

S-043/Stat-151
Analysis for Clustered and Longitudinal Data
(Multilevel & Longitudinal Models)

Lectures 5.3
Longitudinal Data with
a binary outcome

A Case Study

Reading:
Chapter 10 of R-H & S. In particular 10.3-10.13

See free online chapter:

http://www.stata-press.com/books/mlmus3_ch10.pdf

Goals for Today

Today we take a nice deep look at a specific case study:
binary outcomes with longitudinal data

In particular we discuss:

- ★ Fitting such models
- ★ Generating and looking at graphs showing the predicted probabilities of individuals across time
- ★ Interpreting coefficients of these models
- ★ Understanding the difference between population and individual trends (and that random effects make fixed effects individual, not population)

Also

- ★ We briefly look at missing data and exploring missing data.

An important skill: Learning from other fields

Just because the data
are from someplace
bizarre, you should
not turn away.

That being said...



Toenail Infection Data (sorry)

378 patients

Randomly assigned to two oral antifungal treatments (terbinafine, itraconazole)

Followed for 7 visits at weeks 0, 4, 8, 12, 24, 36, 48

Outcome: Dichotomization of onycholysis (separation of nail from bed)

```
>  
> toes = read.dta( "toenail.dta" )  
> head( toes )  
  patient outcome treatment month visit  
1       1        1      1 0.00     1  
2       1        1      1 0.86     2
```

Bird's Eye View of the Analysis

The Randomized Controlled Trial (RCT) gives us two groups that should be the same at baseline.

We watch these two groups evolve over time.

Individuals all have their own trajectories, but we are looking at the aggregate trends of the groups.

We want to see if they evolve *differently* for the different treatments.

Research Questions & Analytic Goals

RQ1: Did the treatment work?

- ★ This is about inference.
- ★ We operationalize it as whether there is a difference in average growth rates in the latent probability of detachment.
- ★ In other words, does the chance of having a detached nail decline faster with one treatment vs another?

RQ2: Did the treatment work well?

- ★ Now we want to measure the relative speed of how these rates change.
- ★ The main struggle here is getting our impacts in an interpretable manner.
 - We can get that in an odds multiplier (treatment A reduces the odds of detachment by # more per month than treatment B)
 - We can also use visualization.

Let's recover by looking at this three-toed sloth's toenails



Looking at Missingness

```
> head( miss )
```

	V1	n
--	----	---

18	XXXXXXX	224
----	---------	-----

16	XXXXX.X	21
----	---------	----

14	XXX.XX	10
----	--------	----

7	XXX....	6
---	---------	---

1	X.....	5
---	--------	---

15	XXXXX..	5
----	---------	---

Not all patients showed up
for all of our sessions to get their
toes examined.

“Monotone missingness” - when you leave and
never come back.

```
> 224 / length( unique( toes$patient ) )
```

```
[1] 0.76
```

Percent of patients with complete data.

How does
this count
patients?

MAR (“Missing at Random”) assumption allows us to
exploit the 24% of patients with some missing data

Supplementary read: RH&S 5.8 for a bit more on missing data

Looking at outcome data

We see strings of 1s and 0s

Lots of autocorrelation
(nearby measures are similar)

Could also think of it as a
growth curve on a **latent**
probability of detachment.

Many possible modeling
paths.

	patient	tx	out
1		1	1110000
2		2	0 001100.
3		3	0 0000001
4		4	0 1000000
5		6	1 1110000
6		7	1 0000111
7		9	1 0000000
8		10	0 0000000
9		11	1 1111000
10		12	0 0100110
11		13	0 1111000
12		15	1 11000.0
13		16	1 0000.10
14		17	0 11110.0
15		18	0 001.000
16		19	1 0000000
17		20	0 000.000

How does model fitting work in the ideal world?

- ★ You specify your model
- ★ You write it down in computer code
- ★ The computer fits your model
- ★ You then have:
 - Point estimates of your parameters
 - Uncertainty estimates of your parameters



Aggregating our Data: Looking at proportions for each visit by treatment combo

```
sumstat = toes %>% group_by( visit,treatment ) %>%  
  summarise( prop = mean(outcome) ,  
             mnmonth=mean(month)  
             n=length(outcome) )  
  
head( sumstat )
```

	visit	treatment	prop	mnmonth	n
1	1	1	0	0.37	0.0 146
2	1	1	1	0.37	0.0 148
3	2	0	0	0.35	1.0 141
4	2	1	1	0.33	1.0 147
5	3	0	0	0.32	2.1 138
6	3	1	1	0.28	2.1 145

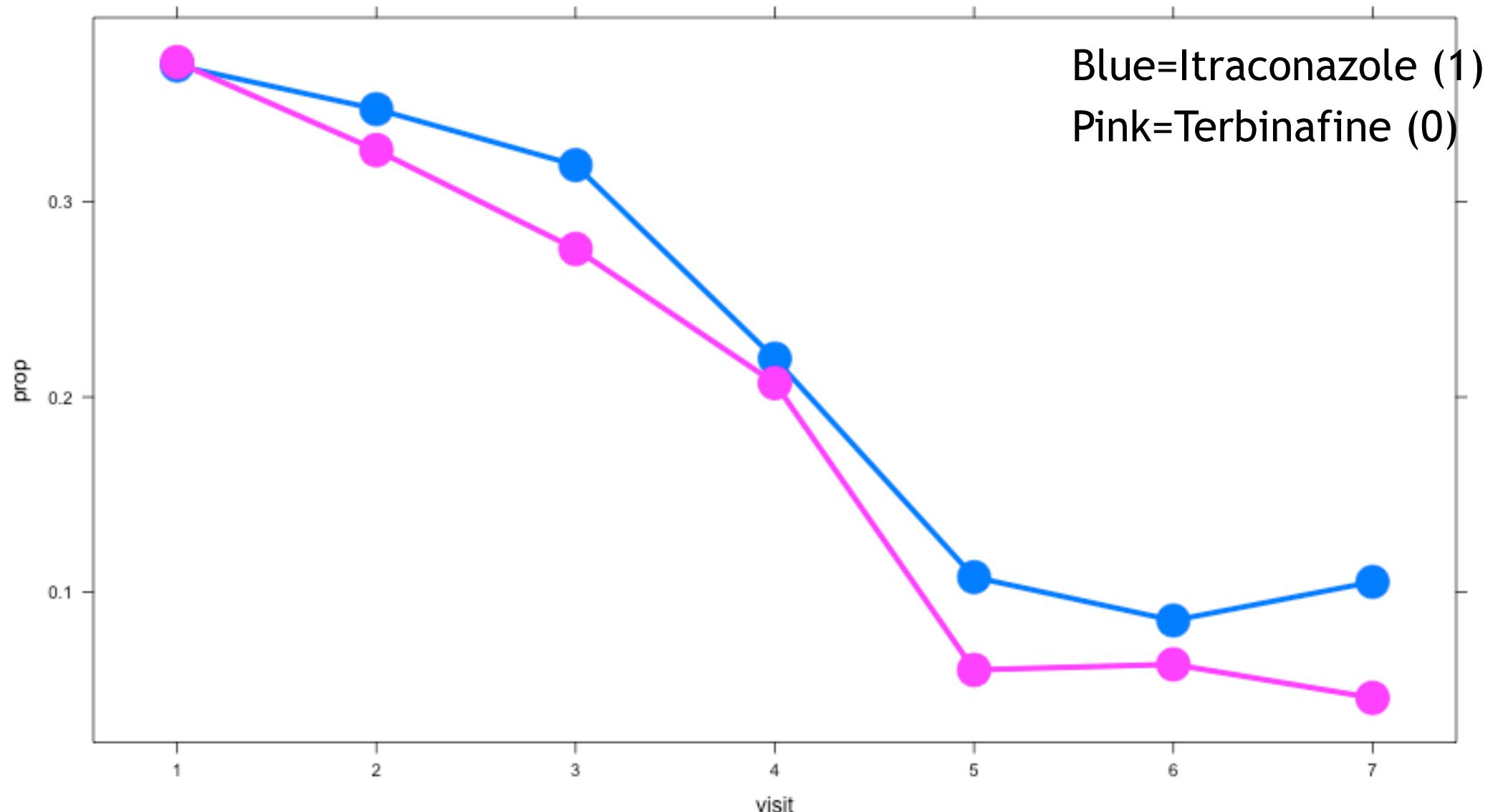
This says: group all our data by visit number and treatment. Within each, get proportion with detachment, average time of visit, and number of observations

The mean(outcome) when outcome is 0/1 variable to get a proportion is a very important trick.

We turn our treatment number into a factor so ggplot thinks of it as categorical rather than continuous

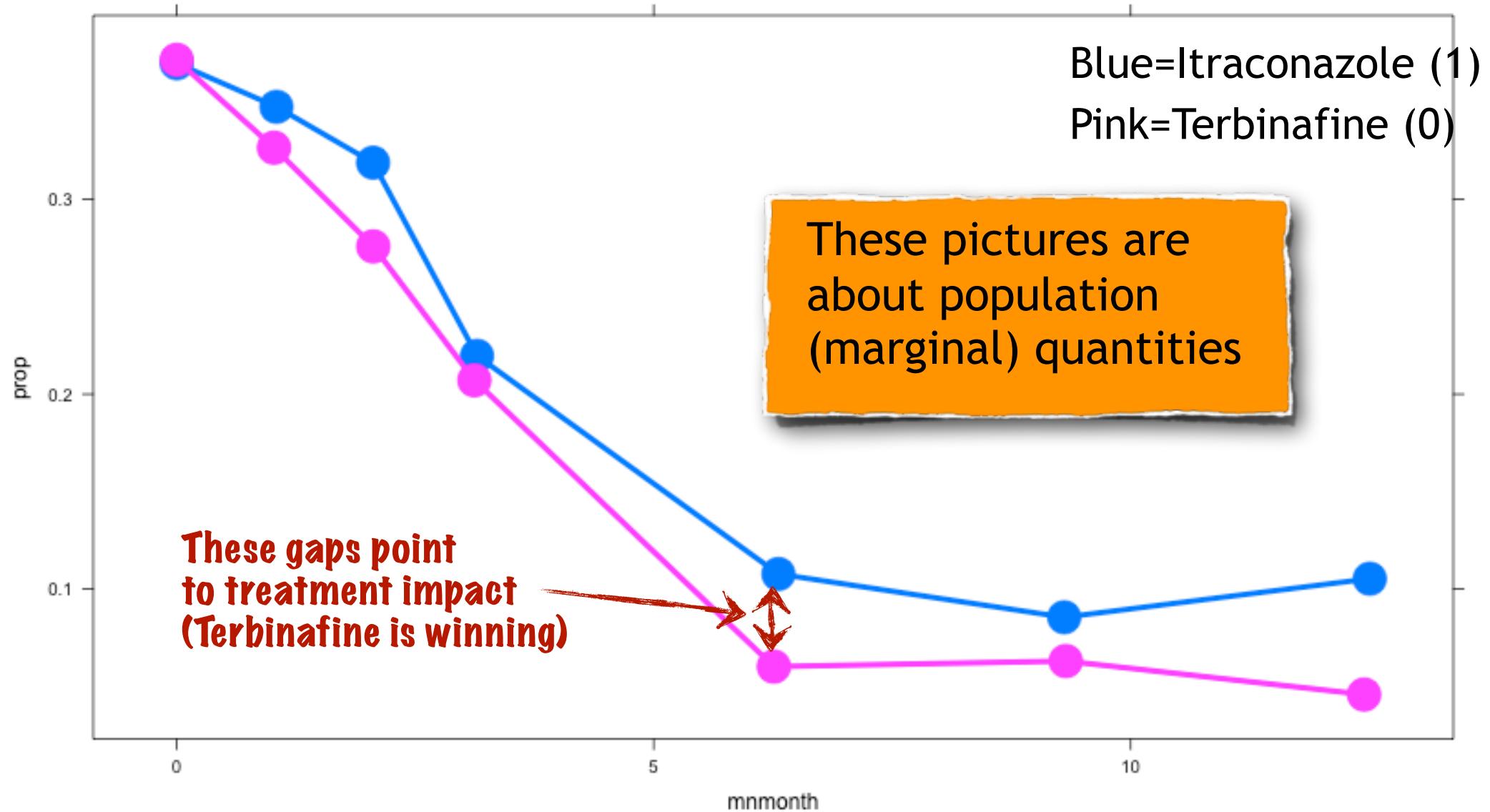
```
> sumstat$treatment = as.factor( sumstat$treatment )  
> ggplot( data=sumstat, aes( x=visit, y=prop, col=treatment ) ) +  
  geom_point() + geom_line()
```

Proportion with detachment



We bin each visit and calculate the marginal proportion with detachment at each visit.

Proportion with detachment



We plot based on mean time for the waves to make x-axis more interpretable. Notice clustering of earlier visits.

Model to confirm a straightforward descriptive analysis

Q:

It looks like treatment worked, but perhaps the pattern of results is just noise.

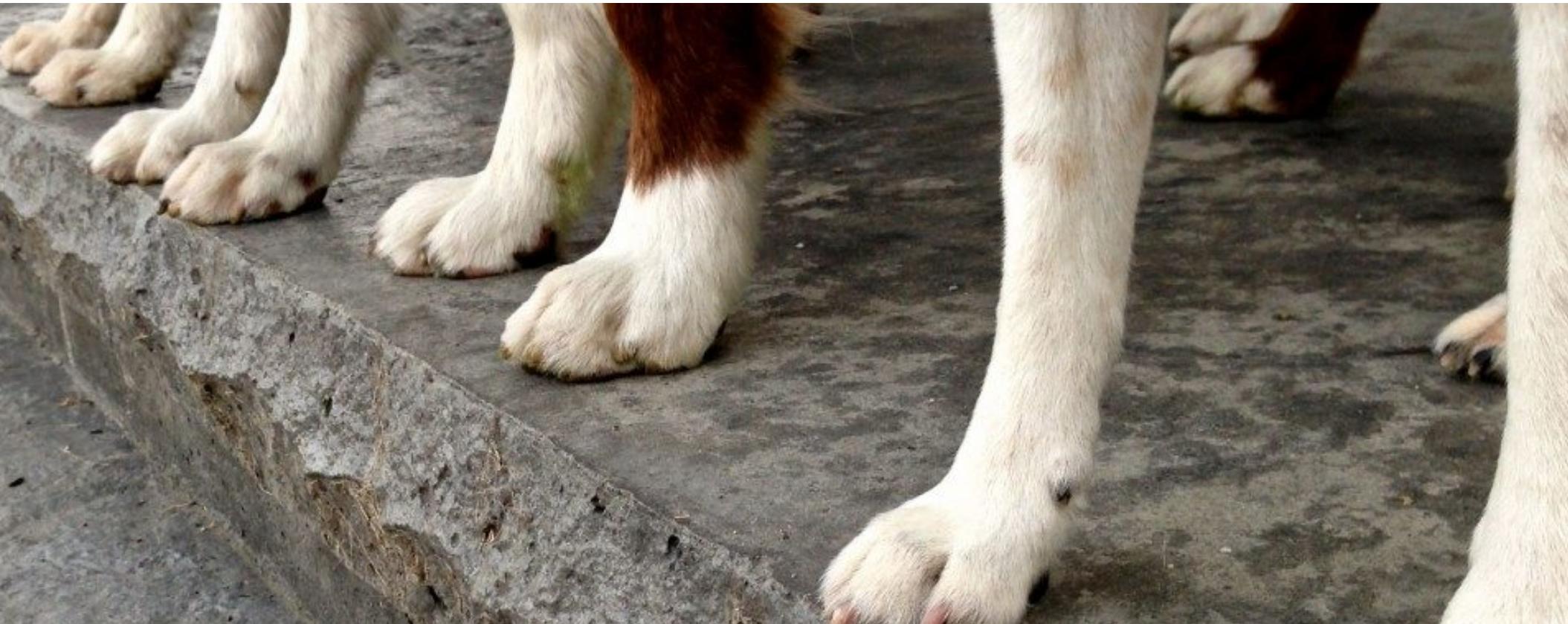
How do we adjust for possible bias due to missing data (differential dropout) and guard against random chance?

A:

Multilevel Modeling!

(But first, we model to give a smoother picture of our data.)

Marginal models: population average effects





Marginal Model (no random effects)

This is a “marginal model.”
It is “population averaged.”

```
> M0 = glm( outcome ~ treatment * month,  
+           family=binomial,  
+           data=toes )
```

```
> display( M0 )
```

	coef.est	coef.se
(Intercept)	-0.56	0.11
treatment	0.00	0.16
month	-0.17	0.02
treatment:month	-0.07	0.04

n = 1908, k = 4

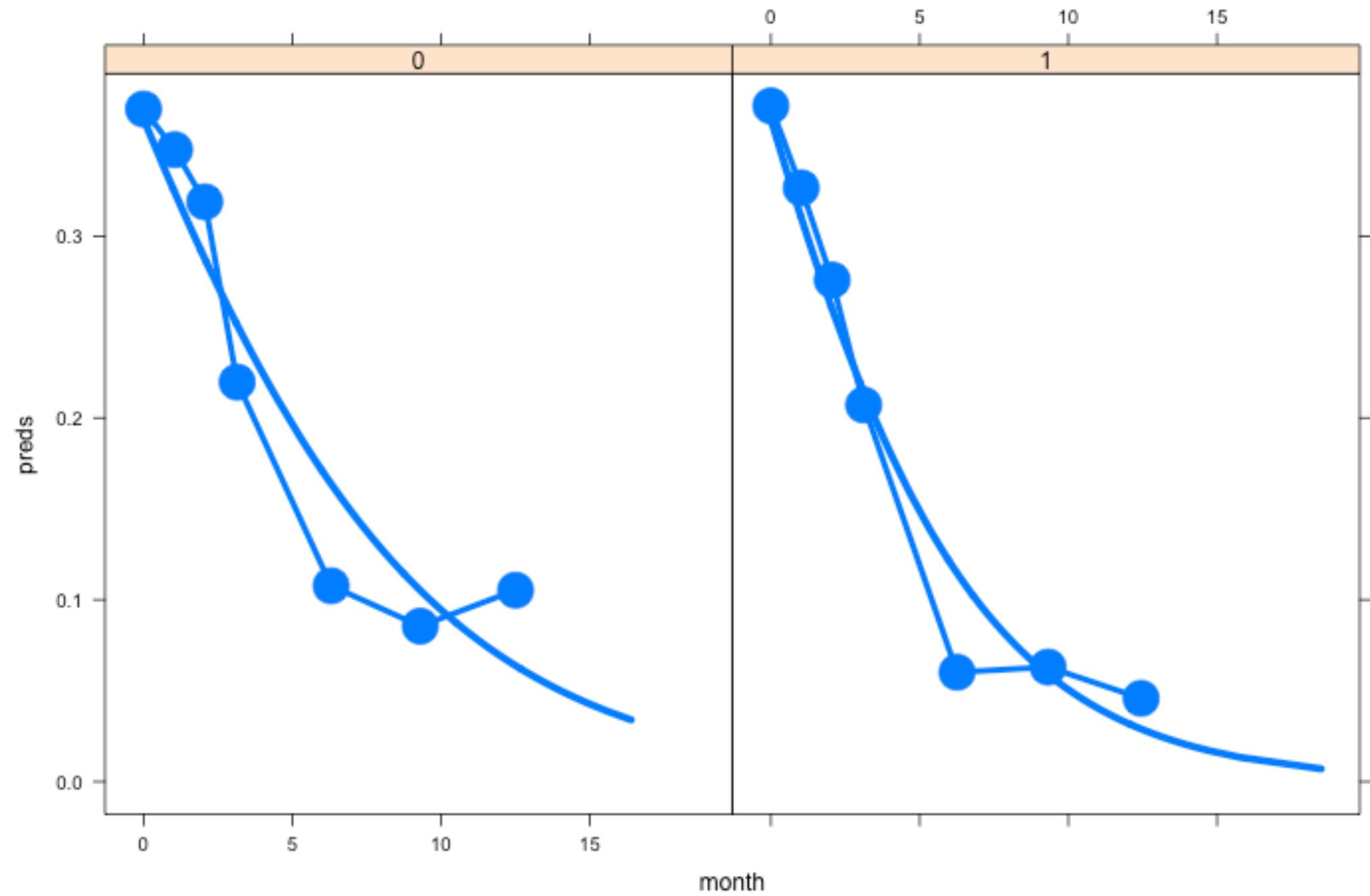
residual deviance = 1816.0, null deviance = 1980.5 (dif

```
> exp( coef( M0 ) )
```

(Intercept)	treatment	month	treatment:month
0.57	1.00	0.84	0.93

The standard errors will
be quite off here,
because we are assuming
1908 independent
observations.

Predicted vs Actual



Two General Strategies

Population Focus:

What we just saw.

- ★ We model the population averaged curves.
- ★ We would need to adjust the standard errors with robust techniques to deal with lack of independence within subjects

Individual Focus:

What comes next.

- ★ We model individual tendencies for toenail separation across time.
- ★ We have a greater dependence on the model.
- ★ We can describe individual variation.

Latent growth models
(Multilevel Modeling)
targeting individual growth curves



Latent growth curves with a random intercept model

$$Y_{ij} \sim \text{Binomial}(1, \pi_{ij}) \equiv \text{Bern}(\pi_{ij})$$

A Bernoulli (coin flip)
is a Binomial with n=1

$$\text{logit}(\pi_{ij}) = \gamma_{00} + \gamma_{01}Z_j + \gamma_{10}\text{Time}_{ij} + \gamma_{11}Z_j\text{Time}_{ij} + u_j$$

$$u_j \sim N(0, \tau_{00})$$



The reduced-form
(marginal) specification

$$Y_{ij} \sim \text{Binomial}(1, \pi_{ij})$$

$$\text{logit}(\pi_{ij}) = \beta_{0j} + \beta_{1j}\text{Time}_{ij}$$

The multilevel
formulation.
(Note: no random slope)

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + u_j$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}Z_j$$

$$u_j \sim N(0, \tau_{00})$$

Logistic is more demanding, assumption-wise

In contrast to linear random effects models, **consistent estimation** requires that the **random part** of model is correctly specified (in addition to fixed part).

Specifically we need

- ★ Correct linear predictor
- ★ Correct link
- ★ Correct specification of which covariates have random coefficients
- ★ Conditional independence of responses given random effects and covariates
- ★ Independence of random effects and covariates (for causal inference)
- ★ Normally distributed random effects



This one less critical



Fitting the Random Intercept Model

```
> M1 = glmer( outcome ~ treatment * month + (1|patient) ,  
  family=binomial ,  
  data=toes )
```

```
>
```

```
>
```

```
> display( M1 )
```

		coef.est	coef.se
(Intercept)	γ_{00}	-2.51	0.76
treatment	γ_{01}	-0.30	0.69
month	γ_{10}	-0.40	0.05
treatment:month		-0.14	0.07
	γ_{11}		

Error terms:

Groups	Name	τ_{00}	Std.Dev.
patient	(Intercept)	4.56	
Residual		1.00	

number of obs: 1908, groups: patient, 294
AIC = 1265.6, DIC = -26
deviance = 615.0



Did our treatment work?



What is the probability of a median treatment individual having an infection 3 months into the study?

What is the probability of a median treatment individual having an infection 3 months into the study?

Our patient variation. What is a range of plausible intercept values?



This is still junk—an artifact of the display() call.



Interpreting the coefficients

	coef.est	coef.se
(Intercept)	-2.51	0.76
treatment	-0.30	0.69
month	-0.40	0.05
treatment:month	-0.14	0.07



Did our treatment work?

This is RQ 1!

Error terms:

Groups	Name	Std. Dev.
patient	(Intercept)	4.56



What is the probability of a median treatment individual having an infection 3 months into the study?



What is the probability of a treated patient with a 1 SD above average proclivity for infection having an infection at 3 months into the study?



Describing treatment impact: Estimated odds ratios and confidence intervals

This begins RQ 2!

```
> exp( fixef( M1 ) )
```

	treatment	month	treatment:month
(Intercept)	0.737	0.670	0.872
0.081			

```
> cis = confint( M1 )
```

This takes awhile

Computing profile confidence intervals ...

```
> cis
```

	2.5 %	97.5 %
.sig01	3.56	7.386
(Intercept)	-5.27	-1.326
treatment	-1.70	1.047
month	-0.50	-0.314
treatment:month	-0.28	-0.003

```
> exp( cis[-1,] )
```

	2.5 %	97.5 %
(Intercept)	0.0051	0.27
treatment	0.1821	2.85
month	0.6070	0.73
treatment:month	0.7576	1.00

The [-1,] drops the first row of our matrix of results.
(This is “slicing” our matrix.)

These are conditional odds, meaning the odds multipliers conditioned on the value of the random effect



Interpreting these values

```
> exp( fixef( M1 ) )  
             (Intercept)      treatment      month  
             0.081          0.737        0.670  
                                         treatment:month  
                                         0.872
```

```
> 0.67 * 0.872  
[1] 0.58
```

The conditional odds of detachment of those in Terbinafine group is multiplied by 0.58 per month (vs. 0.67 for Intraconazole)

```
> 100 * (0.67*0.872 - 1)  
[1] -42
```

```
> 100 * (0.67 - 1)  
[1] -33
```

We see a 42% decrease in the percentage change in estimated odds of detachment each month for the Terbinafine group (vs 33% for Intra.)

population view
vs.
individual view





With logistic models, random effects make fixed effects conditional

Fixed effect model (population averaging)

```
> coef( M0 )
  (Intercept)      treatment       month treatment:month
  -0.55663        -0.00058      -0.17031        -0.06722
> exp( coef( M0 ) )
  0.573            0.999          0.843          0.935
```

Random intercept model

```
> fixef( M1 )
  (Intercept)      treatment       month treatment:month
  -2.51          -0.30          -0.40          -0.14
> exp( fixef( M1 ) )
  0.0813         0.7372         0.6705         0.8719
```

Random intercept model has more extreme coefficients.

The random intercept model is *conditioning* (giving subject-specific or conditional probabilities)

The fixed effect model is averaging over different individuals (giving population-averaged or marginal probabilities).

Comparing population vs. subject

Month:

- ★ Subject-specific is 0.67: for a specific patient, we reduce our predicted odds by 33% per month.
- ★ Population average is 0.84: the odds of having onycholysis *among the patients* decreases by 16% per month.

Treatment:

- ★ Treatment is between-subject, yet for random intercept model we interpret within-subject. This requires serious model dependence (independence of random effect and coefficient in particular)

A mathematical view of population averaging from a random intercept model

For each subject, we have an individual growth curve to predict the probability of detachment at each time point
We could average these across subjects (i.e., the random intercept distribution) to get predicted population averages:

$$\Pr(y_{ij} = 1 | x_{2j}, x_{3ij})$$

Average individual chance of detachment given x_{2j} (treatment) and x_{3ij} (month)

$$= \int \Pr(y_{ij} = 1 | x_{2j}, x_{3ij}, \zeta_j) \phi(\zeta_j; 0, \psi) d\zeta_j$$

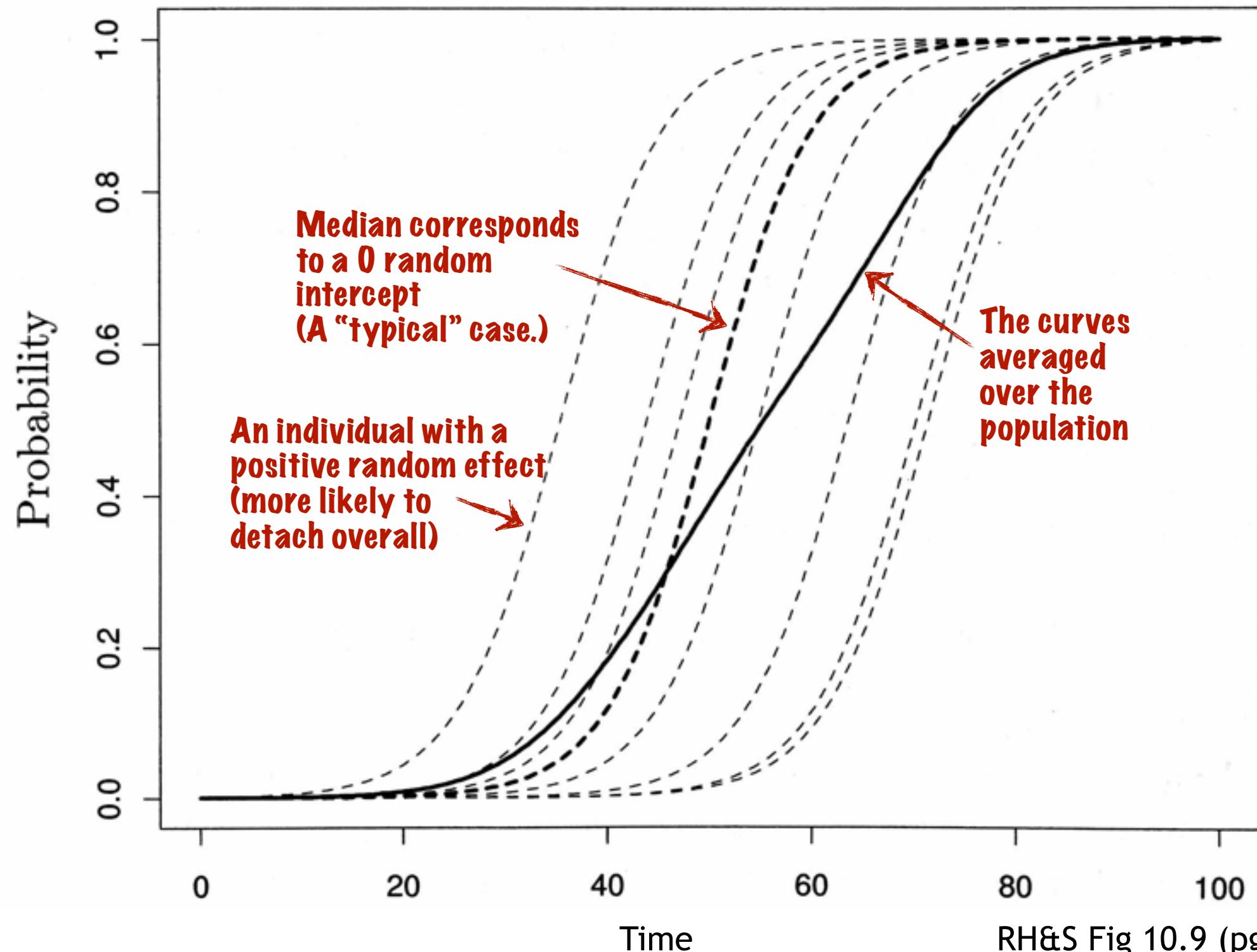
Is the average of everyone's chances at that month with that treatment

$$= \int \frac{\exp(\beta_1 + \beta_2 x_{2j} + \beta_3 x_{3ij} + \beta_4 x_{2j} x_{3ij} + \zeta_j)}{1 + \exp(\beta_1 + \beta_2 x_{2j} + \beta_3 x_{3ij} + \beta_4 x_{2j} x_{3ij} + \zeta_j)} \phi(\zeta_j; 0, \psi) d\zeta_j$$

$$\neq \frac{\exp(\beta_1 + \beta_2 x_{2j} + \beta_3 x_{3ij} + \beta_4 x_{2j} x_{3ij})}{1 + \exp(\beta_1 + \beta_2 x_{2j} + \beta_3 x_{3ij} + \beta_4 x_{2j} x_{3ij})}$$

The chance of the median individual. This is not the average of individuals.

is NOT the same as...



RH&S Fig 10.9 (pg 531)³⁰

Latent Response Formulation

$$Y_{ij} = \begin{cases} 1 & : Y_{ij}^* > 0 \\ 0 & : Y_{ij}^* \leq 0 \end{cases}$$

Our observed outcome
is a threshold on some
latent, unobserved but
continuous, response

$$Y_{ij}^* = \beta_{0j} + \beta_{1j} Time_{ij} + \epsilon_{ij}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01} Z_j + u_j$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} Z_j$$

$$\epsilon_{ij} \sim \text{logit}$$

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



We now have a level-1
residual, but it has a special
logistic distribution.

A probit model would put a
normal distribution on
these.

Why latent responses?

You can now calculate the correlation of your latent responses within subject:

$$\rho \equiv \text{Cor}(Y_{ij}^*, Y_{i'j}^*) = \frac{\sigma^2}{\sigma^2 + \pi^2/3}$$

(The *conditional intraclass correlation*)

You can also see why the population average is flatter than the subject. (See RH&S Chapter 10 text for this discussion.)

Confidence Intervals for Odds Ratios

Generate confidence intervals on the linear model

$$\hat{\beta} \pm z_{0.975} \widehat{SE}(\hat{\beta})$$

Exponentiate to get the odds ratio

$$\exp(\hat{\beta} - z_{0.975} \hat{\beta}), \exp(\hat{\beta} + z_{0.975} \widehat{SE})$$

You can't then shift to probability since the probabilities depend on the other coefficients and change.



Empirical Bayes Predictions

Notice these are constant since we only had a random intercept.

```
> head( coef( M1 )$patient )  
 (Intercept) treatment month treatment:month  
 1           2.30      -0.3     -0.4          -0.14  
 2           0.32      -0.3     -0.4          -0.14  
 3          -0.70      -0.3     -0.4          -0.14  
 4          -0.69      -0.3     -0.4          -0.14  
 6           1.80      -0.3     -0.4          -0.14  
 7           1.91      -0.3     -0.4          -0.14  
>  
> mosaic::favstats( ranef( M1 )$patient[,1] )  
 min   Q1 median   Q3 max mean   sd    n missing  
 -1.3 -1  -0.97  3.9  10  1.4  3 294 0
```

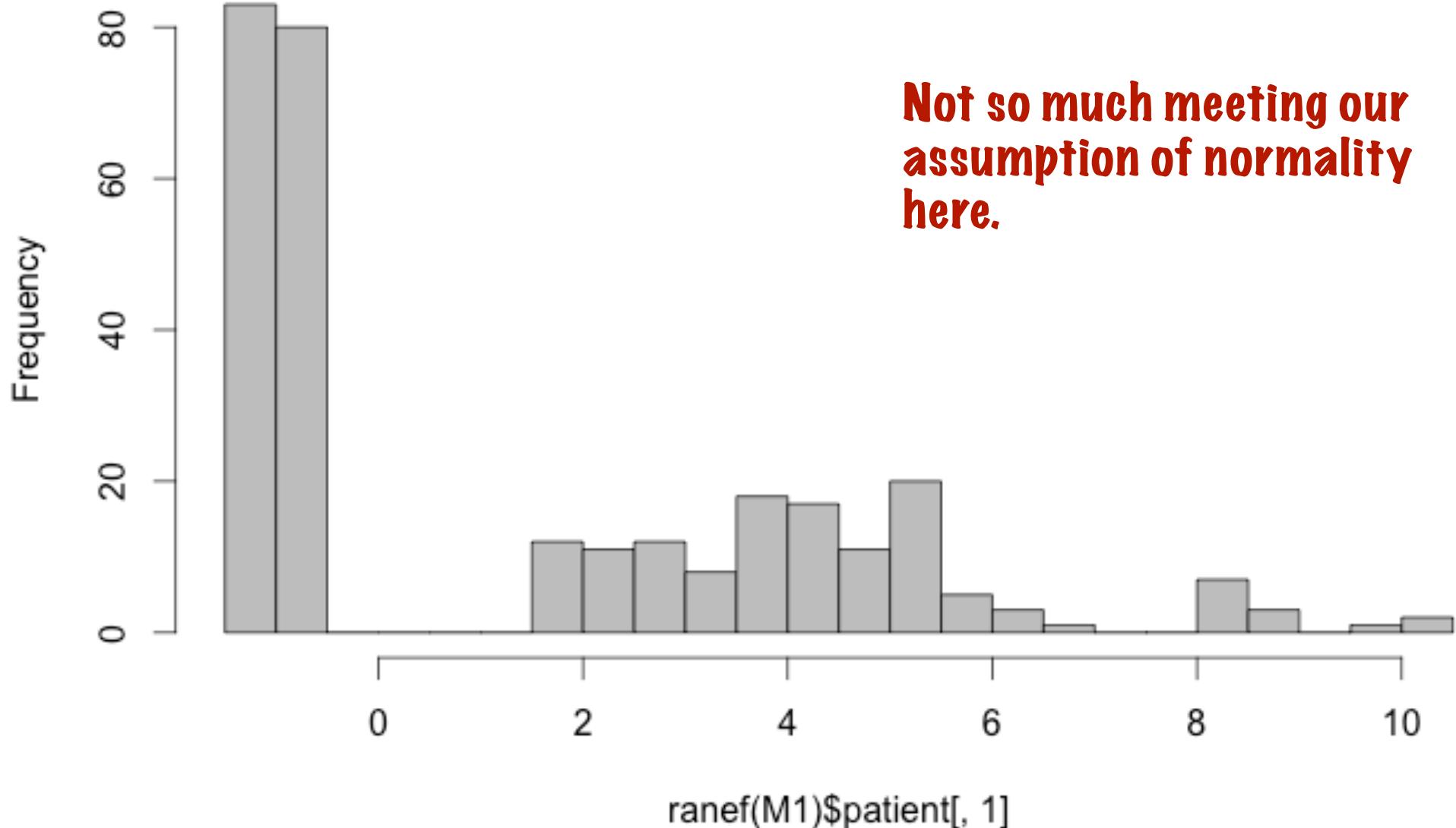


These are our estimated individual random effects (including the fixed intercept) for the patients.

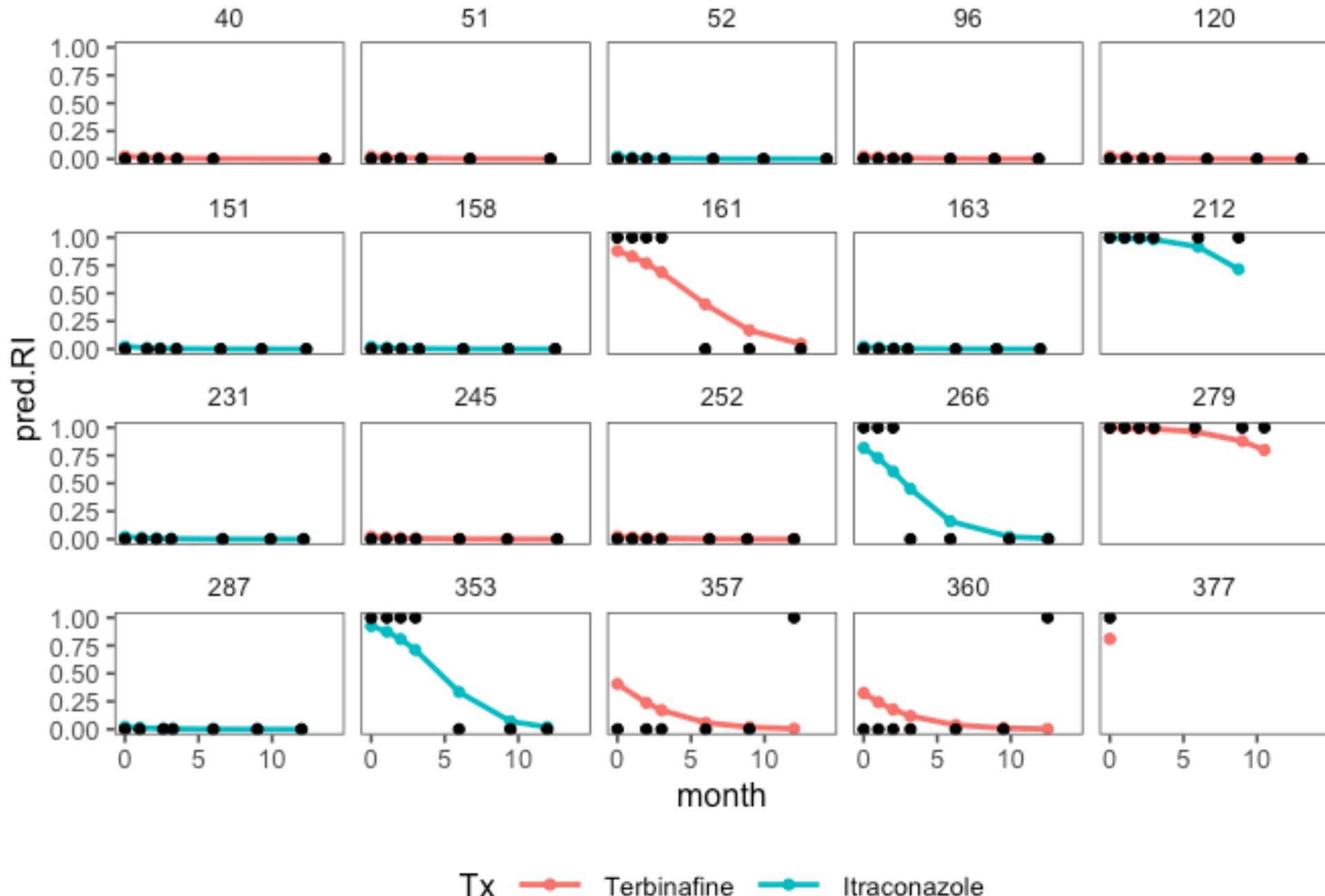
mosaic is a package that has this very useful favstats() function.
But don't load the full package.

Histogram of estimated random intercepts

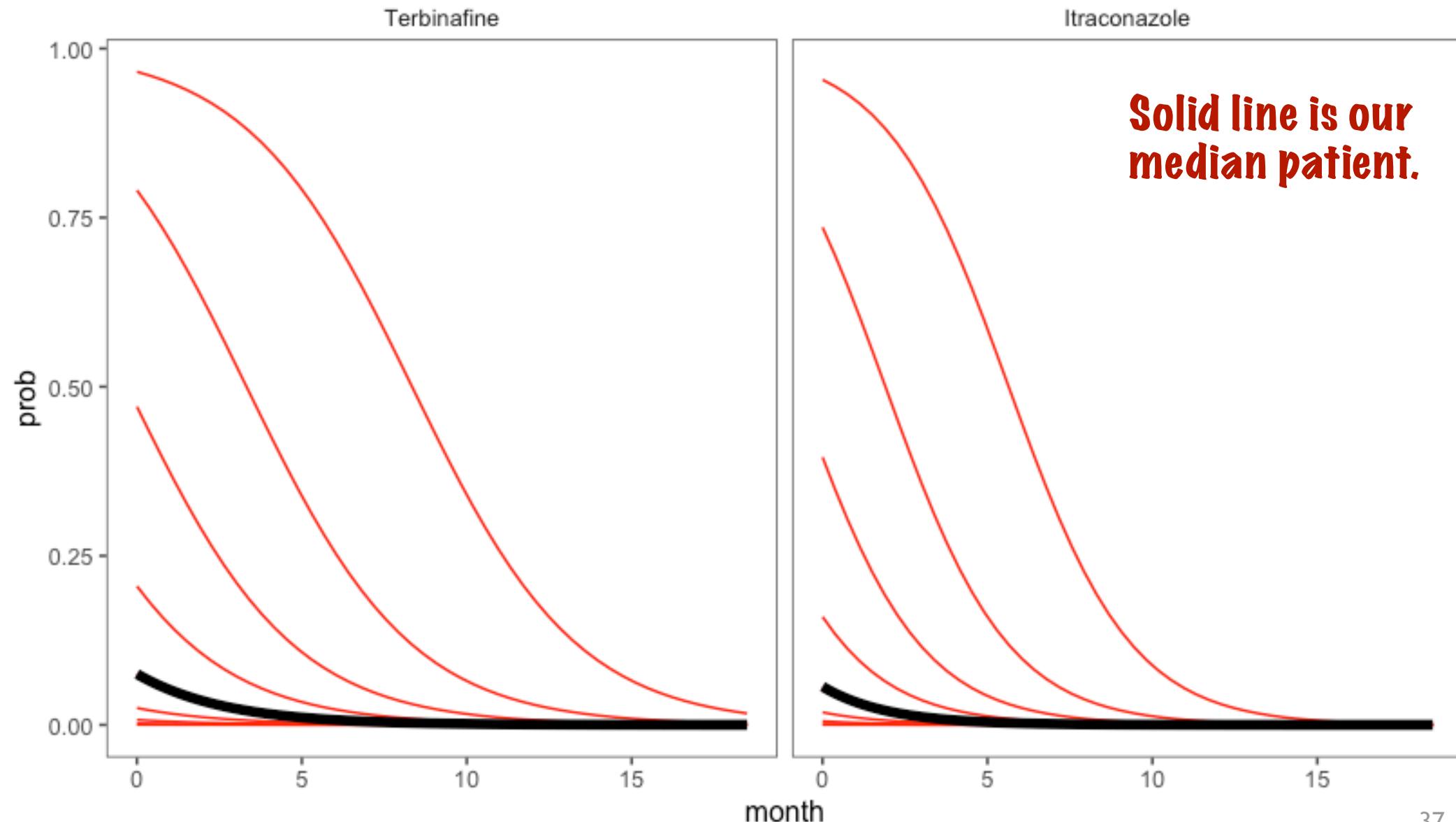
Histogram of ranef(M1)\$patient[, 1]



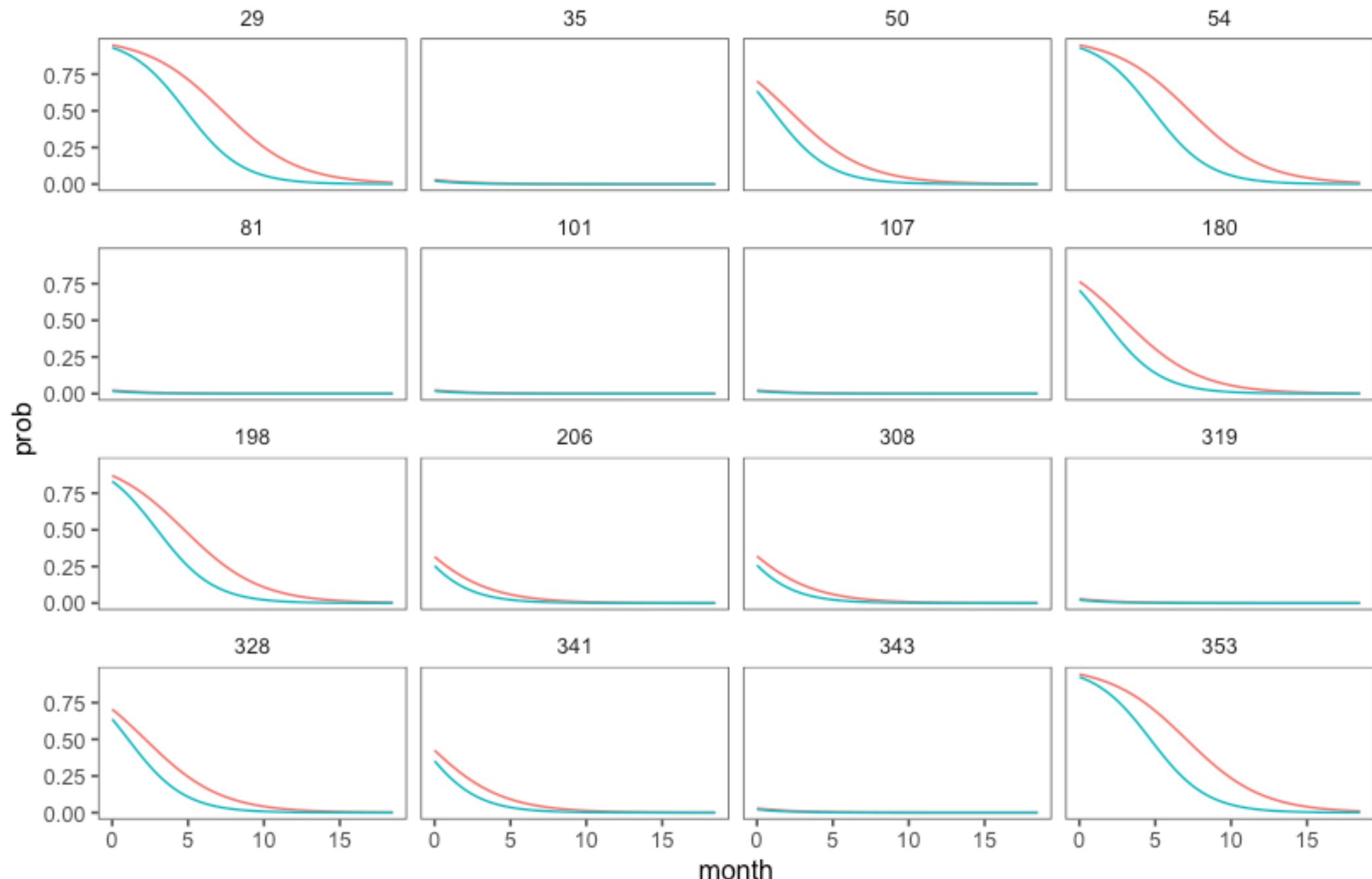
Individual Growth Curves and Raw Data



Predicted curves for different random intercepts (we are looking at deciles)



For 16 patients, the predicted curves for both treatments
(this shows impact of the random intercepts)



Recap



What we did today

Check-In

<http://cs179.org/lec53>

We looked at longitudinal data with a binary outcome.

In particular, this case study really underscores individual trend vs. population trend. These are not the same!

Interpretation of logistic models is tricky. Some good options we have:

- ★ exponentiate fixed effects and interpret as odds multipliers
- ★ make graphs of individuals and populations across time and describe variation and trends (my fav!)