Prediction in Multilevel Logistic Regression

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

Fall North American Stata Users Group meeting
San Francisco, November 2008

Outline

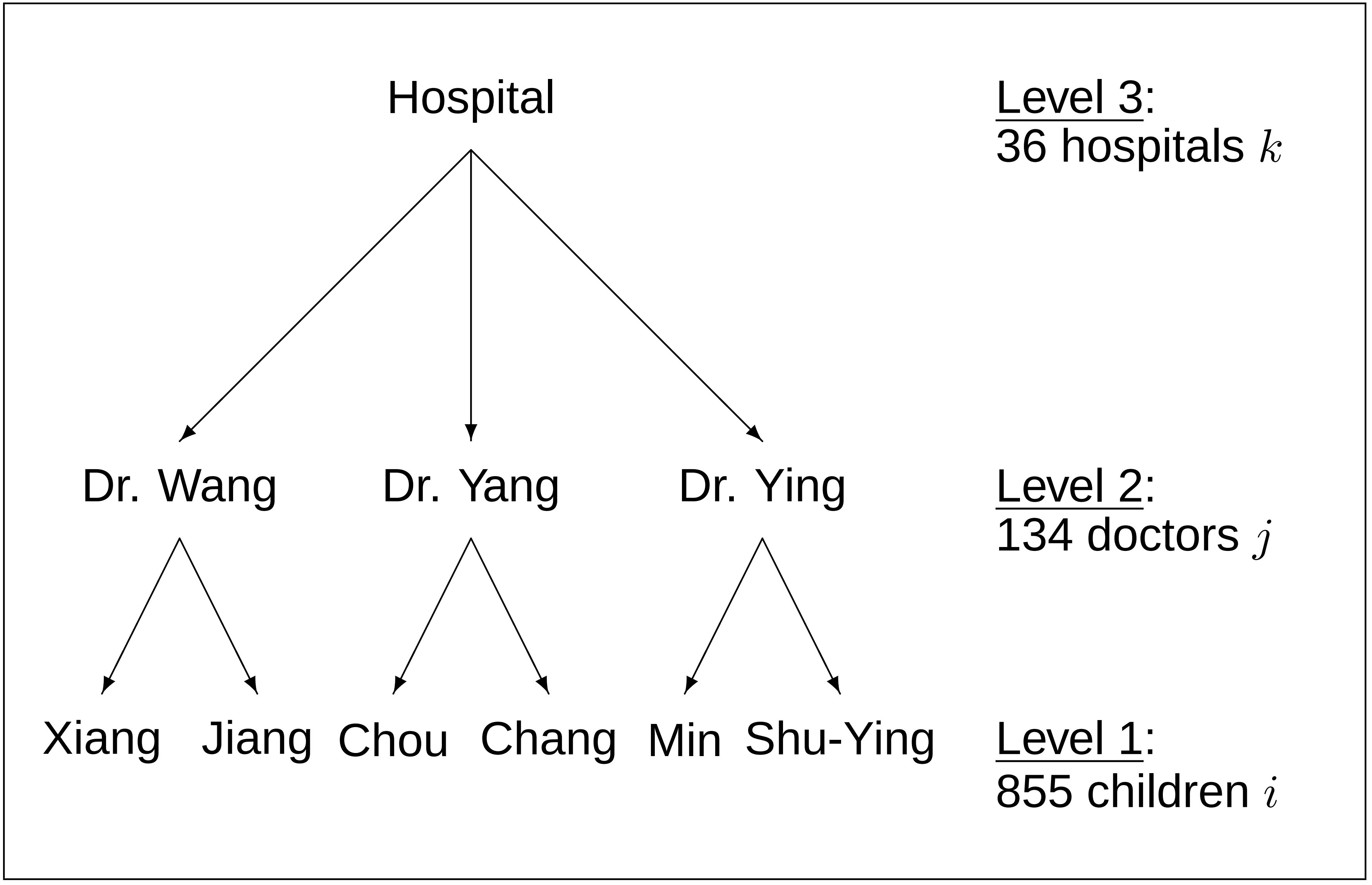
- Application: Abuse of antibiotics in China
- Three-level logistic regression model
- Prediction of random effects
  - Empirical Bayes (EB) prediction
  - Standard errors for EB prediction and approximate CI
- Prediction of response probabilities
  - Conditional response probabilities
  - Posterior mean response probabilities (different types)
  - Population-averaged response probabilities
- Concluding remarks

Abuse of antibiotics in China

- Acute respiratory tract infection (ARI) can lead to pneumonia and death if not properly treated
- Inappropriate frequent use of antibiotics was common in China in 1990's, leading to drug resistance
- In the 1990's the WHO introduced a program of case management for children under 5 with ARI in China
- Data collected on 855 children i (level 1) treated by 134 doctors j (level 2) in 36 hospitals k (level 3) in two counties (one of which was in the WHO program)
- **Response variable:** Whether antibiotics were prescribed when there were no clinical indications based on medical files

Reference: Min Yang (2001). *Multinomial Regression*. In Goldstein and Leyland (Eds). *Multilevel Modelling of Health Statistics*, pages 107-123.

Three-level data structure



Variables

- Response variable y_{ijk}
  - Antibiotics prescribed without clinical indications (1: yes, 0: no)
- 7 covariates $\mathbf{x}_{ijk}$
  - Patient level i
    - [Age] Age in years (0-4)
    - [Temp] Body temperature, centered at 36°C
    - [Paymed] Pay for medication (yes=1, no=0)
    - [Selfmed] Self medication (yes=1, no=0)
    - [Wrdiag] Failure to diagnose ARI early (yes=1, no=0)
  - Doctor level j
    - [DRed] Doctor's education
      (6 categories from self-taught to medical school)
  - Hospital level k
    - [WHO] Hospital in WHO program (yes=1, no=0)

Three-level random intercept logistic regression

- Logistic regression with random intercepts for doctors and hospitals

$$\text{logit}[\Pr(y_{ijk} = 1|\mathbf{x}_{ijk}, \zeta_{jk}^{(2)}, \zeta_{k}^{(3)})] = \mathbf{x}'_{ijk}\boldsymbol{\beta} + \zeta_{jk}^{(2)} + \zeta_{k}^{(3)}$$

  - Level 3: $\zeta_k^{(3)}|\mathbf{x}_{ijk} \sim N(0, \psi^{(3)})$
    independent across hospitals
    $\psi^{(3)}$ is residual between-hospital variance
  - Level 2: $\zeta_{jk}^{(2)}|\mathbf{x}_{ijk}, \zeta_k^{(3)} \sim N(0, \psi^{(2)})$
    independent across doctors, independent of $\zeta_k^{(3)}$
    $\psi^{(2)}$ is residual between-doctor, within-hospital variance

- `gllamm` command:

```
gllamm abuse age temp Paymed Selfmed Wrdiag DRed WHO, ///
   i(doc hosp) link(logit) family(binom) adapt
```

Maximum likelihood estimates

	No covariates		Full model		
Parameter	Est	(SE)	Est	(SE)	(OR)
β_0 [Cons]	0.87	(0.14)	1.52	(0.46)	
β_1 [Age]			0.14	(0.07)	1.15
β_2 [Temp]			−0.72	(0.10)	0.49
β_3 [Paymed]			0.38	(0.30)	1.46
β_4 [Selfmed]			−0.65	(0.21)	0.52
β_5 [Wrdiag]			1.97	(0.20)	7.18
β_6 [DRed]			−0.20	(0.10)	0.82
β_7 [WHO]			−1.26	(0.32)	0.28
$\psi^{(2)}$	0.20		0.14		
$\psi^{(3)}$	0.36		0.19		
Log-likelihood	−512.14		−415.76		

using `gllamm` with adaptive quadrature

Distributions of random effects and responses

- Vector of all random intercepts for hospital k

$$\boldsymbol{\zeta}_{k(3)} \equiv (\zeta_{1k}^{(2)}, \ldots, \zeta_{J_k k}^{(2)}, \zeta_k^{(3)})'$$

- Random effects distribution [Prior distribution]

$$\varphi(\boldsymbol{\zeta}_{k(3)}), \quad \varphi(\zeta_{jk}^{(2)}), \quad \varphi(\zeta_k^{(3)}), \quad \text{all (multivariate) normal}$$

- Conditional response distribution of all responses $\mathbf{y}_{k(3)}$ for hospital k, given all covariates $\mathbf{X}_{k(3)}$ and all random effects $\boldsymbol{\zeta}_{k(3)}$ for hospital k [Likelihood]

$$f(\mathbf{y}_{k(3)}|\mathbf{X}_{k(3)}, \boldsymbol{\zeta}_{k(3)}) = \prod_{\substack{\text{all docs } j \\ \text{in hosp } k}} \; \prod_{\substack{\text{all patients } i \\ \text{of doc } j}} f(y_{ijk}|\mathbf{x}_{ijk}, \zeta_{jk}^{(2)}, \zeta_k^{(3)})$$

Posterior distribution

- Use Bayes theorem to obtain posterior distribution of random effects given the data:

$$\omega(\boldsymbol{\zeta}_{k(3)}|\mathbf{y}_{k(3)},\mathbf{X}_{k(3)}) = \frac{\varphi(\boldsymbol{\zeta}_{k(3)})f(\mathbf{y}_{k(3)}|\mathbf{X}_{k(3)},\boldsymbol{\zeta}_{k(3)})}{\int\varphi(\boldsymbol{\zeta}_{k(3)})f(\mathbf{y}_{k(3)}|\mathbf{X}_{k(3)},\boldsymbol{\zeta}_{k(3)})\mathrm{d}\boldsymbol{\zeta}_{k(3)}}$$

$$\propto \varphi(\zeta_k^{(3)})\prod_j \varphi(\zeta_{jk}^{(2)})\prod_i f(y_{ijk}|\mathbf{x}_{ijk},\zeta_{jk}^{(2)},\zeta_k^{(3)})$$

- Denominator, marginal likelihood contribution for hospital k, simplifies

$$\int\varphi(\zeta_k^{(3)})\prod_j\left[\int\varphi(\zeta_{jk}^{(2)})\prod_i f(y_{ijk}|\mathbf{x}_{ijk},\zeta_{jk}^{(2)},\zeta_k^{(3)})\,\mathrm{d}\zeta_{jk}^{(2)}\right]\mathrm{d}\zeta_k^{(3)}$$

Empirical Bayes prediction of random effects

- Empirical Bayes (EB) prediction is mean of posterior distribution

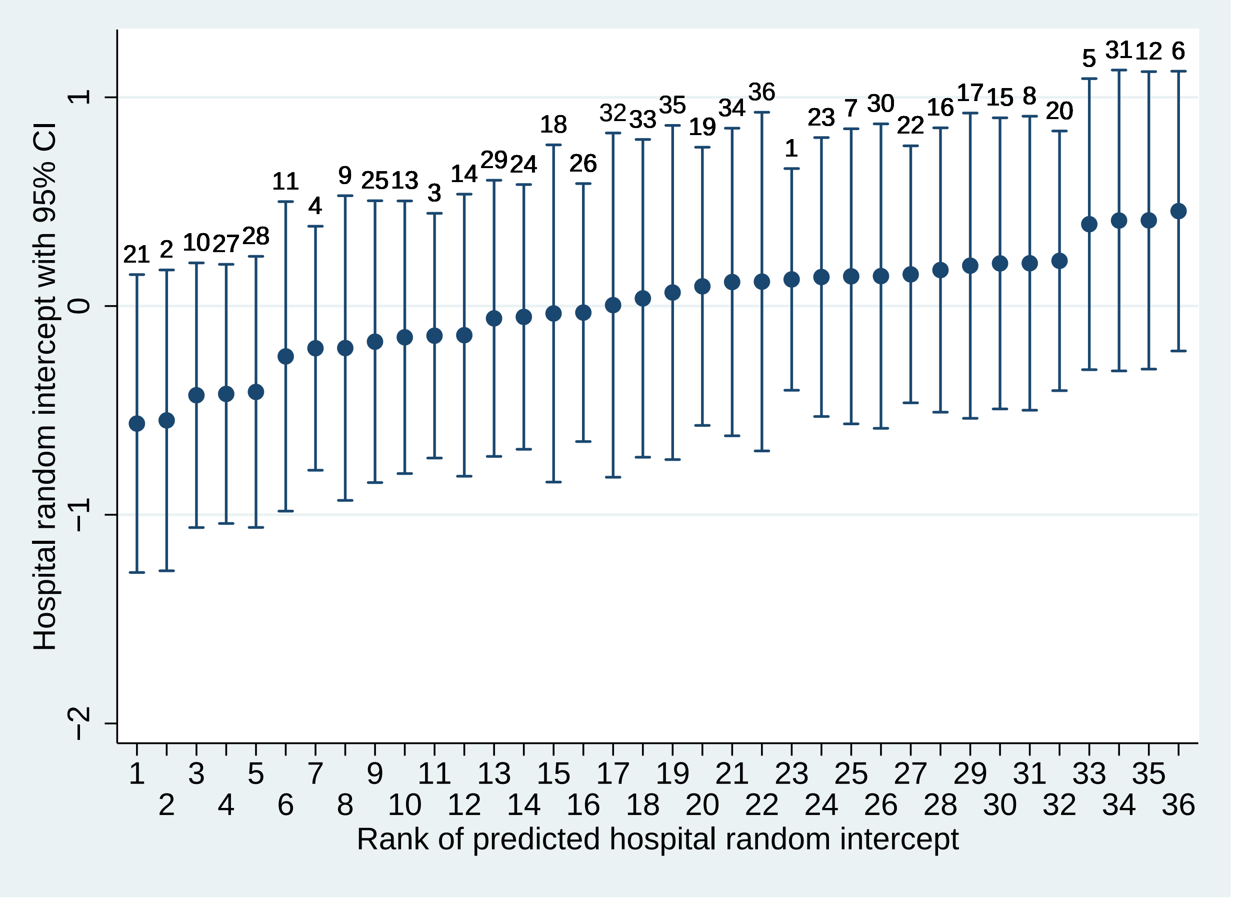
$$\widetilde{\boldsymbol{\zeta}}_{k(3)} = \int \boldsymbol{\zeta}_{k(3)}\, \omega(\boldsymbol{\zeta}_{k(3)}|\mathbf{y}_{k(3)}, \mathbf{X}_{k(3)})\, \mathrm{d}\boldsymbol{\zeta}_{k(3)}$$

- Standard error of EB is standard deviation of posterior distribution
- Using `gllapred` with the u option

```
gllapred eb, u
```

  - `ebm1` contains $\widetilde{\zeta}_{jk}^{(2)}$
  - `ebs1` contains SE($\widetilde{\zeta}_{jk}^{(2)}$)
  - `ebm2` contains $\widetilde{\zeta}_{k}^{(3)}$
  - `ebs2` contains SE($\widetilde{\zeta}_{k}^{(3)}$)
- For approximately normal posterior, use Wald-type interval, e.g., for hospital k, 95% CI is $\widetilde{\zeta}_{k}^{(3)} \pm 1.96$ SE($\widetilde{\zeta}_{k}^{(3)}$)

Confidence intervals for hospital random effects



- $\widetilde{\zeta}_k^{(3)} \pm 1.96\,\text{SE}(\widetilde{\zeta}_k^{(3)})$
- Identify the good and bad with caution

Predicted probability for patient of hypothetical doctor

- Predicted **conditional probability** for hypothetical values $\mathbf{x}^0$ of the covariates and ζ^0 of the random intercepts

$$\widehat{\text{Pr}}(y = 1|\mathbf{x}^0, \boldsymbol{\zeta}^0) \;=\; \frac{\exp(\mathbf{x}^{0\prime}\widehat{\boldsymbol{\beta}} + \zeta^{(2)0} + \zeta^{(3)0})}{1 + \exp(\mathbf{x}^{0\prime}\widehat{\boldsymbol{\beta}} + \zeta^{(2)0} + \zeta^{(3)0})}$$

- If $\zeta^{(2)0} + \zeta^{(3)0} = 0$, median of distribution for $\zeta_{jk}^{(2)} + \zeta_{k}^{(3)}$, then predicted conditional probability is median probability
  - Analogously for other percentiles
- Using `gllapred` with `mu` and `us()` option:

```
replace age = 2 /* etc.: change covariates to x⁰ */
generate zeta1 = 0
generate zeta2 = 0
gllapred probc, mu us(zeta)
```

Predicted probability for new patient of existing doctor in existing hospital

- **Posterior mean probability** for new patient of existing doctor j in hospital k

$$\widetilde{\text{Pr}}_{jk}(y=1|\mathbf{x}^0) = \int \widehat{\text{Pr}}(y=1|\mathbf{x}^0, \boldsymbol{\zeta}_{k(3)})\, \omega(\boldsymbol{\zeta}_{k(3)}|\mathbf{y}_{k(3)}, \mathbf{X}_{k(3)})\, \mathrm{d}\boldsymbol{\zeta}_{k(3)}$$

  - Invent additional patient i^*jk with covariate values $\mathbf{x}_{i^*jk} = \mathbf{x}^0$
  - Make sure that invented observation does not contribute to posterior $\omega(\boldsymbol{\zeta}_{k(3)}|\mathbf{y}_{k(3)}, \mathbf{X}_{k(3)})$

$$\omega(\boldsymbol{\zeta}_{k(3)}|\mathbf{y}_{k(3)}, \mathbf{X}_{k(3)}) \propto \varphi(\zeta_k^{(3)}) \prod_j \varphi(\zeta_{jk}^{(2)}) \prod_{i \neq i^*} f(y_{ijk}|\mathbf{x}_{ijk}, \zeta_{jk}^{(2)}, \zeta_k^{(3)})$$

  - Cannot simply plug in EB prediction $\widetilde{\boldsymbol{\zeta}}_{k(3)}$ for $\boldsymbol{\zeta}_{k(3)}$

$$\widetilde{\text{Pr}}_{jk}(y=1|\mathbf{x}^0) \;\neq\; \widehat{\text{Pr}}(y=1|\mathbf{x}^0, \boldsymbol{\zeta}_{k(3)} = \widetilde{\boldsymbol{\zeta}}_{k(3)})$$

Prediction dataset: One new patient per doctor

Data (ignore gaps)

id	doc	hosp	abuse
1	1	1	0
2	1	1	1
3	2	2	0
4	3	2	1
5	3	2	1

Data with invented observations

id	doc	hosp	abuse
1	1	1	0
2	1	1	1
.	1	1	.
3	2	2	0
.	2	2	.
4	3	2	1
5	3	2	1
.	3	2	.

- Response variable **abuse** must be missing for invented observations
- Use required value of **doc**
- Can invent several patients per doctor

Prediction dataset:
One new patient per doctor (continued)

Data with invented observations

				terms for posterior		
id	**doc**	**hosp**	**abuse**	hospital	doctor	patient
1	**1**	**1**	**0**	$\varphi(\zeta_1^{(3)})$	$\varphi(\zeta_{11}^{(2)})$	$f(y_{111}\|\zeta_1^{(3)}, \zeta_{11}^{(2)})$
2	**1**	**1**	**1**			$f(y_{211}\|\zeta_1^{(3)}, \zeta_{11}^{(2)})$
.	**1**	**1**	**.**			1
3	**2**	**2**	**0**	$\varphi(\zeta_2^{(3)})$	$\varphi(\zeta_{22}^{(2)})$	$f(y_{322}\|\zeta_2^{(3)}, \zeta_{22}^{(2)})$
.	**2**	**2**	**.**			1
4	**3**	**2**	**1**		$\varphi(\zeta_{32}^{(2)})$	$f(y_{432}\|\zeta_2^{(3)}, \zeta_{32}^{(2)})$
5	**3**	**2**	**1**			$f(y_{532}\|\zeta_2^{(3)}, \zeta_{32}^{(2)})$
.	**3**	**2**	**.**			1

- Using `gllapred` with `mu` and `fsample` options:

```
gllapred probd, mu fsample
```

Predicted probability for new patient of new doctor in existing hospital

- Posterior mean probability for new patient of new doctor in existing hospital k

$$\widetilde{\Pr}_k(y=1|\mathbf{x}^0) = \int \widehat{\Pr}(y=1|\mathbf{x}^0, \boldsymbol{\zeta}^*_{k(3)})\, \omega(\boldsymbol{\zeta}^*_{k(3)}|\mathbf{y}_{k(3)}, \mathbf{X}_{k(3)})\, \mathrm{d}\boldsymbol{\zeta}^*_{3(k)}$$

  - Invent additional observation i^*j^*k with covariates in $\mathbf{x}_{i^*j^*k} = \mathbf{x}^0$
  - $\boldsymbol{\zeta}^*_{k(3)} = (\zeta^{(2)}_{j^*k}, \boldsymbol{\zeta}'_{k(3)})'$
  - Make sure that invented doctor **but not invented patient** contribute to posterior $\omega(\boldsymbol{\zeta}^*_{k(3)}|\mathbf{y}_{k(3)}, \mathbf{X}_{k(3)})$

$$\omega(\boldsymbol{\zeta}^*_{k(3)}|\mathbf{y}_{k(3)}, \mathbf{X}_{k(3)}) \propto \overbrace{\varphi(\zeta^{(2)}_{j^*k})}^{\text{Prior}}\, \omega(\boldsymbol{\zeta}_{k(3)}|\mathbf{y}_{k(3)}, \mathbf{X}_{k(3)})$$

Prediction dataset: One new doctor and patient per hospital

Data (ignore gaps)

id	doc	hosp	abuse
1	1	1	0
2	1	1	1
3	2	2	0
4	3	2	1
5	3	2	1

Data with invented observations

id	doc	hosp	abuse
1	1	1	0
2	1	1	1
.	0	1	.
3	2	2	0
4	3	2	1
5	3	2	1
.	0	2	.

- Response variable **`abuse`** must be missing for invented observations
- Use unique (for that hospital) value of **`doc`**
- Can invent several new docs which can all have the same value of **`doc`**

Prediction dataset: One new doctor and patient per hospital (continued)

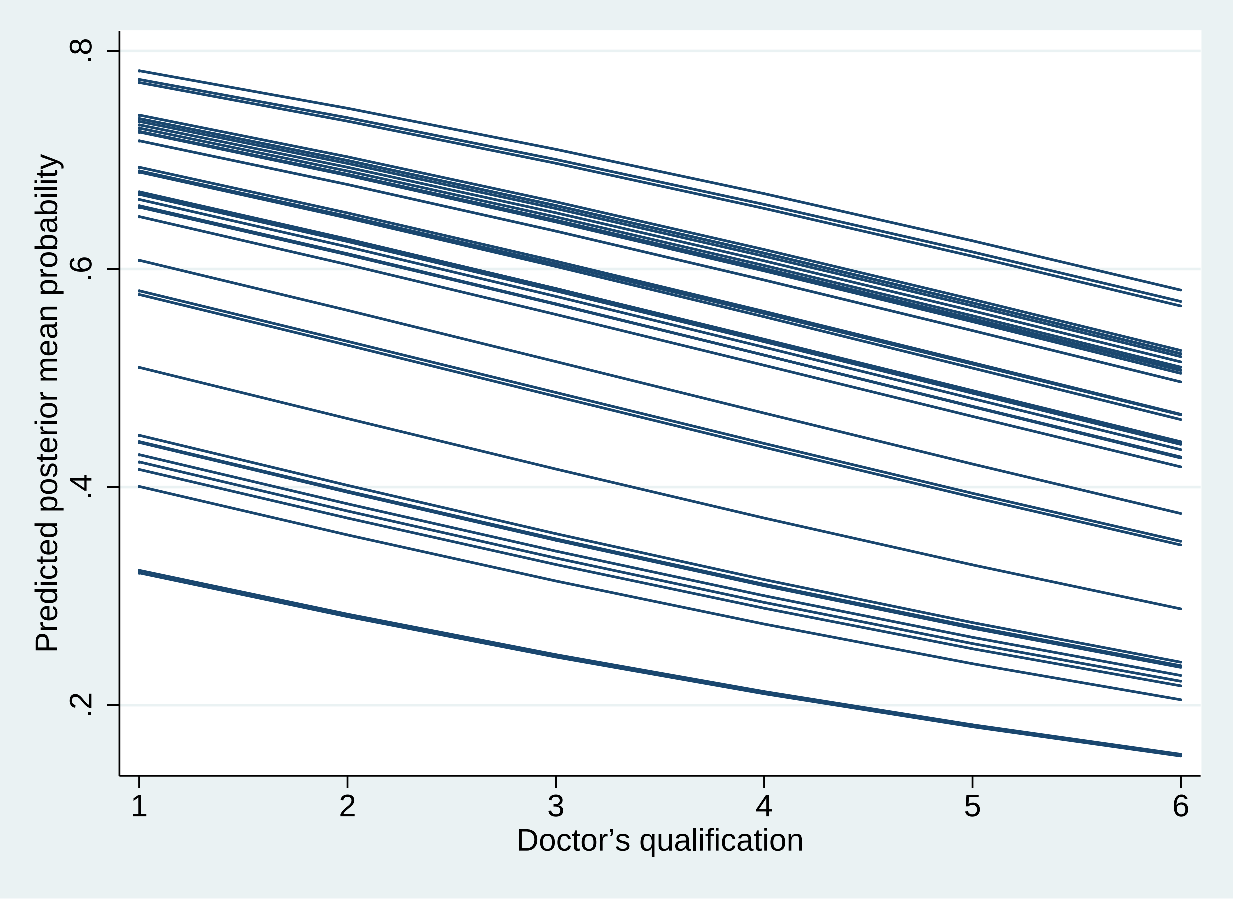
Data with invented observations

				terms for posterior		
id	**doc**	**hosp**	**abuse**	hospital	doctor	patient
1	**1**	**1**	**0**	$\varphi(\zeta_1^{(3)})$	$\varphi(\zeta_{11}^{(2)})$	$f(y_{111}\|\zeta_1^{(3)}, \zeta_{11}^{(2)})$
2	**1**	**1**	**1**			$f(y_{211}\|\zeta_1^{(3)}, \zeta_{11}^{(2)})$
.	**0**	**1**	**.**		$\varphi(\zeta_{01}^{(2)})$	1
3	**2**	**2**	**0**	$\varphi(\zeta_2^{(3)})$	$\varphi(\zeta_{22}^{(2)})$	$f(y_{322}\|\zeta_2^{(3)}, \zeta_{22}^{(2)})$
4	**3**	**2**	**1**		$\varphi(\zeta_{32}^{(2)})$	$f(y_{432}\|\zeta_2^{(3)}, \zeta_{32}^{(2)})$
5	**3**	**2**	**1**			$f(y_{532}\|\zeta_2^{(3)}, \zeta_{32}^{(2)})$
.	**0**	**2**	**.**		$\varphi(\zeta_{02}^{(2)})$	1

- Using `gllapred` with `mu` and `fsample` options:

```
gllapred probh, mu fsample
```

Example: Predicted probability for new patient of new doctor in existing hospital

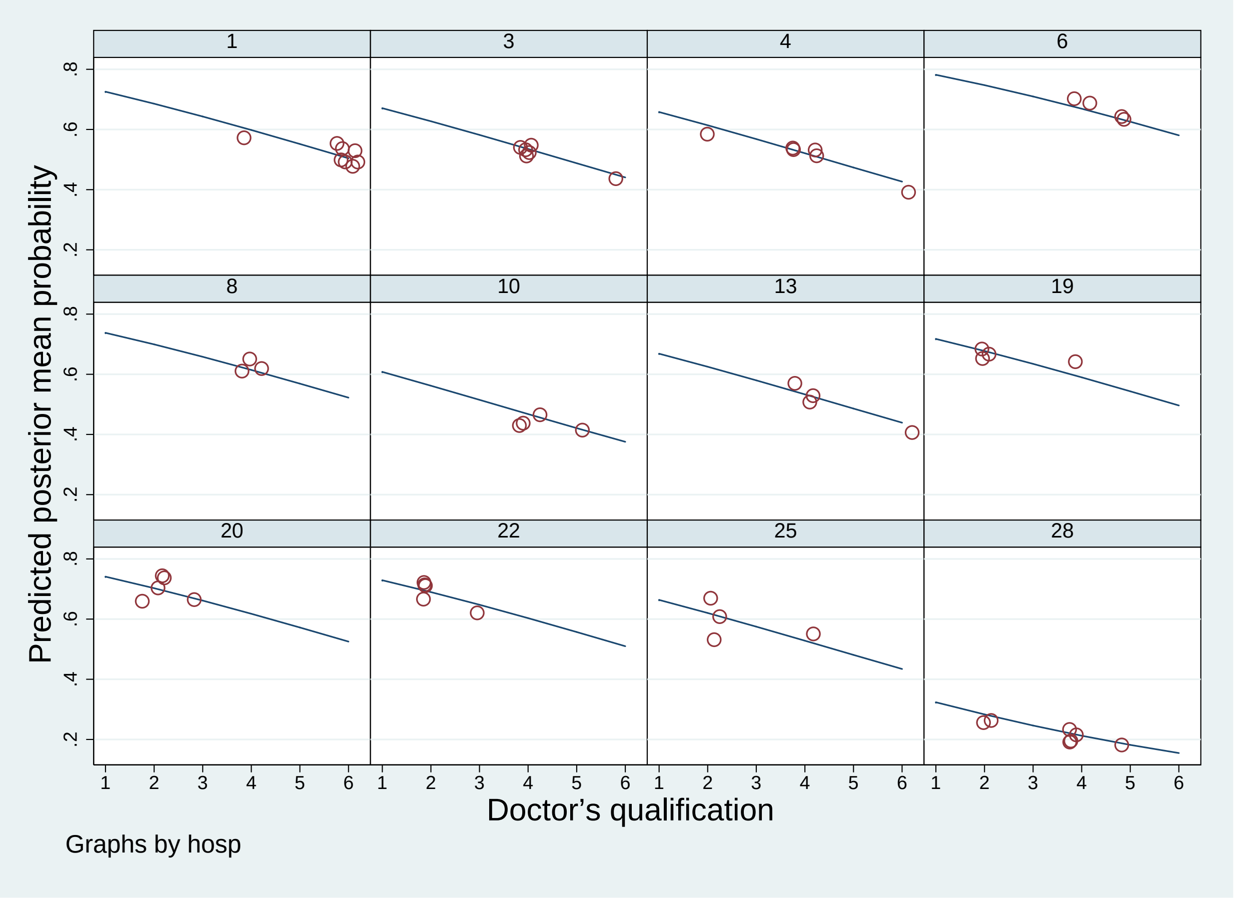


- Each curve represents a hospital

  For each hospital: 6 new doctors with [DRed] $= 1, 2, 3, 4, 5, 6$

  For each doctor: 1 new patient with [Age] $= 2$, [Temp] $= 1$ ($37°C$), [Paymed] $= 0$, [Selfmed] $= 0$, [Wrdiag] $= 0$

Example: Predicted probability for new patient of existing doctor in existing hospital



- 12 of the hospitals, with curves as in previous slide
- Dots represent doctors with [DRed] as observed

  For each doctor: predicted probability for 1 new patient with [Age] $= 2$, [Temp] $= 1$, [Paymed] $= 0$, [Selfmed] $= 0$, [Wrdiag] $= 0$

Predicted probability for new patient of new doctor in new hospital

- Population-averaged or **marginal probability**:

$$\overline{\text{Pr}}(y=1|\mathbf{x}^0) \quad = \quad \int \widehat{\text{Pr}}(y=1|\mathbf{x}^0, \zeta_{jk}^{(2)}, \zeta_k^{(3)})\, \varphi(\zeta_{jk}^{(2)}), \varphi(\zeta_k^{(3)})\, \mathrm{d}\zeta_{jk}^{(2)} \mathrm{d}\zeta_k^{(3)}$$

- Cannot plug in means of random intercepts

$$\overline{\text{Pr}}(y=1|\mathbf{x}^0) \quad \neq \quad \widehat{\text{Pr}}(y=1|\mathbf{x}^0, \zeta_{jk}^{(2)}=0, \zeta_k^{(3)}=0)$$

mean ≠ median

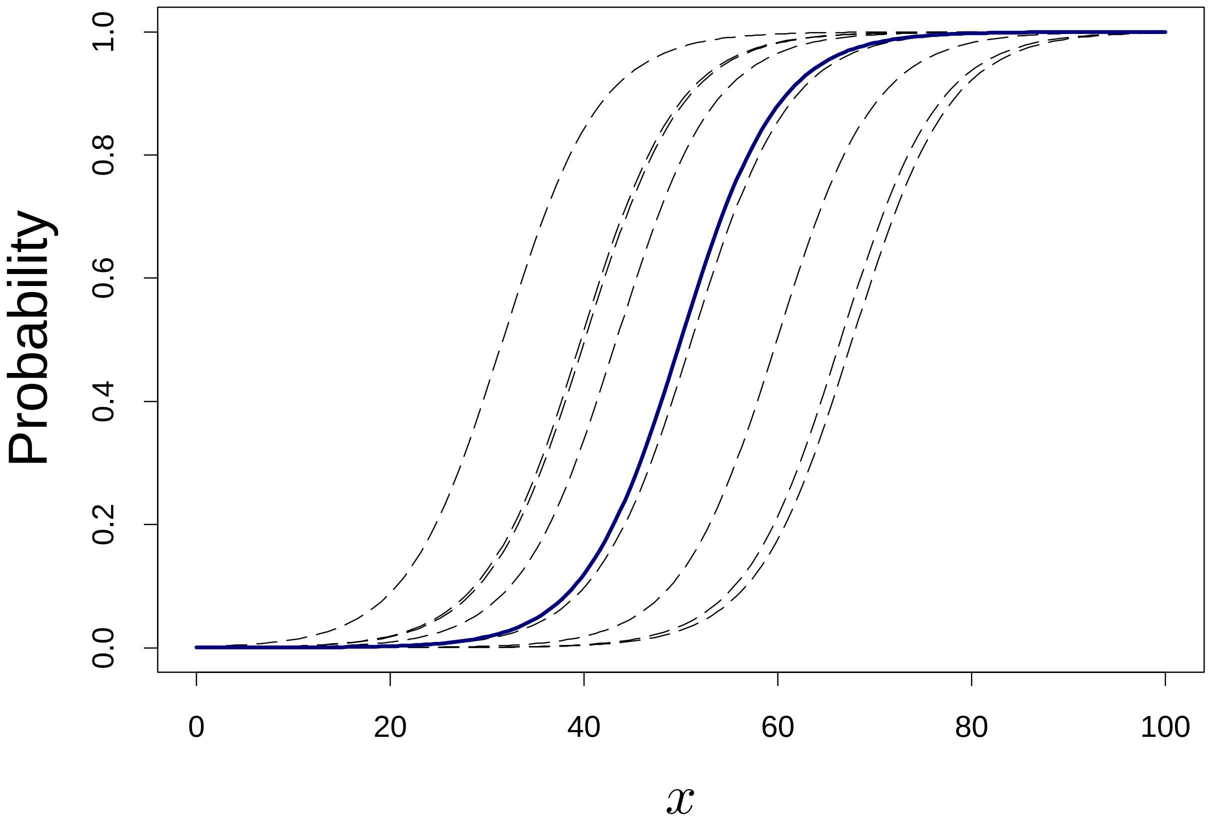
- Using `gllapred` with the `mu` and `marg` options:

```
gllapred prob, mu marg fsample
```

- Confidence interval, by sampling parameters from the estimated asymptotic sampling distribution of their estimates

```
ci_marg_mu lower upper, level(95) dots
```

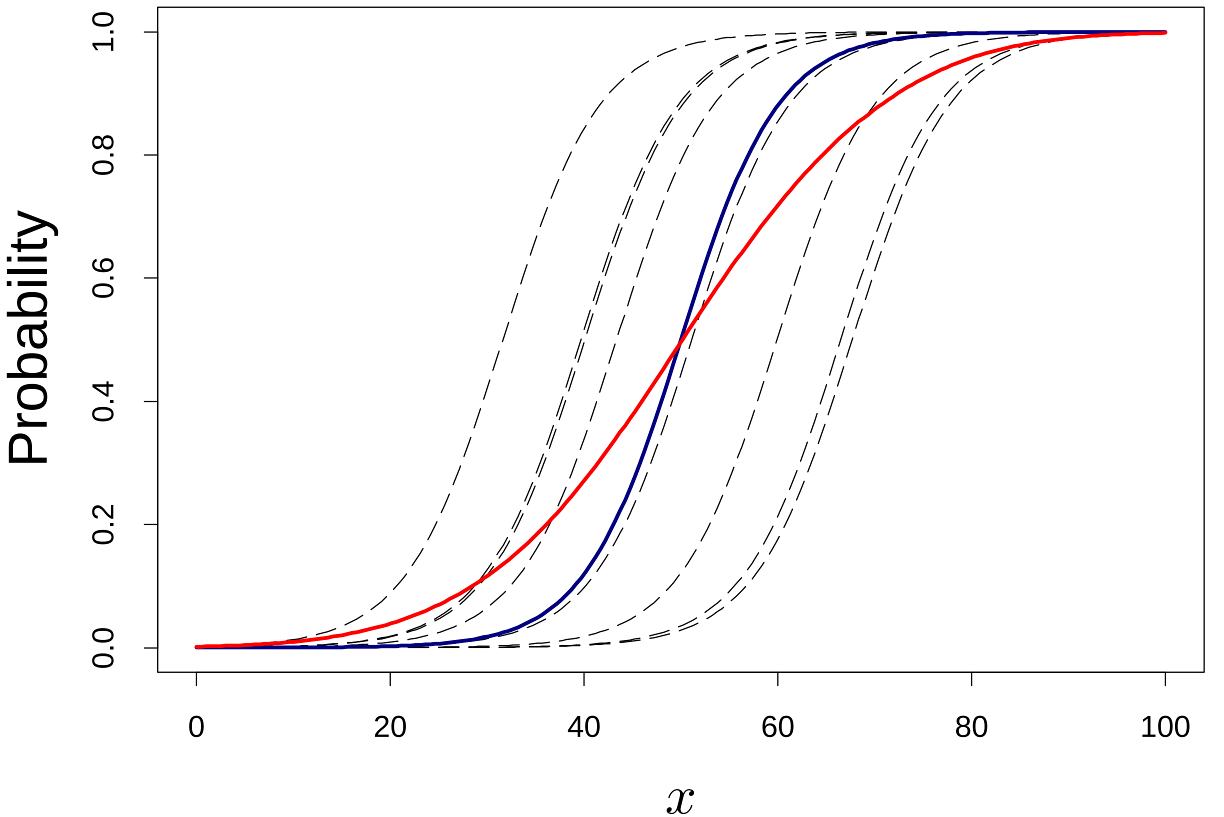
Illustration Cluster-specific: versus population averaged probability



- - - - - cluster-specific (random sample)

——— median

Illustration Cluster-specific: versus population averaged probability

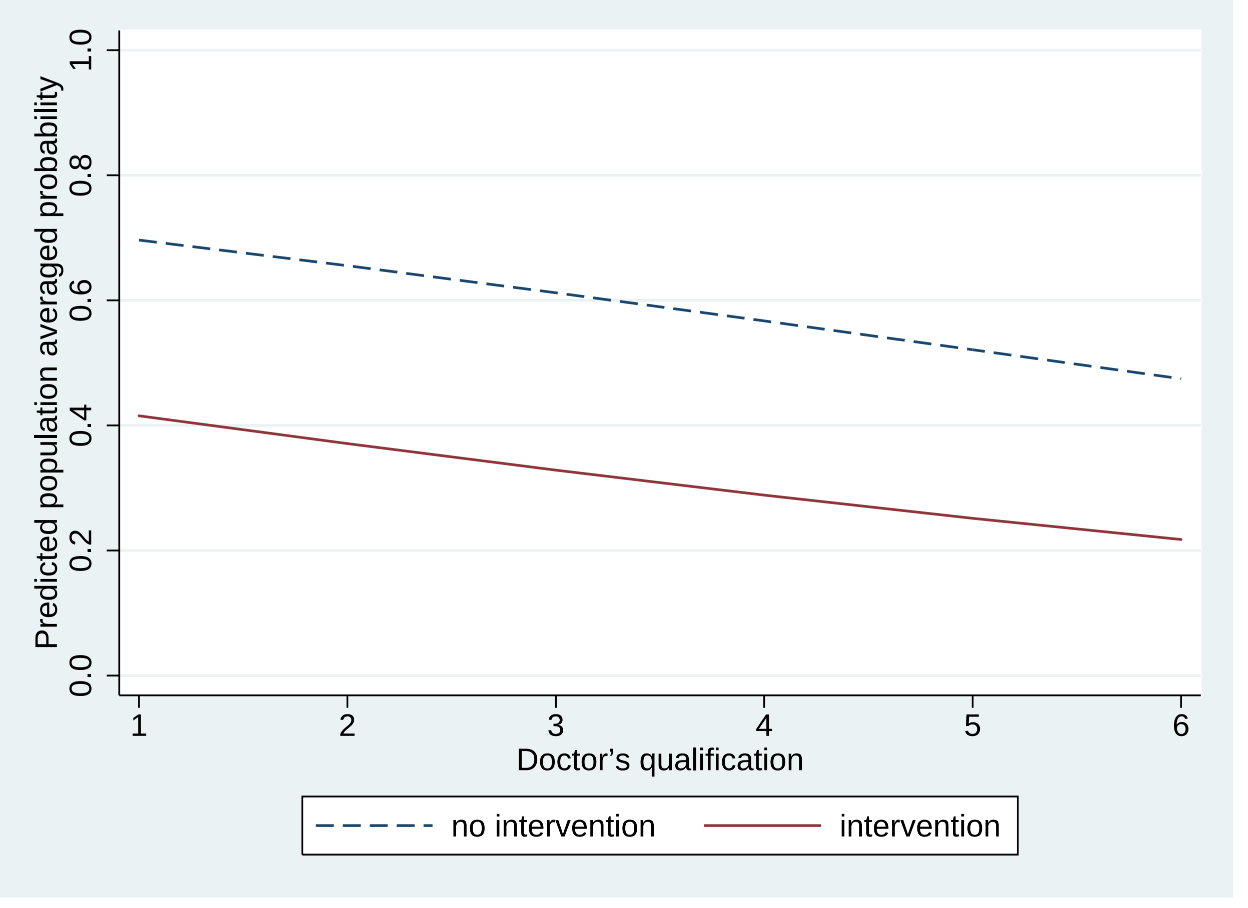


- - - - - cluster-specific (random sample)

—— median

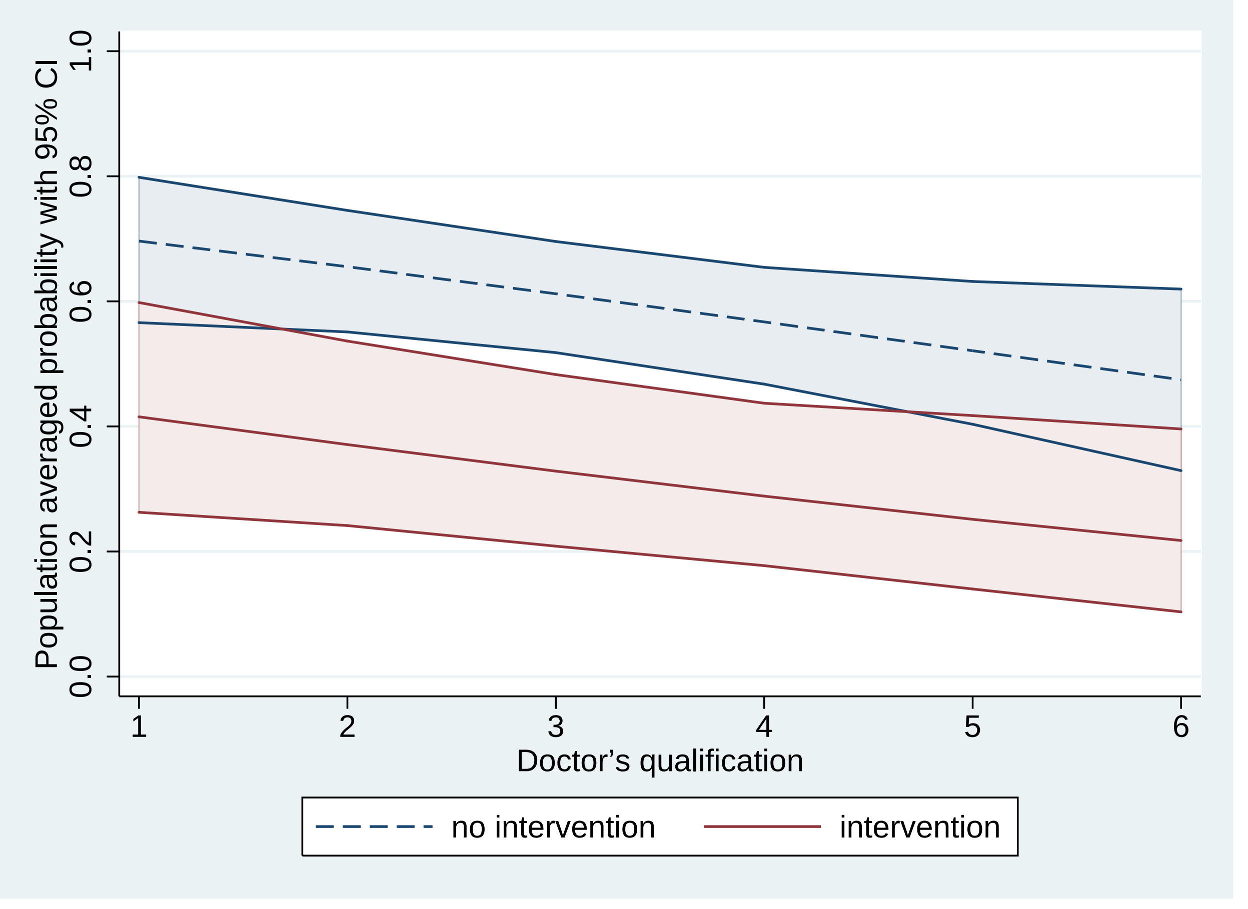
—— population averaged

Example: Predicted probability for new patient of new doctor in new hospital



- Same patient covariates as before
-

Example: Predicted probability for new patient of new doctor in new hospital

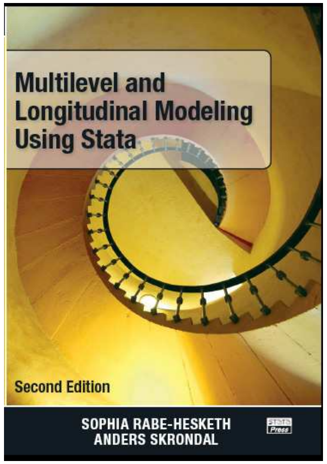


- Same patient covariates as before
- Confidence bands represent parameter uncertainty

Concluding remarks

- Discussed:
  - Empirical Bayes (EB) prediction of random effects and CI using `gllapred`, ignoring parameter uncertainty
  - Prediction of different kinds of probabilities using `gllapred` after careful preparation of prediction dataset
  - Simulation-based CI for predicted marginal probabilities using new command `ci_marg_mu`
- Methods work for any GLLAMM model, including random-coefficient models and models for ordinal, nominal or count data
- Assumed normal random effects distribution
  - EB predictions not robust to misspecification of distribution
  - Could use nonparametric maximum likelihood in `gllamm`, followed by same `gllapred` and `ci_marg_mu` commands

References

- 

Rabe-Hesketh, S. and Skrondal, A. (2008). *Multilevel and Longitudinal Modeling Using Stata (2nd Edition)*. College Station, TX: Stata Press.
- Skrondal, A. and Rabe-Hesketh, S. (2009). Prediction in multilevel generalized linear models. *Journal of the Royal Statistical Society, Series A*, in press.
- Rabe-Hesketh, S., Skrondal, A. and Pickles, A. (2005). Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics* 128, 301-323.