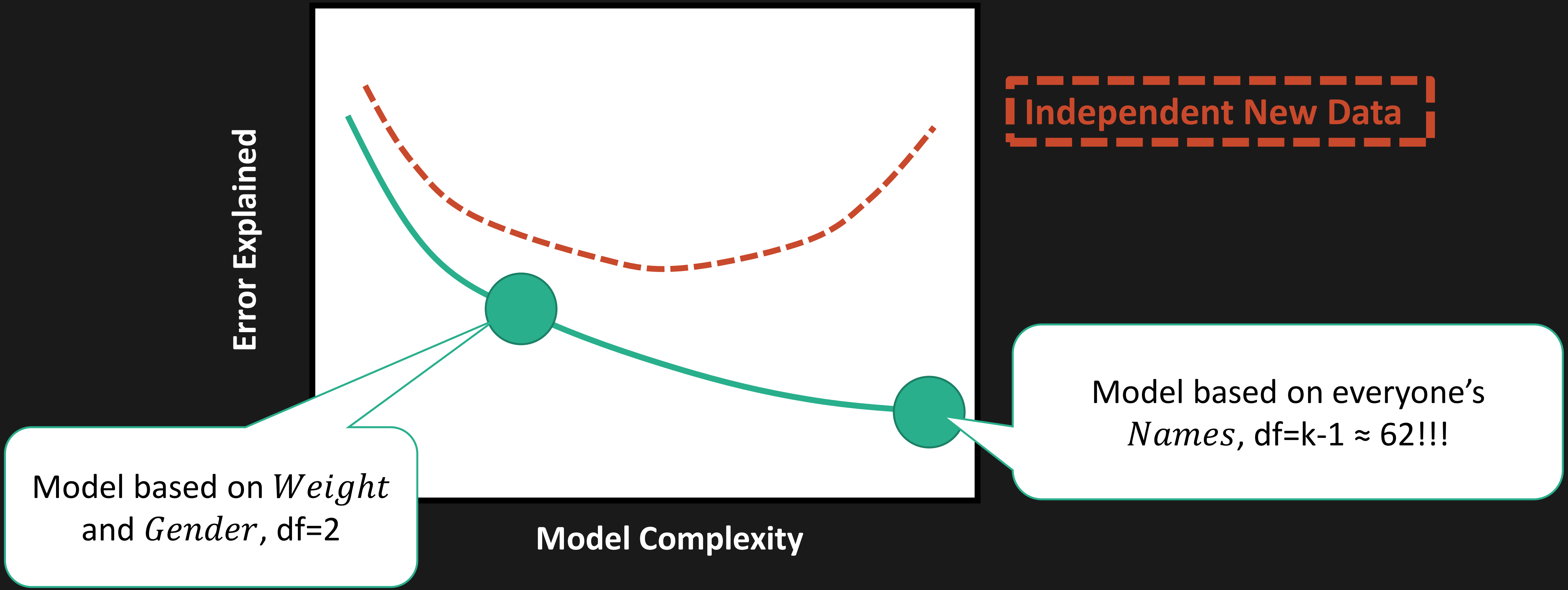
Overfitting

- **Example:** Predicting the height of everyone in the workshop.



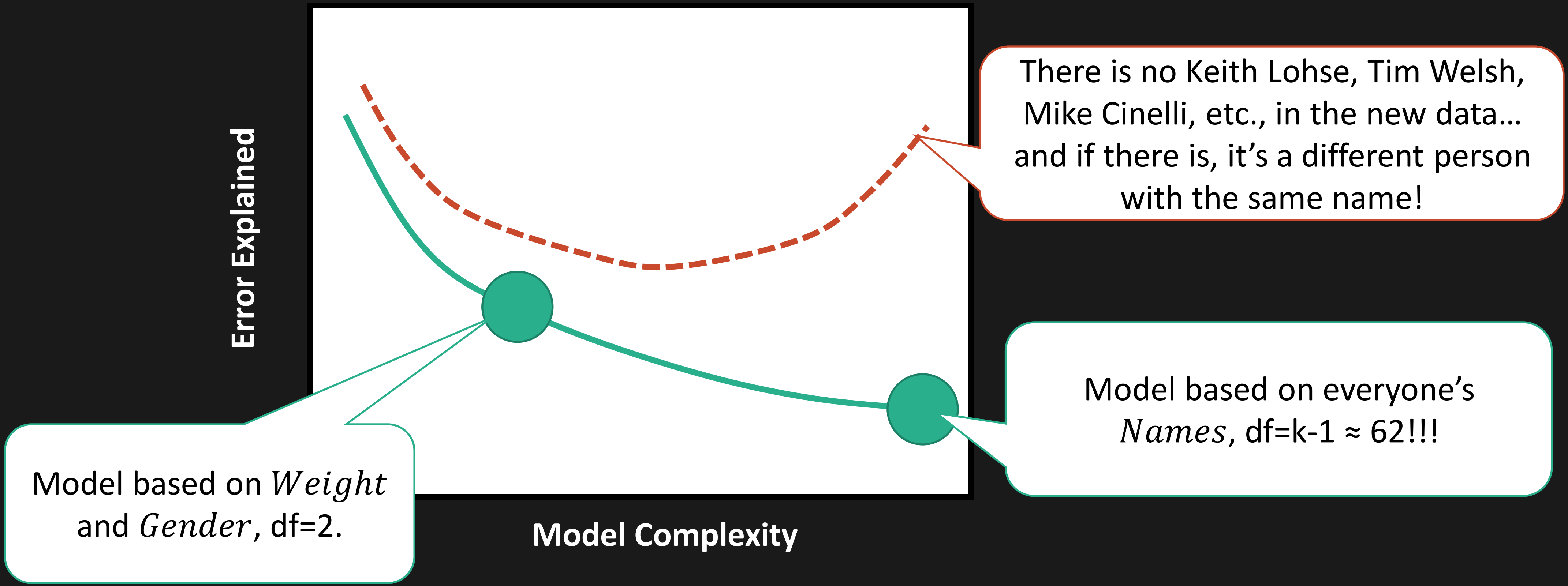
Overfitting

- **Example:** Predicting the height of everyone in the workshop.



Overfitting

- **Example:** Predicting the height of everyone in the workshop.



Overfitting

- Admittedly, that is a silly example, but the principle still holds.

- We don’t want to embrace a 5th order polynomial just because it explains more of the variance.

- Fortunately, we have some tools at our disposal to avoid overfitting.

Akaike's Information Criterion (AIC)

- As with OLS Regression, we have the general formula:
  - $DATA \;=\; MODEL \;+\; ERROR$

- With mixed-effect models using REML/ML estimate, we no longer use the sum of squared errors to quantify error. Instead we use the deviance.
  - $Deviance = -2 * \log(Likelihood)$
  - $Deviance = N * log(2\pi\sigma_\epsilon^2) + (1/\sigma_\epsilon^2) * (\sum(\epsilon_i^2))$

Akaike's Information Criterion (AIC)

- This formula is kind of scary, but there are a few things you need to notice about it:
  - $Deviance = N * log(2\pi\sigma_{\epsilon}^{2}) + \left(\frac{1}{\sigma_{\epsilon}^{2}}\right) * \left(\sum(\epsilon_{i}^{2})\right)$

- **First**, deviance is based on residual error (ϵ), so ***lower deviance means better model fit.***

Akaike's Information Criterion (AIC)

- This formula is kind of scary, but there are a few things you need to notice about it:
  - $Deviance = N * log(2\pi\sigma_{\epsilon}^{2}) + \left(\frac{1}{\sigma_{\epsilon}^{2}}\right) * \left(\sum(\epsilon_{i}^{2})\right)$

- **First**, deviance is based on residual error (ϵ), so ***lower deviance means better model fit.***

- **Second**, notice that size our sample shows up in two different places. ***If we want to compare models based on their deviance, those models need to be based on the same amount of data.***

Akaike's Information Criterion (AIC)

- Now, the logic behind the AIC is pretty complicated, but if you have the deviance, the formula is pretty simple:
    - $AIC = Deviance + 2(k)$
    - where k=the number of parameters.

- Thus, the AIC imposes penalty on additional parameters in order to reduce overfitting.
    - In fact, the AIC produces the best estimate of the long-run predictive deviance.

- As a starting point, we will select models based on the smallest AIC.
    - But it can get more complicated than that (corrections for small samples, etc).

Akaike's Information Criterion

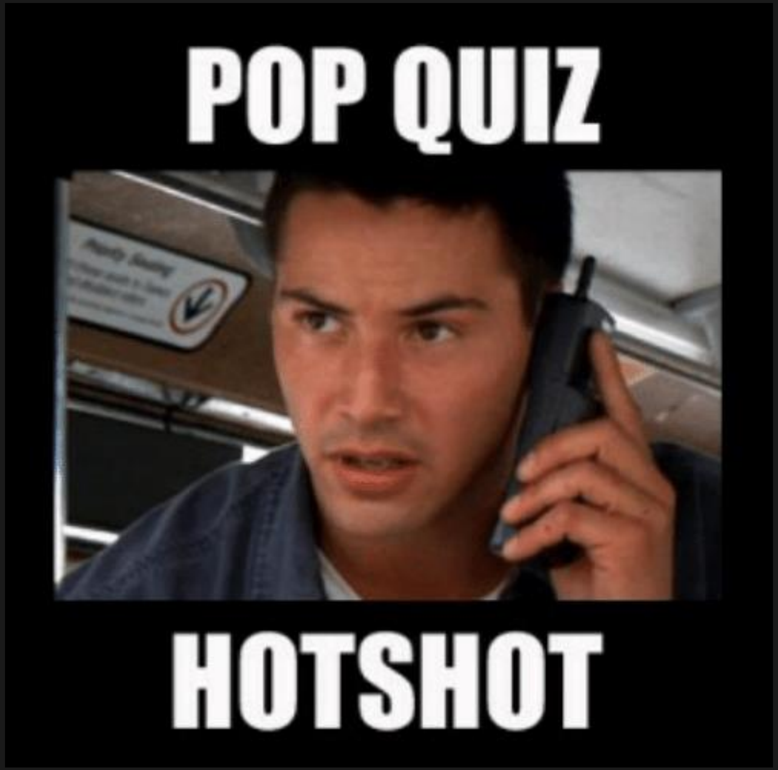
Model	K	Log(Likelihood)	Deviance	AIC	χ^2	Df	$P(<\chi^2)$
Random Intercepts	3	-2929.9	5859.9	5865.9	NA	NA	NA
Linear Slopes	6	-2406.5	4812.9	4824.9	1047.0	3.0	<0.001
Quadratic Slopes	10	-2033.2	4066.4	4086.4	746.5	4.0	<0.001
Cubic Slopes	15	-1907.1	3814.1	3844.1	252.3	5.0	<0.001

It looks like the cubic slopes model gives us the lowest AIC despite the 15 degrees of freedom.

Akaike's Information Criterion

Model	K	Log(Likelihood)	Deviance	AIC	χ^2	Df	$P(<\chi^2)$
Random Intercepts	3	-2929.9	5859.9	5865.9	NA	NA	NA
Linear Slopes	6	-2406.5	4812.9	4824.9	1047.0	3.0	<0.001
Quadratic Slopes	10	-2033.2	4066.4	4086.4	746.5	4.0	<0.001
Cubic Slopes	15	-1907.1	3814.1	3844.1	252.3	5.0	<0.001

But why are there 15 degrees of freedom!?!?



Akaike's Information Criterion

Model	K	Log(Likelihood)	Deviance	AIC	χ^2	Df	P(<χ^2)
Random Intercepts	3	-292					
Linear Slopes	6						
Quadratic Slopes	10						
Cubic Slopes	15	-19					

```
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  subID    (Intercept)  90.95    9.537
##  Residual             176.15   13.272
## Number of obs: 720, groups:  subID, 40
##
## Fixed effects:
##             Estimate Std. Error     df t value Pr(>|t|)
## (Intercept)   44.904      1.587 40.000    28.3   <2e-16 ***
```

Akaike's Information Criterion

Model	K	Log(Likelihood)	Deviance	AIC	χ^2	Df	$P(<\chi^2)$
Random Intercepts	3	-292					
Linear Slopes	6						
Quadratic Slopes	10						
Cubic Slopes	15	-19					

```
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  subID    (Intercept) 46.48    6.818
##           year.0      46.85    6.844    0.37
##  Residual             35.48    5.956
## Number of obs: 720, groups:  subID, 40
##
## Fixed effects:
##             Estimate Std. Error     df t value Pr(>|t|)
## (Intercept)   25.855      1.165 39.979   22.20   <2e-16 ***
## year.0        25.367      1.195 39.981   21.23   <2e-16 ***
```

Akaike's Information Criterion

Model	K	Log(Likelihood)	Deviance	AIC	χ^2	Df	$P(<\chi^2)$
Random Intercepts	3	-292					
Linear Slopes	6	-2					
Quadratic Slopes	10						
Cubic Slopes	15	-19					

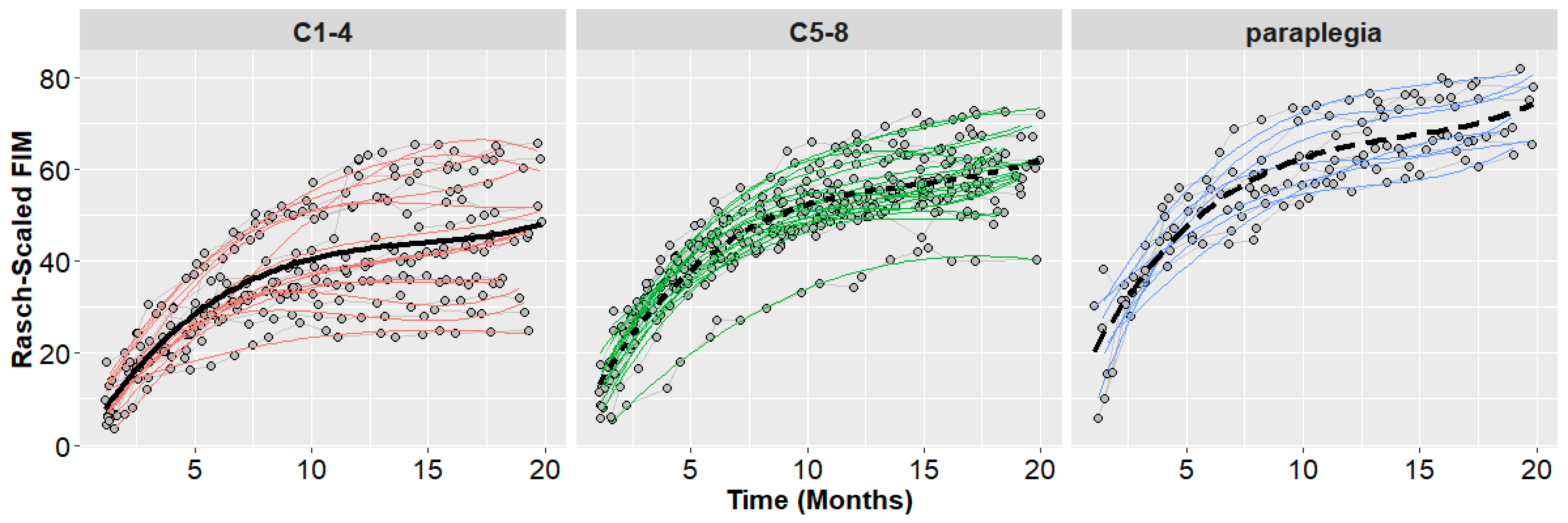
```
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  subID    (Intercept)  48.03    6.931
##           year.0      234.10   15.300   -0.02
##           year.0_sq    35.94    5.995    0.02 -0.94
##  Residual              10.57    3.251
## Number of obs: 720, groups:  subID, 40
##
## Fixed effects:
##             Estimate Std. Error     df t value Pr(>|t|)
## (Intercept)   15.969      1.152 39.907   13.86   <2e-16 ***
## year.0        64.492      2.648 40.391   24.36   <2e-16 ***
## year.0_sq    -25.786      1.170 41.270  -22.03   <2e-16 ***
```

Akaike's Information Criterion

Model	K	Log(Likelihood)	Deviance	AIC	χ^2	Df	$P(<\chi^2)$
Random Intercepts	3	-292					
Linear Slopes	6	-24					
Quadratic Slopes	10						
Cubic Slopes	15						

```
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  subID    (Intercept)  43.87    6.624
##           year.0      626.89   25.038   -0.05
##           year.0_sq   921.27   30.352    0.01 -0.88
##           year.0_cu   137.88   11.742    0.04  0.76 -0.98
##  Residual               6.59    2.567
## Number of obs: 720, groups:  subID, 40
##
## Fixed effects:
##              Estimate Std. Error     df t value Pr(>|t|)
## (Intercept)   11.921      1.111  39.984  10.731 2.44e-13 ***
## year.0        96.527      4.473  40.075  21.580  < 2e-16 ***
## year.0_sq    -78.115      5.740  39.993 -13.610  < 2e-16 ***
## year.0_cu     22.770      2.295  40.418   9.923 2.15e-12 ***
```

Making the Conditional Model



- How does functional independence differ between groups over time?

Making the Conditional Model

- Conceptually,
  - main-effects of linear, quadratic, cubic time and group, plus the interactions of these factors.

- Mathematically,

$$
\begin{aligned}
&y_{ij} \\
&= (\beta_0 + U_{0j}) + (\beta_1 + U_{1j})(T_{1_{ij}}) + (\beta_2 + U_{2j})(T_{1\ ij}^2) + (\beta_3 + U_{0j})(T_{1\ ij}^3) \\
&+ \beta_4\left(G_{1_{ij}}\right) + \beta_5\left(G_{2_{ij}}\right) + \beta_6(TxG_1) + \beta_7(TxG_2) + \beta_8(T^2xG_1) + \beta_9(T^2xG_2) \\
&+ \beta_{10}(T^3xG_1) + \beta_{11}(T^3xG_2)
\end{aligned}
$$

Making the Conditional Model

- Conceptually,
  - main-effects of linear, quadratic, cubic time and group, plus the interactions of these factors.

- Mathematically,

$$y_{ij} = (\beta_0 + U_{0j}) + (\beta_1 + U_{1j})(T_{1_{ij}}) + (\beta_2 + U_{2j})(T_1^2{}_{ij}) + (\beta_3 + U_{0j})(T_1^3{}_{ij}) + \beta_4\left(G_{1_{ij}}\right) + \beta_5\left(G_{2_{ij}}\right) + \beta_6(TxG_1) + \beta_7(TxG_2) + \beta_8(T^2xG_1) + \beta_9(T^2xG_2) + \beta_{10}(T^3xG_1) + \beta_{11}(T^3xG_2)$$

Making the Conditional Model

- Conceptually,
  - main-effects of linear, quadratic, cubic time and group, plus the interactions of these factors.

- Mathematically,

$$
\begin{aligned}
y_{ij} &= (\beta_0 + U_{0j}) + (\beta_1 + U_{1j})(T_{1_{ij}}) + (\beta_2 + U_{2j})(T_1^2{}_{ij}) + (\beta_3 + U_{0j})(T_1^3{}_{ij}) \\
&+ \beta_4\left(G_{1_{ij}}\right) + \beta_5\left(G_{2_{ij}}\right) + \beta_6(TxG_1) + \beta_7(TxG_2) + \beta_8(T^2xG_1) + \beta_9(T^2xG_2) \\
&+ \beta_{10}(T^3xG_1) + \beta_{11}(T^3xG_2)
\end{aligned}
$$

Making the Conditional Model

- Conceptually,
  - main-effects of linear, quadratic, cubic time and group, plus the interactions of these factors.

- Mathematically,

$$
\begin{aligned}
y_{ij} &= (\beta_0 + U_{0j}) + (\beta_1 + U_{1j})(T_{1_{ij}}) + (\beta_2 + U_{2j})(T^2_{1\ ij}) + (\beta_3 + U_{0j})(T^3_{1\ ij}) \\
&+ \beta_4\left(G_{1_{ij}}\right) + \beta_5\left(G_{2_{ij}}\right) + \beta_6(TxG_1) + \beta_7(TxG_2) + \beta_8(T^2xG_1) + \beta_9(T^2xG_2) \\
&+ \beta_{10}(T^3xG_1) + \beta_{11}(T^3xG_2)
\end{aligned}
$$

Making the Conditional Model

- Conceptually,
  - main-effects of linear, quadratic, cubic time and group, plus the interactions of these factors.

- Programatically in R using lme4,

```
lmer(rasch_FIM~
# Fixed Effects
1+year.0*AIS_grade+year.0_sq*AIS_grade+year.0_cu*AIS_grade+
# Random Effects
(1+year.0+year.0_sq+year.0_cu|subID),
data=DAT2, REML=FALSE,
control=lmerControl(optimizer="bobyqa“, optCtrl=list(maxfun=5e5)))
```

Making the Conditional Model

- Conceptually,
  - main-effects of linear, quadratic, cubic time and group, plus the interactions of these factors.

- Programatically in R using lme4,

```
lmer(rasch_FIM~
# Fixed Effects
1+year.0*AIS_grade+year.0_sq*AIS_grade+year.0_cu*AIS_grade+
# Random Effects
(1+year.0+year.0_sq+year.0_cu|subID),
data=DAT2, REML=FALSE,
control=lmerControl(optimizer="bobyqa“, optCtrl=list(maxfun=5e5)))
```

Interpreting the Conditional Model

Model	K	Log(Likelihood)	Deviance	AIC	χ^2	Df	$P(<\chi^2)$
Random Intercepts	3	-2929.9	5859.9	5865.9	NA	NA	NA
Linear Slopes	6	-2406.5	4812.9	4824.9	1047.0	3.0	<0.001
Quadratic Slopes	10	-2033.2	4066.4	4086.4	746.5	4.0	<0.001
Cubic Slopes	15	-1907.1	3814.1	3844.1	252.3	5.0	<0.001
Conditional Model	23	-1886.24	3772.48	3818.49	41.65	8.0	<0.001

The omnibus test of the conditional model is statistically significant compared to the best fitting unconditional model (Wald Test of the change in deviance).

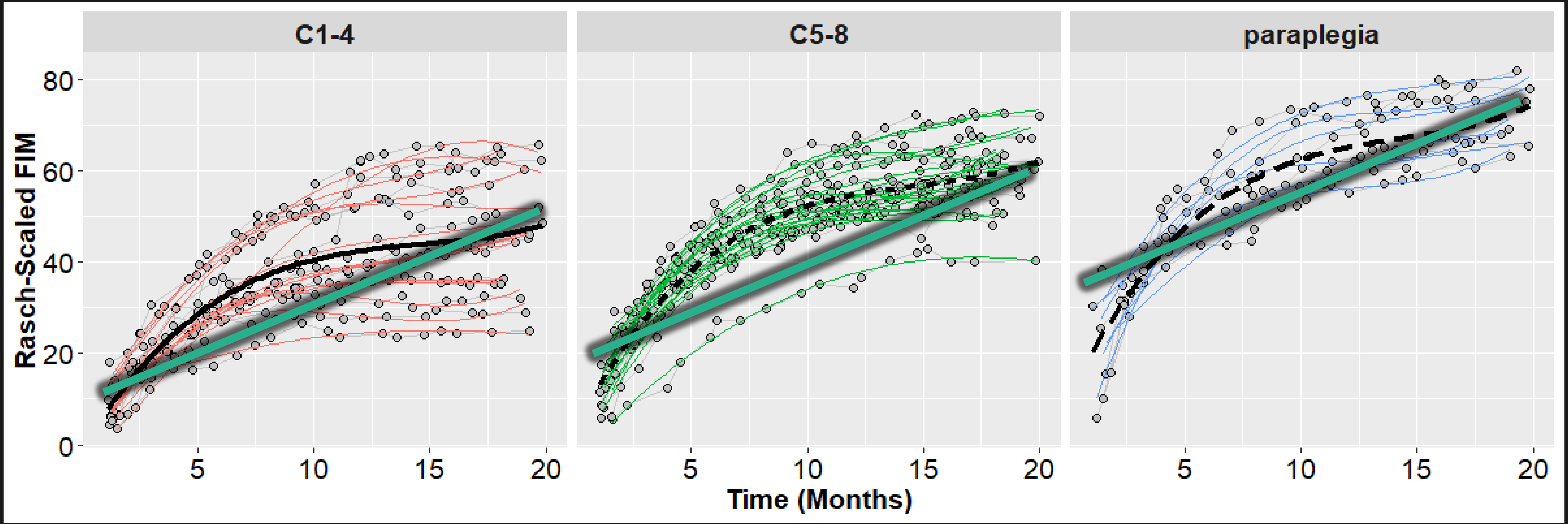
And the conditional model produces the lowest AIC, suggesting we are not overfitting the data.

Interpreting the Conditional Model

Analysis of Deviance Table (Type III Wald Chi-Square Tests)

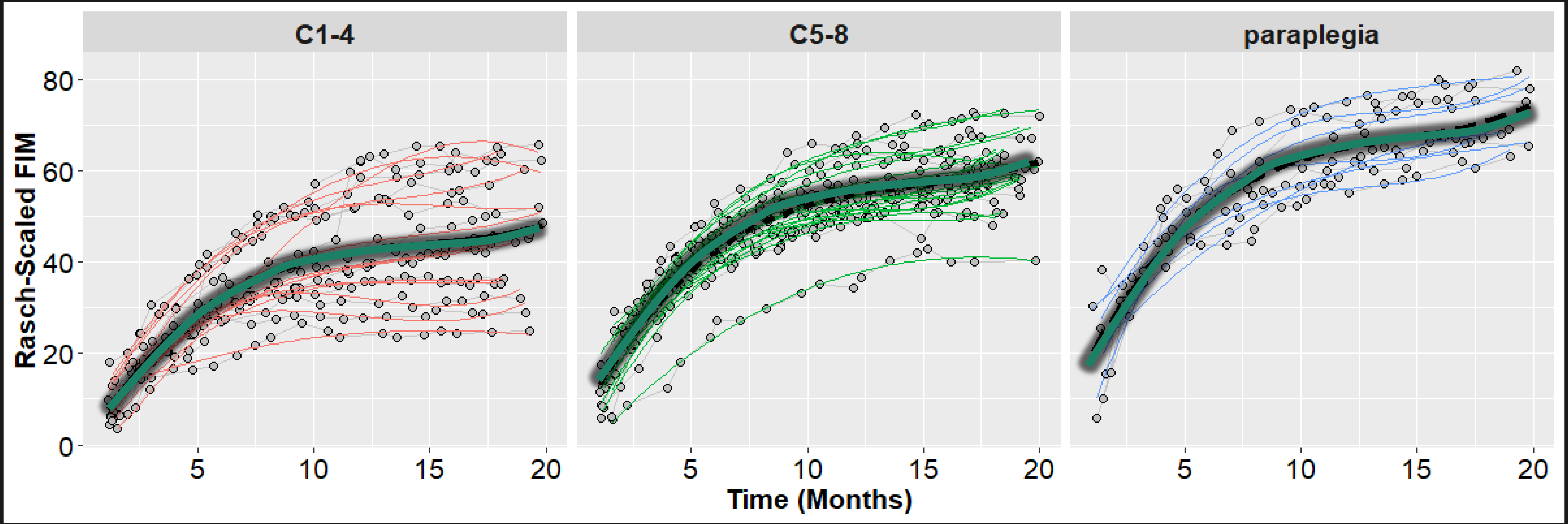
Model	$\chi 2$	Df	$P(>\chi 2)$
Intercept	23.3	1	<0.001
Time	140.2	1	<0.001
Group	28.6	2	<0.001
$Time^2$	50.4	1	<0.001
$Time^3$	24.4	1	<0.001
Time x Group	4.8	2	0.089
$Time^2$ x Group	2.5	2	0.290
$Time^3$ x Group	1.98	2	0.371

Looking within the model, we can see the omnibus F-/χ^2-tests of our main-effects and interactions.
Here I am showing the χ^2-tests based on a Type III calculation of the change in deviance.
Although we have statistically significant main-effects, we do not find statistically significant interactions.



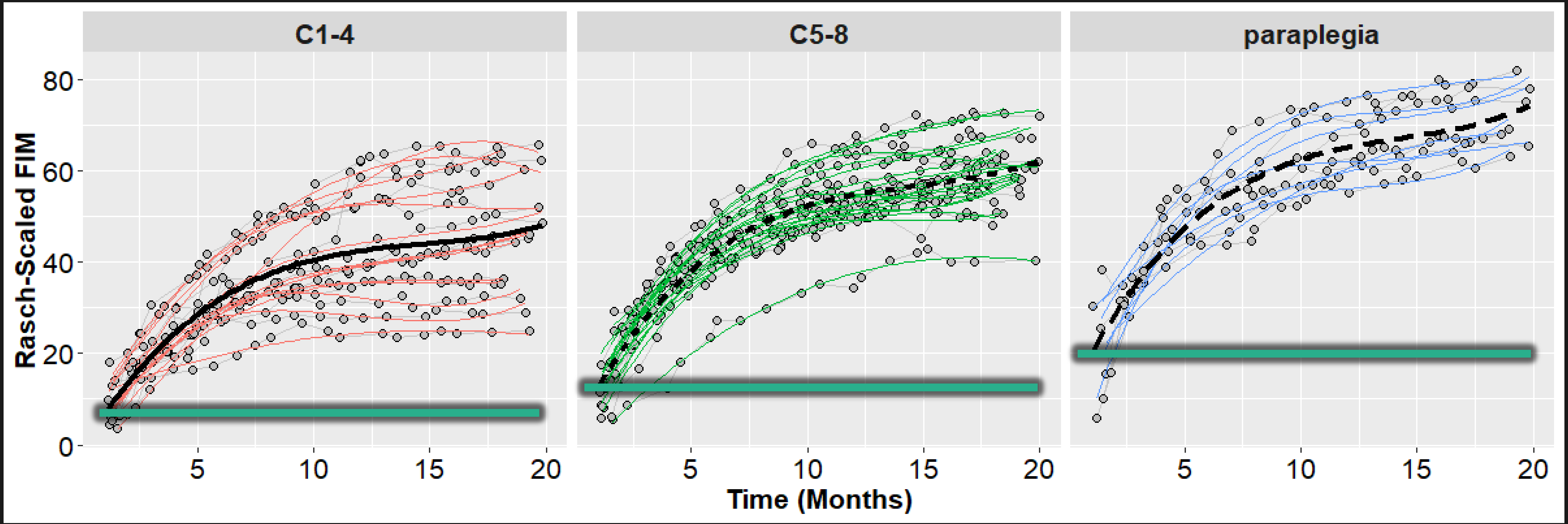
Model	χ2	Df	P(>χ2)
Intercept	23.3	1	<0.001
Time	140.2	1	<0.001
Group	28.6	2	<0.001
Time2	50.4	1	<0.001
Time3	24.4	1	<0.001
Time x Group	4.8	2	0.089
Time2 x Group	2.5	2	0.290
Time3 x Group	1.98	2	0.371

On average, people tended to change linearly over time.



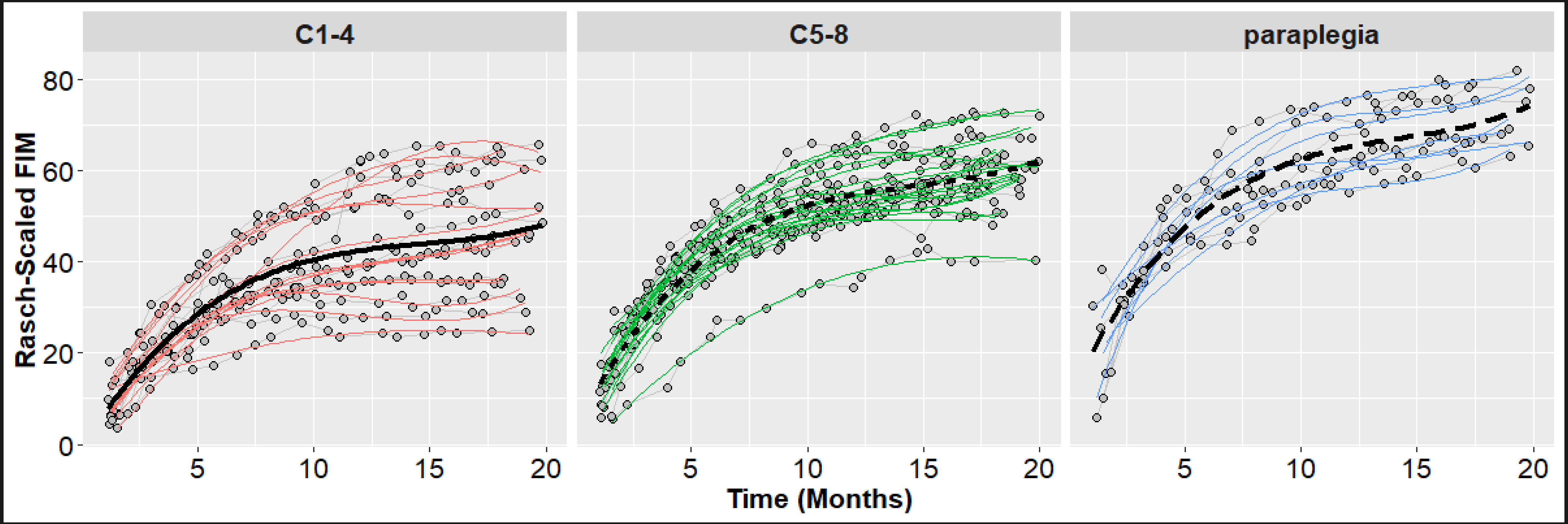
Model	χ2	Df	P(>χ2)
Intercept	23.3	1	<0.001
Time	140.2	1	<0.001
Group	28.6	2	<0.001
Time2	50.4	1	<0.001
Time3	24.4	1	<0.001
Time x Group	4.8	2	0.089
Time2 x Group	2.5	2	0.290
Time3 x Group	1.98	2	0.371

However, this change was not simply linear, there were significant quadratic and cubic aspects to this curve.



Model	χ2	Df	P(>χ2)
Intercept	23.3	1	<0.001
Time	**140.2**	**1**	**<0.001**
Group	28.6	2	<0.001
Time2	**50.4**	**1**	**<0.001**
Time3	**24.4**	**1**	**<0.001**
Time x Group	4.8	2	0.089
Time2 x Group	2.5	2	0.290
Time3 x Group	1.98	2	0.371

The groups also reliably differed in their intercepts, i.e., baseline independence differed between groups.



Model	χ2	Df	P(>χ2)
Intercept	23.3	1	<0.001
Time	**140.2**	**1**	**<0.001**
Group	**28.6**	**2**	**<0.001**
Time2	**50.4**	**1**	**<0.001**
Time3	**24.4**	**1**	**<0.001**
Time x Group	**4.8**	**2**	**0.089**
Time2 x Group	**2.5**	**2**	**0.290**
Time3 x Group	**1.98**	**2**	**0.371**

Although the change over time was not equal in these groups, the differences were not large enough to obtain unusual χ^2-values under the null-hypothesis.

Thus, while it is possible these groups do differ in how they change over time, we did not find compelling evidence to that effect in this sample.

Steps to Building Longitudinal Models

- ~~**Unconditional Models:** we fit a series of models to determine how our dependent variable changes over time.~~
- ~~**Model Comparison:** we can statistically evaluate our candidate models to determine a reasonable shape for change over time and how this differs from person to person.~~
- ~~**Conditional Model:** adding fixed effects of interest, we can see which variables affect where people start and which variables affect how people change over time.~~
- ~~**Model Comparison/Hypothesis Testing:** in an applied context, we are often interested in hypothesis tests of this conditional fixed-effects.~~
  - ~~But note that you can test hypotheses about random-effects as well.~~
- **Regression Diagnostics and Assumption Checks:** As with OLS regression, we have a number of assumptions about our residuals (and random-effects).
  - We can check for outliers, influence and collinearity.
  - We will also often encounter “convergence” warnings in the ML estimation.

Future Things to Learn About

- **Regression Diagnostics and Assumption Checks:** As with OLS regression, we have a number of assumptions about our residuals (and random-effects).
  - **Assumptions (similar to OLS Regression):**
    - Approximately normal distributions of random-effects and residual errors.
    - Approximate homogeneity of the residuals.
    - Approximate linearity of the relationship between predictor and response.
    - Data must be MCAR or MAR (controlling for related effect). Data MNAR will lead to biased model estimates.
  - **Diagnostics (similar to OLS Regression):**
    - Influence (leverage, outliers, Cook's distance) can now be checked at the level of individual data-points (residuals) and participants (random-effects).