

HIGHER ORDER FACTOR SCORES FOR THE QLQ-C30

Chad Gundy, Peter Fayers,
Mogens Groenvold, Morten Petersen, Neil Scott,
Mirjam Sprangers, Galina Velikova, and Neil Aaronson

on behalf of
the EORTC and the Quality of Life Cross-Cultural Analysis Group

Rome, April 2010

AIMS

Define/fit a (higher order) measurement model summarizing the 15 dimensional profile generated by the 30-item EORTC QLQ-C30 questionnaire.

Factor (or component) scores may then be calculated on the basis of this model.

Why “higher order” factor scores?

- 15 dimensions can be unwieldy for clinicians, patients, and researchers
- Reduces complexity of analyses (multiple end-points) and problems with multiple testing
- Frequent requests from users
- The competition is doing it!
- Theory formation –relating 15 dimensions in a meaningful way- may be a way forward.

Data (1)

QLQ-C30 v.3 data for cancer patients (at pre-treatment) selected from 38,000 respondents in 100+ studies from various:

- countries,
- primary disease sites,
- disease stages, and
- treatment phases.

Mean age= 60 years, 55% male, approx. evenly divided between stages I-III vs. IV. Items all highly skewed.

Data (2)

30% simple, random sample drawn
(the remainder left for testing measurement equivalence
and cross-validation), leaving:

2700 patients for model fitting.

*Sufficient sample size, even for “robust”
estimation methods!*

Methods

Confirmatory Factor Analysis (CFA)/
Structural Equation Modeling(SEM)

Modeling Issues

- Estimates for single-item scale error variance
- Exclude FI scale; QL scale retained, yet not subsumed
- ‘robust’ weighted least squares estimator (WLSMV in Mplus v4.2)
- Treatment of missing data (pair-wise deletion)
- Ordinal items (polychoric correlations)
- Correction for sample clustering
- ‘standard’ QLQ-C30 model architecture

i1 | i2 | i3 | i4 | i5

PF

i6 | i7

RF

i20 | i25

CF

i21 | i22 | i23 | i24

EF

i26 | i27

SF

i29 | i30

QL

i10
i12
i18

FA

i14
i15

NV

i9
i19

PA

“Standard” Model

Dy

i8

Sl

i11

Ap

i13

Co

i16

Di

i17

~~Fi~~

i28

QLQ-C30 Higher Order Factor Models (1) (reflective measurement)

- 1) *Standard QLQ-C30*
- 2) *2D PhysicalHealth-MentalHealth*
- 3) *2D PhysicalBurden/MentalFunction*
- 4) *2D Symptom Burden & Function*
- 5) *1D HRQL*

QLQ-C30 Higher Order Factor Models (2)

Hybrid Reflective/Formative measurement

6) 2D formative *Burden & Function*
(free weights)

7) 2D formative *Burden & Function*
(fixed weights)

i1 | i2 | i3 | i4 | i5

PF

i6 | i7

RF

i20 | i25

CF

i21 | i22 | i23 | i24

EF

i26 | i27

SF

i29 | i30

QL

i10
i12
i18

FA

i14
i15

NV

i9
i19

PA

1) "Standard" Model

Dy

i8

Sl

i11

Ap

i13

Co

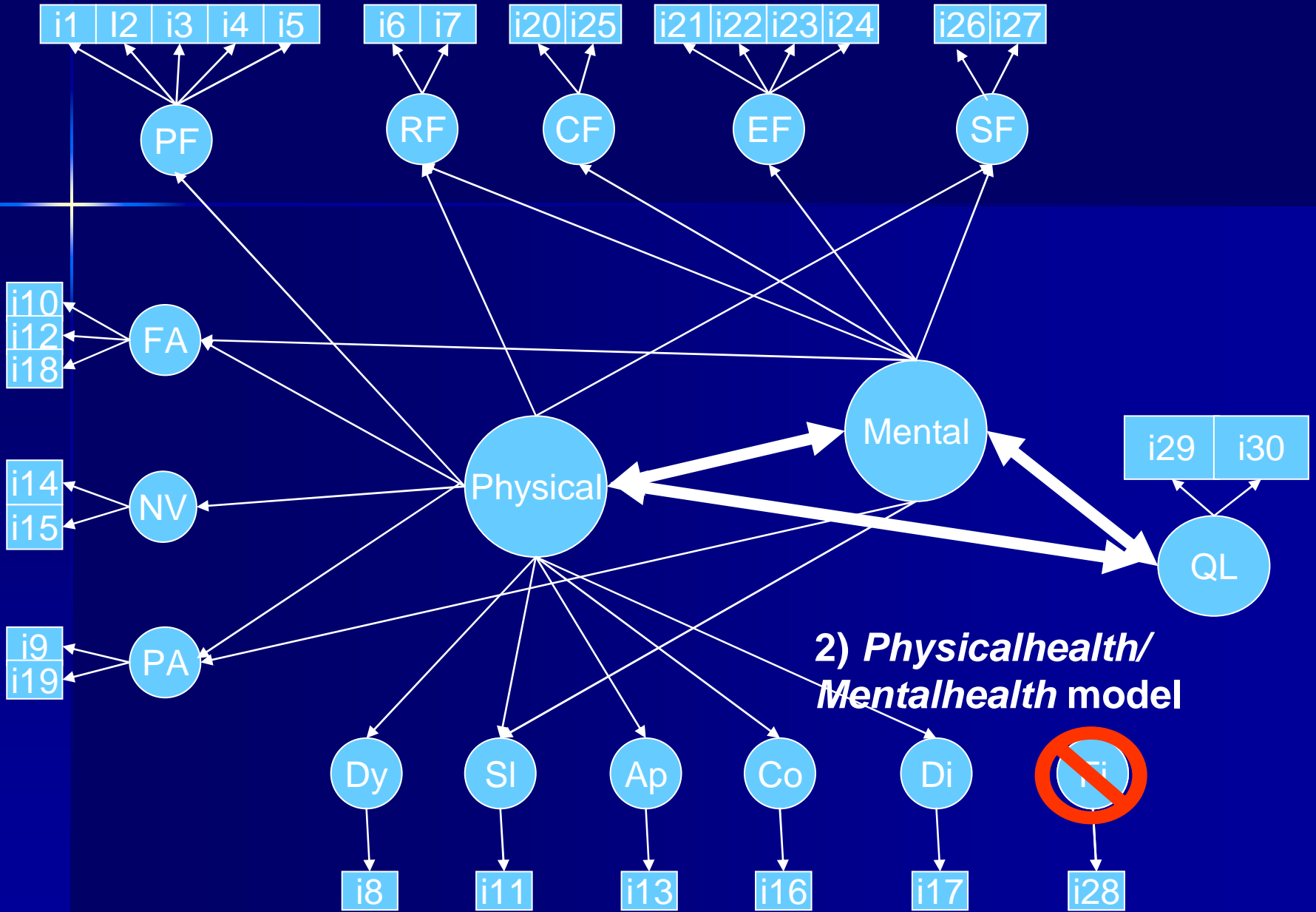
i16

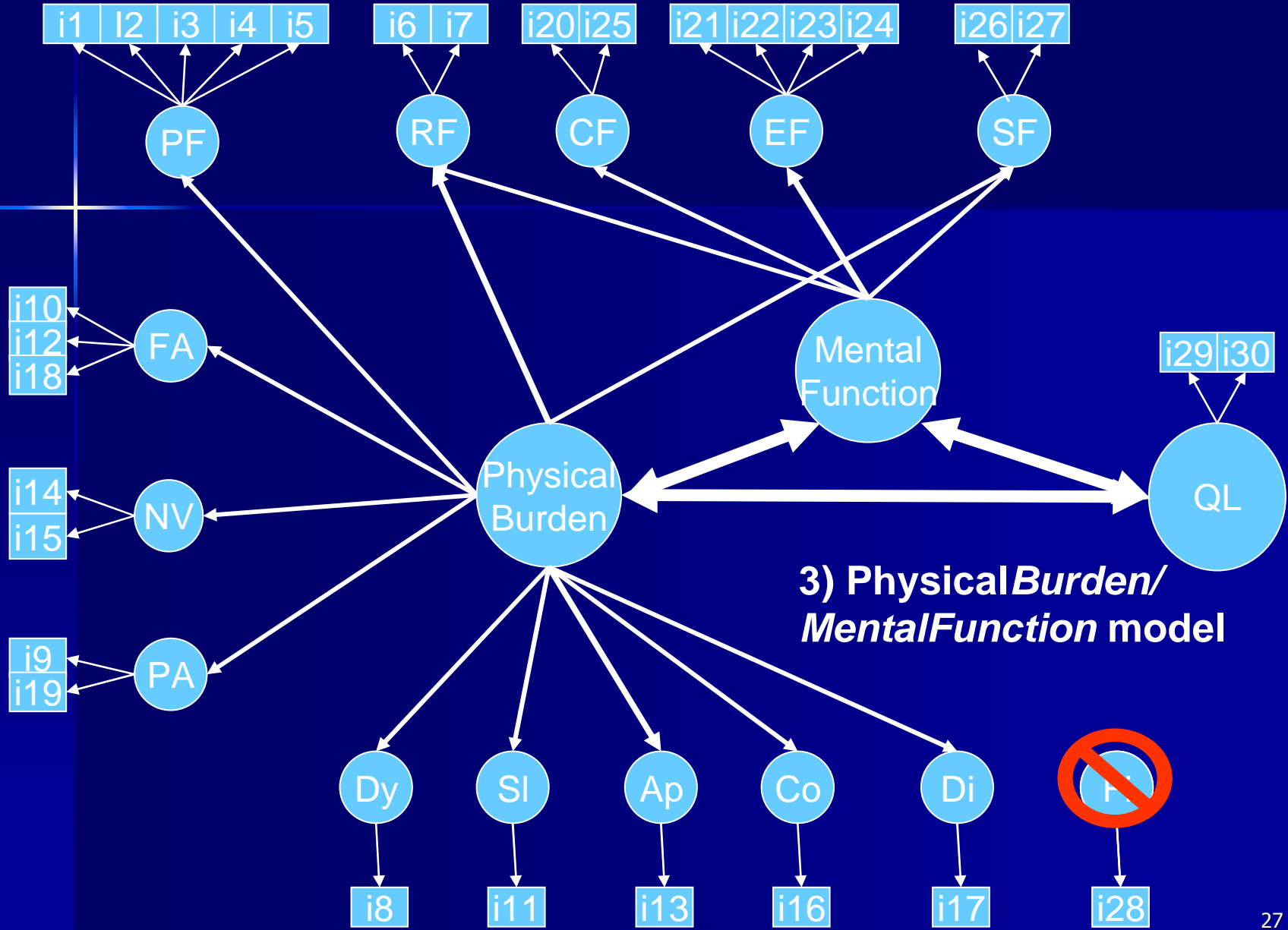
Di

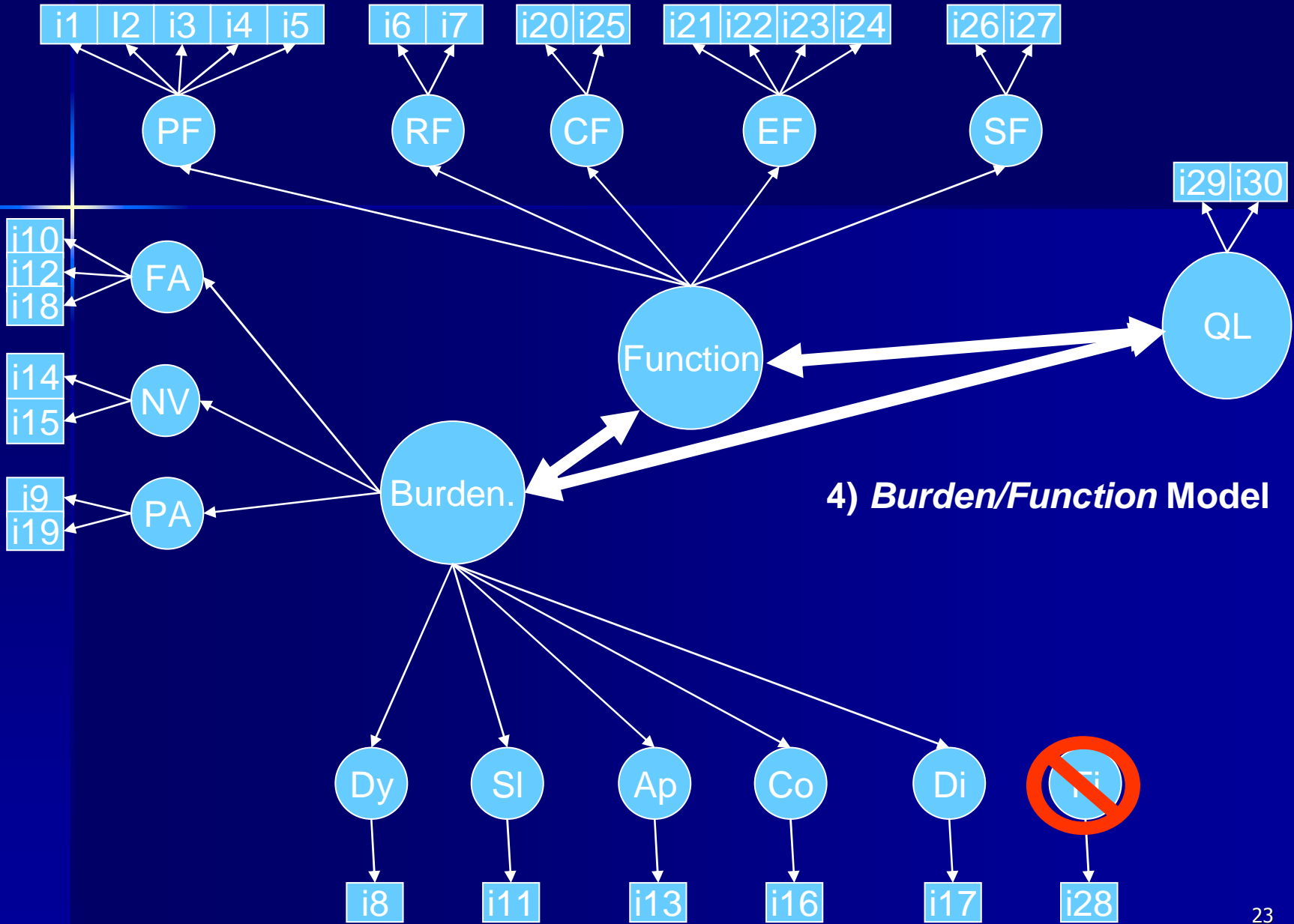
i17

~~Fi~~

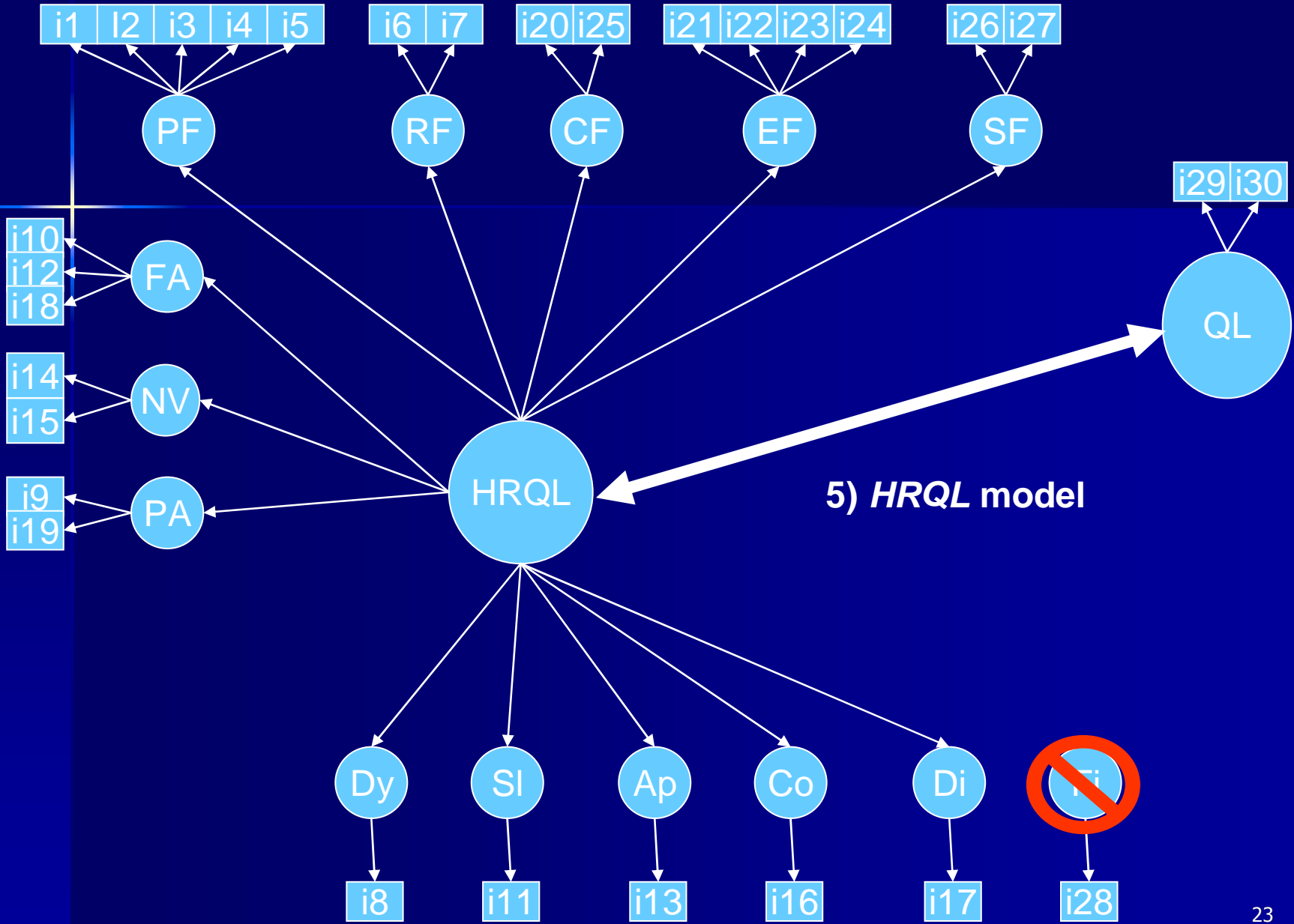
i28

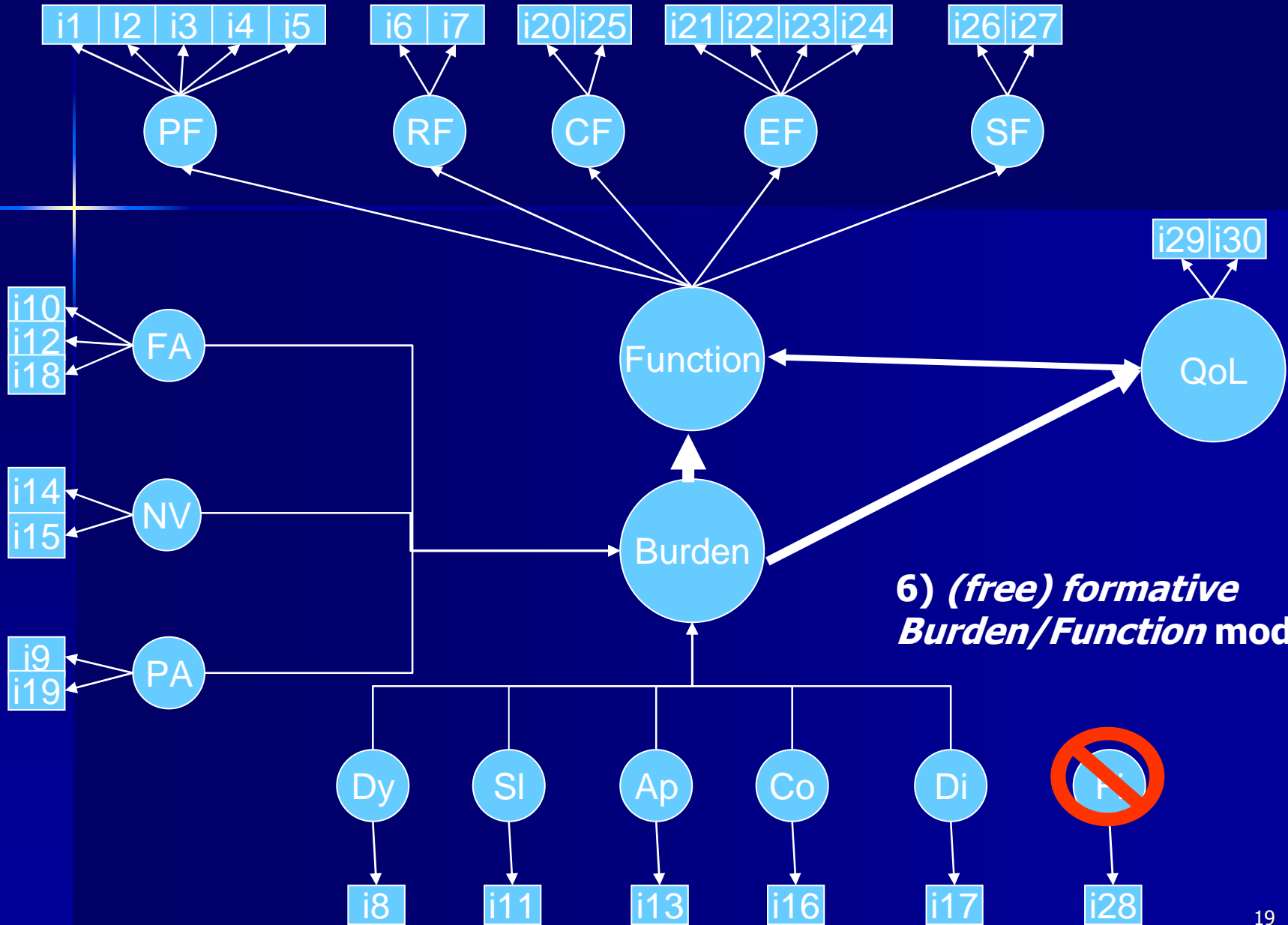




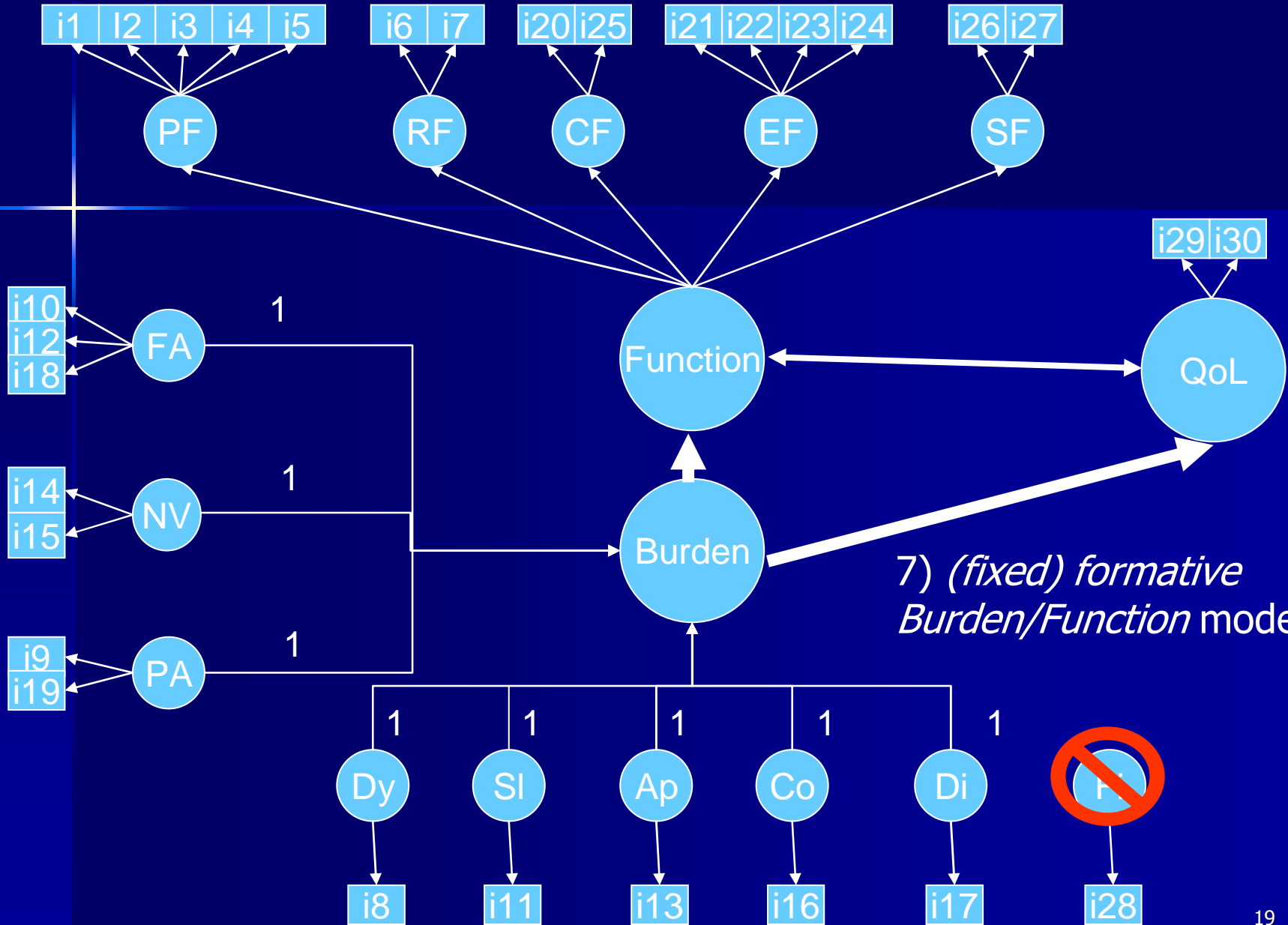


4) Burden/Function Model



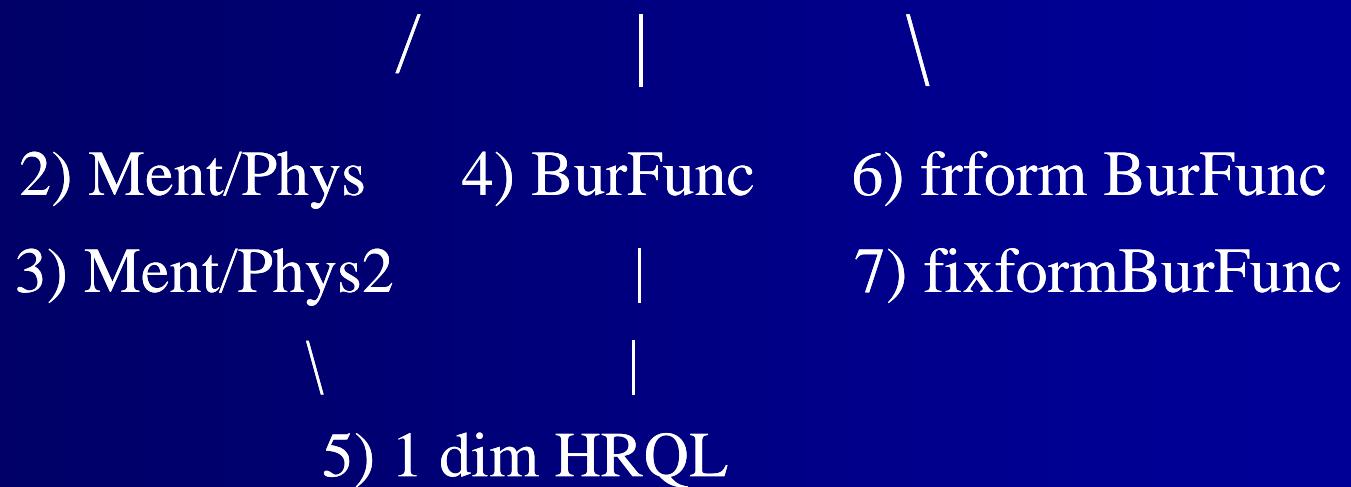


**6) (free) formative
Burden/Function model**



Three branches of nested models

1) Standard



RESULTS (1)

Fit Indices/Tests for Various Models

Model	χ^2	df	CFI/TLI	RMSEA
1) Standard Model	114	19	0.97/0.99	0.043
2) PhysicalHealth, MentalHealth	184	22	0.95/0.98	0.052
3) PhysicalBurden,MentalFunction	191	20	0.95/0.98	0.056
4) Burden, Function	232	20	0.93/0.97	0.063
5) HRQL	235	20	0.93/0.97	0.063
6) (free) formative Burden & Function	231	20	0.93/0.97	0.063
7) (fixed) formative Burden &Function	253	20	0.93/0.97	0.066

RESULTS (2)

- Correlations between higher order factor in 2D models approx. 0.70 for *Mental-Physical* models. >0.9 for other 2D models.
- Every step downwards -in a branch of models- results in a significant deterioration in fit.
- Models in different branches cannot be directly compared by means of a chi-square difference test.

RESULTS (3)

- Variance “explained” (i.e., R^2) by *Mental/Physical* model is superior to *Burden/Function* model for PF,EF,CF, and SL.
- Burden (in the *free formative Burden/Function* model) is determined only by FA. Other weights are n.s., or inconsequential.

RESULTS (4): Cross validation

- Cross validation of the *Physical health/Mental health* model
(using a second 10% holdout sample)
revealed no significant difference.
($\chi^2=4.9$, $df=8$, $p=0.77$).

Discussion/Conclusions (1)

- The *Standard* model has a “good” fit to the data.
- There is room for improvement, even for the *Standard* model.

Discussion/Conclusions (2)

- *Mentalhealth/Physicalhealth* model is the best simplification of the QLQ-C30 for the first branch of the models studied here.
- *Burden/Function* model has the best fit of the second branch.
- *(free weight) formative Burden-Function* model is the best performing model of the third (formative) branch.

Discussion/Conclusions (3)

- All three models are an “adequate” fit to the data, yet this also comes at a cost: a significant deterioration in fit w.r.t. the *Standard* model.

Discussion/Conclusions (4)

- Cannot directly compare models in different branches.
- However, the *Mentalhealth/ Physicalhealth* model has slightly better fit statistics, and provides a better “explanation” (i.e., R²) of several 1st order latents (i.e., PF,EF,CF, and SL) than does the *Burden/Function* model.

Discussion/Conclusion (5)

- The (free weight) *formative Burden-Function* model ignores all of the symptom scales, with the exception of FAtigue.
- Such a simplification may be viewed as being overly zealous, and tends to limit the usefulness of this model (and this branch).

Discussion/Conclusion (6)

- The *Mentalhealth/Physicalhealth* model seems to be the best choice.

Next steps

- measurement equivalence over time
- measurement equivalence over exogenous variables (e.g., age, gender, stage)
- known groups validity of higher order vs. profile scores
- predictive validity (in longitudinal trials)
- publication of algorithm

Thank you for your attention!

Suggestions?

Remarks?

email: c.gundy@nki.nl



***** FATAL ERROR**

THERE IS NOT ENOUGH MEMORY SPACE TO RUN THE PROGRAM ON THE CURRENT INPUT FILE. THE ANALYSIS REQUIRES 14 DIMENSIONS OF INTEGRATION RESULTING IN A TOTAL OF **2.9193E+17** INTEGRATION POINTS. THIS MAY BE THE CAUSE OF THE MEMORY SHORTAGE. YOU CAN TRY TO FREE UP SOME MEMORY BY CLOSING OTHER APPLICATIONS THAT ARE CURRENTLY RUNNING. ANOTHER SUGGESTION IS CLEANING UP YOUR HARD DRIVE BY DELETING UNNECESSARY FILES.

RESULTS (3) Std. Regression Weights for best models in each branch

First Order Factors	Physical/Mental (model # 2)			Burden/Function (model #4)			(free wgt.) formative Burden/Function (model #6)		
	Physical	Mental	R2	Burden	Function	R2	(free) formative Burden	Function	R2
PF	.81\$.65		.77*	.59		.77\$.59
RF	.82*	.12*	.82		.89\$.80		.89*	.80
EF		.75\$.56		.64*	.41		.64*	.41
CF		.90*	.80		.79*	.62		.79*	.62
SF	.38*	.51*	.68		.82*	.67		.82*	.67
FA	.81*	.21*	.93	.98\$.96	.86\$		na
NV	.66*		.44	.65*		.42	.05		na
PA	.59*	.25*	.61	.79*		.63	.13*		na
DY	.81*		.66	.80*		.65	.04		na
SL	<.01	.80*	.64	.76*		.58	.07*		na
AP	.85*		.72	.84*		.71	-.09		na
CO	.74*		.55	.73*		.54	.02		na
DI	.62*		.38	.62*		.38	-.02		na

Extra Requirements: Generalizable Results

- Measurement Equivalence over (patient and language) groups: identical item thresholds, factor structures, and factor loadings

...essential in order to make comparisons over groups: if equivalence (partially) fails, then some comparisons between groups may not be justifiable...

- Cross Validation
- Predictive Validity

Data

Split into 3 analysis groups:

- a 30% sample for model fitting

- a 10% sample for cross validation

and

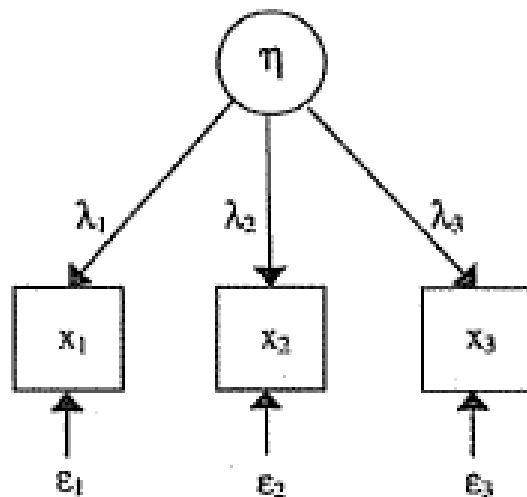
- a 60% sample for examination of measurement equivalence (not yet used!)

Modelling

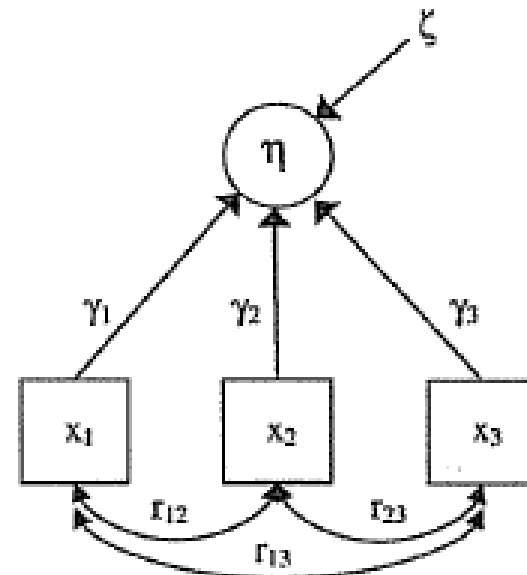
- The *Standard* model
- Reflective/Formative measurement
- Nesting
- !!??!!

Reflective and Formative Models

Panel 1: Reflective measurement model



Panel 2: Formative measurement model



Nestedness of QLQ-C30 Factor Models

- Nested models:
 - one model adds restrictions to another –more general- model
- Useful for directly comparing the fit of two models
- In the present case, there are two lines of nested models, which only be directly compared within their own line

RESULTS

Chi-squared Difference Testing between various (nested) models (branch 2)

Model	$\Delta \chi^2$ wrt Model 7a)	df	$\Delta \chi^2$ wrt Model 7b)	df
1) Bi-factor	316	16	341	16
2) Standard (13 Latents) & QL	296	14	343	15
7a) formative symptom Burden, Function, & QL (free)	n.a.		189	5
7b) formative symptom Burden, Function & QL (fixed)			n.a.	

RESULTS

Cross-validation

- Cross-validation of Mental/Physical model in a second, independent sub-sample (using multi-group CFA) showed no significant difference between samples.
(chi-square=4.9, df=8, p=0.77)

Discussion/Conclusion

- It is the question whether these results also obtain in “on-treatment” groups.