Slide 1

Multilevel modelling of ordered and unordered categorical responses

Sophia Rabe-Hesketh
Institute of Psychiatry
King's College London

Joint work with
Anders Skrondal, Norwegian Institute of Public Health, Oslo
and
Andrew Pickles, The University of Manchester

Institute of Child Health
London, February 2003

Slide 2

Outline

- Ordered categorical responses
  - Single-level models
  - Multilevel models
    - Two-level random intercept models
    - Cluster-specific and population average effects
    - Multilevel random coefficient models
  - Estimation methods
  - Analysis of cluster randomized trial of sex education
- Unordered categorical responses
  - Single-level models
  - Multilevel models
  - Analysis of abuse of antibiotics in young children in China

Slide 3

I. Ordered categorical responses

- Only a small number of responses or 'categories' are possible, y_s, $s = 1, \cdots, S$
- The categories are ordered,

$$y_1 < y_2 < \cdots < y_S, \quad \text{we'll set } y_s = s$$

- Examples:
  - Severity of symptoms (e.g. pain): 'none', 'moderate', 'severe'
  - Frequency of symptoms: 'never', 'occasionally', 'nearly every day', 'every day'
  - Response to treatment: 'progressive disease', 'no change', 'partial remission', 'complete remission'
  - Diagnosis: 'non-autistic', 'PDD-NOS' (pervasive development disorder), 'autistic'
  - Test results (e.g. breathing): 'normal', 'borderline', 'abnormal'
  - Satisfaction with treatment: 'very dissatisfied', 'dissatisfied', 'satisfied', 'very satisfied'

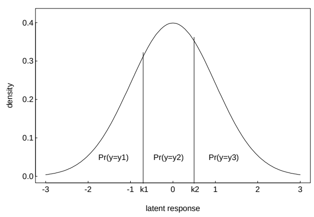
Slide 4

Latent response models

- Underlying the observed ordinal response y_i for subject i is a latent (unobserved) continous response y_i^*.
- A threshold model determines the observed response:

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

- For a probit model, $y_i^* \sim N(0,1)$, with $S = 3$ categories:



- The latent response is modelled as a linear regression without an intercept (for identification)

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

Slide 5

Generalized linear models

- The cumulative probabilities are modeled as

$$\Pr(y_i > s) = F(\boldsymbol{\beta}'\mathbf{x}_i - \kappa_s),\ \ s = 1, \cdots, S-1,$$

giving *cumulative* models. F is the inverse *link function*

- Proportional odds model (logit link):

$$\Pr(y_i > s) = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_i - \kappa_s)}{1 + \exp(\boldsymbol{\beta}'\mathbf{x}_i - \kappa_s)}$$

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

  – The exponentiated regression coefficients can be interpreted as odds ratios for high versus low scores regardless of cut-point (proportional odds):

$$\frac{\Pr(y_i > s)/[1 - \Pr(y_i > s)]}{\Pr(y_j > s)/[1 - \Pr(y_j > s)]} = \exp(\boldsymbol{\beta}'\{\mathbf{x}_i - \mathbf{x}_j\})$$

- Cumulative models are equivalent to latent response models:

Model	Link F^{-1}	Distribution of ϵ_i	Variance of ϵ_i
Proportional odds	logit	logistic	$\pi^2/3$
Ordinal probit	probit	standard normal	1
Compl. log-log	Compl. log-log	Gumbel	$\pi^2/6$

Slide 6

Two-level random intercept models

- Subjects i nested in clusters j (e.g. hospitals). Include a random intercept u_j for clusters in the latent response model

$$y^*_{ij} = \boldsymbol{\beta}'\mathbf{x}_{ij} + u_j + \epsilon_{ij},\ \ u_j \sim N(0, \tau^2),\ \ u_j \text{ indep. of } \epsilon_{ij}.$$

- The total residual $\xi_{ij} = u_j + \epsilon_{ij}$ has variance

$$\text{var}(\xi_{ij}) = \begin{cases} \tau^2 + 1 & \text{for probit models} \\ \tau^2 + \pi^2/3 & \text{for logit models} \end{cases}$$

- The covariance between the total residuals ξ_{ij} and $\xi_{i'j}$ of two subjects in the same cluster is τ^2 and the *intraclass correlation* is

$$\rho \equiv \text{Cor}(\xi_{ij}, \xi_{i'j}) = \begin{cases} \tau^2/(\tau^2 + 1) & \text{for probit models} \\ \tau^2/(\tau^2 + \pi^2/3) & \text{for logit models} \end{cases}$$

- The latent responses for two units in the same cluster are conditionally independent given the random intercept:

$$\text{Cor}(y^*_{ij}, y^*_{i'j}|\mathbf{x}_{ij}, u_j) = 0$$

If we do not condition on the random intercept, the correlation is the intraclass correlation

$$\text{Cor}(y^*_{ij}, y^*_{i'j}|\mathbf{x}_{ij}) = \rho$$

Slide 7

Cluster-specific versus population average effects

- For a probit model, the 'marginal' or population average response probabilities are

$$\begin{aligned}\Pr(y_{ij} > s) &= \Pr(y_i^* > \kappa_s) = \Pr(\boldsymbol{\beta}'\mathbf{x}_{ij} + \xi_{ij} > \kappa_s)\\ &= \Pr(-\xi_{ij} \leq \boldsymbol{\beta}'\mathbf{x}_{ij} - \kappa_s)\\ &= \Pr\left(\frac{\xi_{ij}}{\sqrt{\tau^2+1}} \leq \frac{\boldsymbol{\beta}'\mathbf{x}_{ij} - \kappa_s}{\sqrt{\tau^2+1}}\right)\\ &= \Phi\left(\frac{\boldsymbol{\beta}'\mathbf{x}_{ij} - \kappa_s}{\sqrt{\tau^2+1}}\right),\end{aligned}$$

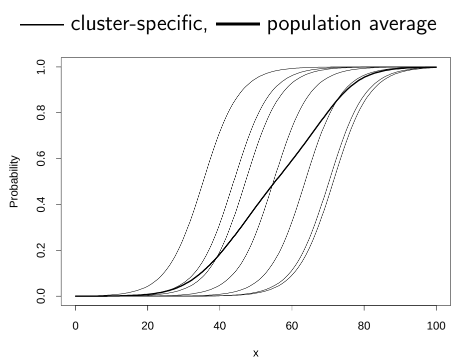
  where $\xi_{ij} = u_j + \epsilon_{ij}$.
  - Therefore, the marginal effects of $\mathbf{x}_{ij}$ are $\boldsymbol{\beta}/\sqrt{\tau^2+1}$. To achieve a given marginal effect, $\boldsymbol{\beta}$ must increase if τ^2 increases.
- The 'conditional' probabilities for a given cluster j are

$$\Pr(y_{ij} > s|u_j) = \Phi\left(\frac{\boldsymbol{\beta}'\mathbf{x}_{ij} + u_j - \kappa_s}{1}\right).$$

  - Therefore, the conditional or cluster-specific effects $\boldsymbol{\beta}$ of $\mathbf{x}_{ij}$ are greater than the marginal or population average effects $\boldsymbol{\beta}/\sqrt{\tau^2+1}$.

Slide 8

Cluster-specific versus population average effects



Slide 9

Multilevel random coefficient models

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

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

  – $u^{(2)}_{0jk}$ and $u^{(3)}_{0k}$ are random intercepts at levels 2 and 3.
  – $u^{(2)}_{1jk}$ and $u^{(3)}_{1k}$ are random coefficients of x_{1ijk}.
  – The random coefficients have zero means and the fixed effect β_1 of x_{1ijk} represents the mean effect.
  – Random effects at the same level are correlated, $(u^{(2)}_{0jk}, u^{(2)}_{1jk})$ is bivariate normal.
- General three-level random coefficient model

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

Slide 10

Estimation

- Estimation of multilevel models with categorical responses, also known as generalised linear mixed models (GLMMs), is not easy because the likelihood does not generally have a closed form.
- Marginal Quasilikelihood (MQL) and Penalized Quasilikelihood (PQL) are approximate methods avaiblable in `MLwiN` and `HLM`.
  – Two versions are available, first and second order (MQL-1,MQL-2,PQL-1,PQL-2), the last being the best.
  – Even PQL-2 sometimes produces biased estimates, particularly when the clusters are small.
  – The methods do not provide a likelihood.
- Maximum likelihood estimation requires evaluation of integrals since the likelihood is marginal with respect to the random effects.
  – Numerical integration using Gauss-Hermite quadrature is used in `MIXOR/MIXNO` (two-level only), aML, `SAS PROC NLMIXED` (two-level only) and `gllamm`.
  – Adaptive quadrature is superior to 'ordinary quadrature' which sometimes doesn't work (e.g. large clusters, counts). This is available in `SAS PROC NLMIXED` (two-level only) and `gllamm`.
  – `HLM` provides a 6th order Laplace approximation for two-level models with dichotomous responses.

Slide 11

Cluster randomized trial of sex education

- Schools were randomised to receive sex education or not
- Assessments pre randomisation, 6 months and 18 months post randomisation
- One outcome is a question relating to 'Contraceptive self-efficacy':
  - "If my partner and I were about to have intercourse without either of us having mentioned contraception, it would be easy for me to produce a condom (if I brought one)"
  - The questions is answered in terms of five ordinal categories: 'not at all true of me', 'slightly true of me', 'somewhat true of me', 'mostly true of me', 'completely true of me'.
- The data are multilevel with responses (level 1) from 1184 pupils (level 2) from 46 schools (level 3).
- Only 570 pupils always responded, 400 responded on some occasions and 114 never responded.

Slide 12

Models

- Occasions t, subjects j, schools k
- Covariates
  - x_{1t} [Time] (0, 1, 3)
  - x_{2jk} [Treat] (yes=1,no=0)
  - x_{3tjk} [Treat] $\times$ [Time]
- Model the probability of exceeding a category s, $s = 1, 2, 3, 4$

  $$\text{logit}[\Pr(y_{tijk} > s)] = \beta_1 x_{1t} + \beta_2 x_{2jk} + \beta_3 x_{3tjk} + u_{jk}^{(2)} + u_k^{(3)} - \kappa_s$$
- Estimation using adaptive quadrature in `gllamm`:

```
gllamm use treat time treat_time, i(id school) /*
   */ link(ologit) family(binom) adapt
```

- Conditional and marginal probabilities:

```
gen u1 = 0
gen u2 = 0
gllapred p_cond, mu us(u) above(2)
gllapred p_marg, mu marg above(2)
```

Slide 13

Estimates

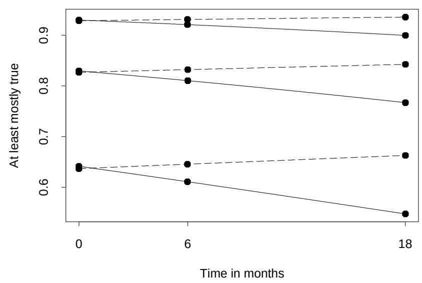
Parameter	Single-level model Est	(SE)	Two-level model Est	(SE)	Three-level model Est	(SE)
β_1 [Time]	-0.12	(0.06)	-0.13	(0.06)	-0.13	(0.06)
β_2 [Treat]	-0.05	(0.14)	-0.02	(0.19)	-0.02	(0.19)
β_3 [Time]×[Treat]	0.17	(0.08)	0.17	(0.09)	0.17	(0.09)
$\text{var}(u_{jk}^{(2)})$	–		2.03	(0.31)	2.03	(0.31)
$\text{var}(u_{k}^{(2)})$	–		–		0.00	(0.00)
κ_1	-3.54	(0.17)	-4.41	(0.23)	-4.41	(0.23)
κ_2	-2.43	(0.13)	-3.15	(0.19)	-3.15	(0.19)
κ_3	-1.18	(0.12)	-1.58	(0.16)	-1.58	(0.16)
κ_4	0.16	(0.12)	0.25	(0.15)	0.25	(0.15)
Log-likelihood	-2531		-2471		-2471	

Slide 14

Conditional relationships

- Probability of responding at least 'mostly true of me' (category 3)
- Relationship between the probability and occasion for the two treatment groups, when the random intercept is -1, 0 and 1:

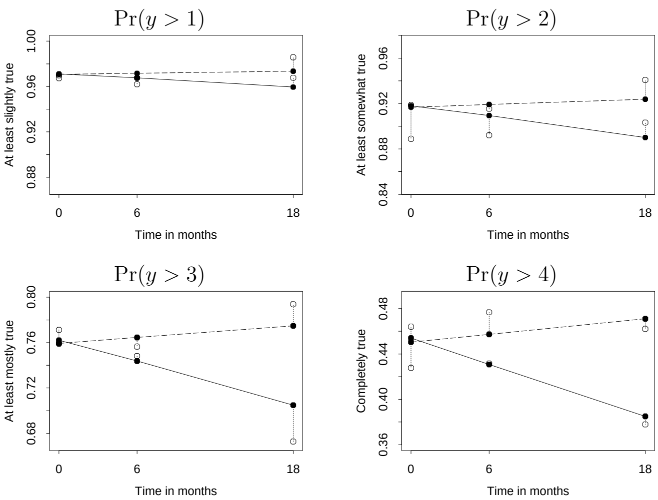
---- intervention group, —— control group



Slide 15

Observed and predicted marginal proportions

---- intervention group, —— control group, • predicted, ○ observed



Slide 16

II. Unordered categorical responses

- Only a small number of responses or 'categories' are possible, a, $a = 1, \cdots, A$.
- The categories cannot be ordered a priori.
- Examples:
  - Treatment decision: 'chemotherapy', 'surgery', 'none'
  - Health insurance choice: 'NHS', 'PPP', etc.
  - Method of birth control: 'pill', 'condom', etc.
  - Diagnosis: 'meningitis', 'influenza', 'common cold'
- Unordered categorical responses often correspond to a 'first choice' among a set of alternatives.

Slide 17

Random Utility Models

- Unobserved 'utility' U_i^a associated with each alternative $a = 1, ..., A$ for unit $i = 1, ..., N$
- Random utility models composed as

$$U_i^a \;=\; V_i^a + \epsilon_i^a$$

  - V_i^a is the linear predictor
  - ϵ_i^a is a residual term (independent over i and a)
- Alternative f is chosen if

$$U_i^f > U_i^g \text{ for all } g \neq f$$

ϵ_i^a independent Gumbel distributed

$$\Updownarrow$$

$$\Pr(f_i) \;=\; \frac{\exp(V_i^f)}{\Sigma_{a=1}^A \exp(V_i^a)}$$

[Conventional multinomial logit]

Slide 18

Covariate effects on the utilities

Linear predictor for unit i and alternative a:

$$V_i^a \;=\; m^a + \mathbf{g}^{a\prime}\mathbf{x}_i + \mathbf{b}'\mathbf{x}_i^a$$

- Covariates and parameters:
  - m^a alternative specific constants
  - $\mathbf{x}_i$ varies over subjects (but not alternatives) and has fixed effects $\mathbf{g}^a$ varying over alternatives
    Examples: Age of subject
  - $\mathbf{x}_i^a$ varies over alternatives (and possibly subjects) and has fixed effects $\mathbf{b}$ not varying over alternatives
    Example: Cost of treatment (could differ between countries)
- Identification:
  - Probability of choosing alternative 1 among alternatives 1, 2 and 3 can be expressed in terms of utility differences $\Pr(U^1 - U^2 > 0 \text{ and } U^1 - U^3 > 0)$
  - Therefore the location of V_i^a is arbitrary:

$$\frac{\exp(V_i^1)}{\Sigma_a \exp(V_i^a)} = \frac{\exp(V_i^1 + c_i)}{\Sigma_a \exp(V_i^a + c_i)}$$

  - Solution: Last alternative S serves as reference alternative, set $m^S = 0$ and $\mathbf{g}^S = \mathbf{0}$.

Slide 19

Multilevel models

- Consider three-level data with patients i (level 1) treated by doctors j (level 2) working in hospitals k (level 3).
- We can include random effects in the linear predictor:

$$\begin{aligned} V_{ijk}^a &= m^a + \gamma_{0jk}^{a(2)} + \gamma_{0j}^{a(3)} \\ &\quad + [\mathbf{g}^{a\prime} + \boldsymbol{\gamma}_{jk}^{a(2)\prime} + \boldsymbol{\gamma}_{k}^{a(3)\prime}]\mathbf{x}_{ijk} \\ &\quad + [\mathbf{b}' + \boldsymbol{\beta}_{jk}^{(2)\prime} + \boldsymbol{\beta}_{k}^{(3)\prime}]\mathbf{x}_{ijk}^a \end{aligned}$$

- Random intercepts: $\gamma_{0jk}^{a(2)}$ and $\gamma_{0k}^{a(3)}$
- Random Coefficients I: $\boldsymbol{\gamma}_{jk}^{a(2)}$ and $\boldsymbol{\gamma}_{k}^{a(3)}$ are alternative specific random coefficients for *subject*-specific covariates $\mathbf{x}_{ijk}$.
- Random Coefficients II: $\boldsymbol{\beta}_{jk}^{(2)}$ and $\boldsymbol{\beta}_{k}^{(3)}$ are random coefficients for *alternative*-specific covariates $\mathbf{x}_{ijk}^a$.

Slide 20

Use and abuse of antibiotics for API*

- Acute respiritory tract infection (API) can lead to pneumonia and death if not properly treated, but inappropriate frequent use of antibiotics can lead to drug resistance.
- In the 1990's the WHO introduced a program of case management for children under 5 in China.
- Doctor's antibiotic prescription was rated as 'abuse' if there were no clinical indicators:
  1. Abuse of serveral antibiotics
  2. Abuse of one antibiotic
  3. Correct use of antibiotics (reference-category)
- Data are multilevel with 2565 children i (level 1) treated by 134 doctors j (level 2) in 36 hospitals k (level 3).
- Covariates ($\mathbf{x}_{ijk}$) include
  – [Age] Age in years (0-5)
  – [Temp] Body temperature, centered at 36^{o}C
  – [Paymed] Pay for medication (yes=1, no=0)
  – [Selfmed] Self medication (yes=1, no=0)
  – [Wrdiag] Wrong diagnosis (yes=1, no=0)
  – [WHO] Hospital in WHO program (yes=1, no=0)
  – [DRed] Doctor's education (self-taught to med. school)

* Thanks to Min Yang for providing the data

Slide 21

Models

- Children i treated by doctors j working in hospitals k
- There are no alternative-specific covariates

$$V^a_{ijk} = m^a + \gamma^{a(2)}_{jk} + \gamma^{a(3)}_{j} + \mathbf{g}^{a\prime}\mathbf{x}_{ijk}$$

- Data:

```
doc child alt choice
1   1     1   0
1   1     2   1
1   1     3   0
1   2     1   0
1   2     2   0
1   2     3   1
```

- Estimation in `gllamm`

```
gen categ1 = alt == 1
gen categ2 = alt == 2
eq c1: categ1
eq c2: categ2
gllamm alt age temp ... , i(doc hosp)    /*
  */ nrf(2 2) eqs(c1 c2 c1 c2)           /*
  */ link(mlogit) family(binom)          /*
  */ expanded(child choice m) basecat(3)
```

Slide 22

Estimates

	Abuse several		Abuse one	
Parameter	Est	(SE)	Est	(SE)
g^a_0 [Cons]	-1.64	(0.49)	0.23	(0.32)
g^a_1 [Age]	0.09	(0.09)	0.19	(0.08)
g^a_2 [Temp]	-0.27	(0.13)	-1.01	(0.12)
g^a_3 [Paymed]	0.91	(0.41)	0.30	(0.31)
g^a_4 [Selfmed]	-0.78	(0.29)	-0.42	(0.24)
g^a_5 [Wrdiag]	1.80	(0.26)	2.08	(0.23)
g^a_6 [WHO]	–		–	
g^a_7 [DRed]	–		–	
Doctor-level variances				
$\mathrm{var}(\gamma^{a(2)}_0)$	0.43	(0.27)	0.51	(0.26)
$\mathrm{Cov}(\gamma^{1(2)}_0, \gamma^{2(2)}_0)$		-0.47 (0.15)		
Hospital-level variances				
$\mathrm{var}(\gamma^{a(3)}_0)$	2.50	(0.93)	0.23	(0.18)
$\mathrm{Cov}(\gamma^{1(3)}_0, \gamma^{2(3)}_0)$		0.68 (0.31)		
Log-likelihood		-730.6		

Slide 23

More Estimates

	Abuse several		Abuse one	
Parameter	Est	(SE)	Est	(SE)
g_0^a [Cons]	-5.72	(0.99)	-0.23	(0.55)
g_1^a [Age]	0.07	(0.09)	0.17	(0.08)
g_2^a [Temp]	-0.27	(0.13)	-0.96	(0.12)
g_3^a [Paymed]	0.92	(0.40)	0.12	(0.32)
g_4^a [Selfmed]	-0.86	(0.29)	-0.49	(0.24)
g_5^a [Wrdiag]	1.85	(0.26)	2.08	(0.23)
g_6^a [WHO]	-2.40	(0.62)	-0.88	(0.33)
g_7^a [DRed]	-0.62	(0.17)	0.08	(0.11)
Doctor-level variances				
$\mathrm{var}(\gamma_0^{a(2)})$	0.46	(0.28)	0.43	(0.22)
$\mathrm{Cov}(\gamma_0^{1(2)}, \gamma_0^{2(2)})$		-0.44 (0.13)		
Hospital-level variances				
$\mathrm{var}(\gamma_0^{a(3)})$	0.88	(0.45)	0.11	(0.12)
$\mathrm{Cov}(\gamma_0^{1(3)}, \gamma_0^{2(3)})$		0.31 (0.20)		
Log-likelihood		-716.2		

Slide 24

References

- "Multilevel logistic regression for polytomous data and rankings". *Psychometrika*, in press. (A.Skrondal & S.Rabe-Hesketh)
- "Generalised multilevel structural equation modeling". *Psychometrika*, in press. (S.Rabe-Hesketh, A.Skrondal & A.Pickles)
- *Generalized latent variable modeling: Multilevel, longitudinal and structural equation models*. Boca Raton, FL: Chapman & Hall/ CRC, to appear. (A.Skrondal & S.Rabe-Hesketh).
- *GLLAMM Manual*. London: Institute of Psychiatry, 2001. (S.Rabe-Hesketh, A.Pickles & A.Skrondal). www.iop.kcl.ac.uk/IoP/Departments/BioComp/programs/gllamm.html
- "Reliable estimation of generalized linear mixed models using adaptive quadrature". *The Stata Journal*, 2: 1-21, 2002. (S.Rabe-Hesketh, A.Skrondal & A.Pickles).