Multilevel Modeling of Ordinal Responses

Sophia Rabe-Hesketh
Graduate School of Education & Graduate Group in Biostatistics
University of California, Berkeley
Institute of Education, University of London

Joint work with Anders Skrondal
Division of Epidemiology
Norwegian Institute of Public Health

Stata Users Group Meeting
Mexico City, April 2009

Outline

- Latent response and generalized linear model formulations
- Two-level random intercept models
- Conditional versus marginal effects
- Example: Recovery after surgery
- Different kinds of predicted probabilities
- Relaxing proportional odds assumption
- Modeling heteroscedasticity at the unit and cluster levels

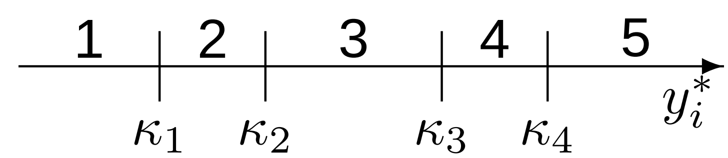
Ordered categorical responses

- Small number of mutually exclusive ‘categories’, $y_i = 1, \ldots, S$
- Categories are ordered, so y_i is **ordinal**
- Examples:
  - Severity of symptom (e.g. pain): none, moderate, severe
  - Frequency of behavior: never, occasionally, daily
  - Agreement (Likert scale):
    disagree strongly, disagree, agree, agree strongly

Latent response models

- Latent (unobserved) continuous response y_i^* underlies observed ordinal response y_i
- Threshold model determines observed response:

$$y_i = \begin{cases} 1 & \text{if} \quad y_i^* \leq \kappa_1 \\ 2 & \text{if} \quad \kappa_1 < y_i^* \leq \kappa_2 \\ \vdots & \vdots \qquad \vdots \\ S & \text{if} \quad \kappa_{S-1} < y_i^* \end{cases}$$



- Latent response modeled as linear regression without intercept (for identification)

$$\begin{aligned} y_i^* &= \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \epsilon_i \\ &= \mathbf{x}_i'\boldsymbol{\beta} + \epsilon_i \end{aligned}$$

Illustration: Latent response formulation ($S = 3$)

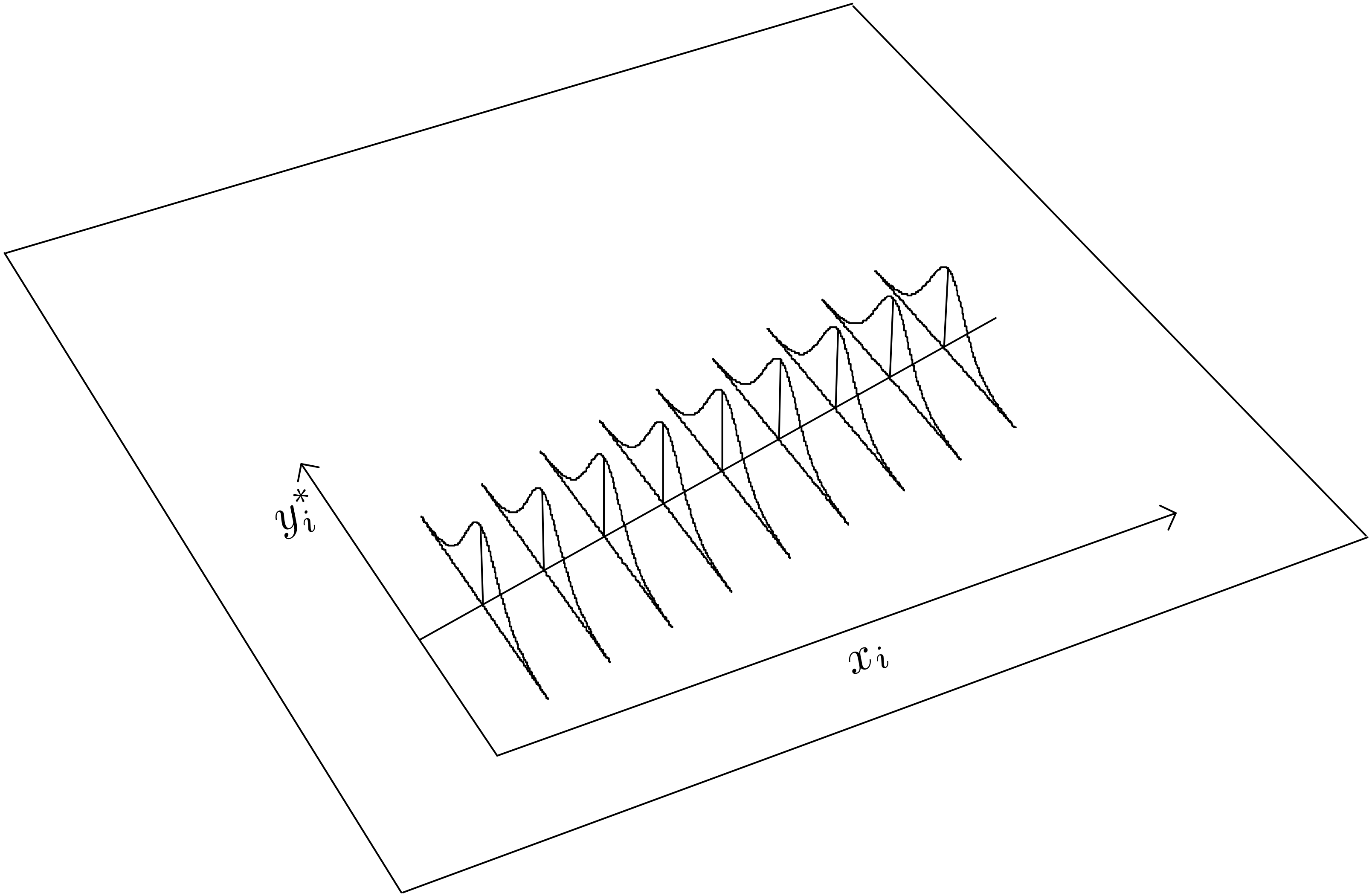


Illustration: Latent response formulation ($S = 3$)

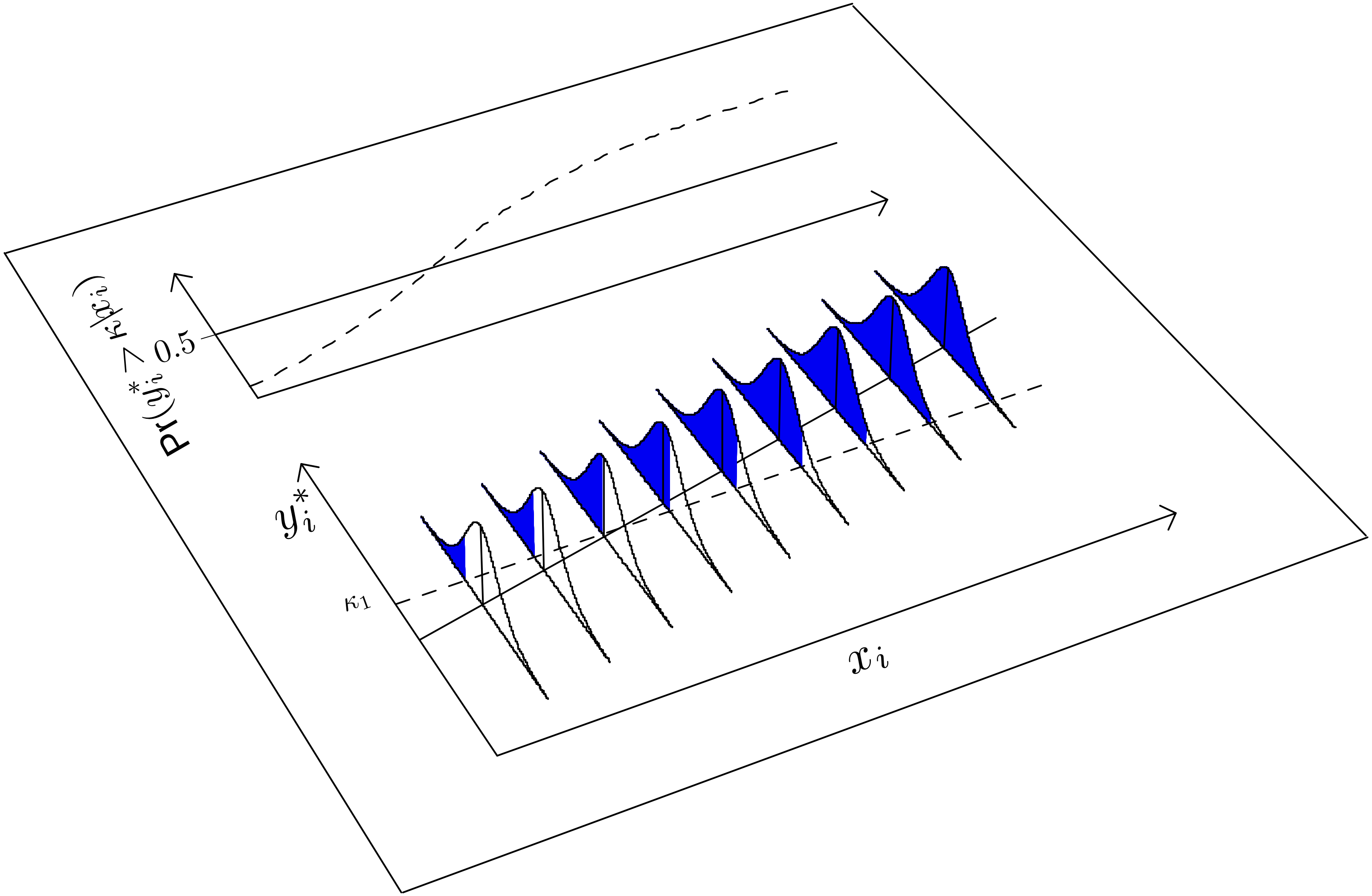
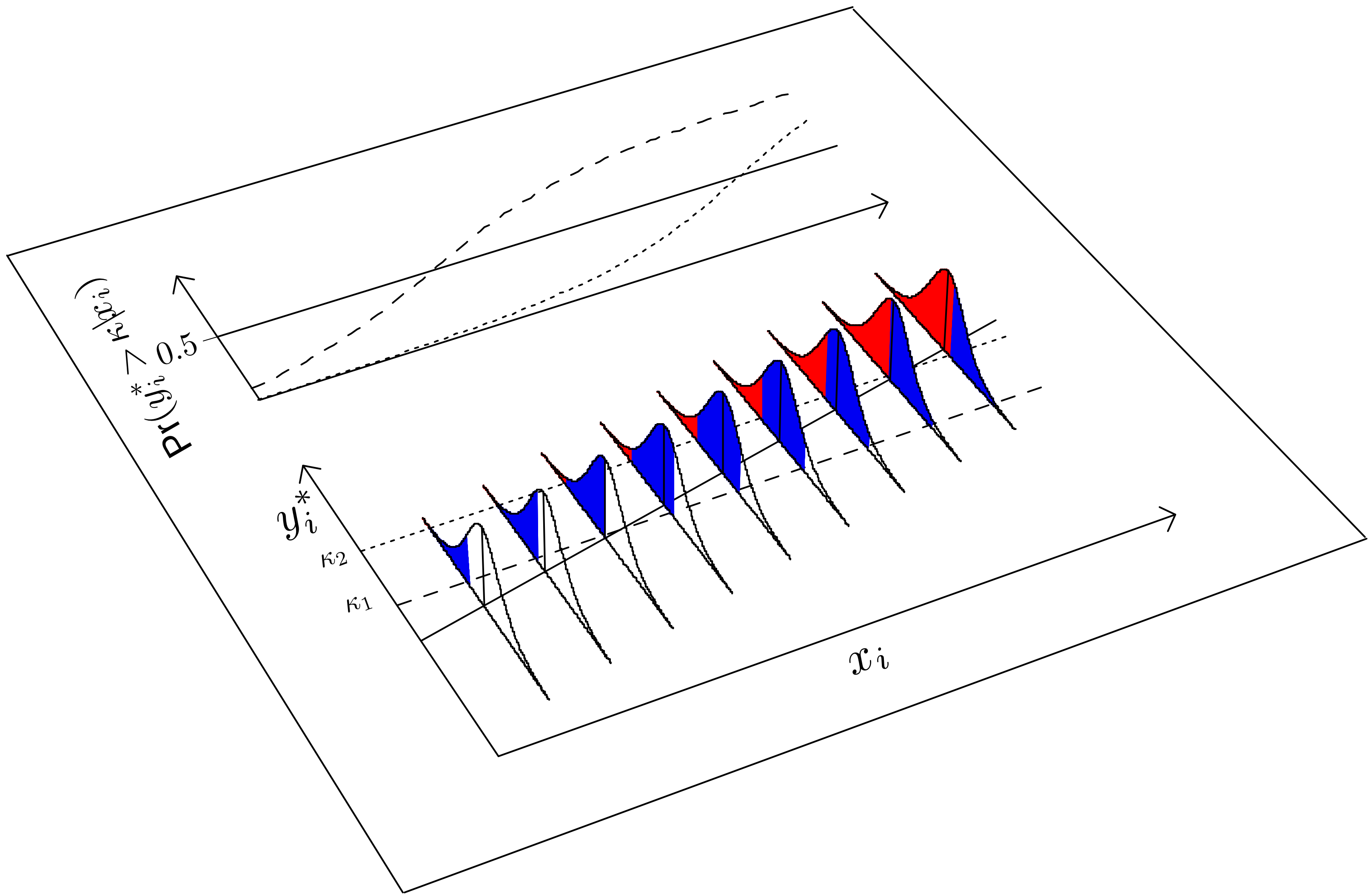


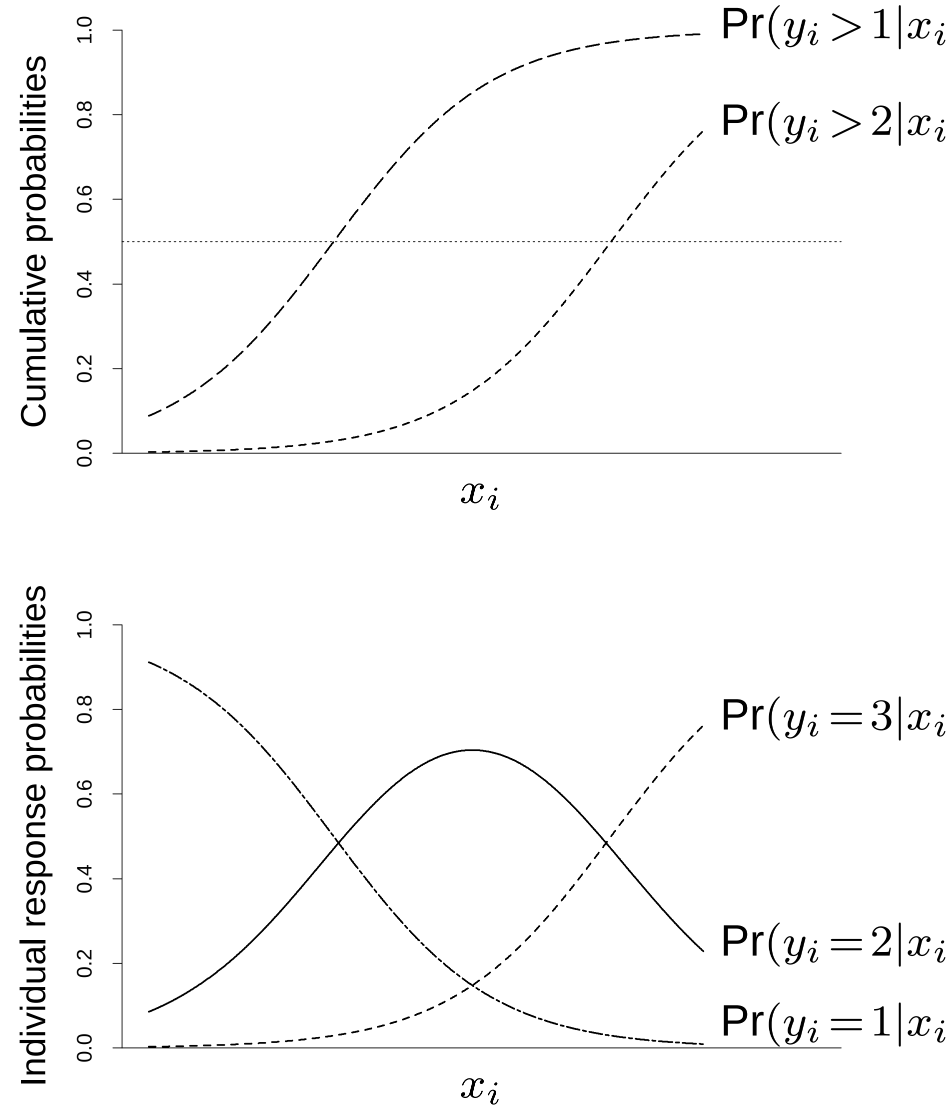
Illustration: Latent response formulation ($S = 3$)



Cumulative versus individual response probabilities

$$
\begin{aligned}
\text{Pr}(y_i > 1) &= \text{Pr}(y_i = 2) + \text{Pr}(y_i = 3) \\
\text{Pr}(y_i > 2) &= \text{Pr}(y_i = 3)
\end{aligned}
$$

$$
\begin{aligned}
\text{Pr}(y_i = 3) &= \text{Pr}(y_i > 2) \\
\text{Pr}(y_i = 2) &= \text{Pr}(y_i > 1) - \text{Pr}(y_i > 2) \\
\text{Pr}(y_i = 1) &= 1 - \text{Pr}(y_i > 1)
\end{aligned}
$$



From latent response to generalized linear model formulations

- Cumulative probabilities:

$$\begin{aligned}\Pr(y_i > s|\mathbf{x}_i) &= \Pr(y_i^* > \kappa_s|\mathbf{x}_i) = \Pr(\mathbf{x}_i'\boldsymbol{\beta} + \epsilon_i > \kappa_s|\mathbf{x}_i)\\ &= \Pr(-\epsilon_i \le \mathbf{x}_i'\boldsymbol{\beta} - \kappa_s|\mathbf{x}_i) = F(\mathbf{x}_i'\boldsymbol{\beta} - \kappa_s)\end{aligned}$$

  - $F(\cdot)$ is the cumulative density function of $-\epsilon_i$
- This is a generalized linear model where
  - $F(\cdot)$ is inverse **link function** and $\mathbf{x}_i'\boldsymbol{\beta} - \kappa_s$ is **linear predictor**

Distribution of $-\epsilon_i$	Var(ϵ_i)	Link $g = F^{-1}$	Model	Stata command
Logistic	$\pi^2/3$	Cumulative Logit	Prop. odds	`ologit`, `gllamm`
Standard normal	1	Probit	Ordinal probit	`oprobit`, `gllamm`
*Gumbel	$\pi^2/6$	Compl. log-log	Compl. log-log	`gllamm`

*Asymmetric (non-equivalent model if scale reversed)

Cumulative logit model and odds ratios

- With a logistic distribution for ϵ_i, a **logit link** (one covariate x_i):

$$\underbrace{F^{-1}}_{g}[\mathrm{Pr}(y_i > s|x_i)] \;=\; \log\underbrace{\left[\frac{\mathrm{Pr}(y_i > s|x_i)}{\mathrm{Pr}(y_i \le s|x_i)}\right]}_{\mathrm{odds}_{(y_i>s|x_i)}} \;=\; \beta x_i - \kappa_s$$

$$\begin{aligned}\beta &= \log[\mathrm{odds}(y_i > s|x_i = a+1)] - \log[\mathrm{odds}(y_i > s|x_i = a)] \\ &= \log\left[\frac{\mathrm{odds}(y_i > s|x_i = a+1)}{\mathrm{odds}(y_i > s|x_i = a)}\right]\end{aligned}$$

- Coefficient β represents increase in log-odds per unit increase in x_i
- Exponentiated coefficient $\exp(\beta)$ represents **odds ratio** per unit increase in x_i
    - Does not depend on $s \Rightarrow$ Proportional odds

Other logit models for ordinal responses

- Adjacent category logit (see Zheng & R-H, Stata Journal 7, 313-333)

$$\log\left[\frac{\Pr(y_i = s+1|\mathbf{x}_i)}{\Pr(y_i = s|\mathbf{x}_i)}\right] \;=\; \mathbf{x}_i'\boldsymbol{\beta} + \kappa_s$$

  Data expansion followed by `clogit` or `gllamm` command

- Continuation ratio logit (see MLMUS2 - CH8)

$$\log\left[\frac{\Pr(y_i = s|\mathbf{x}_i)}{\Pr(y_i > s|\mathbf{x}_i)}\right] \;=\; \mathbf{x}_i'\boldsymbol{\beta} + \kappa_s$$

  Data expansion followed by `logit` or `gllamm` command

- Stereotype model

$$\log\left[\frac{\Pr(y_i = s|\mathbf{x}_i)}{\Pr(y_i = 1|\mathbf{x}_i)}\right] \;=\; \mathbf{x}_i'\boldsymbol{\beta}\alpha_s + \kappa_s$$

  `slogit` command

Two-level random intercept models

- Units i (e.g. subjects) nested in clusters j (e.g. schools)
- Include random intercept ζ_j for clusters

$$y_{ij}^* = \mathbf{x}_{ij}'\boldsymbol{\beta} + \zeta_j + \epsilon_{ij}, \quad \zeta_j|\mathbf{x}_{ij} \sim N(0,\psi)$$

$$\epsilon_{ij}|\mathbf{x}_{ij} \sim \begin{cases} N(0,1) & \text{for probit models} \\ \text{Logistic} & \text{for logit models} \end{cases}$$

  - ζ_j independent of ϵ_{ij}
  - ζ_j and ϵ_{ij} independent across clusters
  - ϵ_{ij} independent across units within clusters

Intraclass correlation of latent responses

- Total residual $\xi_{ij} = \zeta_j + \epsilon_{ij}$ has variance

$$\text{Var}(\xi_{ij}|\mathbf{x}_{ij}) \; = \begin{cases} \psi + 1 & \text{for probit models} \\ \psi + \pi^2/3 & \text{for logit models} \end{cases}$$

- Covariance between total residuals ξ_{ij} and $\xi_{i'j}$ of two subjects in same cluster is ψ and **intraclass correlation** is

$$\text{Cor}(\xi_{ij}, \xi_{i'j}|\mathbf{x}_{ij}, \mathbf{x}_{i'j}) \; = \begin{cases} \psi/(\psi + 1) & \text{for probit models} \\ \psi/(\psi + \pi^2/3) & \text{for logit models} \end{cases}$$

Conditional versus marginal effects

- For probit model, **conditional** or cluster-specific probabilities are

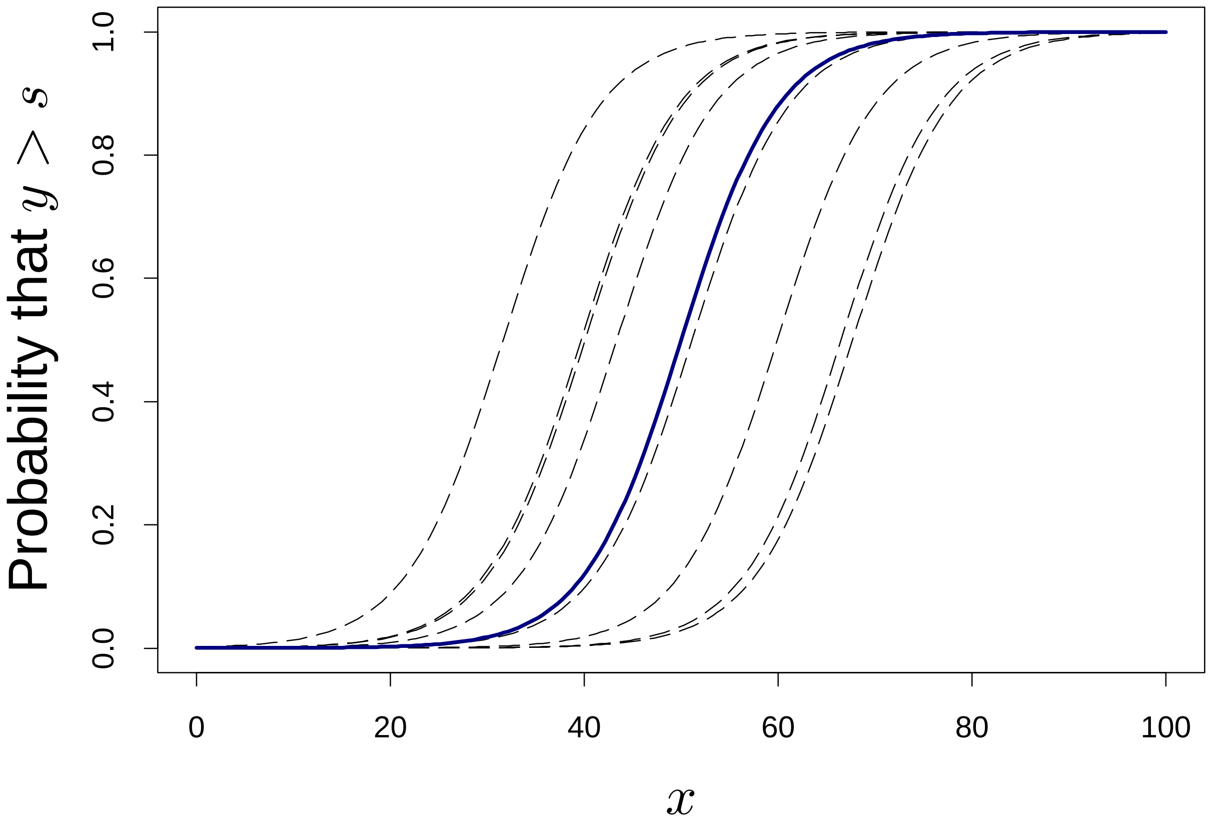
$$\text{Pr}(y_{ij} > s|\zeta_j, \mathbf{x}_{ij}) \;=\; \Phi(\mathbf{x}'_{ij}\boldsymbol{\beta} + \zeta_j - \kappa_s)$$

- **Marginal** or population-averaged response probabilities are

$$\begin{aligned}
\text{Pr}(y_{ij} > s|\mathbf{x}_{ij}) &= \text{Pr}(y^*_{ij} > \kappa_s) = \text{Pr}(\mathbf{x}'_{ij}\boldsymbol{\beta} + \xi_{ij} > \kappa_s) \\
&= \text{Pr}(-\xi_{ij} \leq \mathbf{x}'_{ij}\boldsymbol{\beta} - \kappa_s) \\
&= \text{Pr}\left(\frac{\xi_{ij}}{\sqrt{\psi+1}} \leq \frac{\mathbf{x}'_{ij}\boldsymbol{\beta} - \kappa_s}{\sqrt{\psi+1}}\right) \\
&= \Phi\left(\frac{\mathbf{x}'_{ij}\boldsymbol{\beta} - \kappa_s}{\sqrt{\psi+1}}\right)
\end{aligned}$$

- Therefore, marginal effects $\boldsymbol{\beta}/\sqrt{\psi+1}$ attenuated compared with conditional effects $\boldsymbol{\beta}$

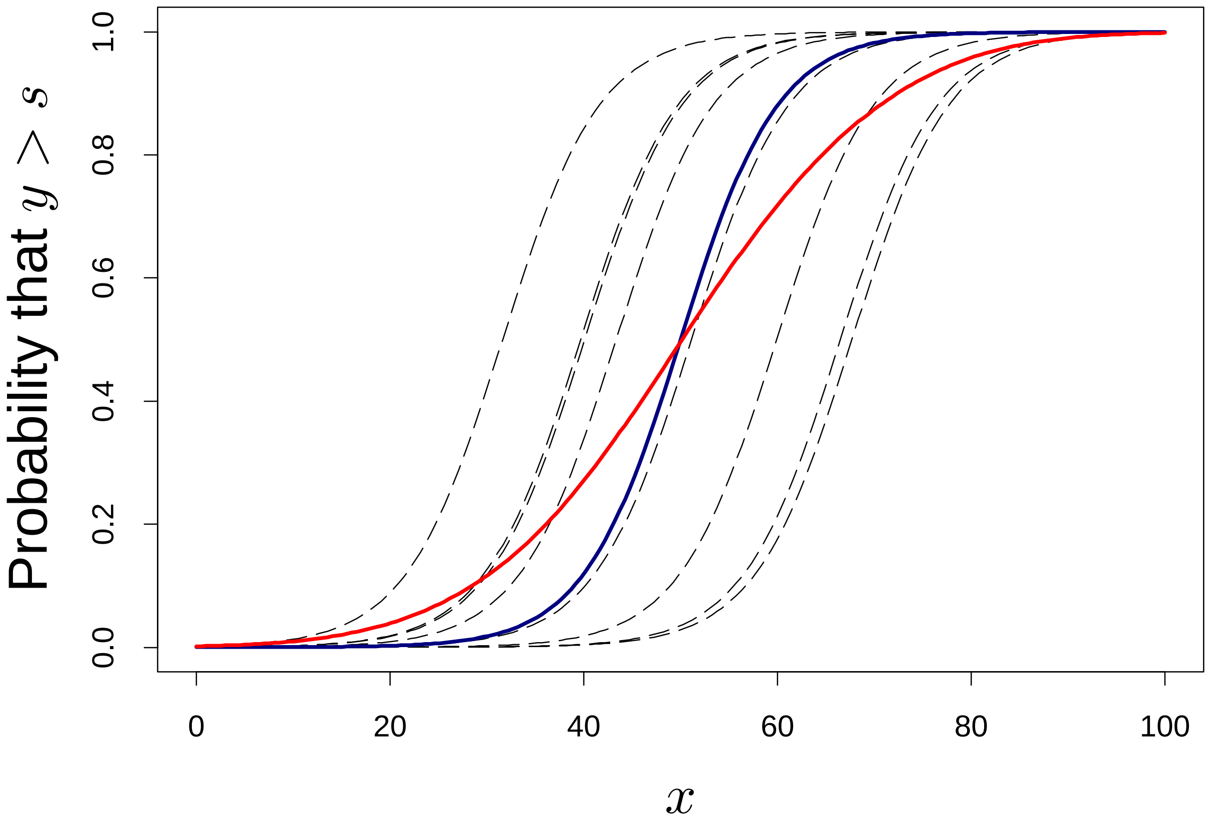
Illustration: Conditional versus marginal relationship



- - - - - conditional or cluster-specific (random sample)

—— median

Illustration: Conditional versus marginal relationship



conditional or cluster-specific (random sample)

median

marginal or population-averaged

Multilevel random coefficient models

- Consider clustered longitudinal data with occasions i (level 1) nested in student j (level 2) in schools k (level 3)
- Example of three-level random coefficient model:

$$y^*_{ijk} \;=\; [\zeta^{(2)}_{0jk} + \zeta^{(3)}_{0k}] + [\beta_1 + \zeta^{(2)}_{1jk} + \zeta^{(3)}_{1k}]x_{1ijk} + \beta_2 x_{2ijk} + \epsilon_{ijk}$$

  - $\zeta^{(2)}_{0jk}$ and $\zeta^{(3)}_{0k}$ are random intercepts at levels 2 and 3
  - $\zeta^{(2)}_{1jk}$ and $\zeta^{(3)}_{1k}$ are random coefficients of x_{1ijk} at levels 2 and 3
  - Random effects at the same level, e.g. $(\zeta^{(2)}_{0jk}, \zeta^{(2)}_{1jk})$, bivariate normal with zero means, independent across units jk
- General three-level random coefficient model

$$y^*_{ijk} \;=\; \mathbf{x}'_{ijk}\boldsymbol{\beta} + \mathbf{z}^{(2)\prime}_{ijk}\boldsymbol{\zeta}^{(2)}_{jk} + \mathbf{z}^{(3)\prime}_{ijk}\boldsymbol{\zeta}^{(3)}_{k} + \epsilon_{ijk}$$

Example: Recovery after surgery

- Study of effect of dosage of anesthetic on postsurgical recovery
- 60 children undergoing surgery were randomized into four dosage groups
- Recovery assessed 0, 5, 15, and 30 minutes after admission to recovery room
  - `score`: recovery score
    1 (no recovery), 2 ,3 ,4 (complete recovery)
  - `id2`: subject identifier
  - `dosage`: dosage of anesthetic in milligram/kilogram
  - `age`: age of child in months
  - `duration`: duration of surgery in minutes
  - `time`: time after surgery in minutes

Two-level model and estimation using gllamm

- Occasions i, child j
- Model the probability of exceeding a category s, $s = 1, 2, 3$

$$\text{logit}[\text{Pr}(y_{ij} > s|\zeta_j, \mathbf{x}_{ij})] \quad = \quad \beta_1 \texttt{dosage}_j + \beta_2 \texttt{age}_j + \beta_3 \texttt{duration}_j$$
$$+ \beta_4 \texttt{time}_i + \zeta_j - \kappa_s$$

- Estimation of two-level model using adaptive quadrature in `gllamm`:

```
gllamm score dosage age duration time, ///
   i(id2) link(ologit) adapt
```

- Without a random intercept (single-level) using `ologit`:

```
ologit score dosage age duration time
```

Maximum likelihood estimates

	Single-level model			Two-level model		
Parameter	Est	(SE)	OR	Est	(SE)	OR
$10\beta_1$ [dosage]	-0.34	(0.22)	0.71	-1.28	(0.92)	0.28
$10\beta_2$ [age]	-0.22	(0.08)	0.80	-0.55	(0.33)	0.57
$10\beta_3$ [duration]	-0.09	(0.03)	0.91	-0.22	(0.14)	0.80
$10\beta_4$ [time]	0.97	(0.12)	2.64	2.35	(0.27)	10.50
$\psi = \text{Var}(\zeta_j)$		–		13.4	(4.12)	
κ_1	-2.13	(0.69)		-6.21	(2.80)	
κ_2	-0.98	(0.68)		-3.43	(2.76)	
κ_3	0.18	(0.68)		-0.73	(2.73)	
Log-likelihood		-283			-222	

Intraclass correlation: 0.80

Predicted probabilities

- Random effects distributions
  - 'Prior' distribution: $g(\zeta_j)$
  - 'Posterior' distribution, given data for person j: $\omega(\zeta_j|\mathbf{y}_j, \mathbf{X}_j)$
- Predicted **marginal** or population averaged probabilities:

$$\text{Pr}(y_{ij} > 2|\mathbf{x}_{ij}) = \int \frac{\text{exp}(\mathbf{x}'_{ij}\widehat{\boldsymbol{\beta}} + \zeta_j - \widehat{\kappa}_2)}{1 + \text{exp}(\mathbf{x}'_{ij}\widehat{\boldsymbol{\beta}} + \zeta_j - \widehat{\kappa}_2)} g(\zeta_j)\text{d}\zeta_j$$

```
gllapred p_marg, marg mu above(2)
```

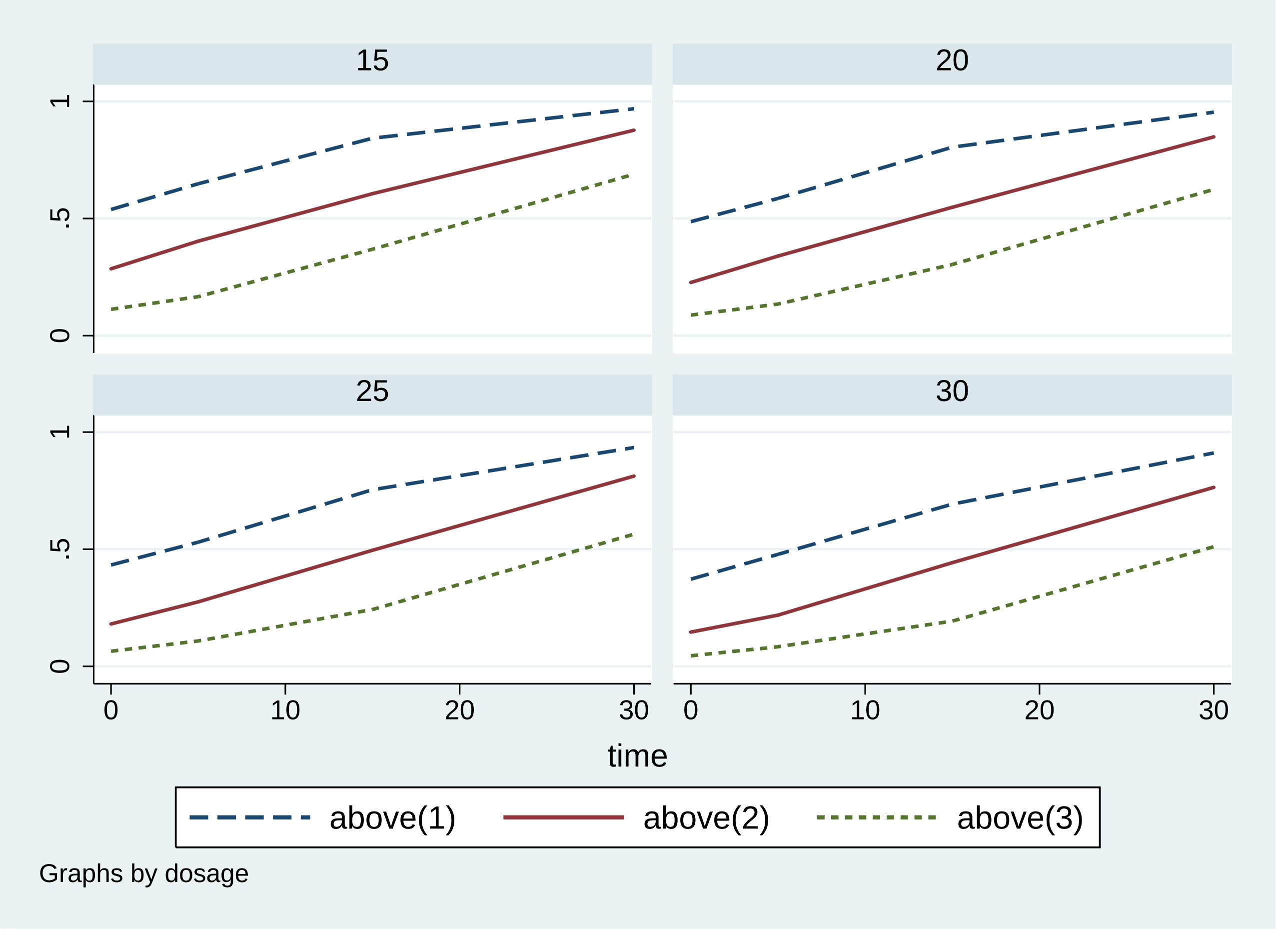
- Predicted **posterior mean** probabilities:

$$\text{E}_\zeta[\text{Pr}(y_{ij} > 2|\zeta_j, \mathbf{x}_{ij})|\mathbf{y}_j, \mathbf{X}_j] = \int \frac{\text{exp}(\mathbf{x}'_{ij}\widehat{\boldsymbol{\beta}} + \zeta_j - \widehat{\kappa}_2)}{1 + \text{exp}(\mathbf{x}'_{ij}\widehat{\boldsymbol{\beta}} + \zeta_j - \widehat{\kappa}_2)} \omega(\zeta_j|\mathbf{y}_j, \mathbf{X}_j)\text{d}\zeta_j$$

```
gllapred p_subj, mu above(2)
```

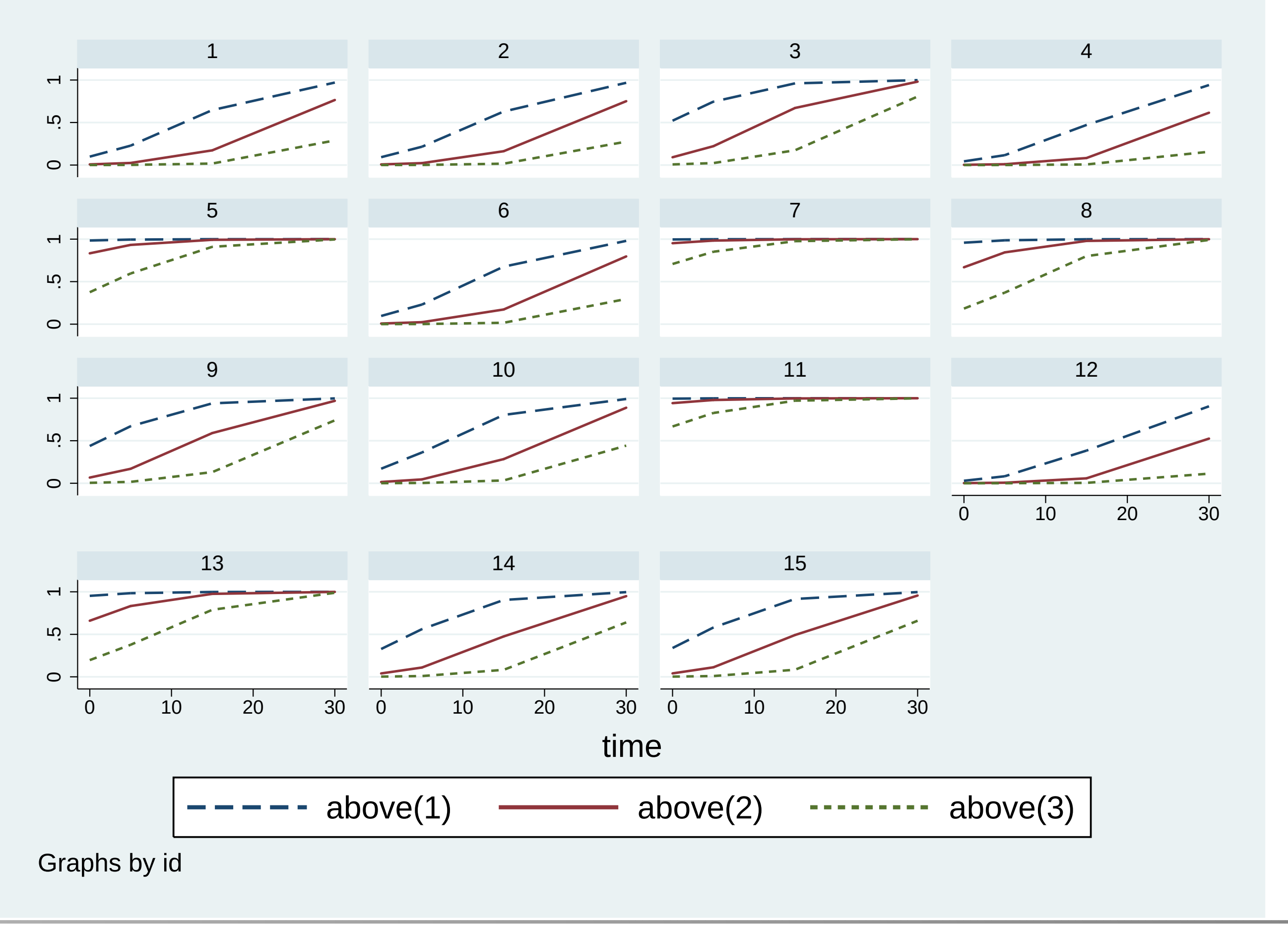
Predicted marginal probabilities

- For four dosage groups (15mg/Kg, 20mg/Kg, 25mg/Kg, 30mg/Kg) over time
- `age`=37months, `duration`=80min



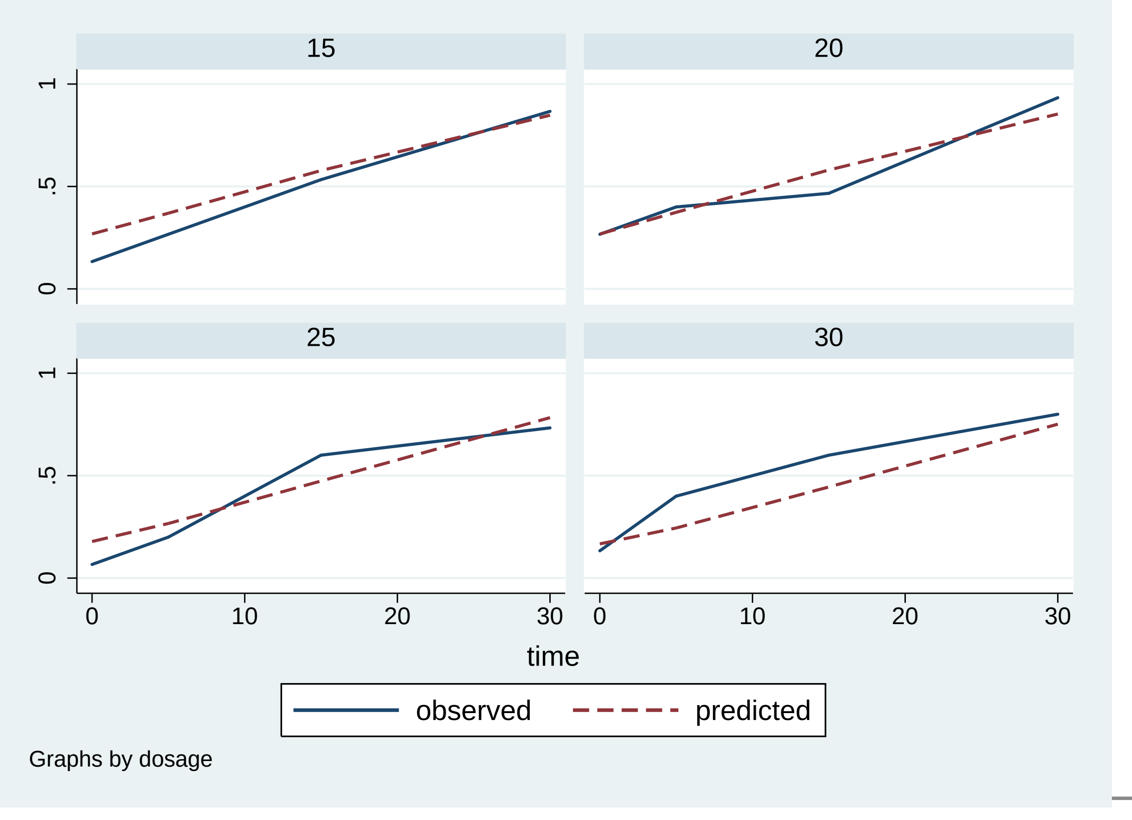
Predicted posterior mean probabilities

- For 15 subjects in dosage group 20mg/Kg over time
- `age`, `duration` as they are in the data



Assessing model fit

- Compare proportions of scores above 2 by `dosage` and `time` with $\Pr(y_{ij} > 2|\texttt{dosage}_j, \texttt{time}_i)$
  - Predict marginal probabilities with covariates as they are in the data $\Rightarrow \Pr(y_{ij} > 2|\texttt{dosage}_j, \texttt{age}_j, \texttt{duration}_j, \texttt{time}_i)$
  - Then obtain average over `age` and `duration` $\Rightarrow \Pr(y_{ij} > 2|\texttt{dosage}_j, \texttt{time}_i)$



Relaxing the proportional odds assumption for dosage

- Instead of $\beta_1\texttt{dosage}_j$:

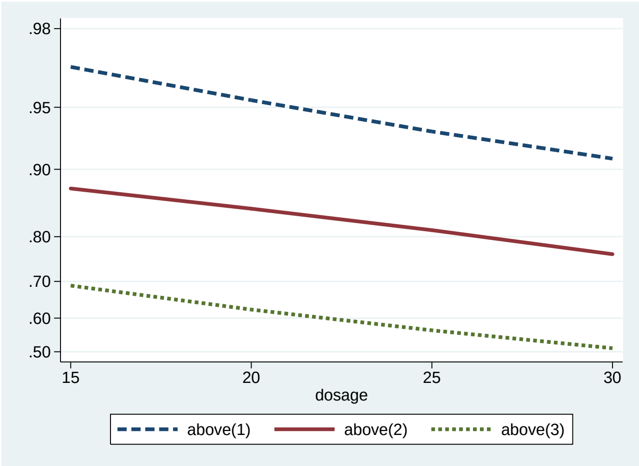
$$\begin{aligned}\text{logit}[\Pr(y_{ij} > s|\zeta_j, \mathbf{x}_{ij})] &= \beta_2\texttt{age}_j + \beta_3\texttt{duration}_j \\ &+ \beta_4\texttt{time}_i + \zeta_j - (\kappa_s + \delta_s\texttt{dosage}_j)\end{aligned}$$

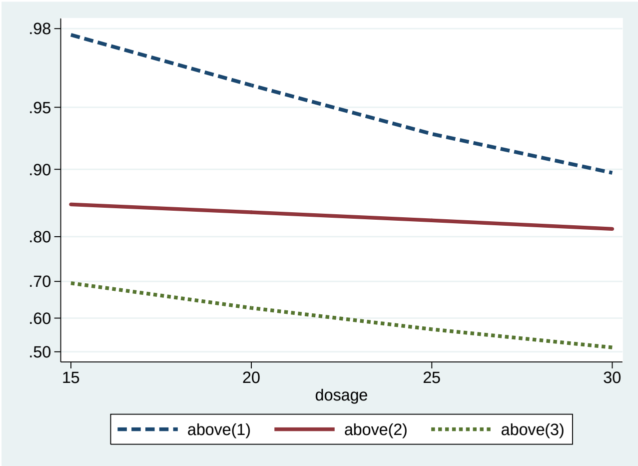
- Non-proportional odds model in `gllamm`:

```
eq thr: dosage
gllamm score age duration time, ///
   i(id2) link(ologit) thresh(thr) adapt
```

Maximum likelihood estimates

	Proportional odds		Non-proportional odds	
s	$\widehat{\beta}_1$	(SE)	$-\widehat{\delta}_s$	(SE)
1	-0.13	(0.09)	-0.20	(0.10)
2	-0.13	(0.09)	-0.05	(0.10)
3	-0.13	(0.09)	-0.13	(0.10)
	Log-lik. = -222		Log-lik. = -218, LR-test, $p = 0.02$	
	Pred. marginal prob. (time=30min)		Pred. marginal prob. (time=30min)	





Modeling heteroscedasticity at level 1

- Scaled ordinal probit link (similar to `hetprob`)

$$\Pr(y_{ij} > s|\mathbf{x}_{ij}, \zeta_j) \;=\; \Phi\left(\frac{\mathbf{x}'_{ij}\boldsymbol{\beta} + \zeta_j - \kappa_s}{\sigma_{ij}}\right)$$

$$\log(\sigma_{ij}) \;=\; \mathbf{z}'_{ij}\boldsymbol{\alpha} = \alpha_2 d_{2j} + \alpha_3 d_{3j} + \alpha_4 d_{4j}$$

```
tabulate dosage, gen(d)
eq het: d2 d3 d4  /* log sd = 0 in one group */
gllamm score dosage age duration time, i(id2) ///
   link(soprobit) s(het) from(a) adapt
```

- LR test: No evidence that level-1 variance depends on dosage group

dosage	15	20	25	30
k	1	2	3	4
$\alpha_k = \log(\sigma)$	0.00	0.06	-0.06	-0.09
σ	1.00	1.06	0.94	0.91

Modeling heteroscedasticity at level 2

- At level 2: Use factor loadings

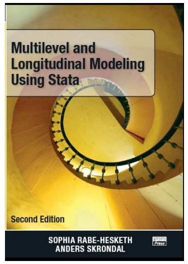
$$\begin{aligned}\text{linear predictor}_s &= \mathbf{x}'_{ij}\boldsymbol{\beta} + \zeta_j \mathbf{z}'_j \boldsymbol{\lambda} - \kappa_s, \qquad \lambda_1 = 1 \\ &= \mathbf{x}'_{ij}\boldsymbol{\beta} + \zeta_j(d_{1j} + \lambda_2 d_{2j} + \lambda_3 d_{3j} + \lambda_4 d_{4j}) - \kappa_s\end{aligned}$$

```
eq load: d1 d2 d3 d4
gllamm score dosage age duration time, i(id2) ///
   eqs(load) link(oprobit) adapt
```

- LR test: No evidence that level-2 variance depends on dosage group

dosage	15	20	25	30
k	1	2	3	4
λ_k	1.00	1.16	1.25	1.70
$\sqrt{\psi}\lambda_k$	1.65	1.91	2.06	2.80

References

- 

“Multilevel and Longitudinal Modeling Using Stata”. Stata Press 2008. (S. Rabe-Hesketh & A. Skrondal)
- “Generalized linear mixed effects models”. In G. Fitzmaurice et al. (Eds.) *Longitudinal Data Analysis.* Chapman & Hall/CRC, pp. 79-106, 2009. (S. Rabe-Hesketh & A. Skrondal)
- “Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects”. *Journal of Econometrics* 128, 301-323, 2005. (S. Rabe-Hesketh, A. Skrondal & A. Pickles)
- “Prediction in multilevel generalized linear models”. *Journal of the Royal Statistical Society, Series A* 172, 659-687, 2009. (A. Skrondal & S. Rabe-Hesketh)
- *GLLAMM Manual.* U.C. Berkeley Division of Biostatistics Working Paper Series. Working Paper 160, 2004. `http://www.bepress.com/ucbbiostat/paper160/` (S. Rabe-Hesketh, A. Skrondal & A. Pickles)
- See also `http://www.gllamm.org`