Lies, damned lies and latent classes: Can factor mixture models allow us to identify the latent structure of common mental disorders?

Rachel McCrea
Mental Health Sciences Unit, University College London
rachel.mccrea.09@ucl.ac.uk

Overview

- Introduction to factor mixture models
- How are they being used in mental health research?
- My research application
- Difficulties with interpreting factor mixture models
- Trying to understand my results
- Lies, damned lies and latent classes?

Extension of latent class analysis

- Factor mixture models = extension of LCA

- Cornerstone of LCA is the assumption of <u>conditional independence</u>
  - Conditional on class membership, all variables should be uncorrelated
  - Observed correlations in the sample should be entirely accounted for by the latent classes

- This rules out severity variation within a class
  - e.g. mild and severe depression severity
  - Additional ‘severity classes’ needed to account for correlations

Factor mixture models

- Factor mixture models relax the assumption of conditional independence within latent classes
  - Allow for severity variations within a class

- Include one or more factors to model correlation structure for the variables in each class
  - Combines LCA and CFA/IRT modelling

- Specification similar to multi-group factor analysis in Mplus
  - Grouping variable unmeasured = latent classes
  - Specify a factor model within each class
  - Can constrain intercepts and factor loadings to be equal
  - Many different specifications possible

FMMs popular for mental health research

- Identifying disorder subtypes
- Exploring diagnostic boundaries (my focus)
  - e.g. anxiety and depressive disorders
  - One multi-faceted distress disorder or several distinct disorders?
- Resolving the 'continuity controversy'
  - Do symptoms vary along continuum with normal functioning?
  - Or do we have a distinct disorder category with <u>objective</u> boundaries? – a 'taxon'

    (may still be severity variation within a taxon)
  - Seen as important for research into causes and treatments

Can we identify the 'true' latent structure?

- Simulation studies suggest it may be possible (Lubke and Neale, 2006)
  - Generated data to fit FA, FMM and LC models (continuous items)
  - Datasets all analysed by each model structure
  - AIC & saBIC usually allowed correct structure to be identified

- Less clear for ordinal data (Lubke and Neale, 2008)
  - BIC best for identifying correct structure
  - Fit indices tended to favour models with <u>too few</u> classes (many category intercepts needed per extra class)

Example – Autism Spectrum Disorder

"Validation of Proposed DSM-5 Criteria for Autism Spectrum Disorder" (Frazier et al., 2012)

- Children with diagnosed ASD and undiagnosed siblings
- Compared LCA, EFA and FMMs
- Chose FMM with 2 classes and 2 dimensions

- Authors' conclusions:
  – Validates DSM-5 proposal for categorical ASD diagnosis with 2 dimensions within it
  – "The presence of an ASD versus non-ASD distinction coheres with data identifying a divergent trajectory of brain development in ASD."

Example 2 – Health anxiety

"Should health anxiety be carved at the joint?"

(Asmundson et al., 2012)

- Used large samples of undergraduate students
- Selected model: FMM with 2 classes
  - 'anxious' and 'nonanxious'

- Authors' conclusion (from the abstract):
  - "Contrary to current conceptualizations […], the FMM results indicate the latent structure of health anxiety to be <u>taxonic rather than continuous</u>."

My application of FMMs

- Aim: to investigate the latent structure of symptoms of common mental disorders

- Data: 3 household surveys of psychiatric morbidity in UK
  - repeated cross-sectional surveys
    (1993, 2000, 2007)

- Combined dataset ~ 22,000 individuals aged 16-64

- Symptoms of CMD measured by standardised interview
  - Community Interview Schedule (Revised) – CIS-R

Structure of CIS-R interview

- 13 sections covering different symptom areas
  - Ordinal score (0-4) for each symptom
  - Based on symptoms from past 7 days only
  - Symptoms not necessarily signs of illness

- Symptoms covered:
  - Somatic symptoms
  - Fatigue
  - Concentration/forgetfulness
  - Sleep
  - Irritability
  - Worry about physical health
  - Depression
  - Worry
  - Anxiety
  - Phobias
  - Panic
  - Compulsions
  - Obsessions

- Unidimensional scale – severity of mental distress

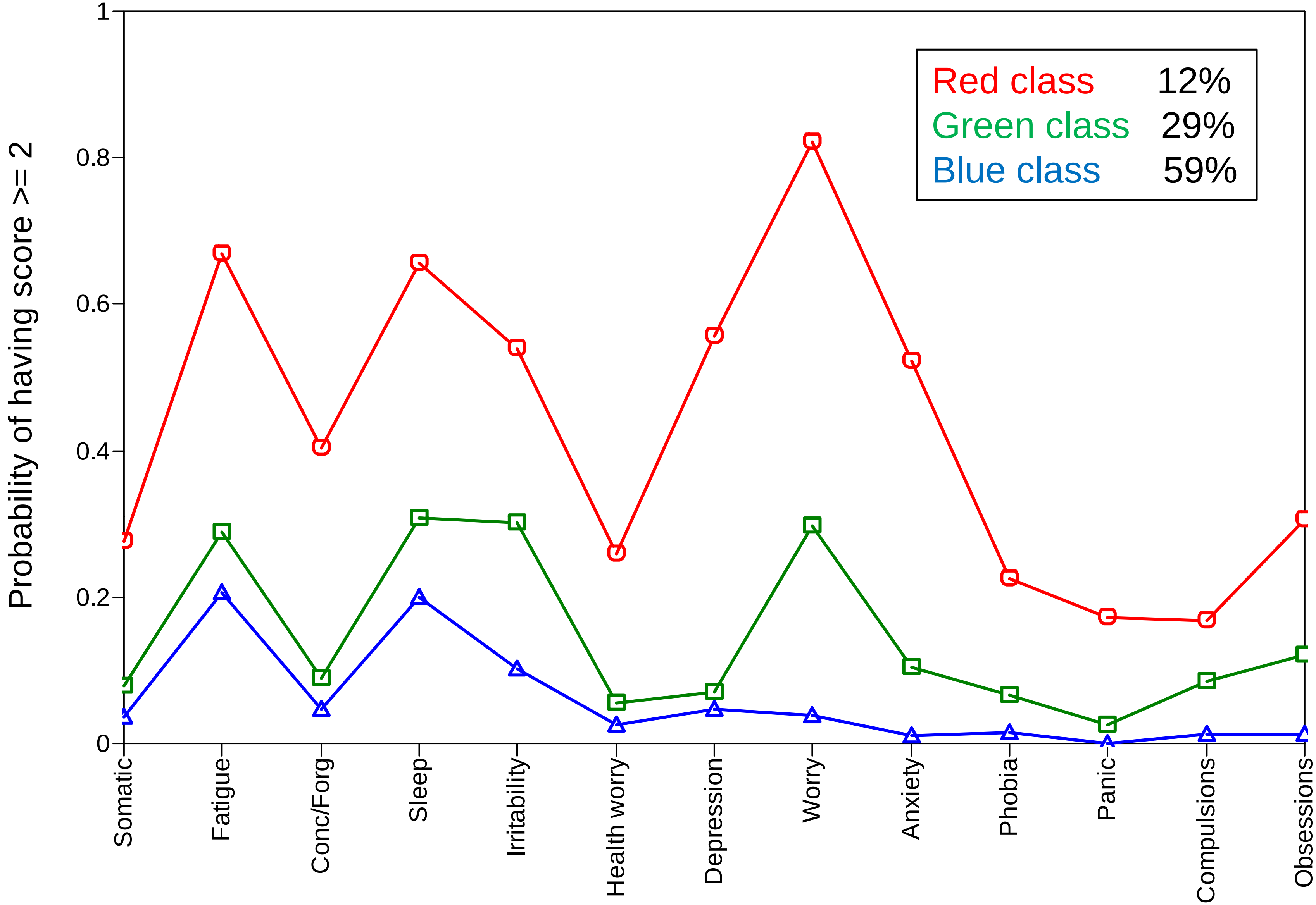
Model comparison from CIS-R data: n=11,230

Model	# p	BIC	Pairs of variables with 'poor fit' (out of 78)
Factor 1f	65	182,531	35
Factor mixture 1f 2c	119	182,015	8
Factor mixture 1f 3c	173	**181,895**	4
Latent class 4c	211	182,961	8

Based on bivariate chi-squares in Tech10

FMMs have same loadings in each class but different item intercepts

Factor mixture model – 1 factor, 3 classes



What do the latent classes mean?

- Do these three latent classes really represent distinct clinical groups in the population?
  - I need to be sure before making claims
  - Class membership probs. – no obvious clinical interpretation

- FMMs allow for a severity dimension within class
  - Can't be simple severity classes
  - If FMM fits better than factor model without classes, surely classes must be real groups?

…not necessarily!

Two roles of latent classes

- Direct role: represent true groups
- Indirect role: approximating a continuous distribution

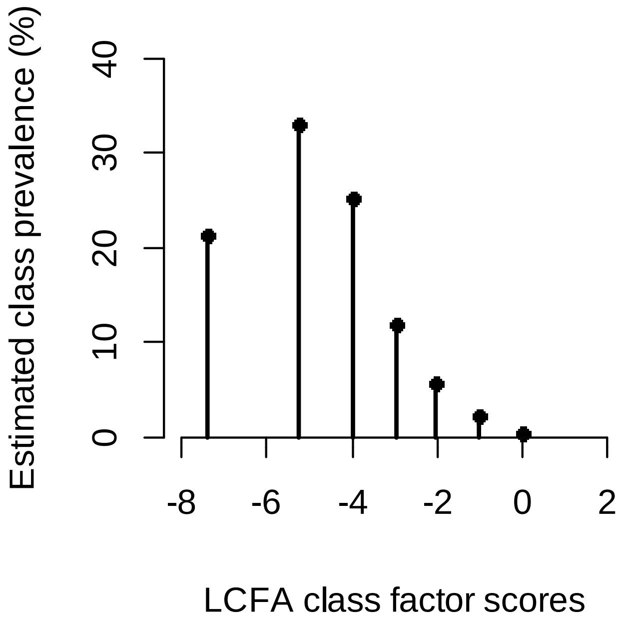
- Situations where a factor mixture model may appear to describe data better <u>in the absence of true groups</u> (Bauer and Curran, 2004)
    - Non-normality of the factor(s)
    - Miss-specification of measurement model
    - Non-linearity (for logistic models: on logit scale)

- Must rule out alternatives to conclude real groups

Are classes just modelling non-normality?

- Standard factor model assumes normally distributed factor
  - Inappropriate for mental health
  - Classes accommodating this?

- Test: Fit 'latent class factor model'
  - Approximates continuous factor distribution with 'located' classes
  - 'Non-parametric' factor analysis

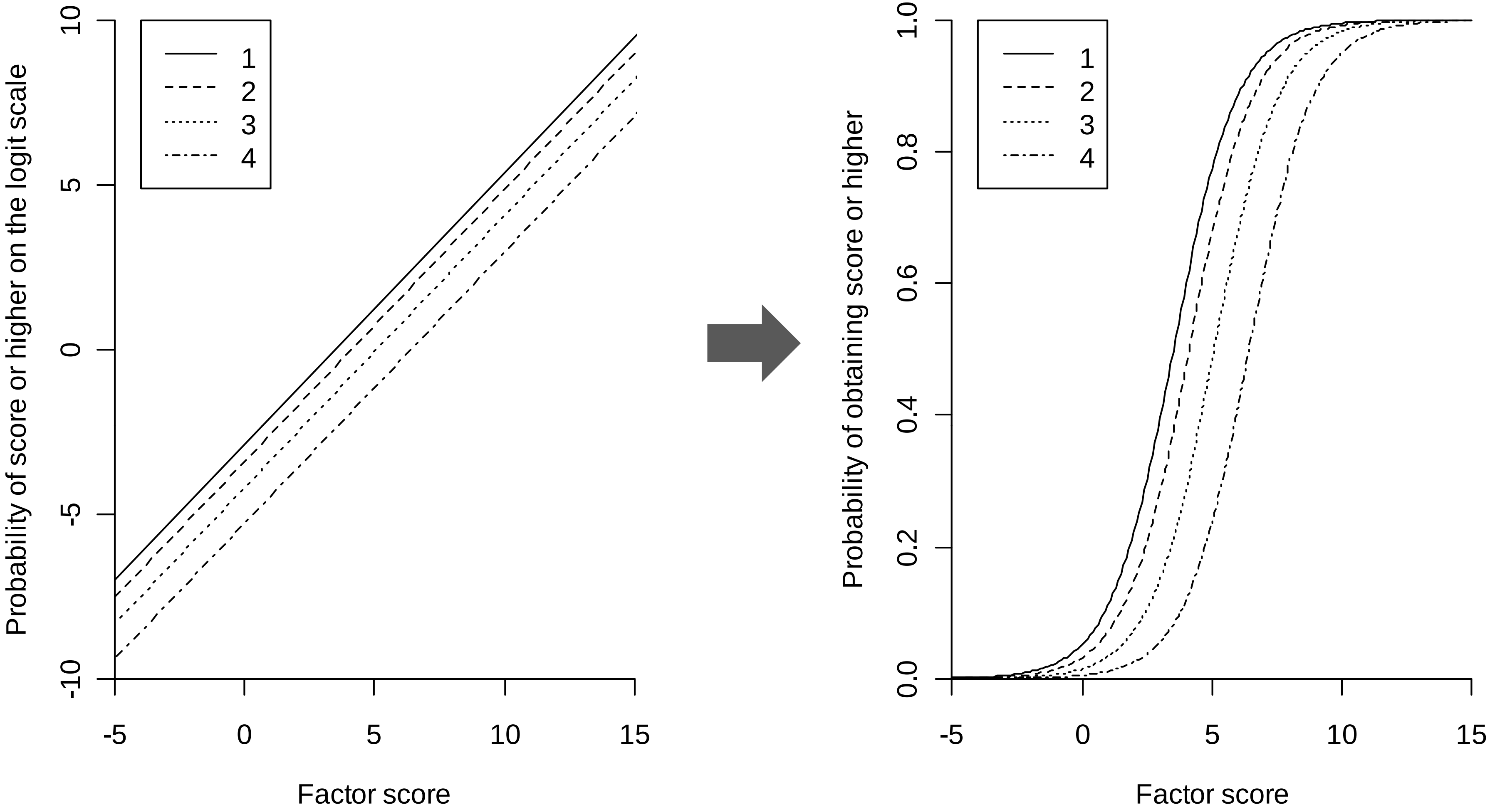


- Result: no improvement on fit of standard factor model
  - 2 and 3 class FMMs still much better
  - Suggests FMMs not just modelling non-normality

What about non-linearity?

- Factor model for ordinal data related to ordinal logistic regression:
  - Cumulative probability model (as in ‘ologit’ in Stata)
  - Assumes <u>linear</u> relationship between probabilities of responses to each variable and the factor <u>on logit scale</u>
  - Assumes <u>proportional odds</u> for each ordered response category (= equal slopes = parallel lines)

Linearity and proportional odds assumptions



Investigating nonlinearity

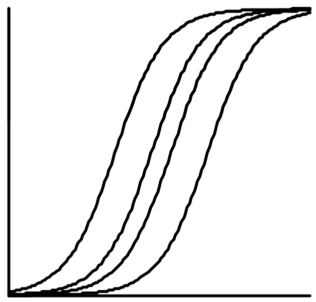
- Maybe the FMM is relaxing the linearity and proportional odds assumptions

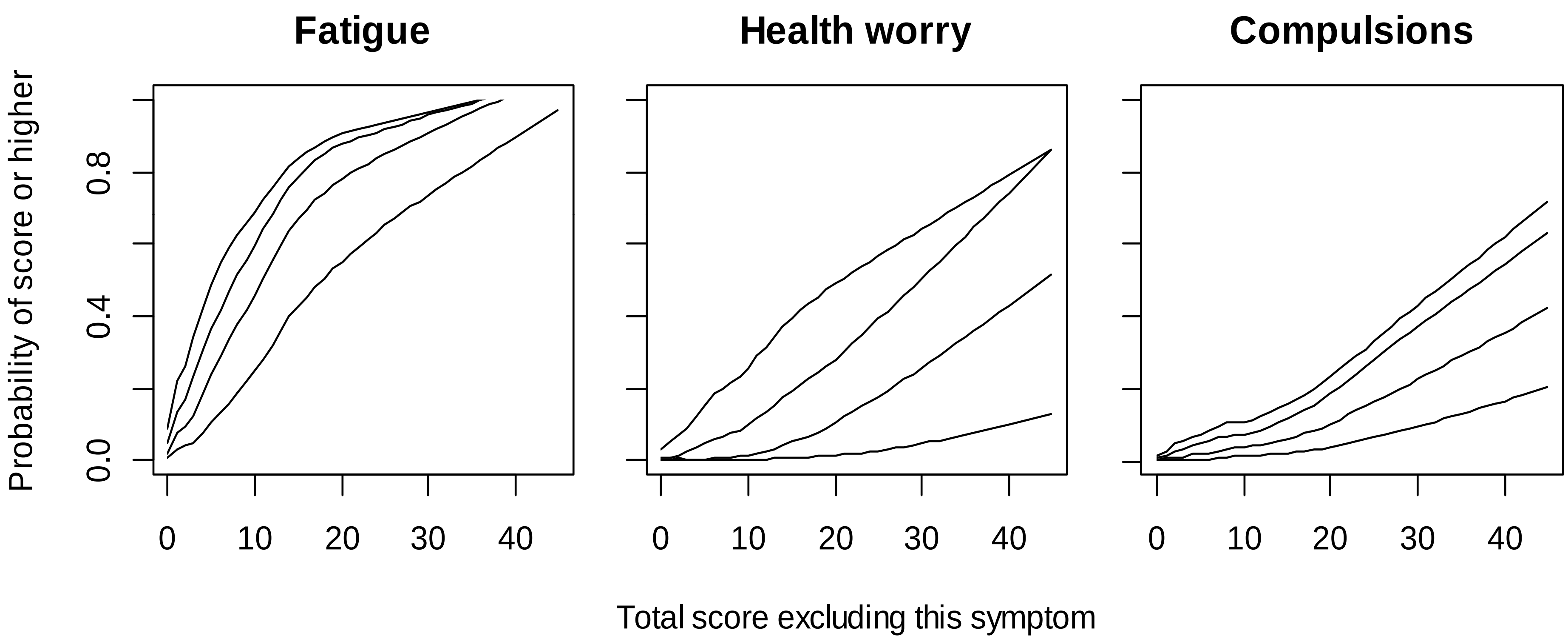
- How to investigate this for ordinal data?
  - No straightforward way to assess true shape of relationship
  - Used a whole suite of approaches
  - All had some limitations, but fairly consistent picture

- Clearest to present:
  - Lowess curves (descriptive: form of non-parametric regression)
  - Used summed item scores to ‘represent’ factor scores

Lowess curves

Describing cumulative probabilities, as in:





Conclusions

- Factor mixture models appear to describe the CIS-R data better than models without classes
- However, evidence of non-linearity (on logit scale) and violations of proportional odds
- Careful examination suggests latent classes are accommodating these violations
  - Class allocations consistent with patterns of non-linearity
  - Implies classes unlikely to represent real groups
  - BUT can’t prove this either way

Conclusions (continued)

- No clear evidence for any ‘disorder classes’
- BUT doesn’t prove that there are <u>no</u> discrete disorders
  - ‘Signal’ drowned out by ‘noise’ from factor model misfit?
  - May be impossible to distinguish dimensions from discrete categories empirically
  - Key discriminating characteristics not measured?
  - Lack of power?
- My view: disappointingly ambiguous conclusions for a <u>very</u> time-consuming exercise

Interpretation of FMMs in the literature

- Some papers mention alternative roles of classes
  - Usually simulation studies or illustration papers
  - Tend to avoid drawing substantive conclusions

- ‘Taxonicity’ of classes often unquestioned in applied psychiatric research papers
  - Papers frequently don’t mention that classes may reflect non-normality or other factor model violations
  - Authors may not be aware

- BUT model violations may be common in mental health
  - Measures often designed as screening tools
  - Items not selected for psychometric properties

Lies, damned lies and latent classes?

- These hybrid mixture models are very complex
  - Huge effort required to develop real understanding
  - Many readers will have to take findings ‘on trust’
  - Reviewers may lack sufficient expertise to spot problems
- FMMs present <u>severe risk</u> of over-interpretation
  - Not a magic bullet for identifying true latent structure
  - Could lead to research blind alleys
- Researchers reporting FMMs <u>must</u> highlight and explore alternative interpretations
  - If not, we should be sceptical of any claims

References

- Asmundson, G.J.G. et al., 2012. Should health anxiety be carved at the joint? A look at the health anxiety construct using factor mixture modeling in a non-clinical sample. *Journal of Anxiety Disorders*, 26(1), pp.246–251.

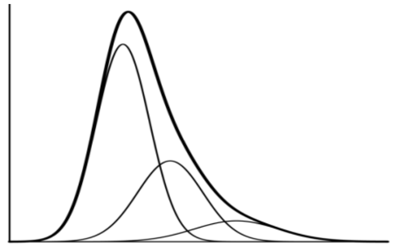
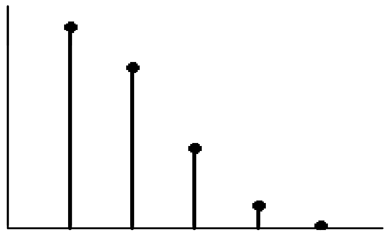
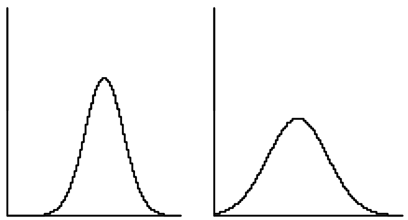
- Bauer, D.J. & Curran, P.J., 2004. The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. *Psychological Methods*, 9(1), pp.3-29.

- Frazier, T.W. et al., 2012. Validation of Proposed DSM-5 Criteria for Autism Spectrum Disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, 51(1), pp.28–40.e3.

- Lubke, G.H. & Neale, M.C., 2006. Distinguishing between latent classes and continuous factors: Resolution by maximum likelihood? Multivariate Behavioral Research, 41(4), pp.499–532.

- Lubke, G.H. & Neale, M.C., 2008. Distinguishing between latent classes and continuous factors with categorical outcomes: Class invariance of parameters of factor mixture models. Multivariate Behavioral Research, 43(4), pp.592–620.

Three main families of factor mixture model

	Measurement invariance? (intercepts/loadings)	**Factor variance within class?**	**Example**
'Semi-parametric factor model' **a.k.a. mixture factor model**	Yes	Yes factor*;	
'Latent class factor model' **a.k.a. non-parametric factor model**	Yes	**No** factor@0;	
'Factor mixture model'	**No / weak**	Yes factor*;	

Mplus code: ‘Latent class factor model’ 1f 4c

```
!This code is for ordinal data
Variable:
   …
   Categorical are …;
   Classes = c(4); ! num. classes

Analysis:
   Estimator = MLR;
   Algorithm = Integration;
   Type = Mixture;
   Starts = 100 50;

Model:
    %OVERALL%
     factor BY somatic;
     factor BY fatigue;
     factor BY concforg;
     factor BY sleep;
     …
```

```
[ c#1*];
[ c#2*];
[ c#3*];

%C#1% !One section for each class
factor BY somatic@1;
factor BY fatigue* (1);
factor BY concforg* (2);
factor BY sleep* (3);
...
factor@0; !Fixed @0 in all classes
[ factor@0 ]; !Fixed in 1 class only

[ somatic$1* ] (16);
[ somatic$2* ] (17);
[ somatic$3* ] (18);
[ somatic$4* ] (19);
[ fatigue$1* ] (20);
[ fatigue$2* ] (21);
```

```
[ fatigue$3*] (22);
[ fatigue$4*] (23);
...

%C#2%
factor BY somatic@1;
factor BY fatigue* (1);
factor BY concforg* (2);
factor BY sleep* (3);
...

factor@0;
[ factor* ]; !Free in other classes

[ somatic$1* ] (16);
[ somatic$2* ] (17);
[ somatic$3* ] (18);
[ somatic$4* ] (19);
[ fatigue$1* ] (20);
[ fatigue$2* ] (21);      ...
```

Mplus code: a ‘Factor mixture model’ 1f 3c

```
!This code is for ordinal data
Variable:
   …
   Categorical are …;
   Classes = c(3);

Analysis:
   Estimator = MLR;
   Algorithm = Integration;
   Type = Mixture;
   Starts = 2000 500; !Need lots

Model:
    %OVERALL%
     factor BY somatic;
     factor BY fatigue;
     factor BY concforg;
     factor BY sleep;
     …
```

```
[ c#1*];
[ c#2*];

%C#1% !One section for each class
factor BY somatic@1;
factor BY fatigue* (1);
factor BY concforg* (2);
factor BY sleep* (3);
...

factor* ;          !Free in all classes
[ factor@0 ]; !Fixed in all classes

[ somatic$1* ] ; !Intercepts can now
[ somatic$2* ] ; !differ between
[ somatic$3* ] ; !classes.
[ somatic$4* ] ;
[ fatigue$1* ] ;
[ fatigue$2* ] ;
```

```
[ fatigue$3*] (22);
[ fatigue$4*] (23);
...

%C#2%
factor BY somatic@1;
factor BY fatigue* (1); !Loadings
factor BY concforg* (2); !still
factor BY sleep* (3);       !equal.
...

factor* ;          !Free
[ factor@0 ]; !Fixed

[ somatic$1* ] ;
[ somatic$2* ] ;
[ somatic$3* ] ;
[ somatic$4* ] ;
[ fatigue$1* ] ;
[ fatigue$2* ] ;      ...
```

Example code for lowess curves in R

```
## This code was written for ordinal data with five categories per item (coded as 0-4)
## Curves estimated separately – may cross inappropriately in regions where data are sparse

library(Hmisc)   ## Package containing the plsmo() function
responses <- read.table("C:/Data/mplusexport.dat", sep=",")

item <- 1   ## This number should be the column number of the item you wish to plot
score4 <- as.numeric(responses[,item]==4)
score3 <- as.numeric(responses[,item]==4|responses[,item]==3)
score2 <- as.numeric(responses[,item]==4|responses[,item]==3|responses[,item]==2)
score1 <- as.numeric(responses[,item]==4|responses[,item]==3|responses[,item]==2|
   responses[,item]==1)
totscores <- rowSums(responses)    ## Assumes there are no other variables in the dataset
restscores <- totscores - responses[,item]

plsmo(restscores, score1, ylab="Probability of score or higher", xlab="Restscore",
     ylim=c(0,1), trim=0, f=0.1)    ## "f=0.1" controls the spikiness of the curve - it can range from 0 to 1.
plsmo(restscores, score2, trim=0, add=T, f=0.1)
plsmo(restscores, score3, trim=0, add=T, f=0.1)
plsmo(restscores, score4, trim=0, add=T, f=0.1)
```