Multivariate response models for global quality of life measures

Annette Kifley

Principal supervisor:	Gillian Heller
Associate supervisors:	Jun Ma
	David Bulger
Adjunct supervisor:	Val Gebski
Contributors/collaborators:	Ken Beath
	NHMRC Clinical Trial Centre

Clinical trial data on health-related QOL

Two studies

1: UBQVG:

Cross-sectional study, 200 cancer patients, to validate UBQ-C questionnaire

2: ANZ0001 study - baseline data only in this talk:

Longitudinal clinical trial

325 patients with advanced breast cancer

Compared oral therapy (intermittent or continuous capecitabine) vs standard intravenous regimen (CMF)

QOL assessments

Global measures from UBQ-C, LASA scales:

- perceived overall health state
- perceived overall QOL
  Both continuous - mark a line from best to worst possible

Subdimensional measures from UBQ-C ± other scales:

- recent health impact on specific physical, social, self-care capabilities
  Ordinal categories: None, Slight, Severe, Can't do
- recent distress levels due to specific symptoms, potential side effects, thoughts, emotions
  Ordinal levels: 0,1,...,10 (none to extreme)

Health-related disability items

Physical:	Walk several blocks
	Climb a flight of stairs
	Undertake vigorous activities
Social:	Usual daily activities
	Social life
	Leisure activities
Self-care:	Wash
	Dress
	Eat/drink
	Go to toilet

Distress items

Shortness of breath	Sadness
Sleeping difficulty	Anxiety
Nausea/vomiting	Unhappy with appearance/weight
Lack of energy	Uncertainty about future
Aches/pains	Anger/resentment
Loss of appetite	Loneliness
Hair loss	Loss of self confidence
Diarrhoea	Feeling dependent
Constipation	Thoughts of chemoRx
Numbness	Unable to concentrate

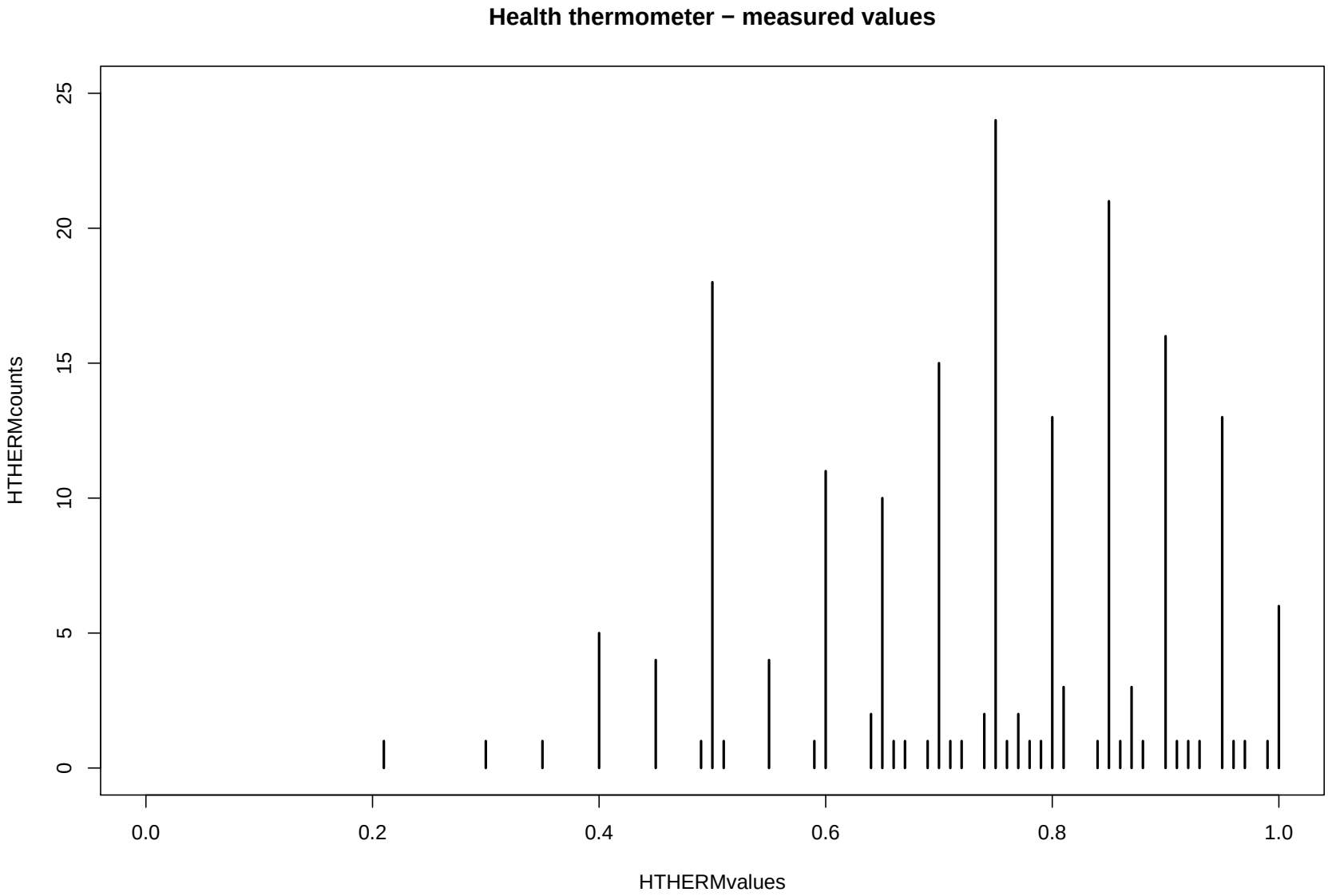
Research question

We want to summarise the available information on global health-related QOL in a way that will facilitate comparisons between treatments in clinical studies

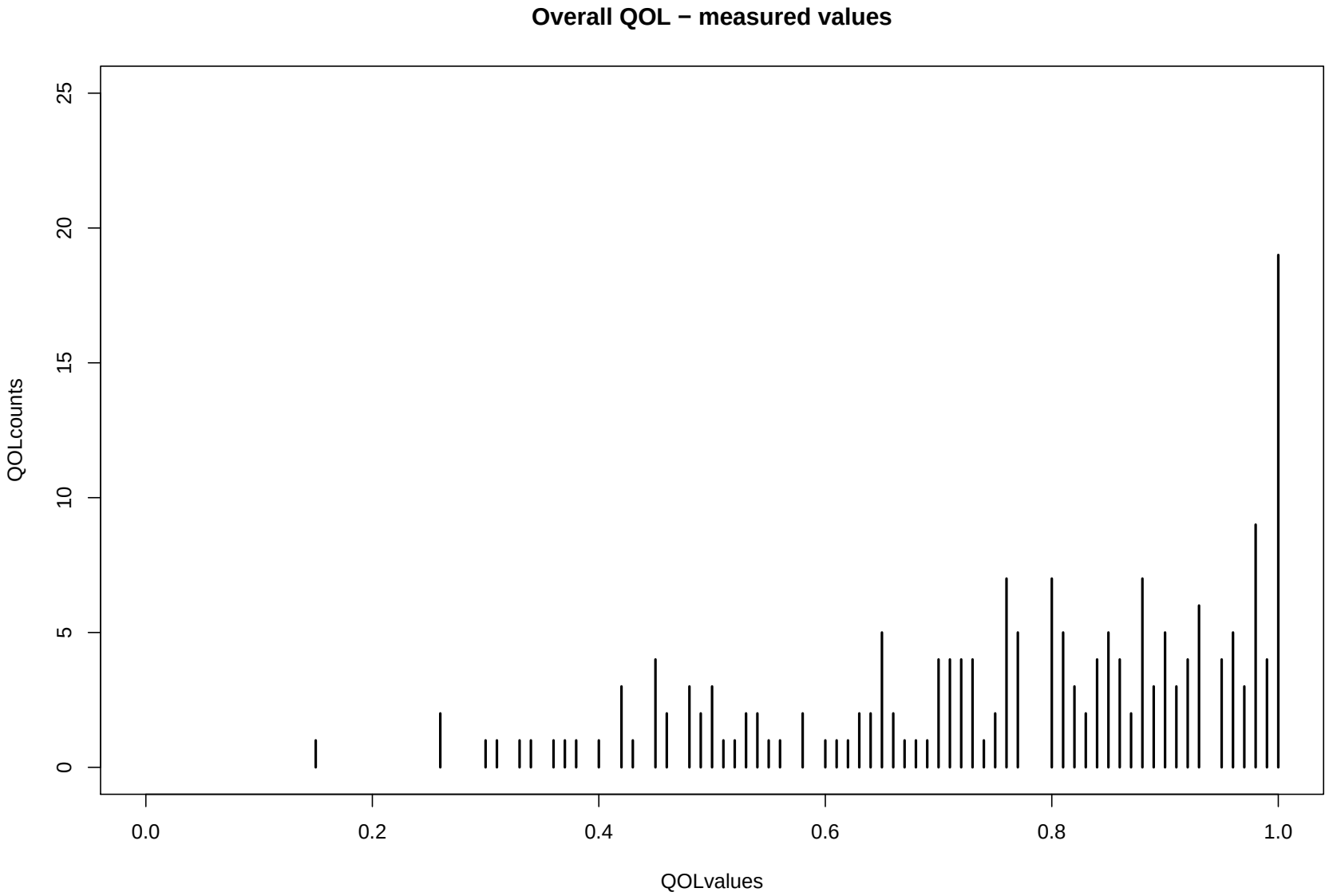
Common methods:

- Select one of the global item scores
- Use the mean of the two global items
- Calculate individual scale scores from subitem measures
- Develop and model a theoretical framework of subdomains of QOL and relate them to overall QOL

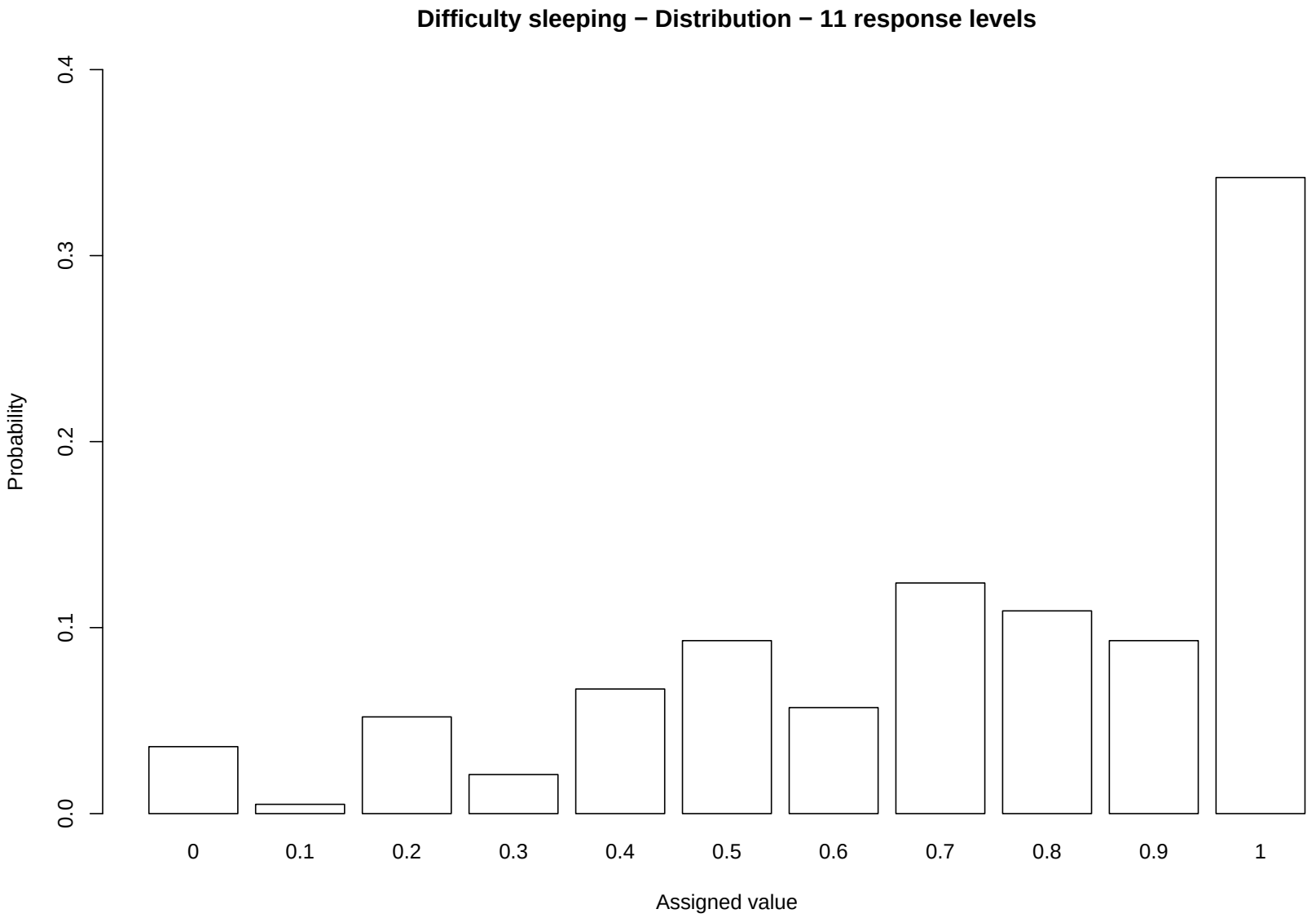
Distribution of global measures



Distribution of global measures



Subitem distns - distress scales



Correlations between items

Items	**UBQVG**	**ANZ0001**
The two global items	ρ =0.62	ρ =0.65
Disabilities	Mostly 0.4-0.6	0.4-0.8
Disabilities and globals	Mostly 0.4-0.6	0.4-0.6
Distresses	Diverse, many uncorrelated pairs	
	A few 0.6-0.8 between psych items	
Distresses and globals	0.05-0.6. Moderate for things like	
	energy, anxiety, confidence, dependence (both)	
	breath, sleep, future (UBQVG)	
	sickness, aches, sadness, appetite (ANZ0001)	

How many dimensions in the data?

Principal components analysis of subdimension item measures:

	% **Var**	
UBQVG		
PC1	33%	Good vs poor overall QOL
PC2	12%	Diff. psych vs phys items
PC3+	6% (PC3) etc	Many small compts explain the rest
ANZ0001		
PC1	38%	
PC2	9%	Diff. disability vs phys vs psych
PC3+	5% (PC3) etc	

Missing responses are common

	UBQVG	ANZ0001
No of participants	200	325
No with complete data	120	209
No with no responses at all	4	16
No with sporadic missing items	72	100
No with completed item responses	5944	9085

Imputation option

Commonly missed items often excluded - doesnt help much

Missingness issues worse in the longitudinal data

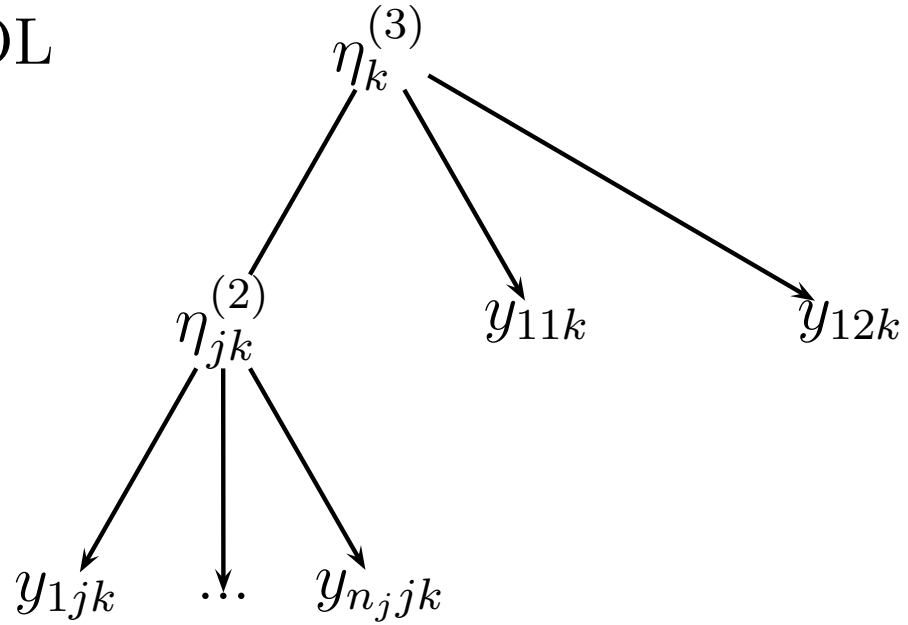
Assessments not compulsory

Multilevel latent variable model for QOL

Level 3: Underlying global QOL

Level 2: Global measures or item groups

Level 1: Specific items (subdimensions)



Derived QOL/Health

Perceived QOL

Perceived Health

Corresponding model formulation

Full factor model:

$$y_{ijk} = \beta_i + \lambda_i^{(2)}\eta_{jk}^{(2)} + \lambda_j^{(3)}\eta_k^{(3)} + \epsilon_{ijk}$$

y item responses, ϵ_{ijk} random error

i items, j global measures or item groups, k subjects

η latent constructs, λ factor loadings

β fixed means/intercepts

Clusters involving direct global measures (j in 1,2) have no loading on $\eta_{jk}^{(2)}$

Initially, MVN errors and latent variables will be assumed
$\eta_{jk}^{(2)} \sim N(0, \psi^{(2)}) \quad \eta_k^{(3)} \sim N(0, \psi^{(3)}) \quad \epsilon_{ijk} \sim N(0, \theta_i)$

Estimation

All models estimated in R using maximum likelihood

Approach as described by Skrondal and Rabe-Hesketh

Marginal likelihood - marginal to all latent variables
Integrals over latent variable distributions approximated using quadrature

NLM function used for optimisation

How many latent constructs?

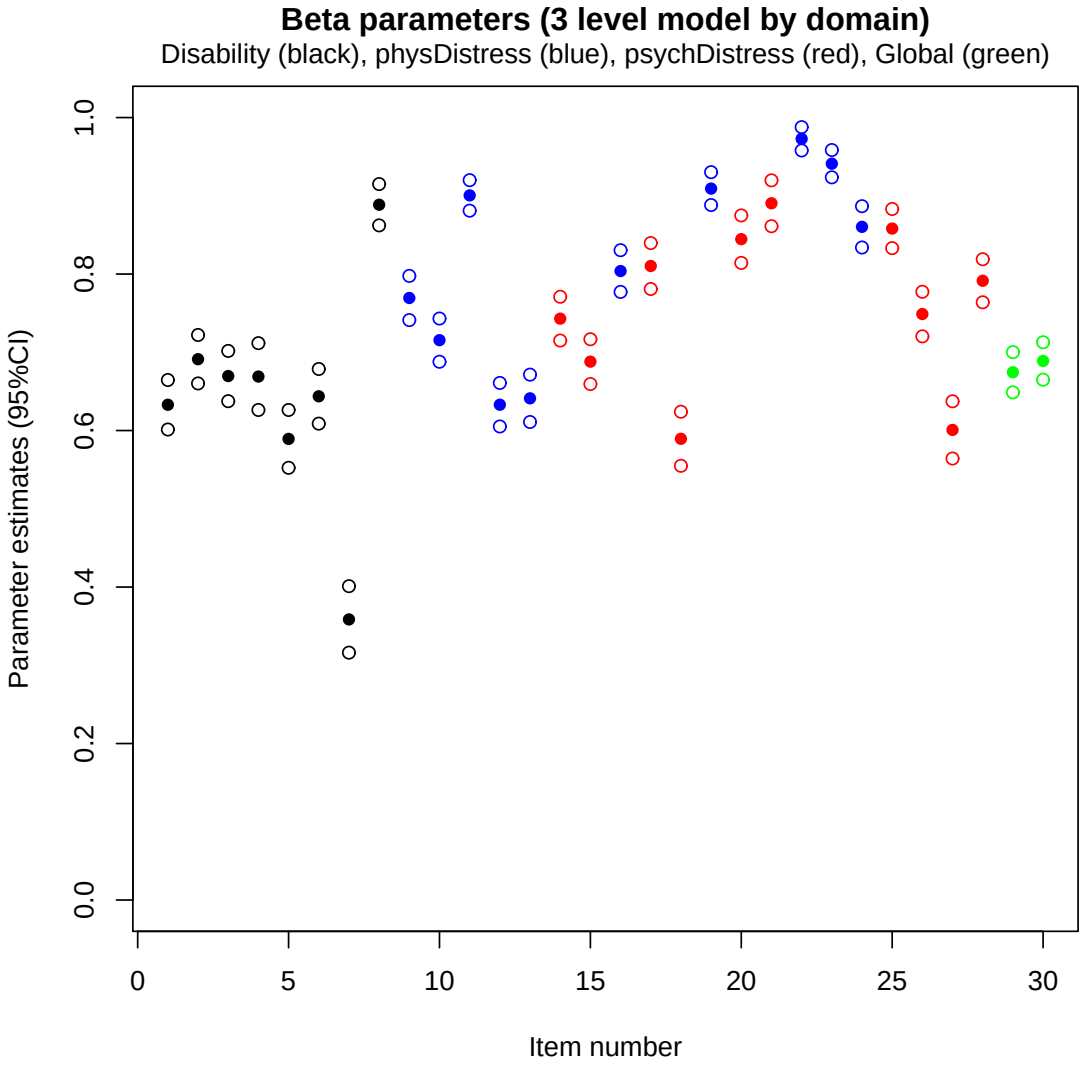
One (2L)	Only $\eta_k^{(3)}$, no intermediate level in the model
	Direct, indirect subdim items both load directly
	Distinguished by factor loadings only
Two (3Lscale)	$\eta_k^{(3)}$ plus a single $\eta_{jk}^{(2)}$
	Single group for all subdim items in a QOL qnnaire
	$\eta_{jk}^{(2)}$ captures within-scale correlation btw subitems
	Sets direct global and indirect subdim items apart
Four (3Ldomain)	$\eta_k^{(3)}$ plus three $\eta_{jk}^{(2)}$
	Items grouped by roughly defined domains:
	Disability, Physical distress, Psychological distress

Model comparisons

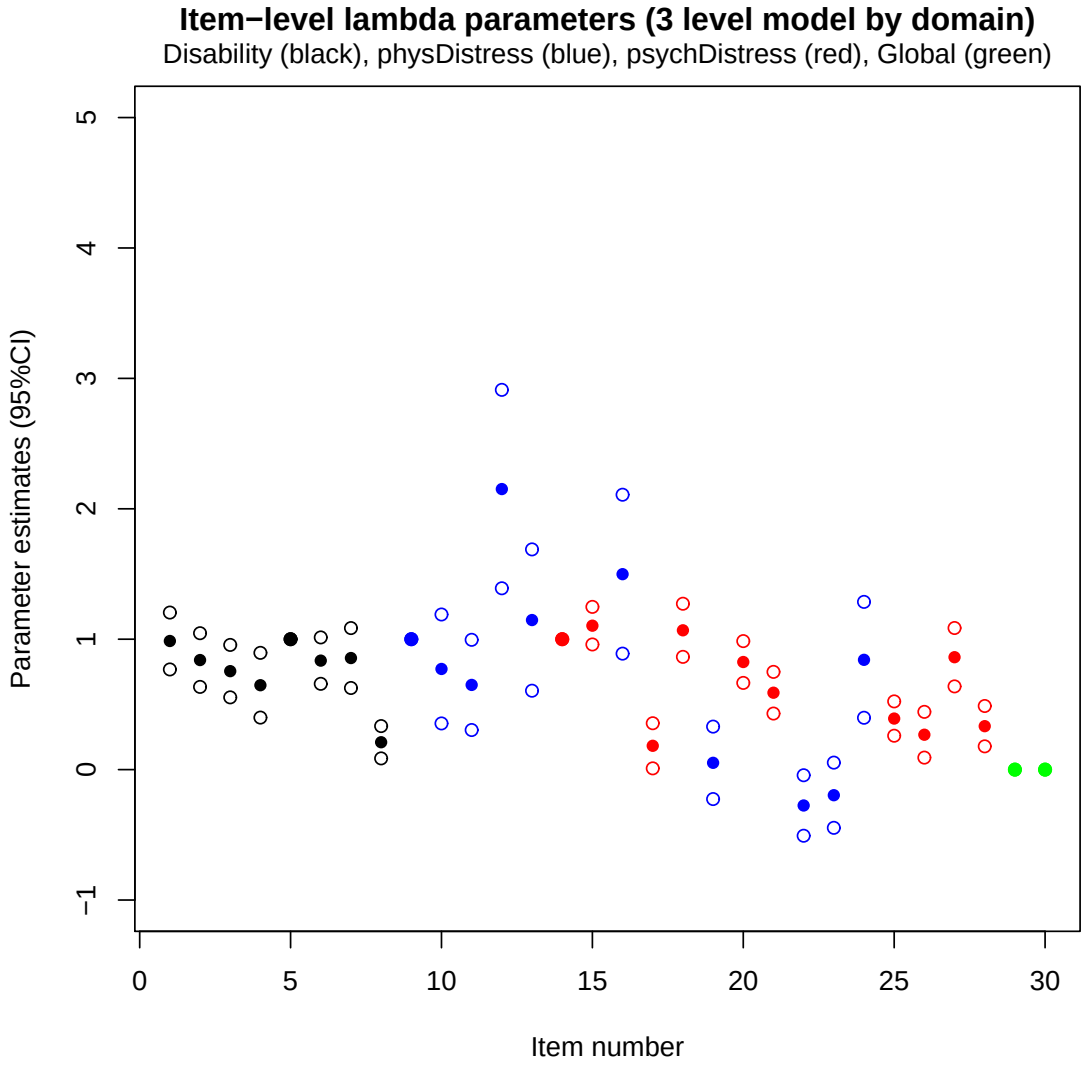
Latent constructs	Npar	LogL	BIC
		UBQVG	
One (Two level model)	92	926	-1366
Two (Three level by scale)	93	1103	-1716
Four (Three level by domain)	95	1153	-1805
		ANZ0001	
One (Two level model)	89	1393	-2276
Two (Three level by scale)	90	1353	-2191
Four (Three level by domain)	92	1632	-2737

A smaller (or more negative) BIC is better

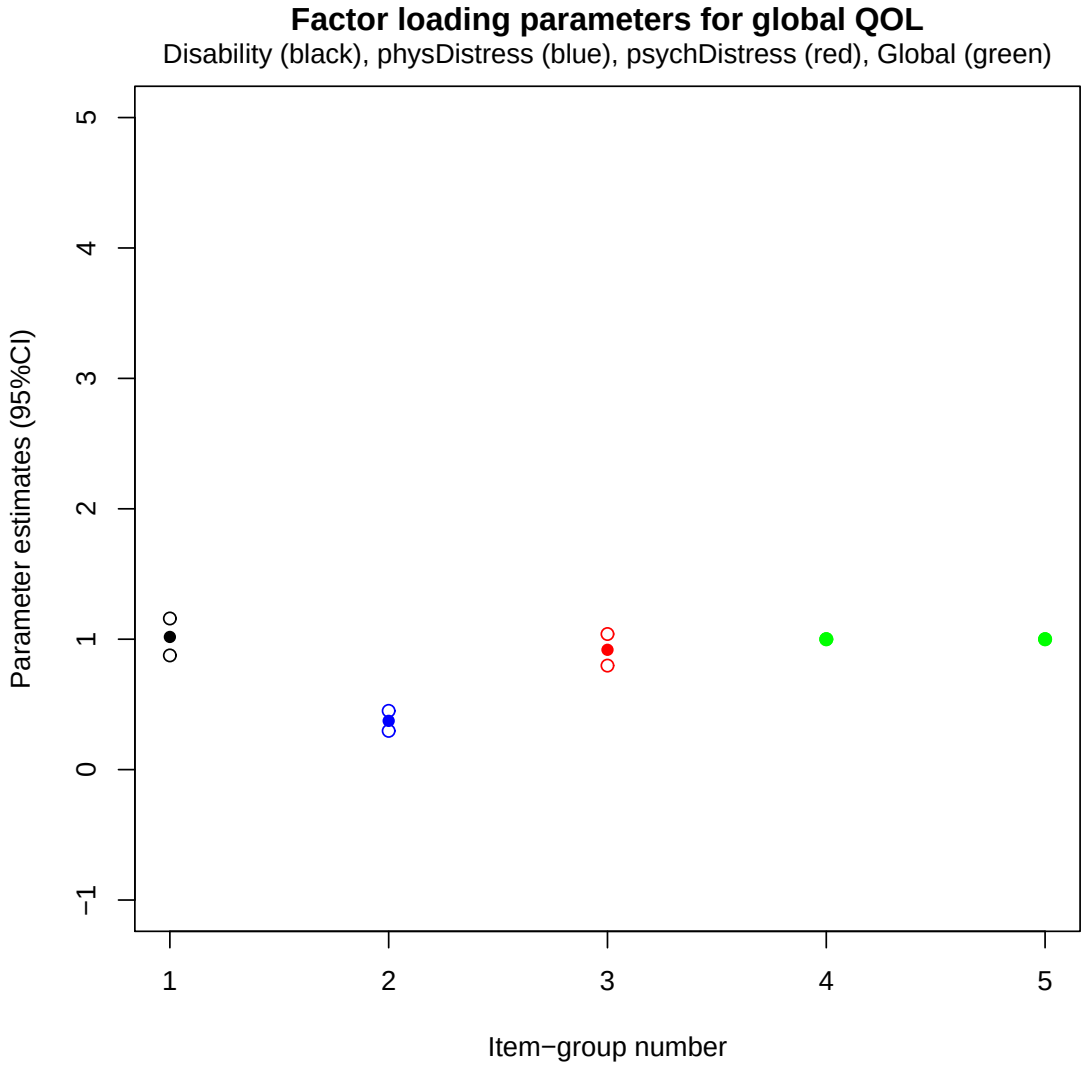
Estimated betas (ANZ)



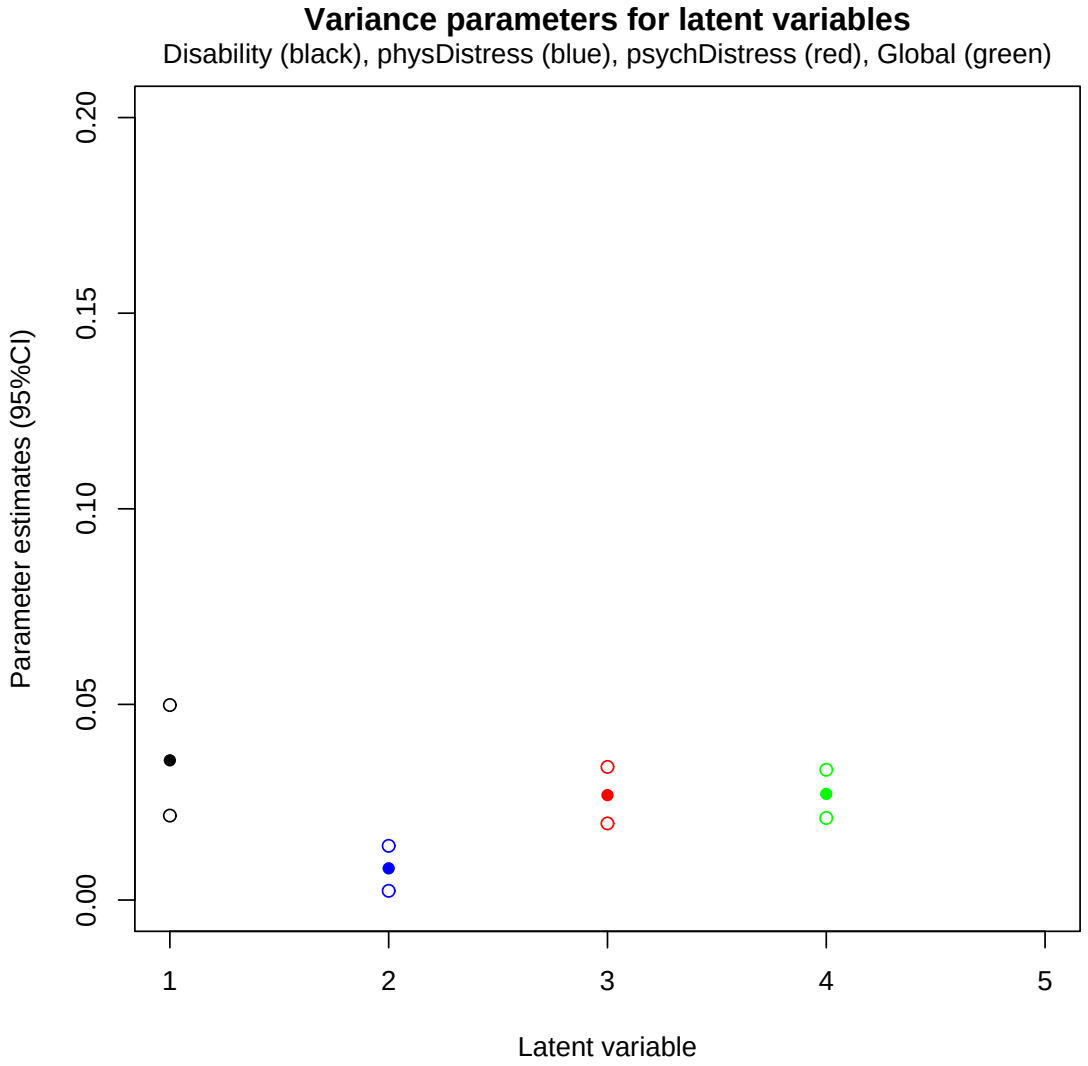
Estimated factor loadings (ANZ)



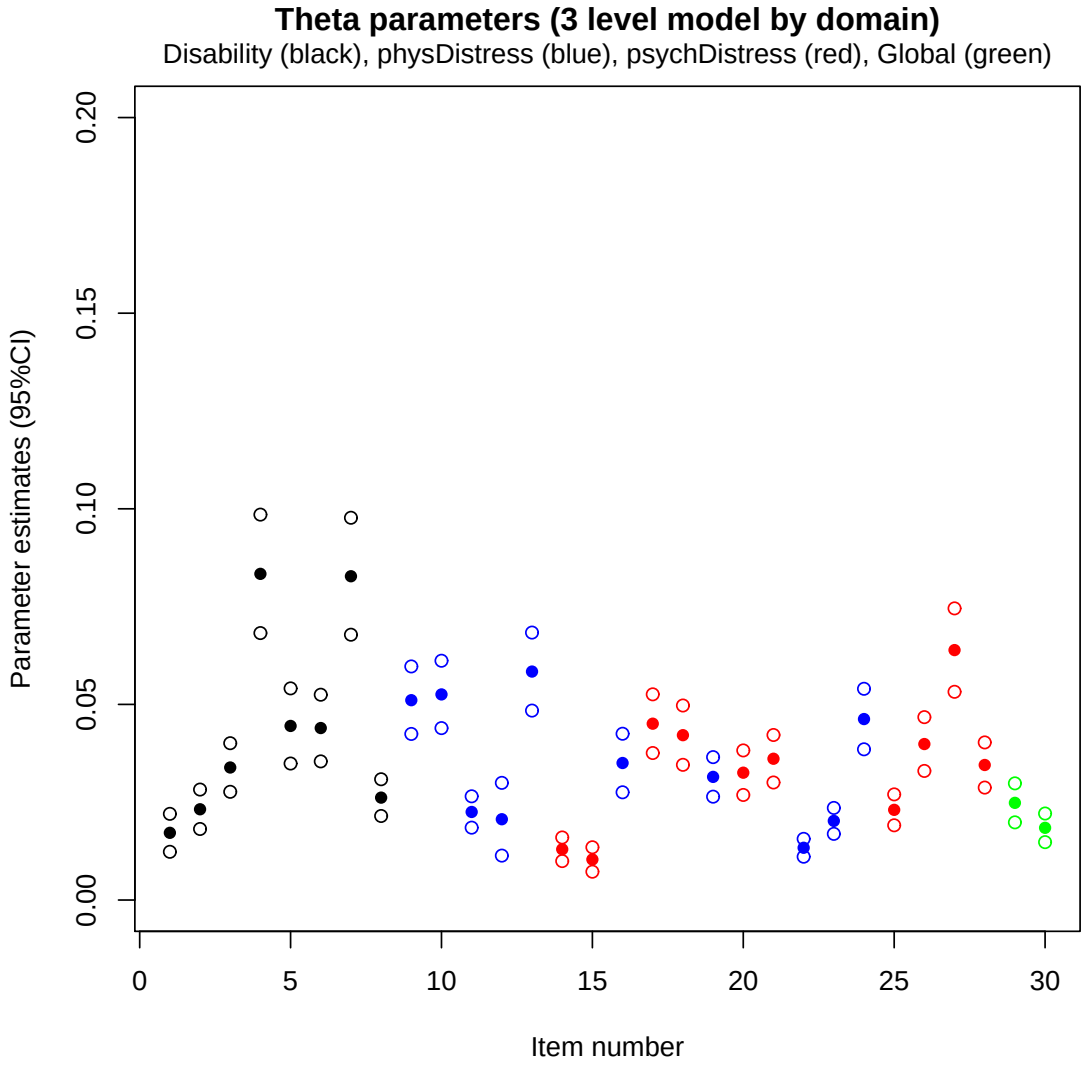
Estimated factor loadings (ANZ)



Estimated latent variable variances (ANZ)



Estimated residual variances (ANZ)



Can we simplify the model?

The full factor model involves a very large number of parameters - can we simplify?

- Many of the item-level parameters are not meaningfully different - however the item groups do not provide a basis for summarising

- Can we constrain some of the variance components to reduce the item-level variation modelled?

Which variance components matter?

	Simple random effect model (p=4)	Full factor model (p=94)
Item means	Common β_0	Item-varying
Item residvars	Common θ	Item-varying
Var($\eta^{(2)}$)	Freely estimated	Freely estimated
Var($\eta^{(3)}$)	Freely estimated	Freely estimated
Loadings:		
Subdims	All 1:1	Free (except 1)
Globals	All 1:1	Possibly free (exc. 1)

Model comparisons

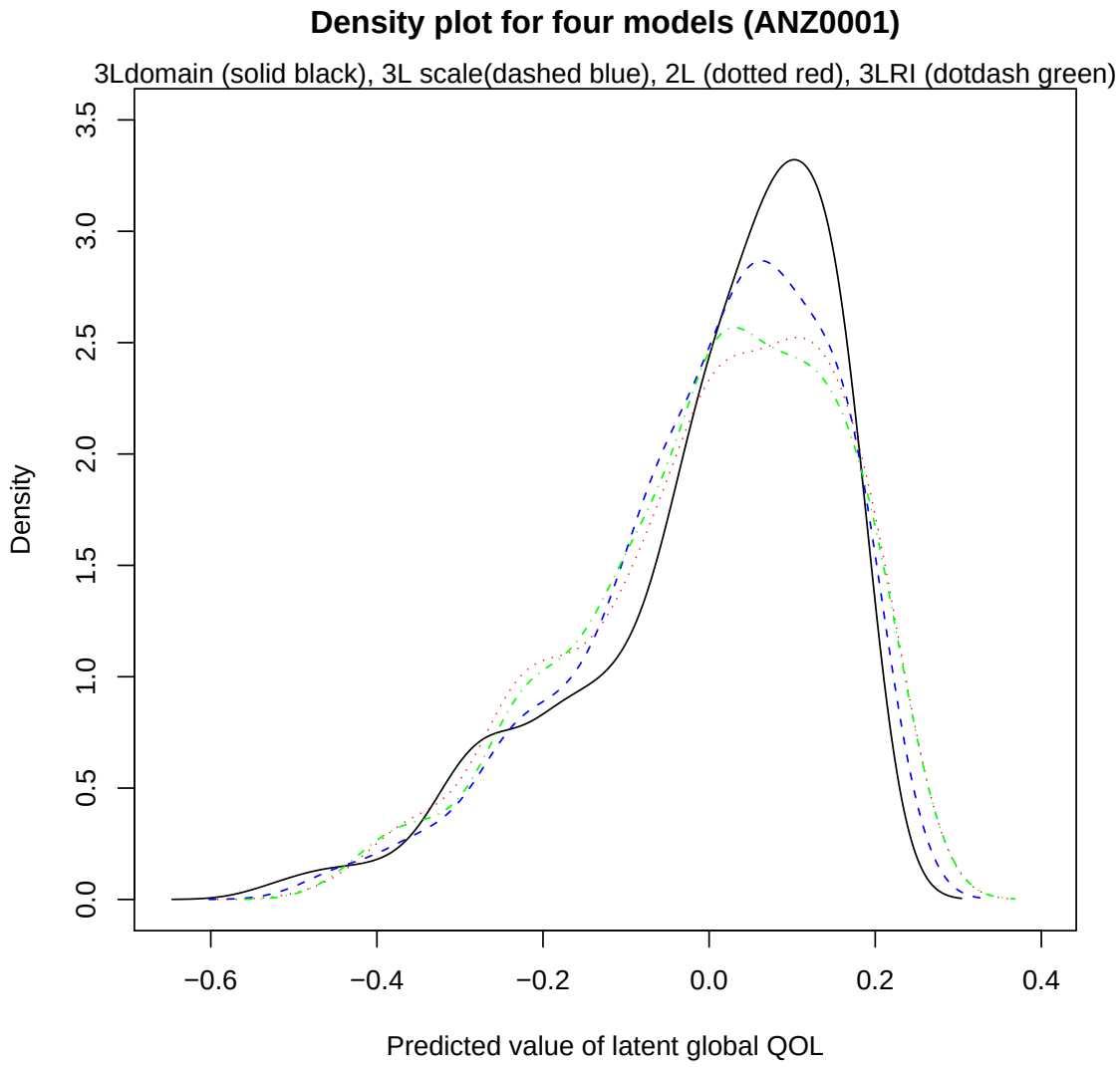
Analysis of UBQVG (196 subjects, 5944 item responses)

Model	Npar	LogL	BIC
A) Simplest (β_0, $\psi^{(2)}$, $\psi^{(3)}$, θ)	4	-144	309
B) As for A exc. item-varying β_i	35	727	-1269
C) As for B exc. free loadings	65	897	-1452
D) As for B exc. free residvars	66	975	-1602
E) Fullest	94	1104	-1712

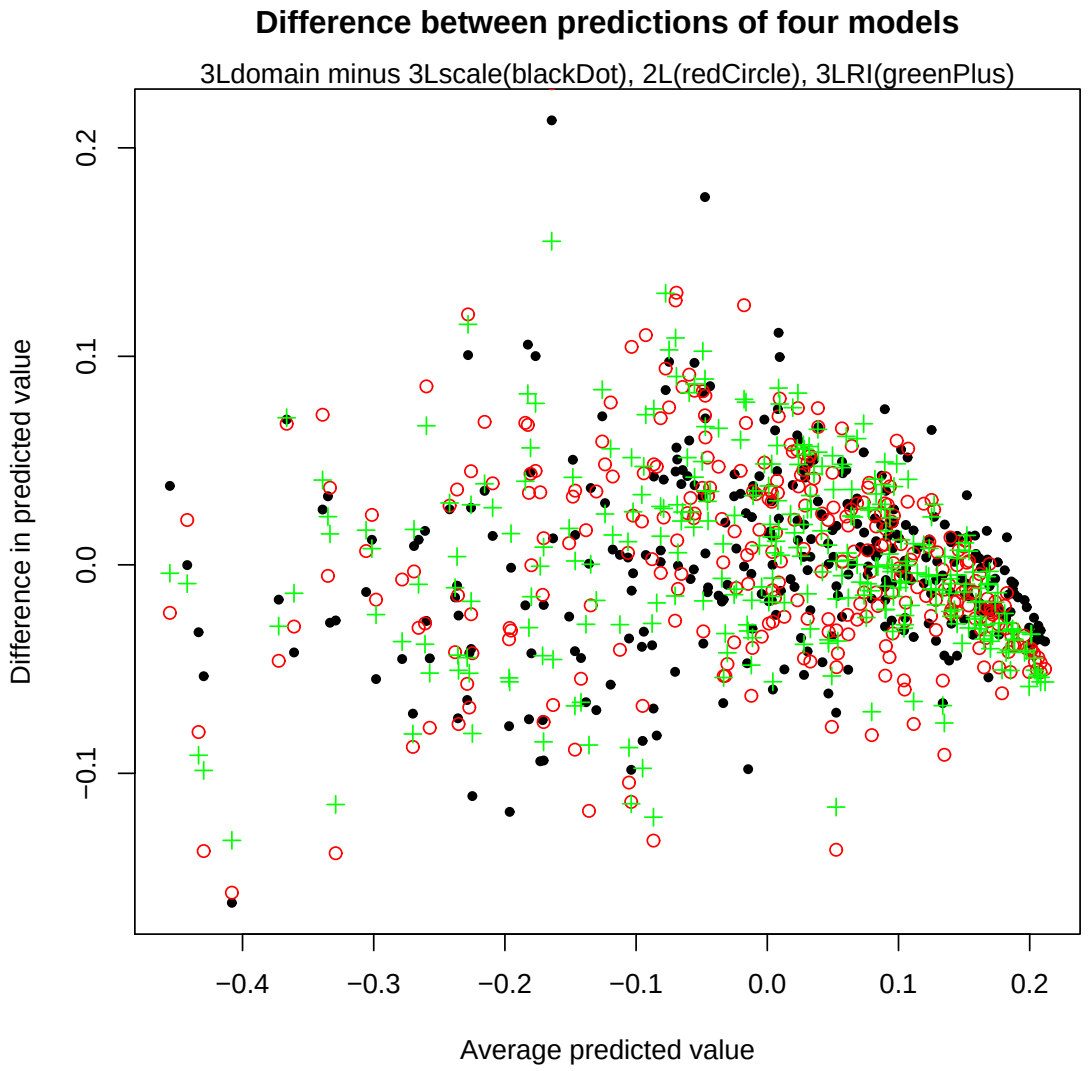
Model A: $\text{Var}(\eta^{(3)}) = 0.017$ (95% CI 0.014, 0.022)

Model E: $\text{Var}(\eta^{(3)}) = 0.016$ (95% CI 0.0084, 0.020)

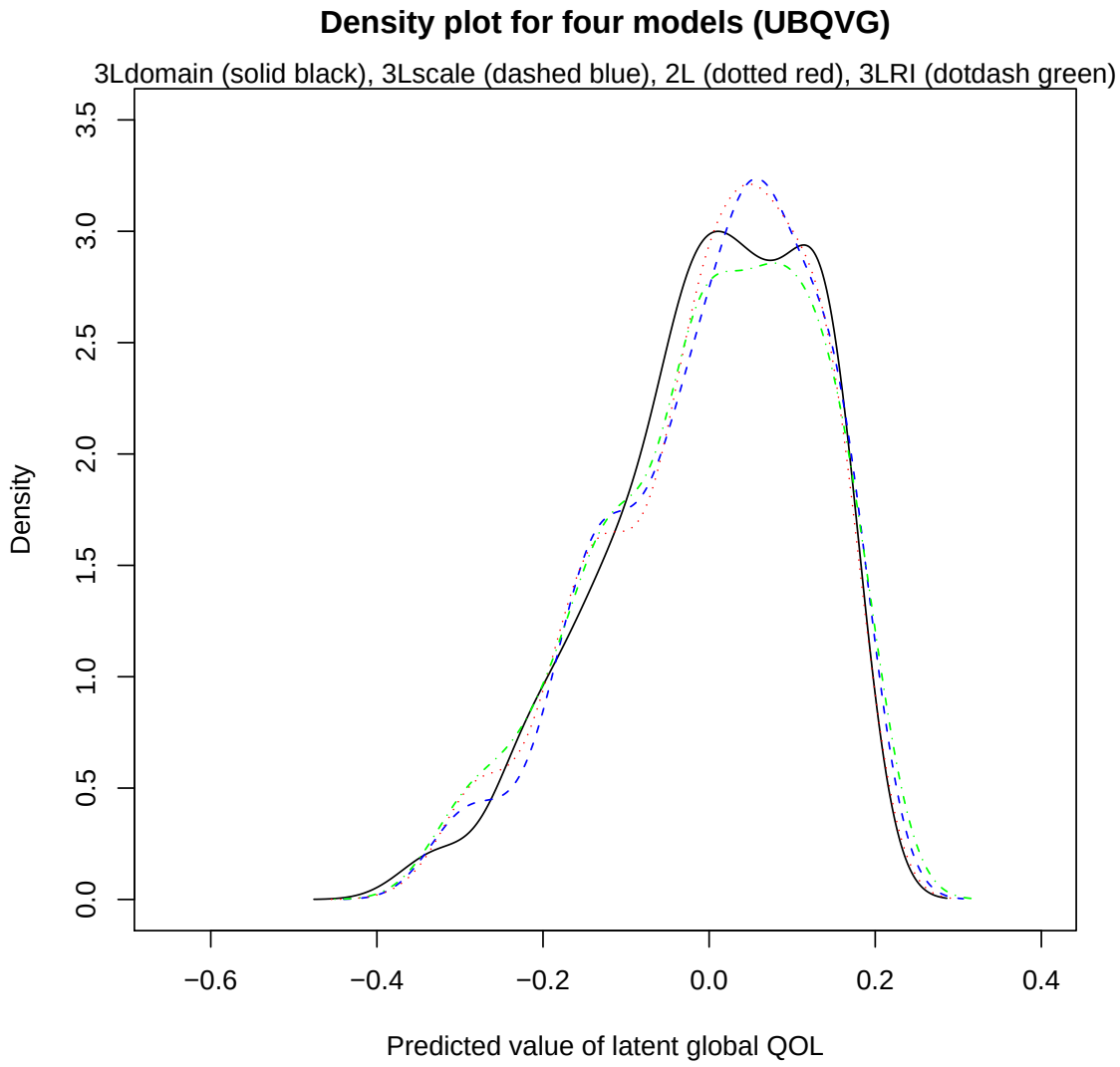
Empirical Bayes estimates of $\eta_k^{(3)}$



Empirical Bayes estimates of $\eta_k^{(3)}$



Empirical Bayes estimates of $\eta_k^{(3)}$



Summary

- We used irregular multilevel latent variable models with a mixture of random and non-random cluster types to accommodate direct and indirect QOL item measures
- Models that delineated QOL domains performed better and captured item correlations better, even for QOL data not focussed around theoretical domains
- Model simplification based on selecting variance components or grouping meaningfully different parameters did not compete or were impractical
- Empirical Bayes predictions of latent global QOL were meaningfully different between models

Thank you