

HIGHER ORDER COMPONENT SCORES FOR THE QLQ-C30:

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

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

Pamplona, September 2009

Why “higher order” component 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.

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

The CCA project has gathered QLQ-C30 data from 38,000 respondents in 100+ studies from various:

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

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!)

Data

However, only those cancer patients using the QLQ-C30 version 3.0 at “pre-treatment” were utilized, 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 (test-retest correlations)
- Exclude FI scale; QL scale retained, yet not subsumed
- Treatment of missing data (pair-wise deletion)
- Ordinal items (polychoric correlations)
- Correction for clustering
- Standard QLQ-C30 model architecture

Modelling

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

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

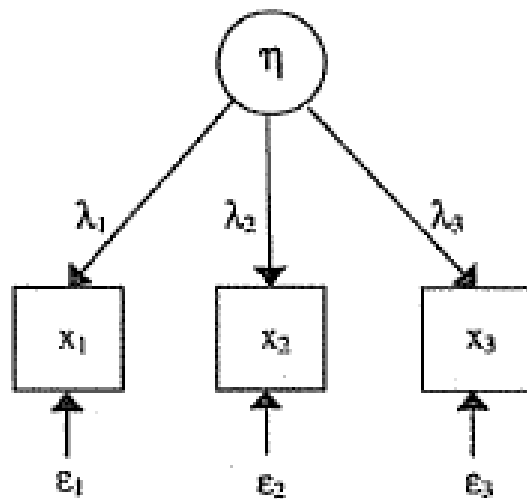
i17

~~Fi~~

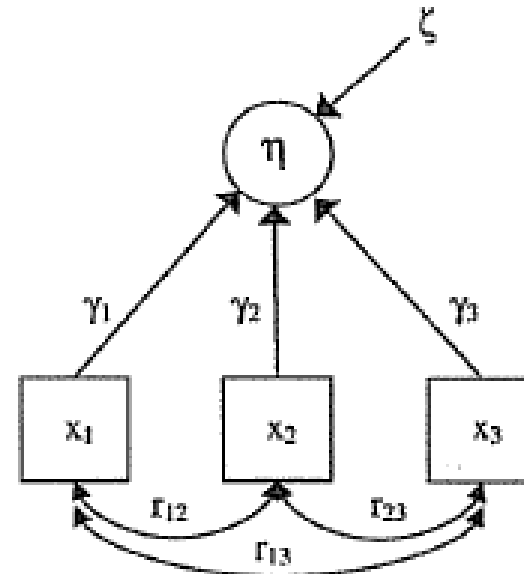
i28

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



***** 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.

QLQ-C30 Factor Models (1)

(reflective measurement)

- 1) *Bi-Factor*
- 2) *Standard QLQ-C30*
- 3) *2D PhysicalHealth-MentalHealth*
- 4) *2D PhysicalBurden/MentalFunction (v.2)*
- 5) *2D Symptom Burden & Function*
- 6) *1D 2nd order HRQL*

QLQ-C30 Factor Models (2)

Hybrid Reflective/Formative measurement

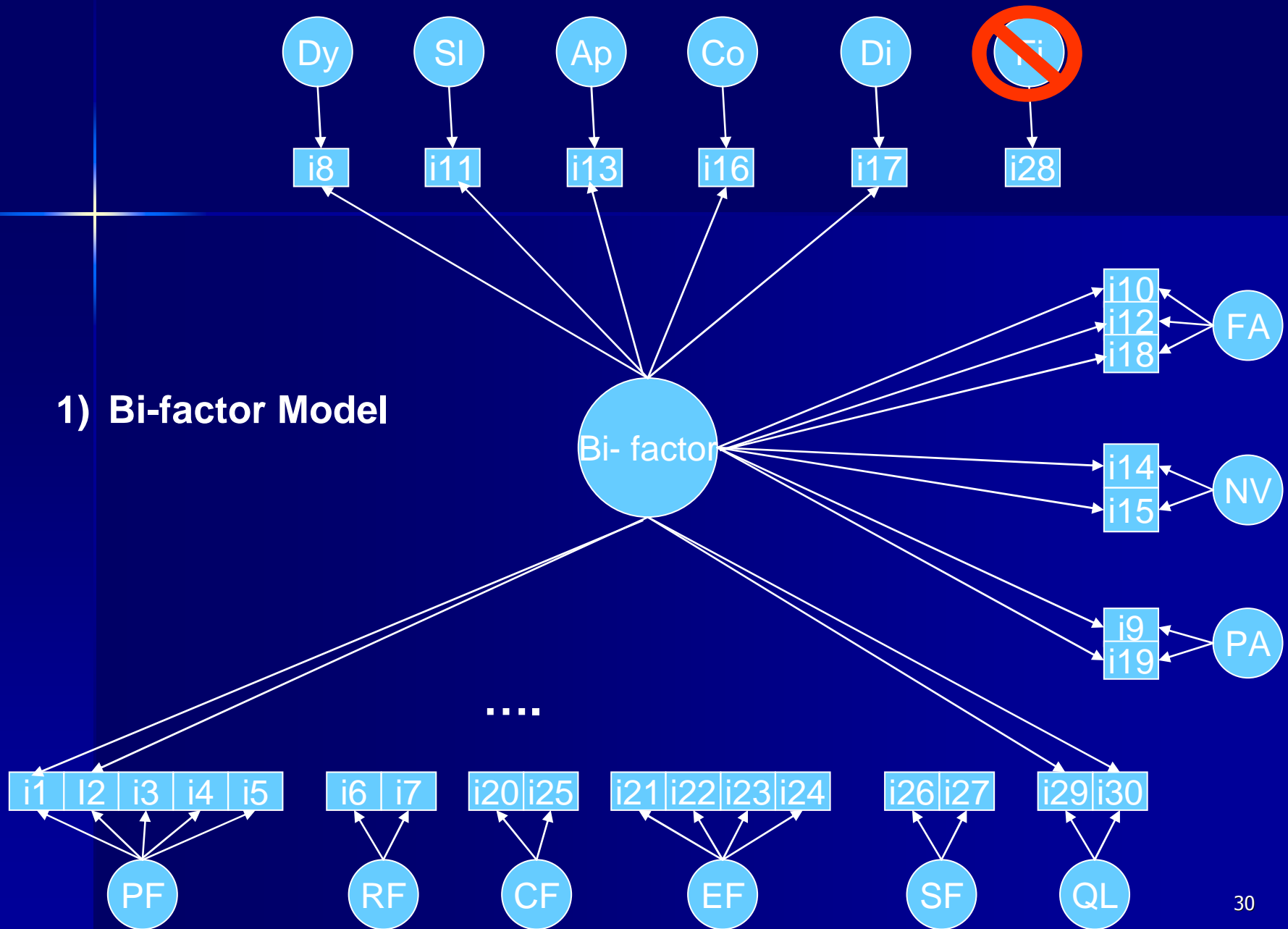
7a) *2D Symptom & Function*

(formative symptoms with free weights)

7b) *2D Symptom & Function*

(formative symptoms with fixed weights)

1) Bi-factor Model



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

2) "Standard" Model

Dy

i8

Sl

i11

Ap

i13

Co

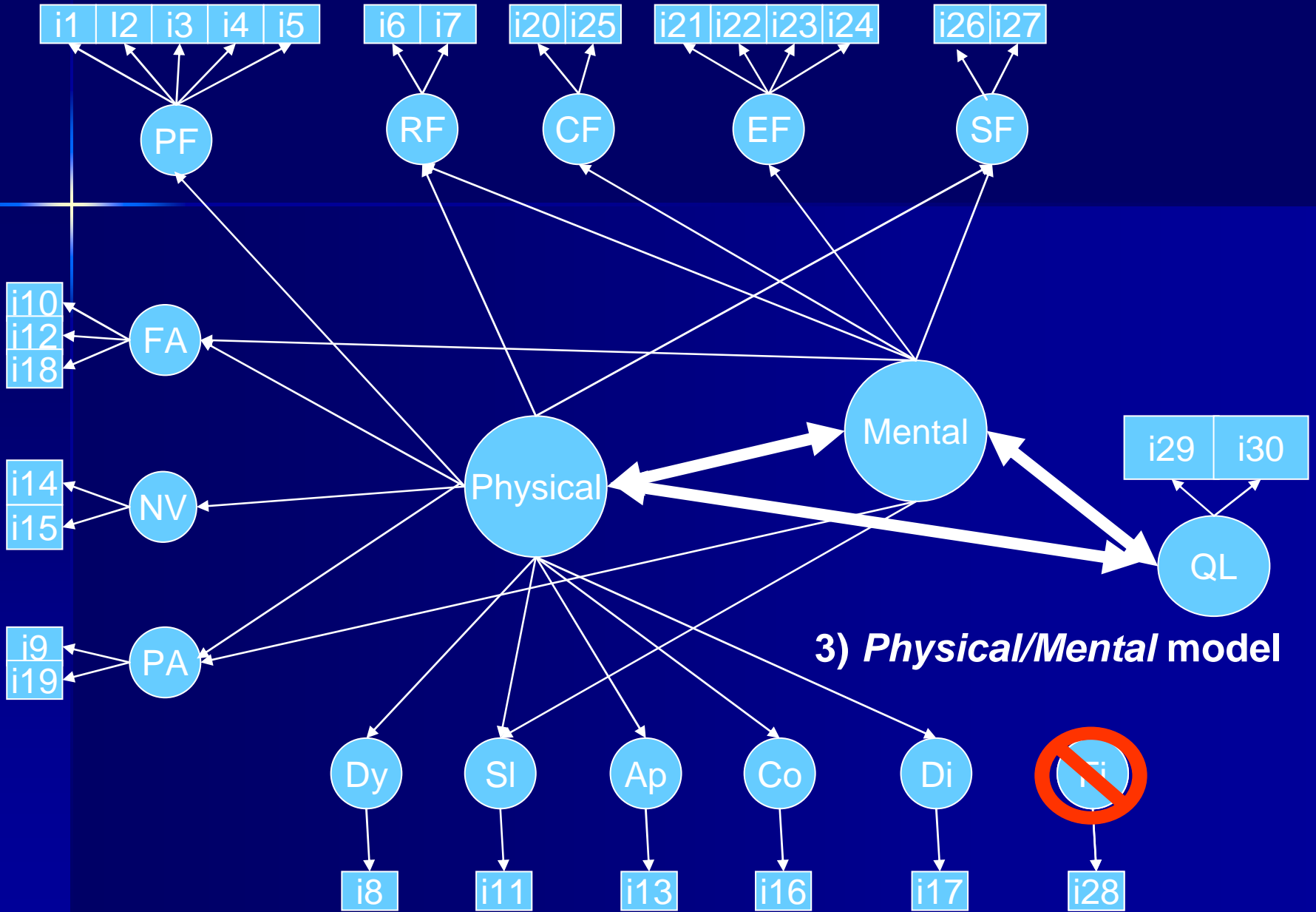
i16

Di

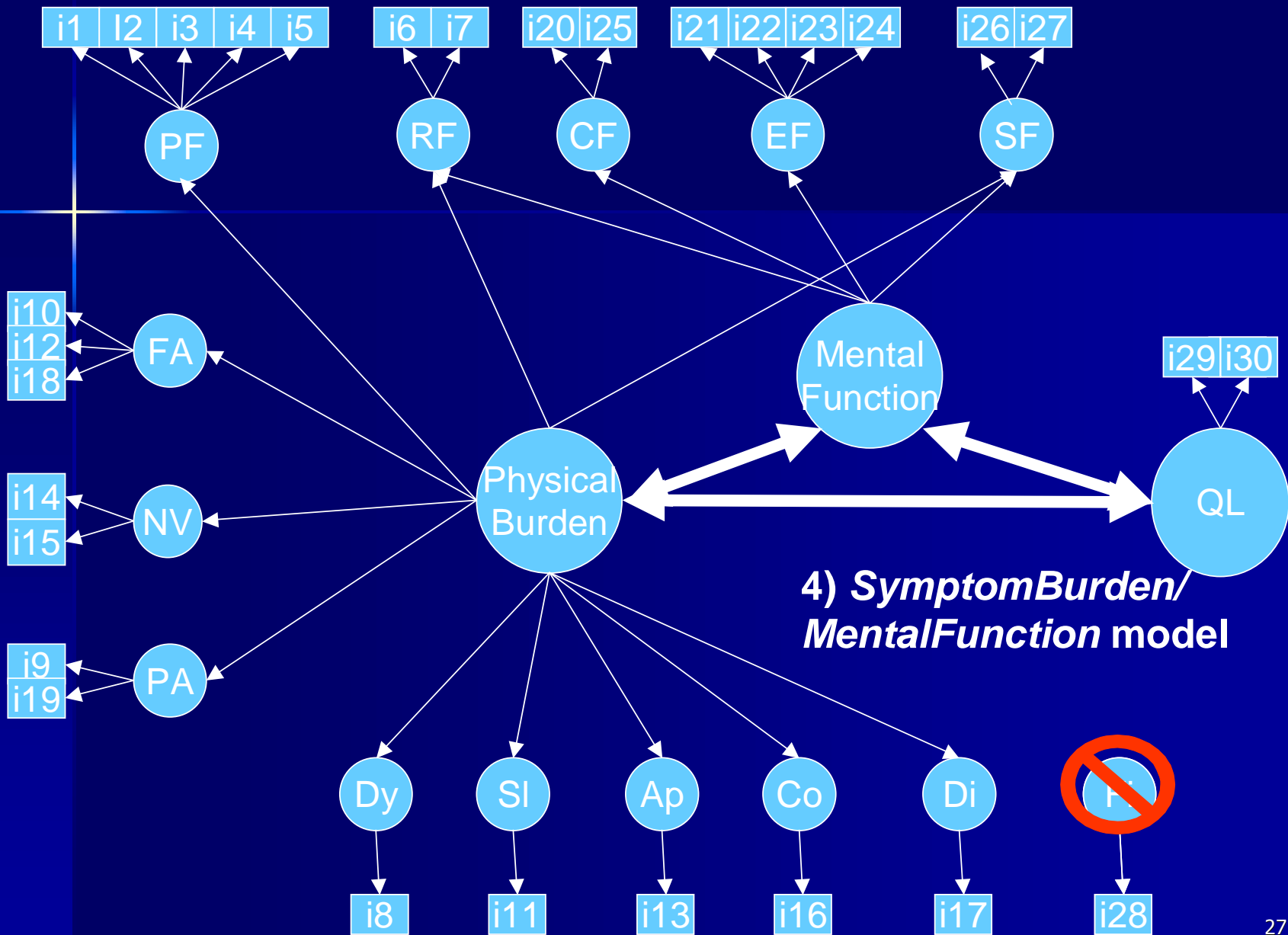
i17

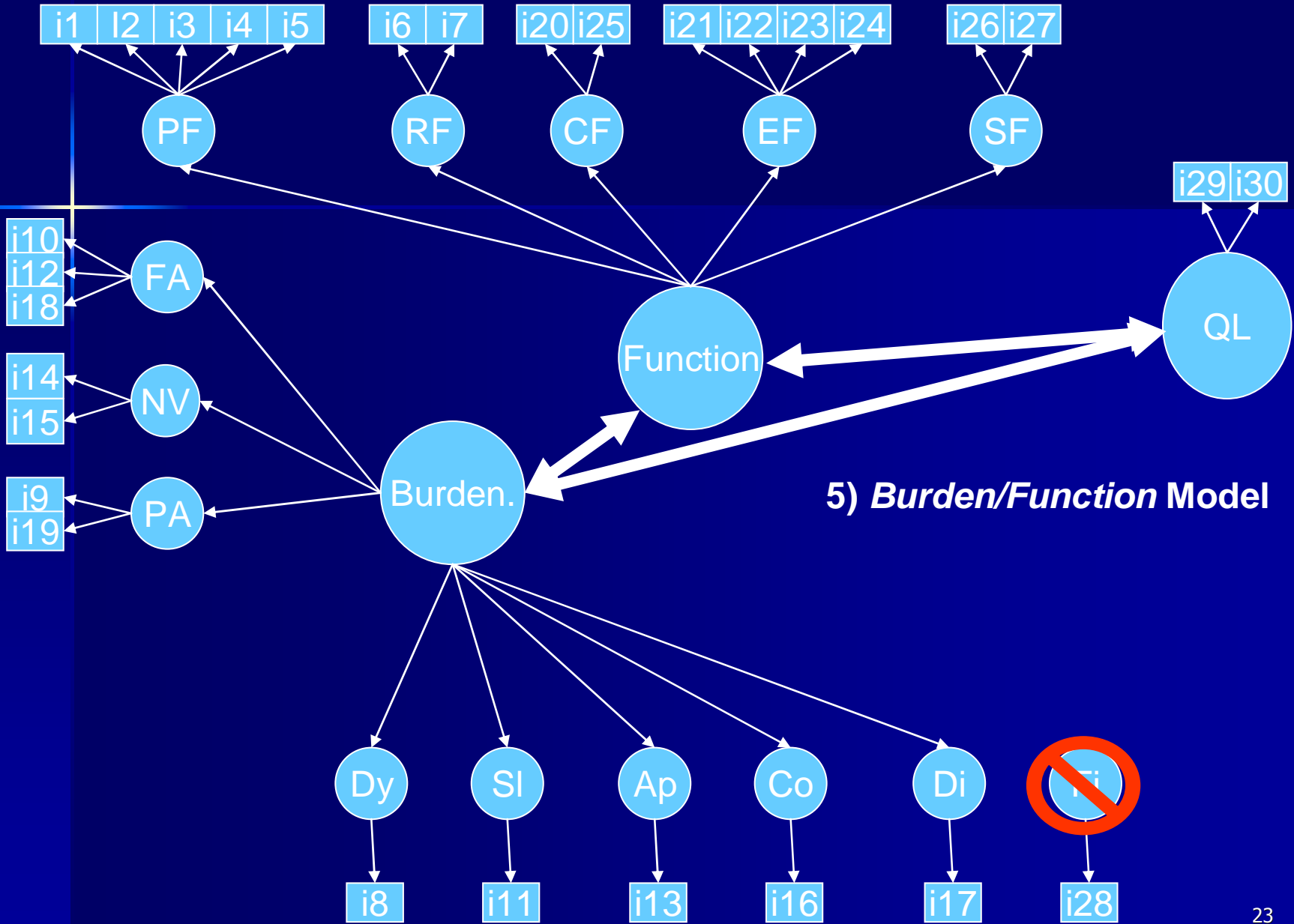
~~Fi~~

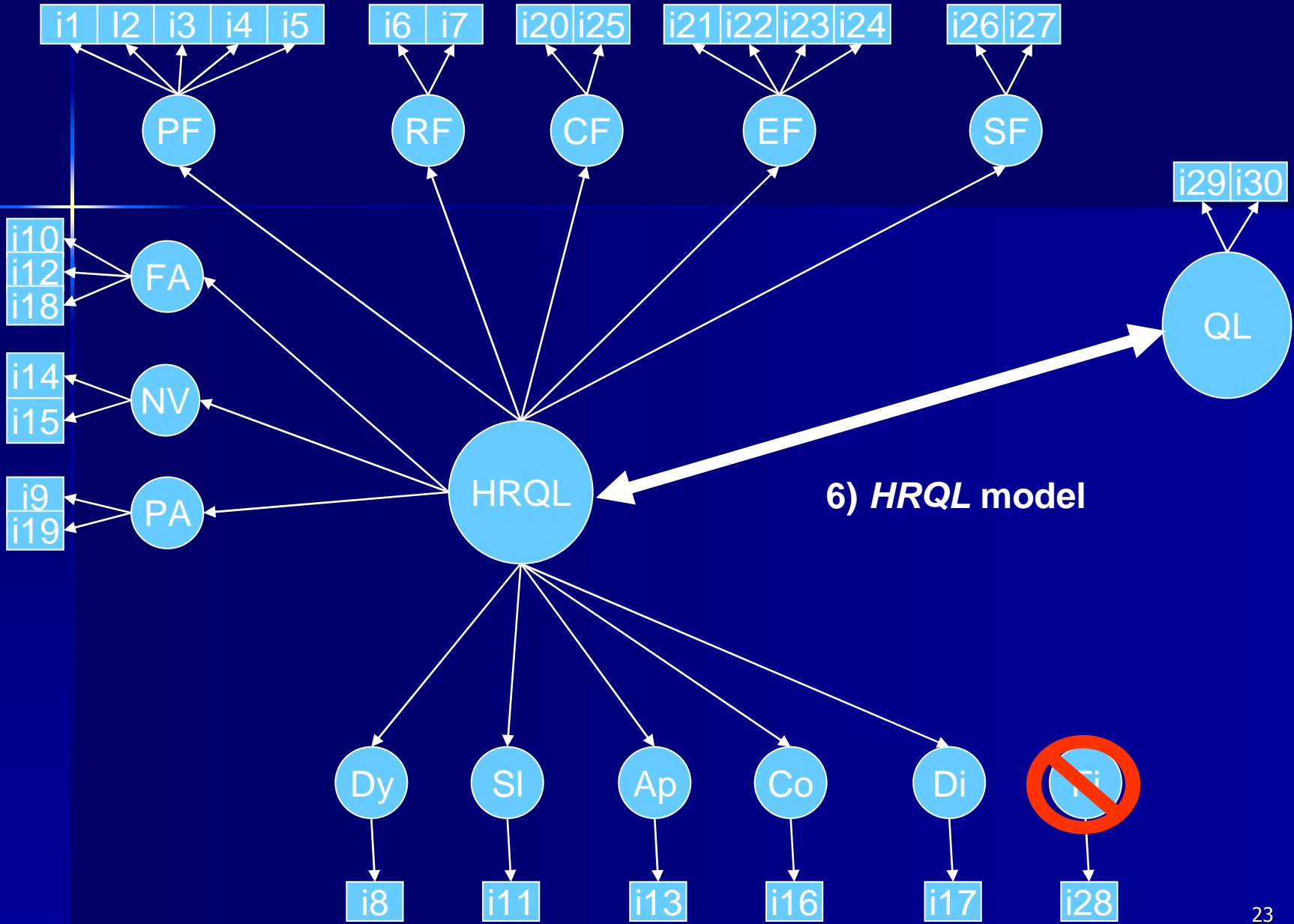
i28

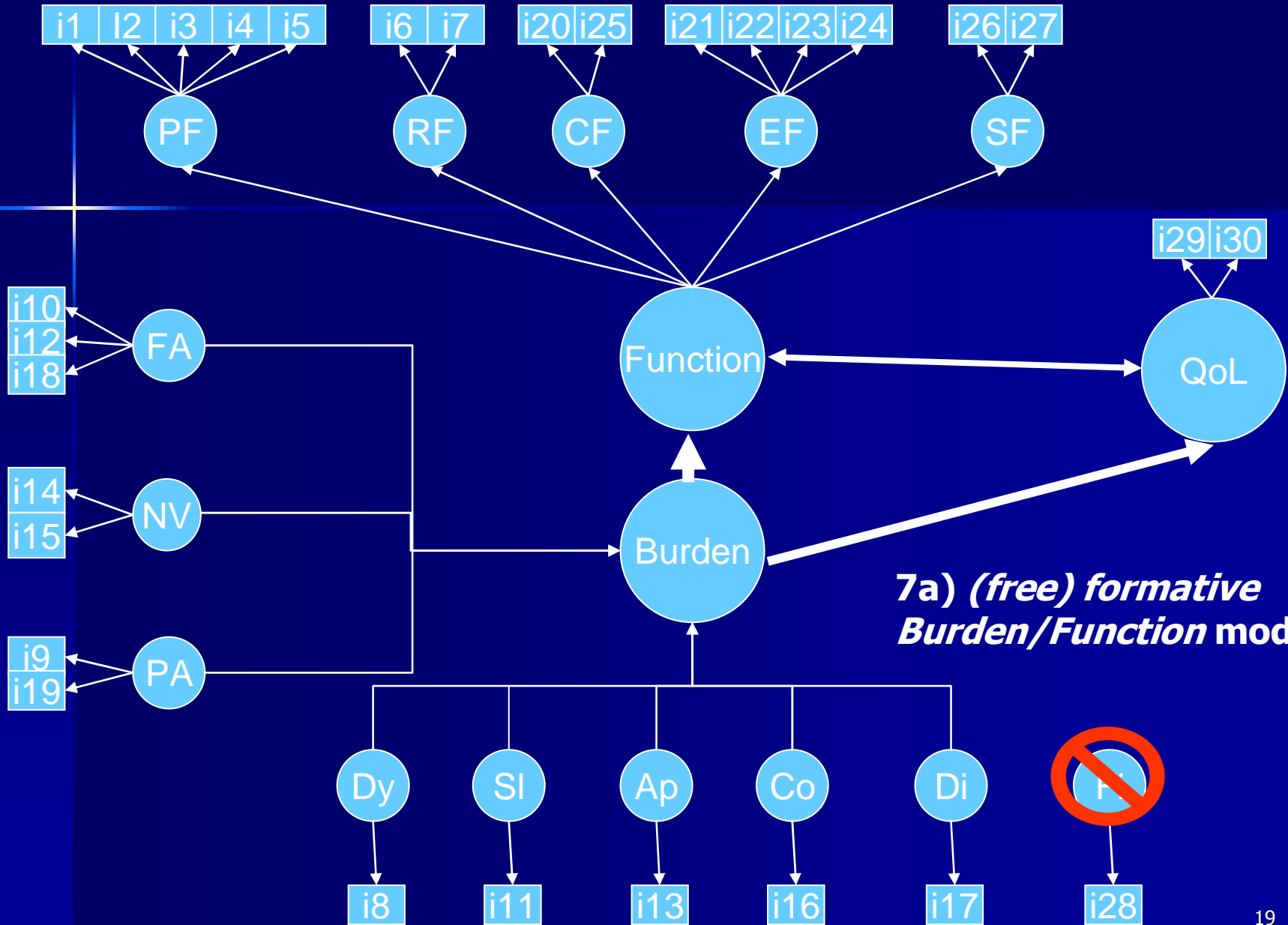


3) *Physical/Mental* model

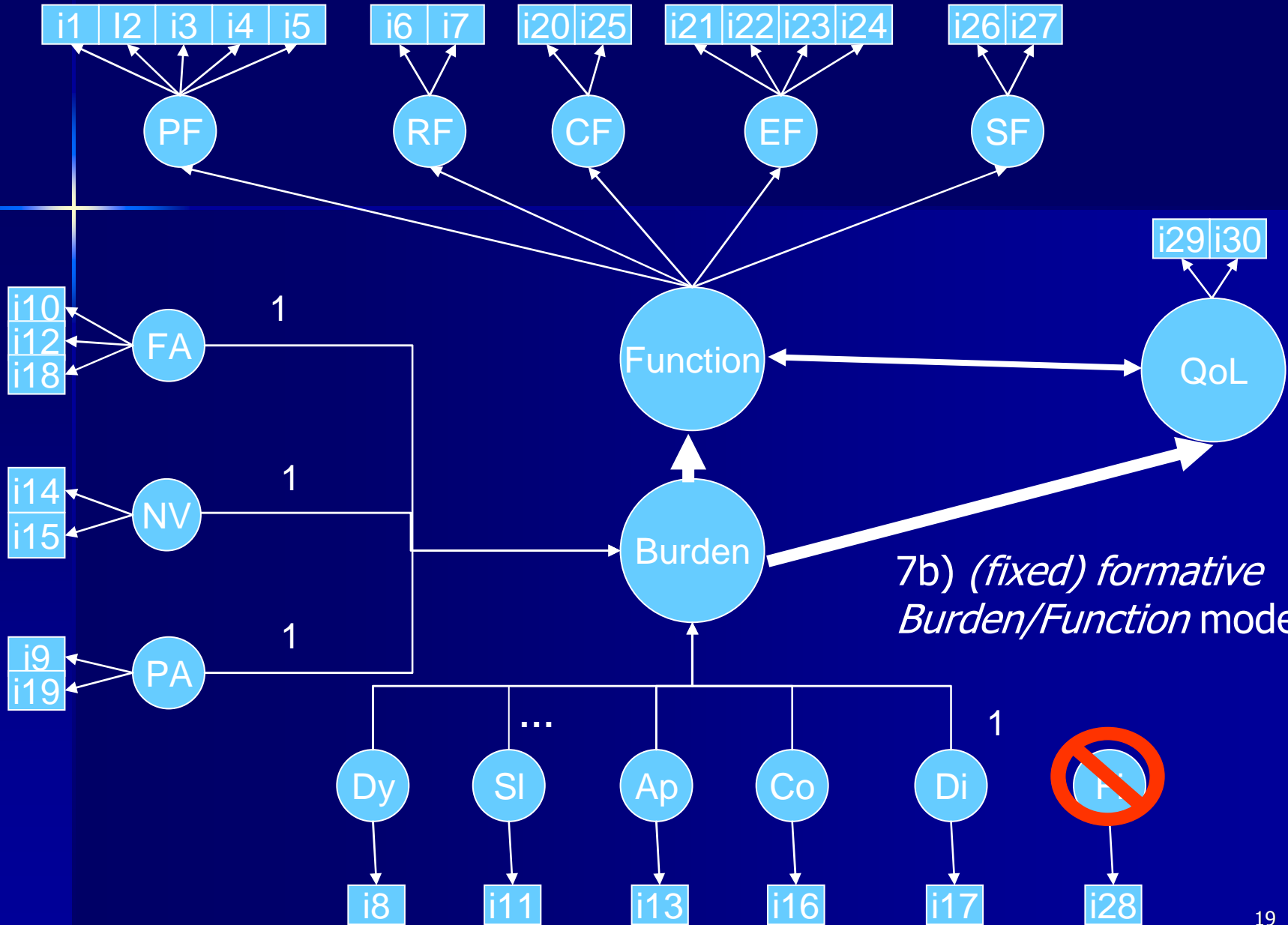








**7a) (free) formative
Burden/Function model**



Two branches of nested models

Bi-factor
Standard

/ \

Ment/Phys

frform BurFunc

Ment/Phys2

fixformBurFunc

BurFunc

1dim HRQL

RESULTS

Fit Indices/Tests for Various Models

Model	χ^2	df	CFI/TLI	RMSEA
1) Bi-factor	68	19	0.99/0.99	0.031
2) Standard Model	114	19	0.97/0.99	0.043
3) PhysicalHealth, MentalHealth	184	22	0.95/0.98	0.052
4) PhysicalBurden,MentalFunction (v.2)	191	20	0.95/0.98	0.056
5) Burden, Function	232	20	0.93/0.97	0.063
6) HRQL	235	20	0.93/0.97	0.063
7a) (free) formative Burden & Function	231	20	0.93/0.97	0.063
7b) (fixed) formative Burden &Function	253	20	0.93/0.97	0.066

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

Std. Regression Weights for best 2 models in each branch

	Physical	Mental	(free) formative Burden	Function
PF	.81#			.77#
RF	.82*	.12*		.89*
EF		.75#		.64*
CF		.90*		.79*
SF	.38*	.51*		.82*
FA	.81*	.21*	.86#	
NV	.66*		.050	
PA	.59*	.25*	.13*	
DY	.81*		.040	
SL	<.01	.80*	.07*	
AP	.85*		-.090	
CO	.74*		.020	
DI	.62*		-.020	

*p<0.05 # unstd weight fixed at 1.0

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/Conclusions

- The *Bi-factor* model and the *Standard* model both have a “good” fit to the data. All other (more restricted) models studied have an “adequate” fit.
- There is room for improvement, even for the *Standard* and *Bi-Factor* models.

Discussion/Conclusions

- The *Mental/Physical* model seems to be the best simplification of the QLQ-C30 for the reflective branch of the models studied here.
- This model is an “adequate” fit to the data, yet this comes at a cost: a significant deterioration in fit w.r.t. the *Standard* model.

Discussion/Conclusions

- The (free weight) *formative Burden-Function* model is the best performing model of the second (formative) branch.
- It is also an "adequate" fit to the data, yet also represents a significant deterioration w.r.t. the *Standard* model.

Discussion/Conclusion

- Strikingly, this last model ignores all of the symptoms, with the exception of Fatigue. Such a simplification may be viewed as being overly zealous, and tends to limit the usefulness of the model (and this branch).

Discussion / Conclusion

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

Next steps

- measurement equivalence over time
- measurement equivalence over groups
- predictive validity (in longitudinal trials)
(4 reactions! Other reactions are welcome!)
- publication of algorithm

Thank you for your attention!

Suggestions?

Remarks?