0:00 / 1:00

Growth models

Professor Andy Field

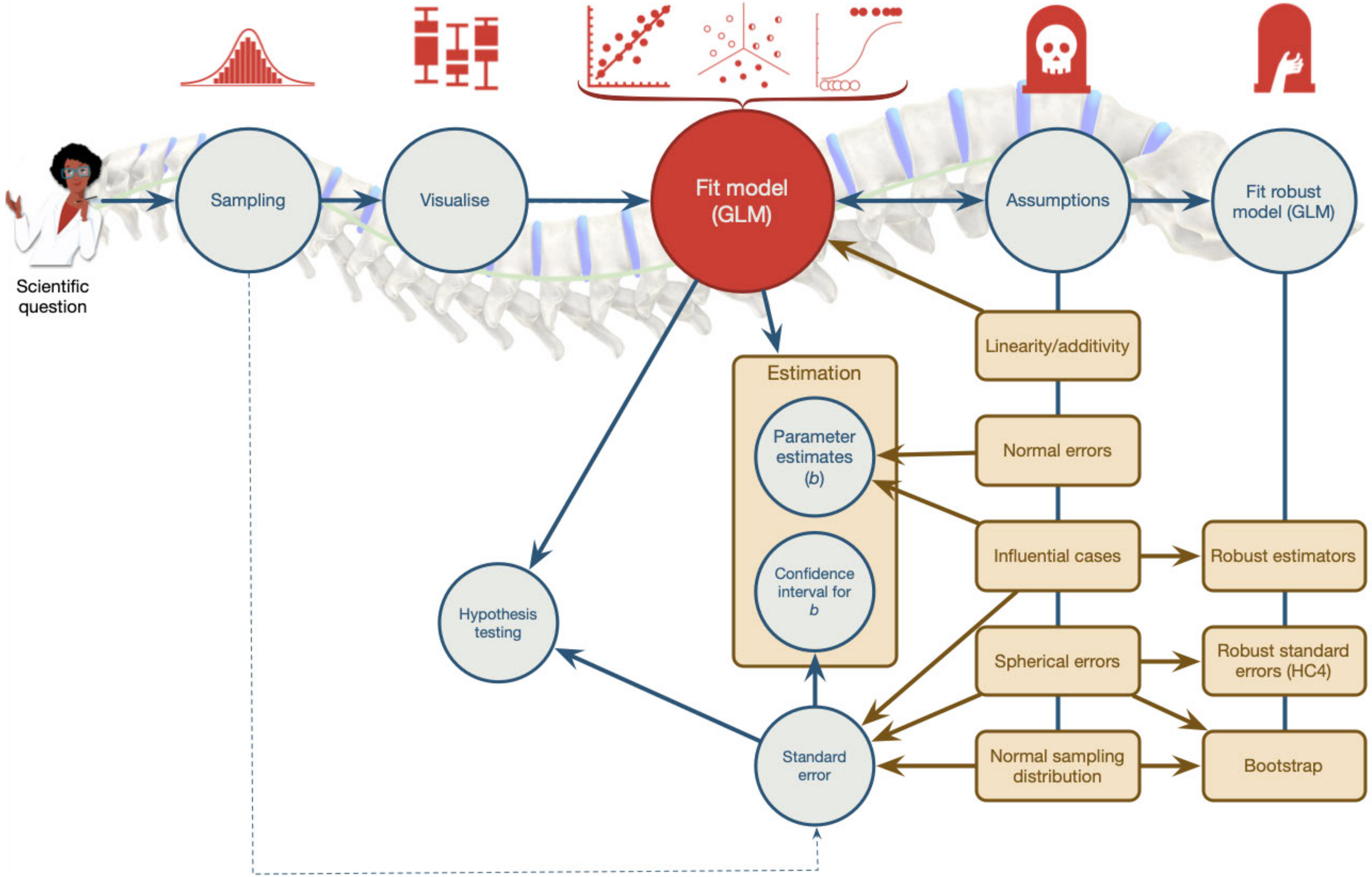
@profandyfield

www.youtube.com/user/ProfAndyField/

www.discoveringstatistics.com

www.milton-the-cat.rocks

www.discovr.rocks


Scientific question
Sampling
Visualise
Fit model (GLM)
Assumptions
Fit robust model (GLM)
Estimation
Parameter estimates (b)
Confidence interval for b
Hypothesis testing
Standard error
Linearity/additivity
Normal errors
Influential cases
Spherical errors
Normal sampling distribution
Robust estimators
Robust standard errors (HC4)
Bootstrap

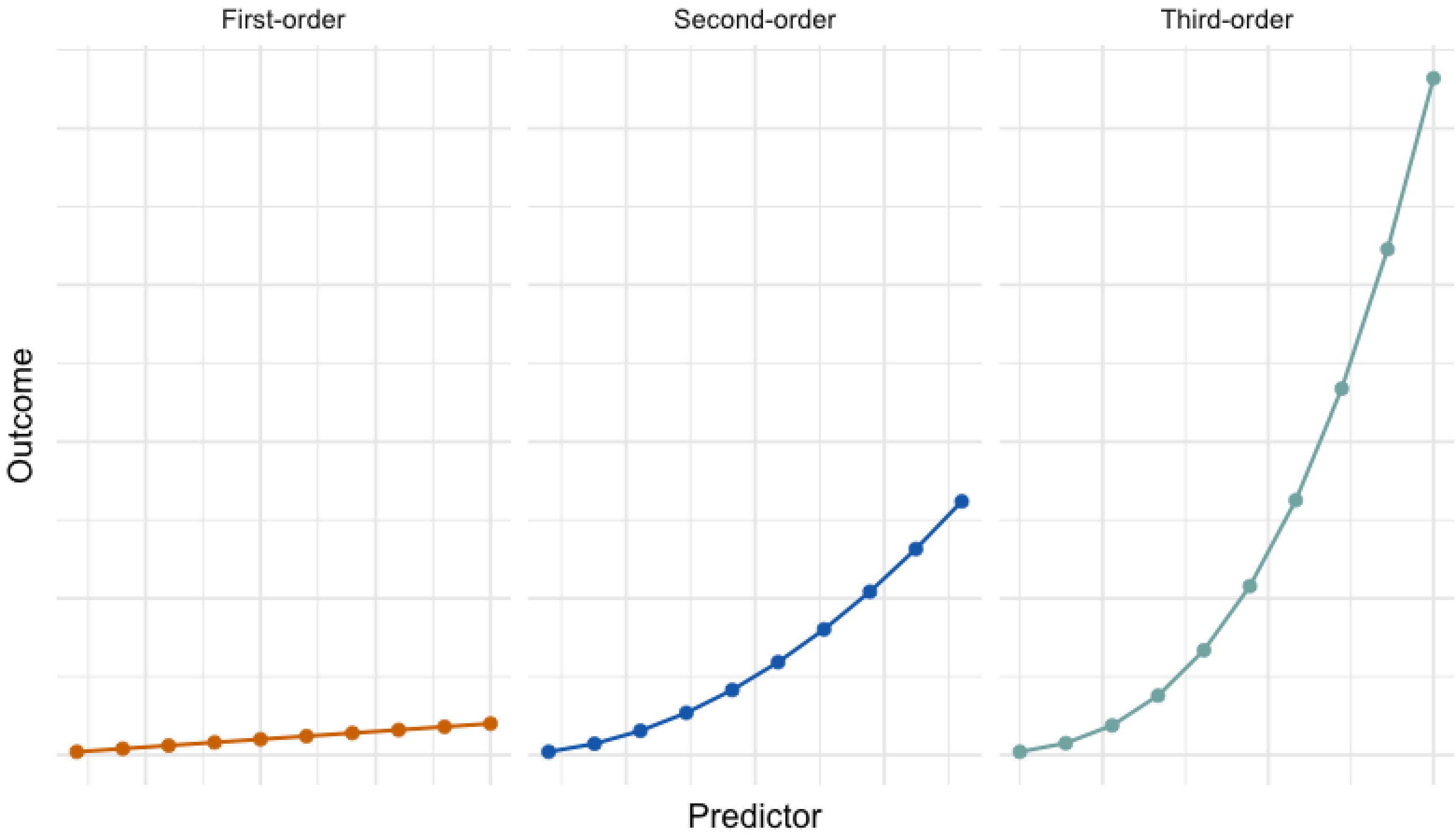
Learning outcomes

- Describe what a growth model is
- Describe what an autoregressive covariance structure is
- Distinguish fixed from random effects
- Be able to interpret a growth model

Growth models

- Growth models look at the rate of change of a variable over time
    - Depression over 8 weeks of treatment
    - Back pain over 10 weeks of physiotherapy
    - Profits over months of the year
    - Radioactive decay

Types of growth curve



A rehabilitative example

- id: Zombie id
- intervention: Was the participant assigned to waiting list (0) or gene therapy (1)?
- time: When was outcome measured: baseline, 1, 6 and 12 month follow up
- time_num: Time since treatment (expressed numerically): 0, 1, 6, 12
- resemblance: Resemblance of their face to their pre-zombie state (0% to 100%).

The data in R

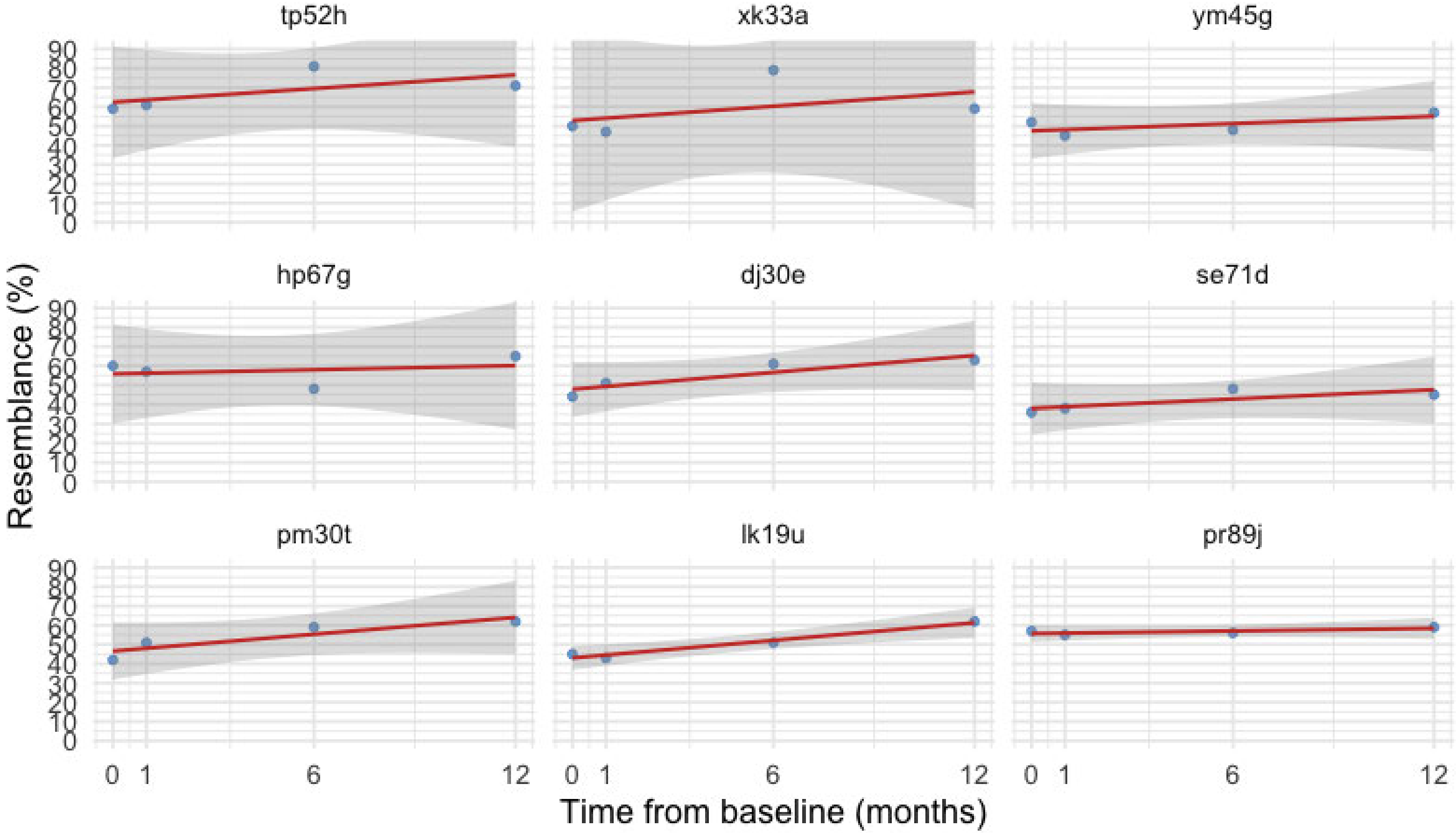
Table 1: Data for zombie rehabilitation

	id	intervention	time	resemblance	time_num
1	tp52h	Gene therapy	Baseline	59	0
2	bm95v	Gene therapy	Baseline	31	0
3	lj12s	Gene therapy	Baseline	35	0
4	jb38i	Wait list	Baseline	54	0
5	xt29u	Wait list	Baseline	49	0
6	ep42w	Gene therapy	Baseline	49	0
7	po11r	Wait list	Baseline	36	0
8	nz76p	Gene therapy	Baseline	48	0
9	xk33a	Gene therapy	Baseline	50	0
10	mp79j	Gene therapy	Baseline	48	0

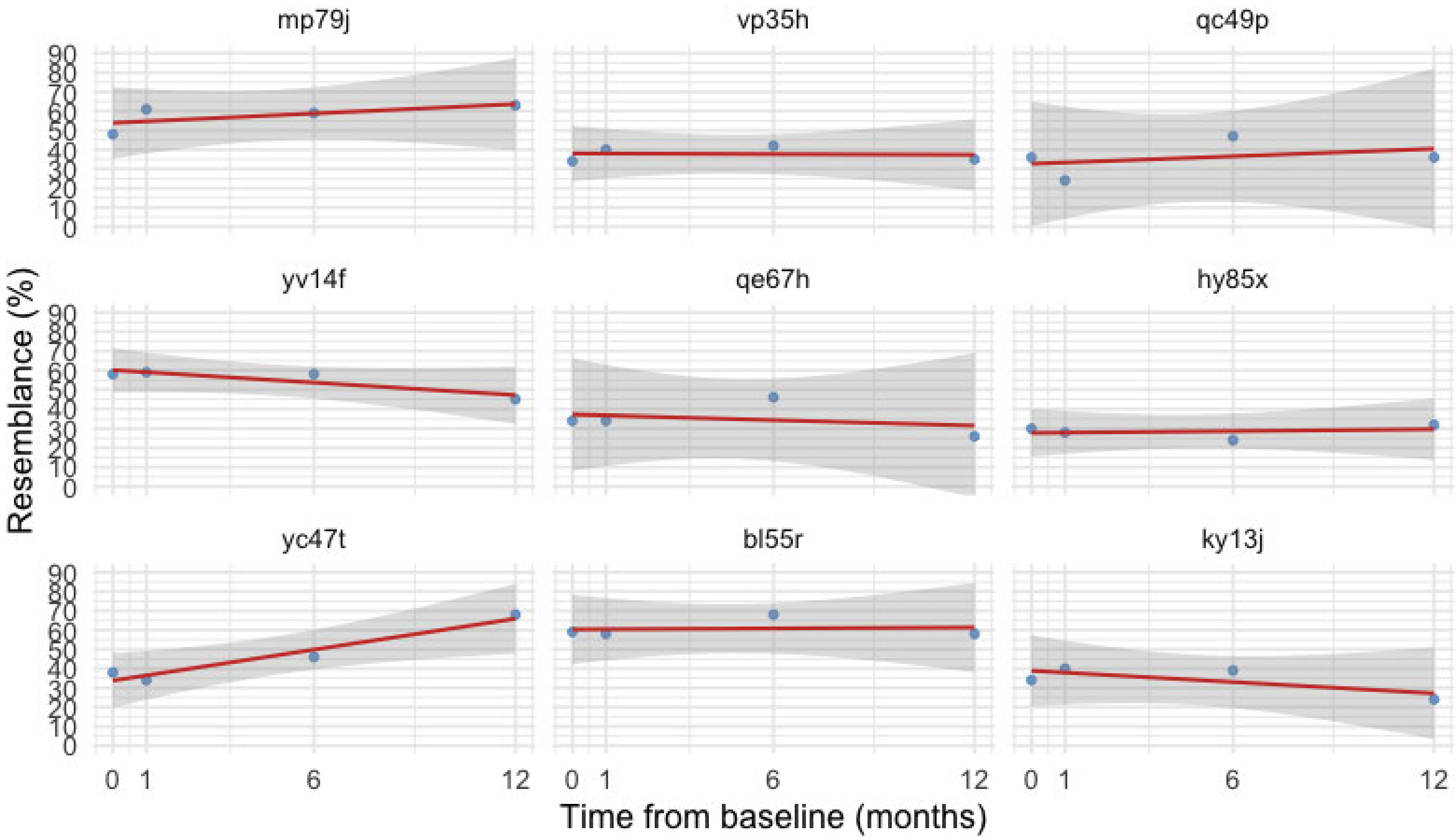
Previous 1 2 3 4 5 ... 57 Next


Level
2
Zombie 1
Zombie 2
Zombie 3
...
Zombie 141
1
59
61
81
71
31
45
33
39
35
41
34
28
...
41
38
48
43

Fixed slopes



Random slopes



Covariance structures

The model has an error matrix

$$\Phi = \begin{pmatrix} \text{var}(e_{t1}) & \text{cov}(e_{t1}, e_{t2}) & \cdots & \text{cov}(e_{t1}, e_{tn}) \\ \text{cov}(e_{t2}, e_{t1}) & \text{var}(e_{t2}) & \cdots & \text{cov}(e_{t2}, e_{tn}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(e_{tn}, e_{t1}) & \text{cov}(e_{tn}, e_{t2}) & \cdots & \text{var}(e_{tn}) \end{pmatrix}$$

$$\Phi_{\text{Default}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Covariance structures

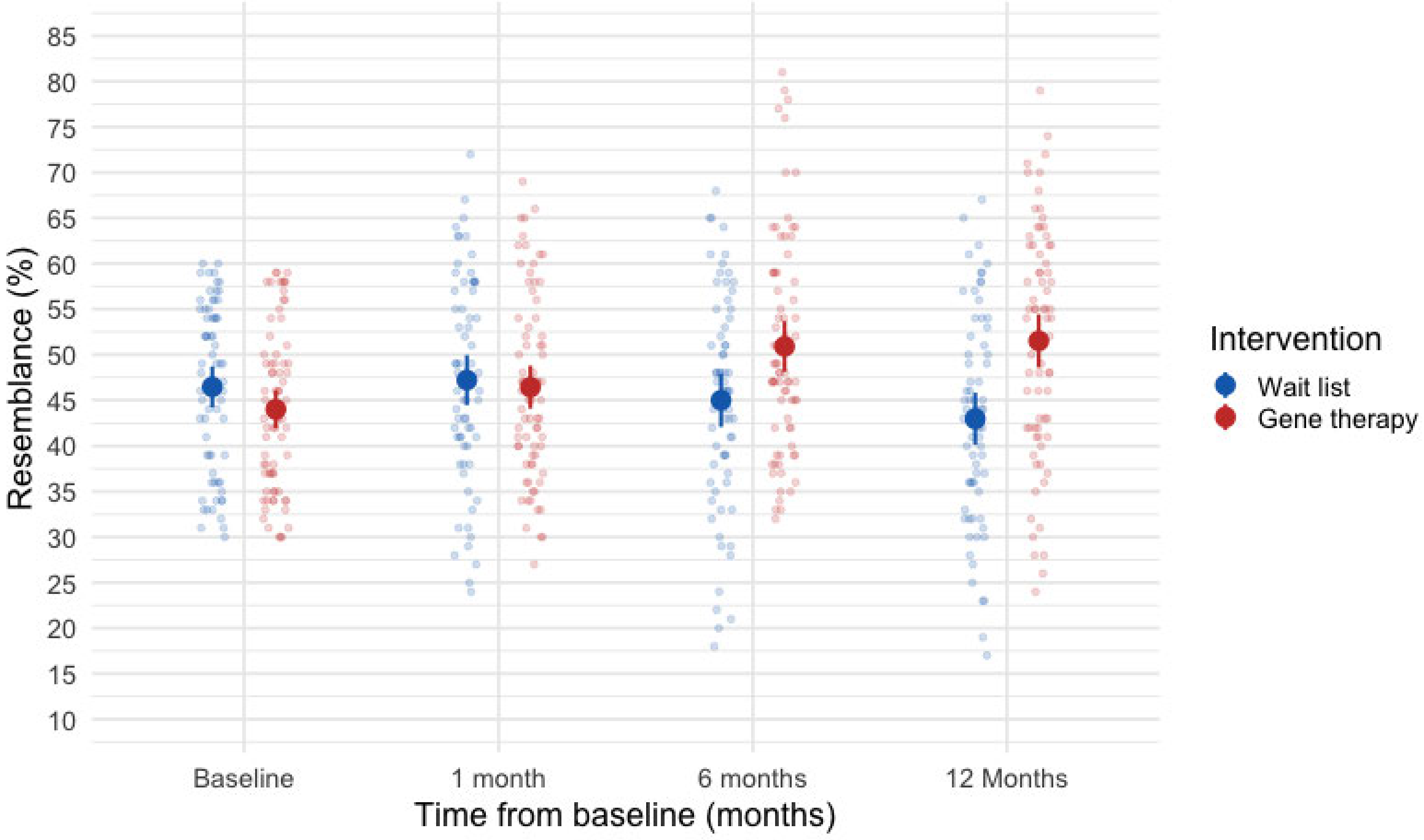
$$\Phi_{\text{Compound symmetry}} = \begin{pmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{pmatrix}$$

$$\Phi_{\text{AR(1)}} = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{pmatrix}$$

The RM-ANOVA approach

- Time is treated as a factor (4 levels)
- Restrictions
  - Effects of time constant across individuals
  - Residuals between time points correlated to the same degree (compound symmetry)
- You end up with a mixed design
  - Predictor 1 (repeated-measures/within-participant)
    - time (baseline, 1 month, 6 months, 12 months)
  - Predictor 2 (independent measures/between group)
    - intervention (wait list, gene therapy)
  - Outcome: resemblance scores (0-100)

RM-ANOVA: how are we treating time?



Traditional approach (RM-ANOVA) using `afex`

```
rehab_afx <- afex::aov_4(resemblance ~ time*intervention + (time|id), data = rehab_tib)
rehab_afx
```

Effect	num Df	den Df	MSE	F	ges	Pr(>F)
intervention	1.0	139.00	366.66	3.01	0.02	.085
time	2.4	333.35	49.85	4.67	0.01	.006
intervention:time	2.4	333.35	49.85	24.35	0.04	<.001

Basically a wrapper for

```
aov(resemblance ~ time*intervention + Error(id), data = rehab_growth_tib) %>%
  summary()
```

Multilevel approach (restricted)

```
rehab_cs <- nlme::lme(resemblance ~ time*intervention, random = ~ 1|id, data = rehab_tib , method
= "ML", correlation = corCompSymm (form = ~ time|id))
anova(rehab_cs)
```

	numDF	denDF	F-value	p-value
(Intercept)	1	417	3379.21	0.00
time	3	417	5.58	0.00
intervention	1	139	3.01	0.08
time:intervention	3	417	24.35	0.00

Limitations of RM-ANOVA for mixed designs

- Time treated unrealistically
  - Doesn't model unequal timepoints
  - Treats time as categorical not continuous
- The effect of time treated as the same across individuals
  - Can’t model variability in individual change over time (i.e. random slopes)
- No flexibility in covariance structure
- Can’t handle missing data
- Can’t model non-linear change
- (But probably better in small samples)

The multilevel approach

$$\widehat{\text{resemblance}}_{ij} = \hat{\pi}_{0i} + \hat{\pi}_{1i}\text{time}_{ij} + e_{ij}$$

$$\hat{\pi}_{0i} = \hat{\gamma}_{00} + \hat{\gamma}_{01}\text{intervention}_i + \hat{\zeta}_{0i}$$

$$\hat{\pi}_{1i} = \hat{\gamma}_{10} + \hat{\gamma}_{11}\text{intervention}_i + \hat{\zeta}_{1i}$$

$i = \text{individual}, \quad j = \text{occasion}$

- $\hat{\gamma}_{00}$ = average baseline resemblance when intervention = 0 (wait list)
- $\hat{\gamma}_{10}$ = average rate of change in resemblance when intervention = 0 (wait list)
- $\hat{\gamma}_{01}$ = baseline difference between wait list and gene therapy
- $\hat{\gamma}_{11}$ = difference in rate of change in resemblance between wait list and gene therapy groups
- $\hat{\zeta}_{0i}$ = deviation of individual's baseline resemblance from group average
- $\hat{\zeta}_{1i}$ = deviation of individual's rate of change in resemblance from group average
- e_{ij} = portion individual's resemblance score that is unpredicted at time *j*

The same model (in composite form)

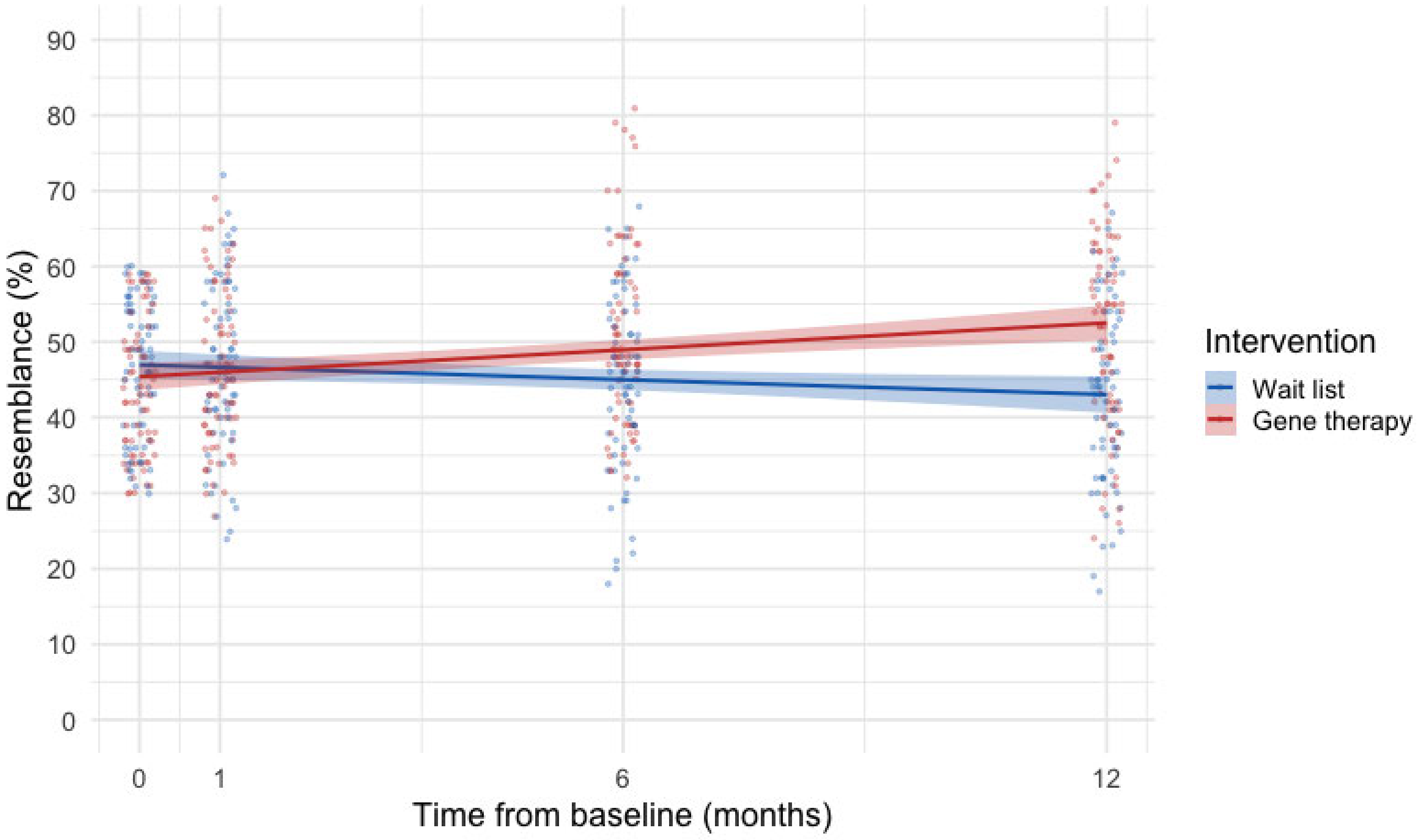
$$\text{resemblance}_{ij} = [\hat{\gamma}_{00} + \hat{\gamma}_{10}\text{time}_{ij} + \hat{\gamma}_{01}\text{intervention}_i + \hat{\gamma}_{11}\text{intervention}_i \times \text{time}_{ij}] + [\hat{\zeta}_{0i} + \hat{\zeta}_{1i}\text{time}_{ij} + e_{ij}]$$

- $\hat{\gamma}_{00}$ = average baseline resemblance when intervention = 0 (wait list)
- $\hat{\gamma}_{10}$ = average rate of change in resemblance when intervention = 0 (wait list)
- $\hat{\gamma}_{01}$ = baseline difference between wait list and gene therapy
- $\hat{\gamma}_{11}$ = difference in rate of change in resemblance between wait list and gene therapy groups
- $\hat{\zeta}_{0i}$ = deviation of individual's baseline resemblance from group average
- $\hat{\zeta}_{1i}$ = deviation of individual's rate of change in resemblance from group average
- e_{ij} = portion individual's resemblance score that is unpredicted at time *j*

Advantages of multilevel framework

- Time treated realistically
  - You can model unevenly spaced or discontinuous measurements
  - Time can be treated continuously
- Model variability in individual change over time (i.e. random slopes)
- Flexibility in covariance structure
  - Fit AR(1) covariance structure
- Can handle missing data
- Can model non-linear change
- (But may not converge in small samples)

First-order growth model



First-order model

```
rehab_rs <- nlme::lme(
  resemblance ~ time_num*intervention,
  random = ~ time_num|id,
  data = rehab_tib,
  method = "ML"
  )
```

Fixed effects

```
anova(rehab_rs)
```

Effect	numDF	denDF	F-value	p-value
(Intercept)	1	421	3466.43	<.001
time_num	1	421	5.73	.017
intervention	1	139	0.22	.640
time_num:intervention	1	421	50.33	<.001

Parameter estimates

Fixed effects

```
broom.mixed::tidy(rehab_rs, effects = "fixed")
```

term	estimate	std.error	df	statistic	p.value
(Intercept)	46.96	1.17	421	40.12	<.001
time_num	-0.33	0.09	421	-3.49	<.001
interventionGene therapy	-1.58	1.62	139	-0.98	.330
time_num:interventionGene therapy	0.92	0.13	421	7.09	<.001

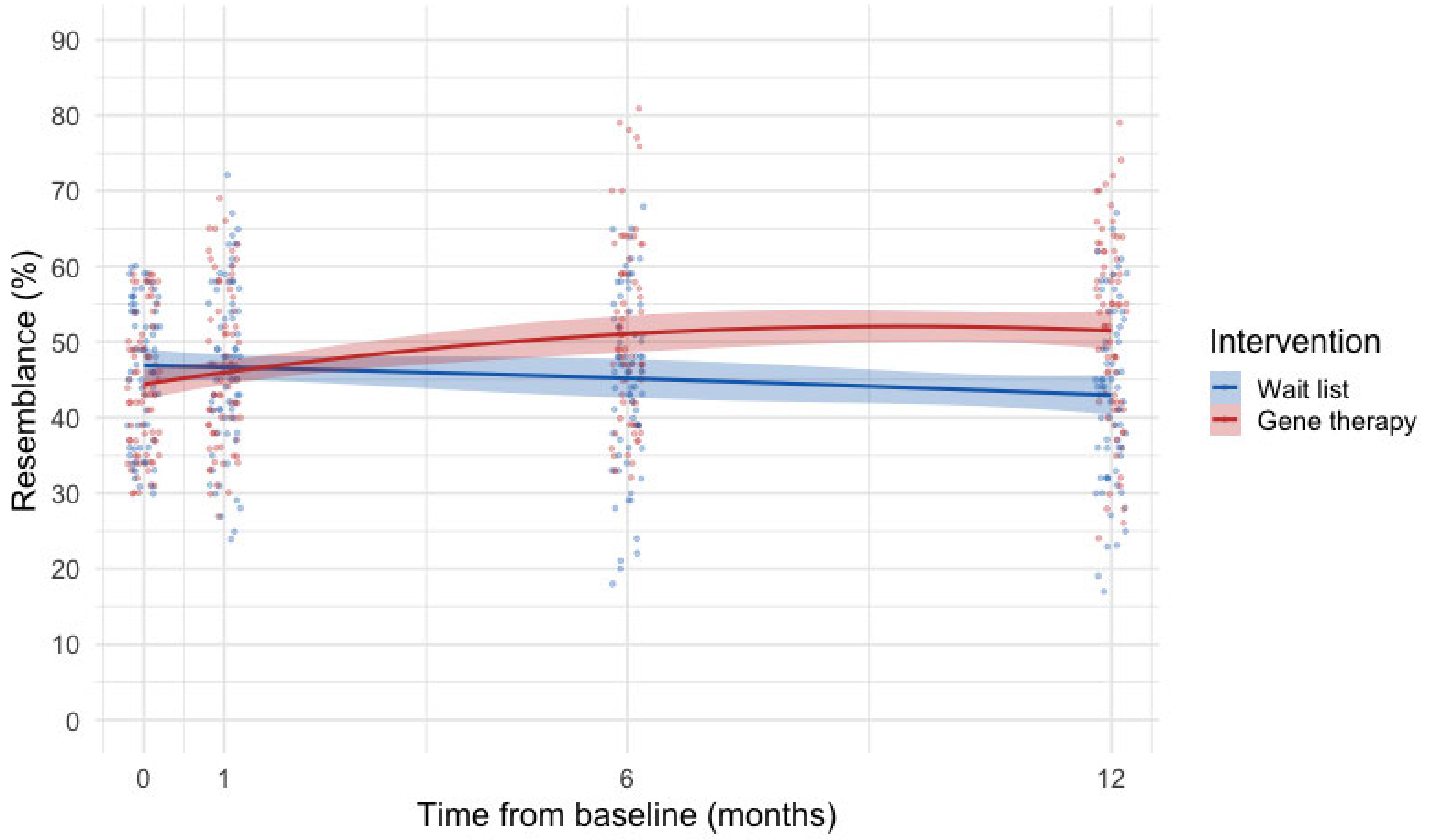
Parameter estimates

Random effects

```
broom.mixed::tidy(rehab_rs, effects = "ran_pars")
```

effect	group	term	estimate
ran_pars	id	sd_(Intercept)	8.612
ran_pars	id	cor_time_num.(Intercept)	0.074
ran_pars	id	sd_time_num	0.464
ran_pars	Residual	sd_Observation	5.837

A second-order growth model



Second-order model

- We add two fixed effects
  - Fixed effect of time^2
    - `poly(time_num, 2)`
  - Fixed effect of the interaction between intervention and time^2
    - `poly(time_num, 2):intervention`
- We change the random effect
  - `random = ~poly(time_num, 2)|id`

Second-order model

```r
rehab_so <- nlme::lme(
  resemblance ~ poly(time_num, 2)*intervention,
  random = ~poly(time_num, 2)|id,
  data = rehab_tib,
  method = "ML"
  )
```

Fixed effects

```r
anova(rehab_so)
```

Effect	numDF	denDF	F-value	p-value
(Intercept)	1	419	3522.27	<.001
poly(time_num, 2)	2	419	7.83	<.001
intervention	1	139	0.07	.785
poly(time_num, 2):intervention	2	419	31.76	<.001

Parameter estimates

Fixed effects

```
broom.mixed::tidy(rehab_so, effects = "fixed")
```

term	estimate	std.error	df	statistic	p.value
(Intercept)	45.40	1.17	419	38.88	<.001
poly(time_num, 2)1	-37.19	10.68	419	-3.48	<.001
poly(time_num, 2)2	-2.45	9.55	419	-0.26	.798
interventionGene therapy	2.80	1.61	139	1.74	.084
poly(time_num, 2)1:interventionGene therapy	104.40	14.74	419	7.08	<.001
poly(time_num, 2)2:interventionGene therapy	-27.82	13.19	419	-2.11	.035

Parameter estimates

Random effects

```r
broom.mixed::tidy(rehab_so, effects = "ran_pars")
```

effect	group	term	estimate
ran_pars	id	sd_(Intercept)	9.215
ran_pars	id	cor_poly(time_num, 2)1.(Intercept)	0.242
ran_pars	id	cor_poly(time_num, 2)2.(Intercept)	-0.229
ran_pars	id	sd_poly(time_num, 2)1	67.005
ran_pars	id	cor_poly(time_num, 2)2.poly(time_num, 2)1	0.383
ran_pars	id	sd_poly(time_num, 2)2	54.587
ran_pars	Residual	sd_Observation	4.667

To sum up ...

- A common form of repeated-measures data comes from longitudinal studies
- Growth models quantify change over time
- Can factor in between-participant measures
- Multilevel models
  - Treat observations as nested within entities
  - Allow you model individual differences in growth
  - Allow you to look at different covariance structures
  - Cope with missing data
  - Can model non-linear growth