Constructing Latent Variable Models using Composite Links

Anders Skrondal

Norwegian Institute of Public Health

London School of Hygiene and Tropical Medicine

Based on joint work with

Sophia Rabe-Hesketh

University of California, Berkeley

Institute of Education, University of London

University of Twente, October 15, 2009

Outline

- Traditional and generalized latent variable models
- Composite links
- Some standard extensions of archetypical IRT models specified using composite links
  - Three-parameter model for guessing
  - Graded response model for ordinal items
- Some non-standard IRT models specified using composite links
  - Unfolding models for attitude and preference items
  - Randomized response models
- Conclusion

Traditional latent variable models

		LATENT VARIABLE(S)	
		Continuous	Categorical
OBSERVED VARIABLE(S)	Continuous	*Common factor model* *Structural equation model* *Linear mixed model* *Covariate measurement error model*	*Latent profile model*
	Categorical	*Latent trait model*	*Latent class model*

[S & RH (2007)]

Generalized latent variable models

- Response model: **Generalized linear model** conditional on latent variables
  - Linear predictor: latent variables as factors or random coefficients varying at different hierarchical levels
  - Links and distributions
- Structural model:
  - Regressions of latent variables on observed variables
  - Regressions of latent variables on other latent variables (at same and/or higher hierarchical levels)
- Distribution of latent variables (disturbances)
  - Continuous
    - Parametric
    - Nonparametric
  - Discrete
  - Continuous and discrete

[S & RH (2007)]

Why generalized latent variable models

- Unifying model framework:
  - Conceptually appealing
  - Encourages specification of models tailormade to research problems by making it easy to combine features of different model types
  - Facilitates unified approach to estimation implemented in single software program

Generalized linear model

1. **Linear predictor** $\nu_i = \mathbf{x}_i'\beta$ for unit i
2. Conditional expectation of response y_i, $\mu_i \equiv \mathrm{E}(y_i|\mathbf{x}_i)$, linked to linear predictor ν_i through **link function** $g(\cdot)$,

$$g(\mu_i) \;=\; \nu_i,$$

   or equivalently,

$$\mu_i \;=\; g^{-1}(\nu_i),$$

   where $g^{-1}(\cdot)$ is **inverse link function**
3. **Conditional distribution** of response, given linear predictor, from exponential family

$$f(y_i|\nu_i) \;=\; \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{\phi_i} + c(y_i, \phi_i)\right\}$$

   - θ_i is canonical or natural parameter (a function of μ_i)
   - ϕ_i is scale or dispersion parameter
   - $b(\cdot)$ and $c(\cdot)$ are functions depending on distribution

Common link functions

- Continuous responses: **Identity link**
  Link and Inverse link: $\mu_i = \nu_i$
- Dichotomous responses: **Logit link**
  Link: $\log[\mu_i/(1-\mu_i)] = \nu_i,$ Inverse link: $\mu_i = \frac{\exp(\nu_i)}{1+\exp(\nu_i)}$
- Dichotomous responses: **Probit link**
  Link: $\Phi^{-1}(\mu_i) = \nu_i,$ Inverse link: $\mu_i = \Phi(\nu_i)$
- Counts: **Log link**
  Link: $\log(\mu_i) = \nu_i,$ Inverse link: $\mu_i = \exp(\nu_i)$

Composite links [Thompson & Baker, *JRSSC*, 1981]

- μ_i linear combination of **several** inverse link functions
- **Simple composite links:**
  Expectation μ_i weighted sum of inverse links with known weights w_{qi}

$$\mu_i = \sum_q w_{qi}\, g_q^{-1}(\nu_{qi})$$

  - ν_{qi} is qth linear predictor for unit i
  - $g_q^{-1}(\cdot)$ an inverse link function
- **Bilinear composite links:**
  Known constants w_{qi} replaced by linear combinations of unknown parameters and constants $\boldsymbol{\alpha}'\mathbf{w}_{qi}$

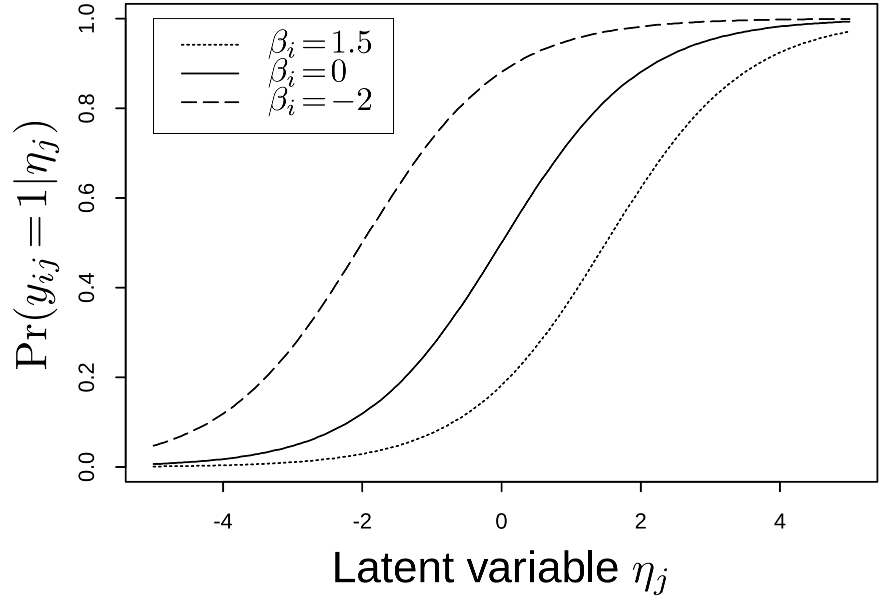
$$\mu_i = \sum_q \boldsymbol{\alpha}'\mathbf{w}_{qi}\, g_q^{-1}(\nu_{qi})$$

- Further extensions discussed in [T & B (1981)]

Archetypical item response theory (IRT) model

- **One parameter logistic (1-PL) model** or **Rasch model** for ‘ability’ testing of persons j using items i

$$\mu_{ij} = \Pr(y_{ij} = 1 \mid \eta_j) \; = \; \frac{\exp(\nu_{ij})}{1 + \exp(\nu_{ij})}, \qquad \nu_{ij} = \beta_i + \eta_j$$



- $-\beta_i$ is ‘difficulty’ of item i
- η_j is the ability of person j (latent variable)

Some standard extensions of archetypical IRT model using composite links

- Some limitations of Rasch model:
  - Probability of positive response goes to 0 if ability goes to $-\infty$
    Not well suited for multiple choice items where examinees can guess right answer
    $\Longrightarrow$ (I.) IRT models with guessing parameters
  - Only for dichotomous responses
    Unsuitable for items 'graded' in ordered categories
    $\Longrightarrow$ (II.) IRT models for ordinal items

I. IRT with guessing (3-PL model)

- If right answer to 'item' can be guessed as with multiple choice questions, use three-parameter logistic (3-PL) model:

$$\Pr(y_{ij}=1|\eta_j) \;=\; c_i + (1-c_i)\frac{\exp(\nu_{ij})}{1+\exp(\nu_{ij})},$$

  where c_i are 'guessing parameters'
  (cf 'natural responsiveness' or 'nonzero background' in quantal response bioassay)

- Fixing $w_{1ij} = c_i$ and $w_{2ij} = 1 - c_i$, get model with composite link:

$$\Pr(y_{ij}=1|\eta_j) \;=\; \overbrace{w_{1ij}}^{c_i}\,\underbrace{g_1^{-1}(1)}_{\text{identity}} + \overbrace{w_{2ij}}^{1-c_i}\,\underbrace{g_2^{-1}(\nu_{ij})}_{\text{logit}}$$

IRT with guessing (cont'd)

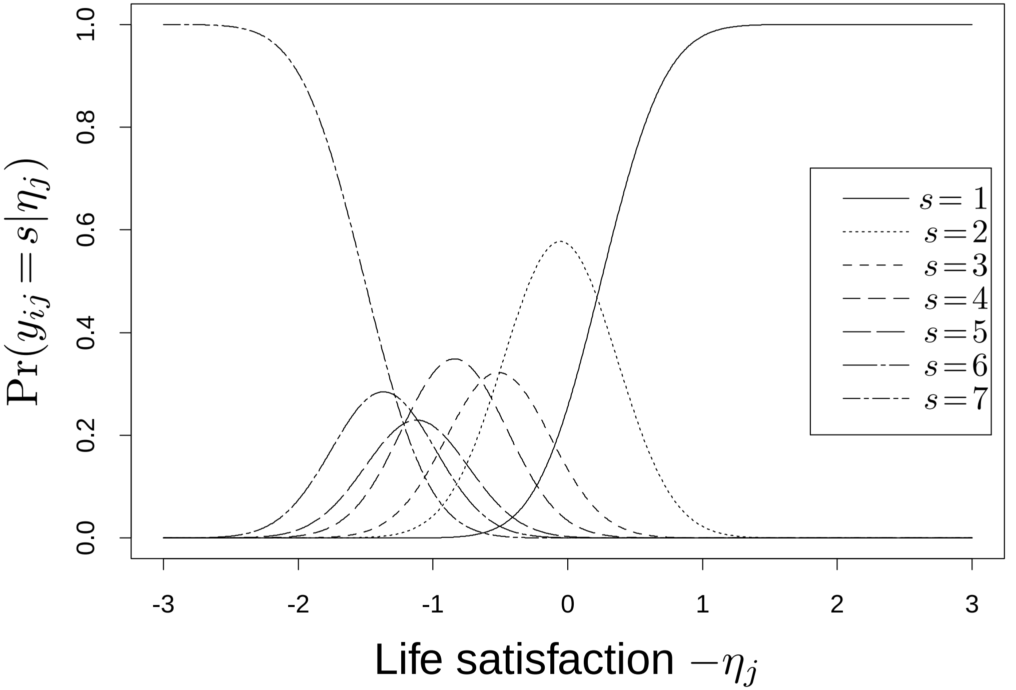
- Estimate guessing parameters in 3-PL using bilinear composite link with
  - $\boldsymbol{\alpha}' = (1, c_1, \ldots, c_I)$
  - $\mathbf{w}'_{1ij} = (0, \mathbf{d}'_i)$ and $\mathbf{w}'_{2ij} = (1, -\mathbf{d}'_i)$, where $\mathbf{d}_i$ is a vector of dimension I (for I items) with a one in position i and zeros elsewhere

II. Samejima's graded response model for ordinal items

- Cumulative models for ordinal items $y_{ij} = s$ with linear predictor

$$\nu_{ij} = \beta_i + \lambda_i \eta_j, \qquad \text{with thresholds } \kappa_{si}$$

- Satisfaction from five areas of life ($i = 1, \ldots, 5$), rated from 1='a very great deal' to 7='none' [S & R-H, 2004: Ch. 10]
  - Conditional response probabilities for [Friend] item:

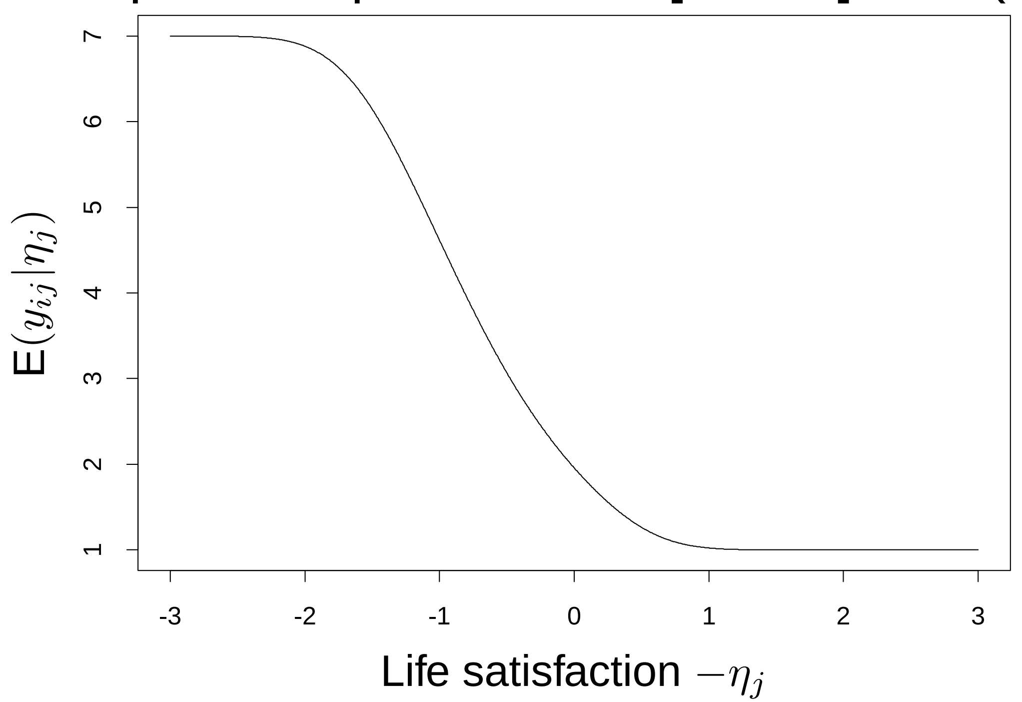


II. Samejima's graded response model for ordinal items

- Cumulative models for ordinal items $y_{ij} = s$ with linear predictor

$$\nu_{ij} = \beta_i + \lambda_i \eta_j, \quad \text{with thresholds } \kappa_{si}$$

- Satisfaction from five areas of life ($i = 1, \ldots, 5$), rated from 1='a very great deal' to 7='none' [S & R-H, 2004: Ch. 10]
  - Conditional response expectation for [Friend] item (monotonic):



Cumulative response models: Composite link

- S ordered response categories $s=1,\ldots,S$

$$\Pr(y_{ij}>s|\eta_j) \;=\; g^{-1}(\nu_{ij}-\kappa_s), \quad s=0,\ldots,S$$

  - κ_s are threshold parameters, $\kappa_0=-\infty$, $\kappa_S=\infty$
  - Inverse link function $g^{-1}(\cdot)$ is cumulative density function such as standard normal, logistic or extreme value
- Composite link for 'difference model':

$$\begin{aligned}\Pr(y_{ij}=s|\eta_j) &= \overbrace{w_{1ij}}^{1}\, g^{-1}(\nu_{ij}-\kappa_{s-1}) + \overbrace{w_{2ij}}^{-1}\, g^{-1}(\nu_{ij}-\kappa_s)\\ &= g^{-1}(\nu_{ij}-\kappa_{s-1}) - g^{-1}(\nu_{ij}-\kappa_s)\end{aligned}$$

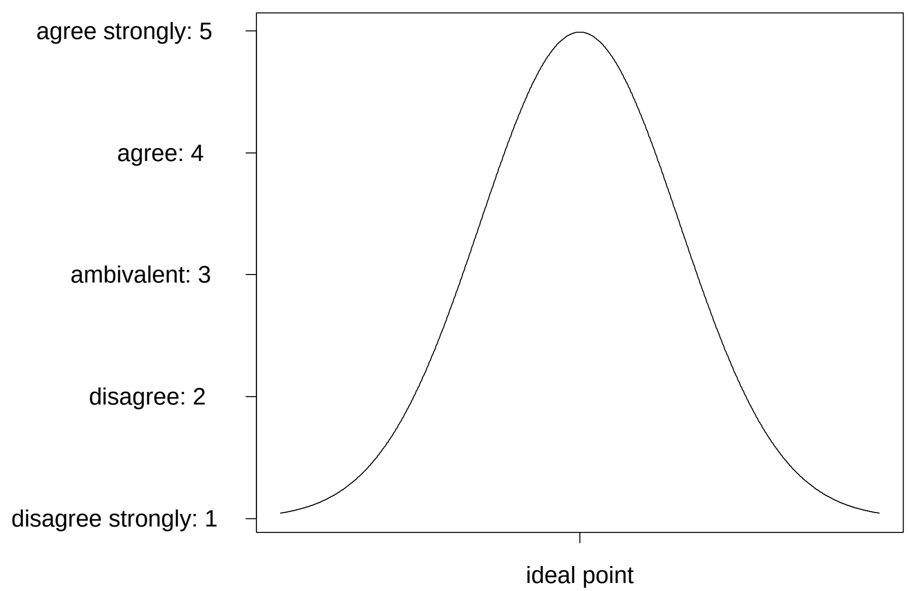
- Advantage that left and right-censoring, or even interval censoring of ordinal response easily accommodated
  - Particularly useful for discrete time survival data
  - Also handles accidental collapsing of categories

Some non-standard IRT models specified using composite links

- ‘Expected value function’ a monotonic function of latent variable in most IRT models
  May be unrealistic for attitude or preference items
  $\Longrightarrow$ III. Unfolding or ideal point models
- Responses to sensitive items may not be truthful
  $\Longrightarrow$ IV. Randomized response models

III. *Unfolding or ideal point models*

- Preference items
- Stages of development
- Agreement with attitude statements
  - e.g. 'Capital punishment seems wrong but is sometimes necessary', expected response given sentiment favoring capital punishment η_j should be unimodal:



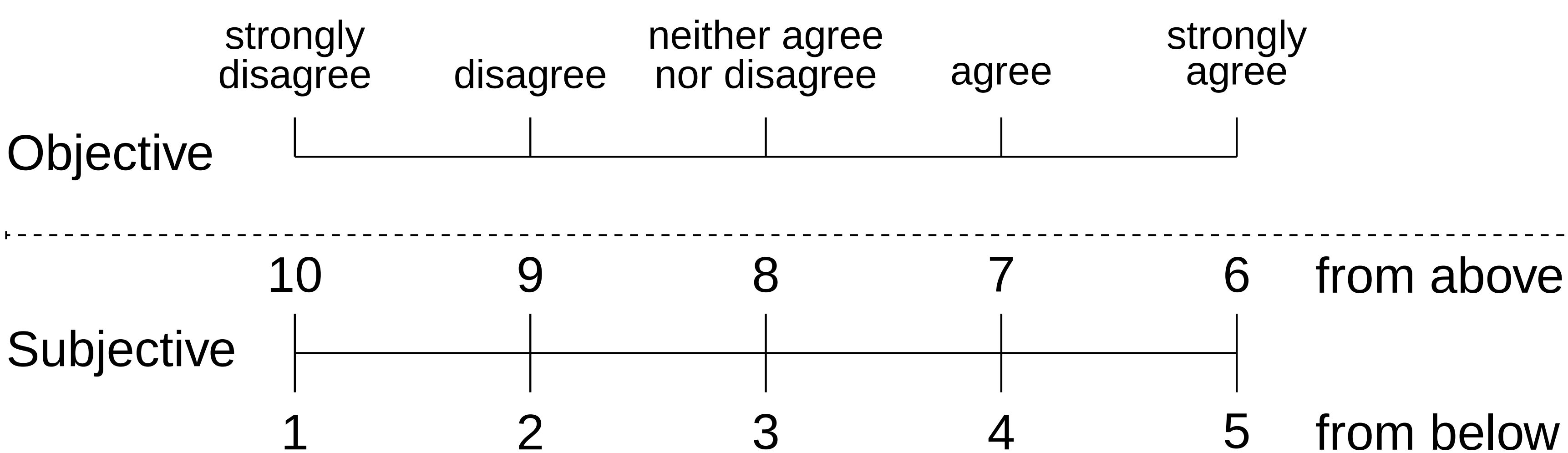
Sentiment favoring capital punishment

Unfolding or ideal point models (cont'd)

- Agreement with attitude statement, rated as 'objective response'

  $y_{ij} = s, \qquad s = 1, \ldots, 5, \quad S = 5$

- Respondent may disagree from above or below, with corresponding unobserved 'subjective responses' z_{ij}
  - From below: latent variable is below position of item, $z_{ij} = s$
  - From above: latent var. is above position of item, $z_{ij} = 2S+1-s$



Unfolding or ideal point models: Composite link

- Since z_{ij} not observed, probability of observed rating y_{ij} equals sum of probabilities of the two possible 'subjective ratings'

$$\Pr(y_{ij}=s|\eta_j) \;=\; \Pr(z_{ij}=s|\eta_j)+\Pr(z_{ij}=2S+1-s|\eta_j)$$

- Cumulative model for subjective ratings

$$\underbrace{g^{-1}(\nu_{ij}-\kappa_{s-1})-g^{-1}(\nu_{ij}-\kappa_s)}_{\Pr(z_{ij}=s|\eta_j)}+\underbrace{g^{-1}(\nu_{ij}-\kappa_{2S-s})-g^{-1}(\nu_{ij}-\kappa_{2S+1-s})}_{\Pr(z_{ij}=2S+1-s|\eta_j)},$$

  where $\nu_{ij}=\beta_i+\lambda_i\eta_j$

- Composite link with 4 components

IV. *Randomized response models*

- Want to estimate probability π of positive response to sensitive question (such as illegal drug use)
- Warner (1965) design:
    - Question selected at random from two possibilities
        1. Positively worded ('Have you *ever* used illegal drugs') with known probability p
        2. Negatively worded ('Have you *never* used illegal drugs') with known probability $1-p$
    - Actual selection unknown to interviewer to secure privacy
- Probability of positive response to item whether positively or negatively worded

$$\Pr(y=1) \ = \ p\pi + (1-p)(1-\pi)$$

- $\widehat{\pi} \ = \ \frac{\widehat{\Pr}(y=1)+p-1}{2p-1}$

Randomized response models (cont'd)

- Combine Warner model with model for π_i with linear predictor ν_i and link function $g(\cdot)$, for instance logit link,

$$\Pr(y_i = 1|\nu_i) \;=\; p\frac{\exp(\nu_i)}{1+\exp(\nu_i)} + (1-p)\left(1 - \frac{\exp(\nu_i)}{1+\exp(\nu_i)}\right)$$

- Specify using composite link

$$\Pr(y_i = 1|\nu_i) \;=\; w_{1i}g_1^{-1}(\nu_{1i}) + w_{2i}g_2^{-1}(\nu_{2i})$$

  - $g_1(\cdot)$ and $g_2(\cdot)$ logit links
  - $w_{1i} = p$ and $w_{2i} = 1-p$
  - $\nu_{1i} = \nu_i = -\nu_{2i}$

- Straightforward to combine with IRT models using linear predictors

$$\nu_{ij} = \beta_i + \lambda_i\eta_j$$

Some other uses of composite links in IRT

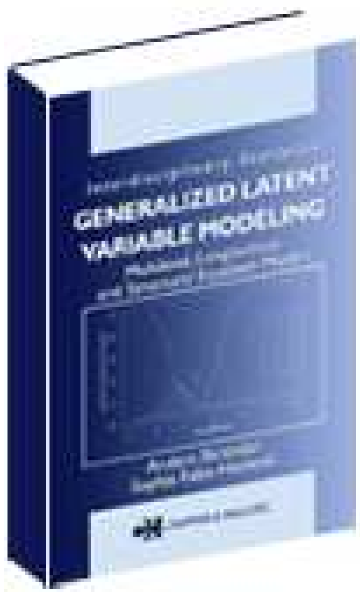
- Missing categorical data
  - Validation data
  - Phase II designs
  - ‘Fused’ cells
- Latent class models
  - Standard single level models
  - Multilevel models
- Models with combinations of discrete and continuous latent variables
- Log-normal latent variables

[RH & S (2007)]

Conclusion

- Using generalized latent variable framework with composite links produces a wide range of novel IRT models, for instance including:
  - Several latent variables
  - Discrete latent variables and NPML
  - Latent variable(s) regressed on same or higher level latent variable(s)
  - Latent variable(s) regressed on covariates
- Easy to implement in software for latent variable modelling
  - Already implemented in gllamm
- Many extensions also for latent variable models used in other areas of statistics

References



- Rabe-Hesketh & Skrondal (2007). Multilevel and latent variable modelling with composite links and exploded likelihoods. *Psychometrika*, 72, 123-140.
- Skrondal & Rabe-Hesketh (2007). Latent variable modelling: A survey. *Scandinavian Journal of Statistics*, 34, 712-745.
- Skrondal & Rabe-Hesketh (2004). *Generalized latent variable modeling: Multilevel, longitudinal and structural equation models*. Chapman & Hall/ CRC
- Rabe-Hesketh, Skrondal & Pickles (2004). Generalized multilevel structural equation modelling. *Psychometrika*, 69, 183-206.
- Skrondal & Rabe-Hesketh (2003). Mulitlevel logistic regression for polytomous data and rankings. *Psychometrika*, 68, 267-287.
- Thompson & Baker (1981). Composite link functions in generalized linear models. *Journal of the Royal Statistical Society (Series C)*, 30, 125-131.

The IRT epidemic

- Is transportation of IRT to medicine necessarily harmless?
- Example: MRC clinical trial of colorectal cancer
- Some example quality of life (QoL) items:
  - Lack of appetite
  - Decreased sexual interest
  - Dry mouth