2016 Brazilian Stata Users Group meeting
São Paulo | December 2

SEM in health sciences:

combining multiscale questionnaires and checking the performance of GOF parameters up to the minimally acceptable sample size

Marcos Antonio Almeida Santos MD, PhD

Tenured Professor at Universidade Tiradentes (UNIT) – Brazil
General physician and cardiologist at Clínica & Hospital São Lucas – Aracaju (SE)
Senior Teaching Assistant in PPCR Course at Harvard T.H.Chan School of Public Health – USA

Disclosures

- Marcos Antonio Almeida Santos has no relevant conflict of interest related to the content of this presentation;

- The views expressed in this presentation do not necessarily reflect the views of the institutions.

Introduction – 1

- In health sciences, relevant issues are handled with complex questionnaires;
- These questionnaires oftentimes present dozens of indicators under Likert scales;
- However, Likert scales can be challenging to curb with an overarching "regression" approach;
- What is more, ordinal in principle, they usually present a skewed distribution, which may remain after algebraic transformation in 20-point or 100-point scales.

Introduction – 2

- The panoply of scales leads to a plethora of criteria of normality;
- To approach several questionnaires at once and, at the same time, to provide reliable measures of association among them, the analysis may rely on the standardization of the coefficients;
- We present a strategy to work with complex stress and QOL questionnaires assembled into an overarching model.

Case-study “situation” – 1

- **Questionnaire WHOQOL-BREF:**

- Quality of life – Developed by the WHO (1996);
- Number of questions: 26;
- Likert scale: scores from 1 to 5: (1 = not at all; 2 = not much; 3 = moderately; 4 = a great deal; 5= completely).
- Negatively phrased items (3): Q3, Q4 and Q26;
- Four Domains + Self-appraisal:
- Physical = mean (Q3r, Q4r, Q10, Q15, Q16, Q17, Q18);
- Psychological = mean (Q5,Q6,Q7,Q11,Q19,Q26r);
- Social relationships = mean(Q20,Q21,Q22);
- Environment = mean (Q8,Q9,Q12,Q13,Q14,Q23,Q24,Q25);
- Self-appraisal = mean (Q1,Q2).
- Scores lately *4 (range: 4-20) or a scale 0-100.

Case-study "situation" – 2

- **Questionnaire ISSL:**

- Inventory of Symptoms of Stress – Lipp
- Number of questions: 53;
- Binary variables (0 or 1);
- Physical = 34; psychological 19;
- Results used as: # positive questions;
- Three Domains:
- Alertness (15 Qs): range 0–15; >3 = yes;
- Resistance+near exhaustion (15 Qs): range 0–15; >6 = yes;
- Exhaustion (23 Qs): range 0–23; >8 = yes.

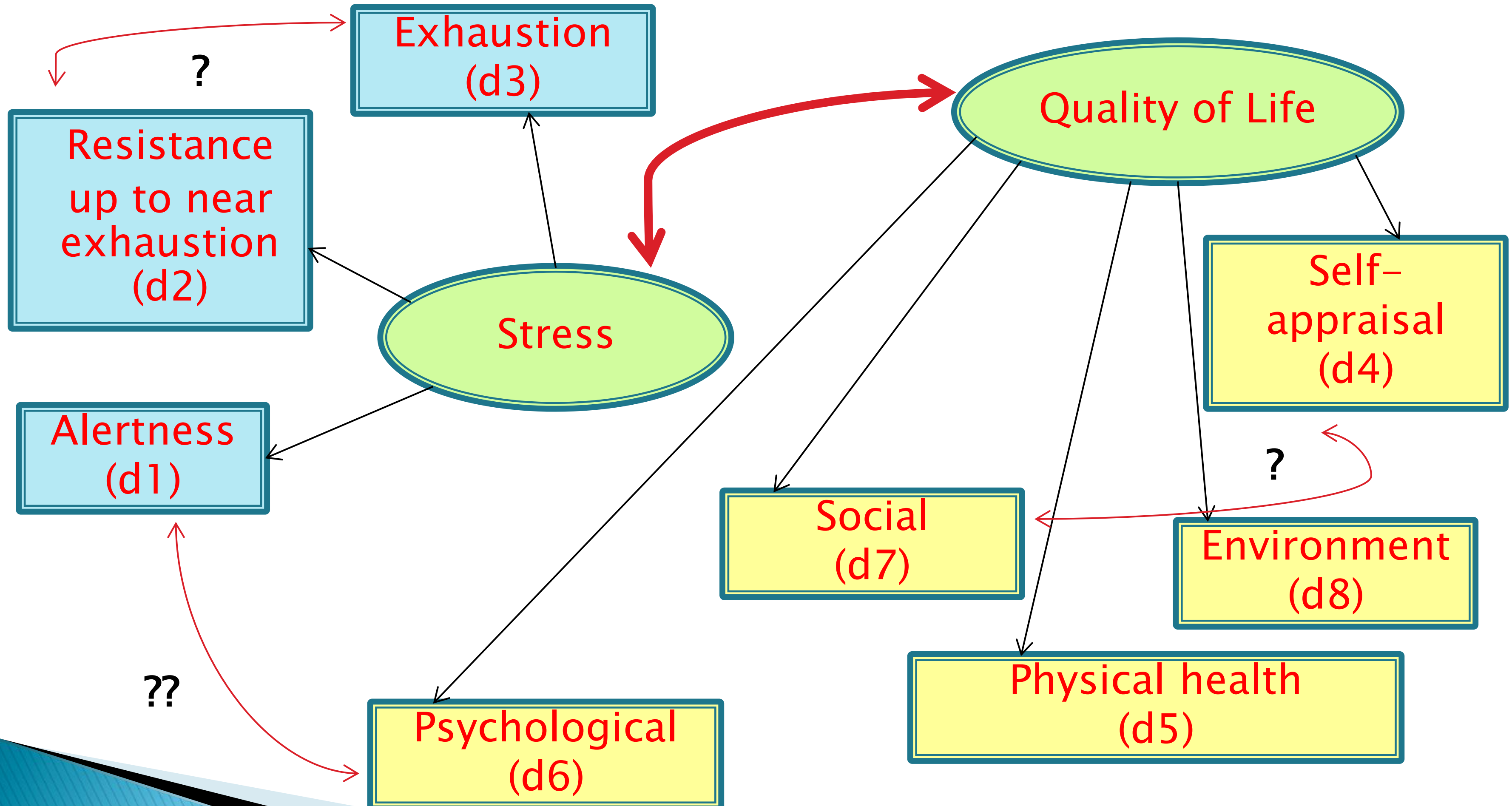
First things, first

- **WHOQOL-BREF: QOL**
- 26 Qs;
- Likert scale (1-5) turned into a 4-20 range;
- Negatively phrased Qs recoded.
- Scale 4-20 selected.
- Parceling in five "independent" domains;*
- But...we aggregate the analysis leaving each domain as an "endogenous variable" associated with the latent variable QOL.

- **ISSL: STRESS**
- 53 Qs;
- Dichotomous variables
- Sum of + answers;
- Scale of similar range;
- Parceling in three domains;*
- But... instead of categorizing QOL according to scores from each domain (binary "yes-no" or prevalent domain), we leave the domains as "reflective indicators" associated with Stress as a latent variable.

*Up to this point, following guidelines of each questionnaire.

Building a graphical scheme



Model design:

- CFA under SEM;
- Two latent variables were created as reflective “exogenous” factors: QOL and stress;
- Parceling: questions from the respective questionnaires were used to create an “aggregate” arrangement, according to the specifications;
- Selection of scales of similar range;
- Thence, the number of loadings was decreased by parceling items by similarity and treating these parceled constructs as “endogenous” variables.

Steps of the analysis

- 1. Parceling, checking severe departs from normality, selecting estimation method (ML);
- 2. Avoiding identification issues: ideally, at least 3 parcelled endogenous variables for each latent one;
- 3. Modeling "full" data (around 600 individuals):
  - a) From a simple model up to a more complex one;
  - b) Checking GOF parameters up to the "best fit";
  - c) Adding variance-covariance terms according to the rationale as well as the modification indices and convergence isssues.
- 4. Re-starting with random sub-samples: checking model's reliability as well as performance of GOF parameters under progressively smaller sample sizes.

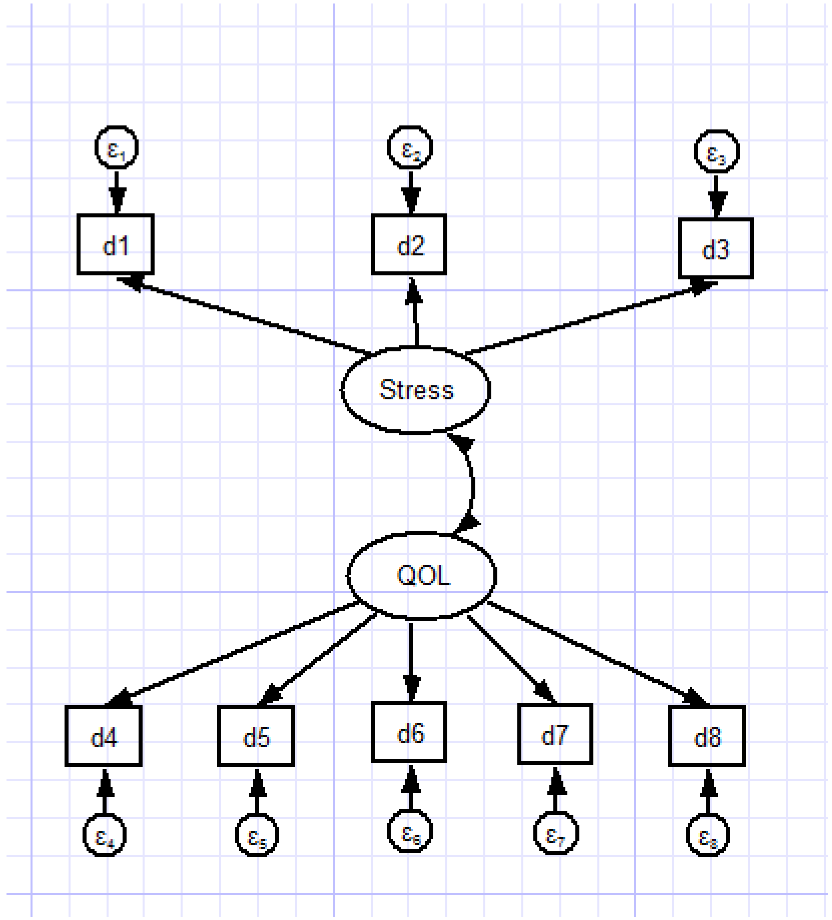
Summary statistics data: ssd

- Immediate set of commands that creates a "compact data set":
- Allows Stata users to reproduce original data;
- Data shared between statisticians or sent to reviewers (since it preserves confidentiality);
- May be applied in the modeling strategy;
- Used to perform GOF tests, etc.
- Warning: it applies to *sem*, but not *gsem*.

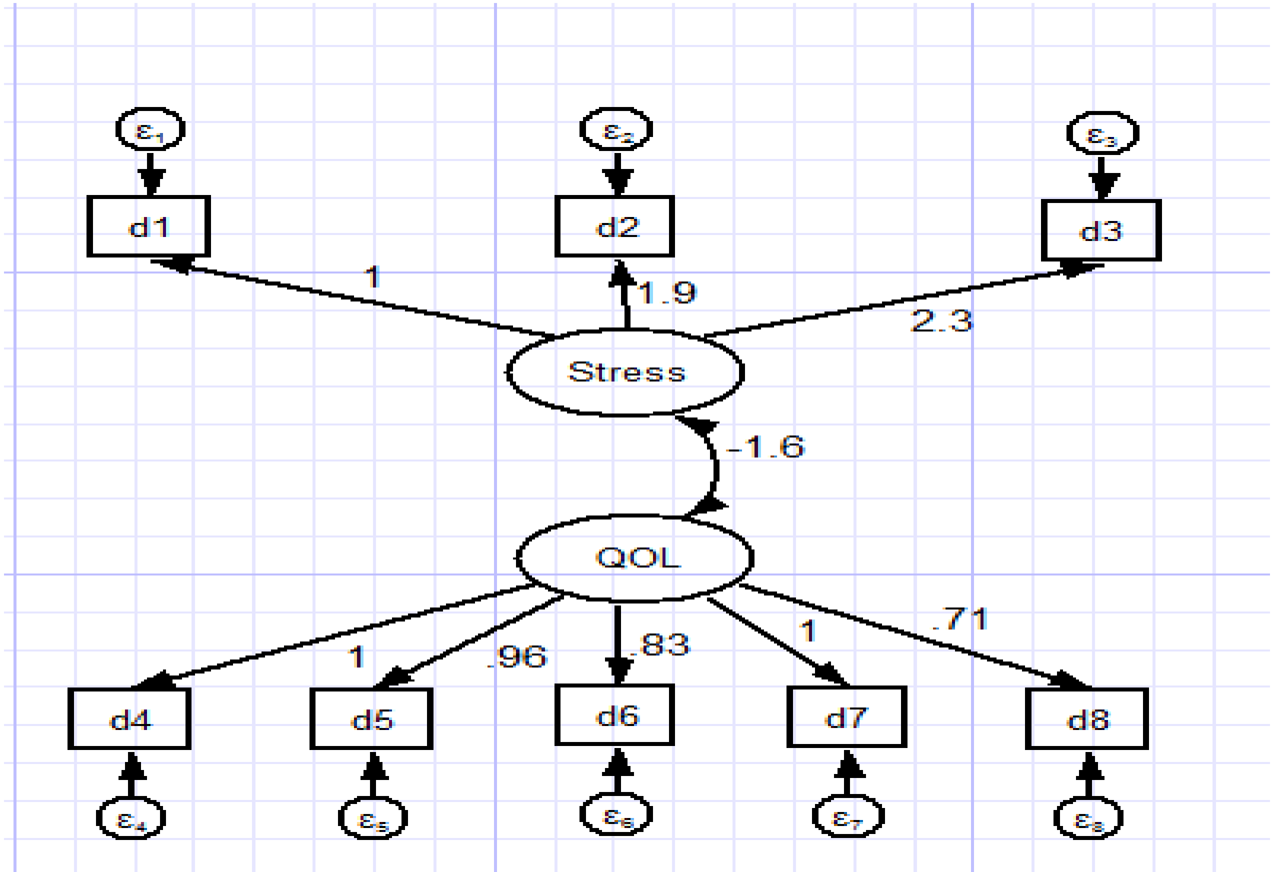
Reproducing data: ssd

```
. ssd init d1 d2 d3 d4 d5 d6 d7 d8
. ssd set observations 597
. ssd set means  2.963149 4.396985 4.574539 14.47236 14.2846 ///
13.75366 14.64992  11.93786
. ssd set sd 2.120208 2.820382  3.512665 2.733951 2.422642 ///
2.813333 3.234396 2.25064
. ssd set correlations 1.0 \ ///
0.5965   1.0000\ ///
0.5870   0.8156   1.0000\ ///
-0.2583  -0.4770  -0.4415   1.0000\ ///
-0.2184  -0.4368  -0.4971   0.5983   1.0000\ ///
-0.0994  -0.2326  -0.2406   0.4364   0.5241   1.0000\ ///
-0.1015  -0.2528  -0.2354   0.4823   0.5033   0.4730   1.0000\ ///
-0.2141  -0.3555  -0.3299   0.4878   0.5233   0.3288   0.4641   1.0000
```

Basic graph: sembuilder



Unstandardized results




```
. sem (Stress -> d1, ) (Stress -> d2, ) (Stress -> d3, ) (QOL -> d4, ) (QOL
-> d5, ) (QOL -> d6, ) (QOL -> d7, ) (QOL -> d8, ), covstruct(_lexogenous,
diagonal) vce(oim) latent(Stress QOL ) cov( Stress*QOL) nocapslatent
```

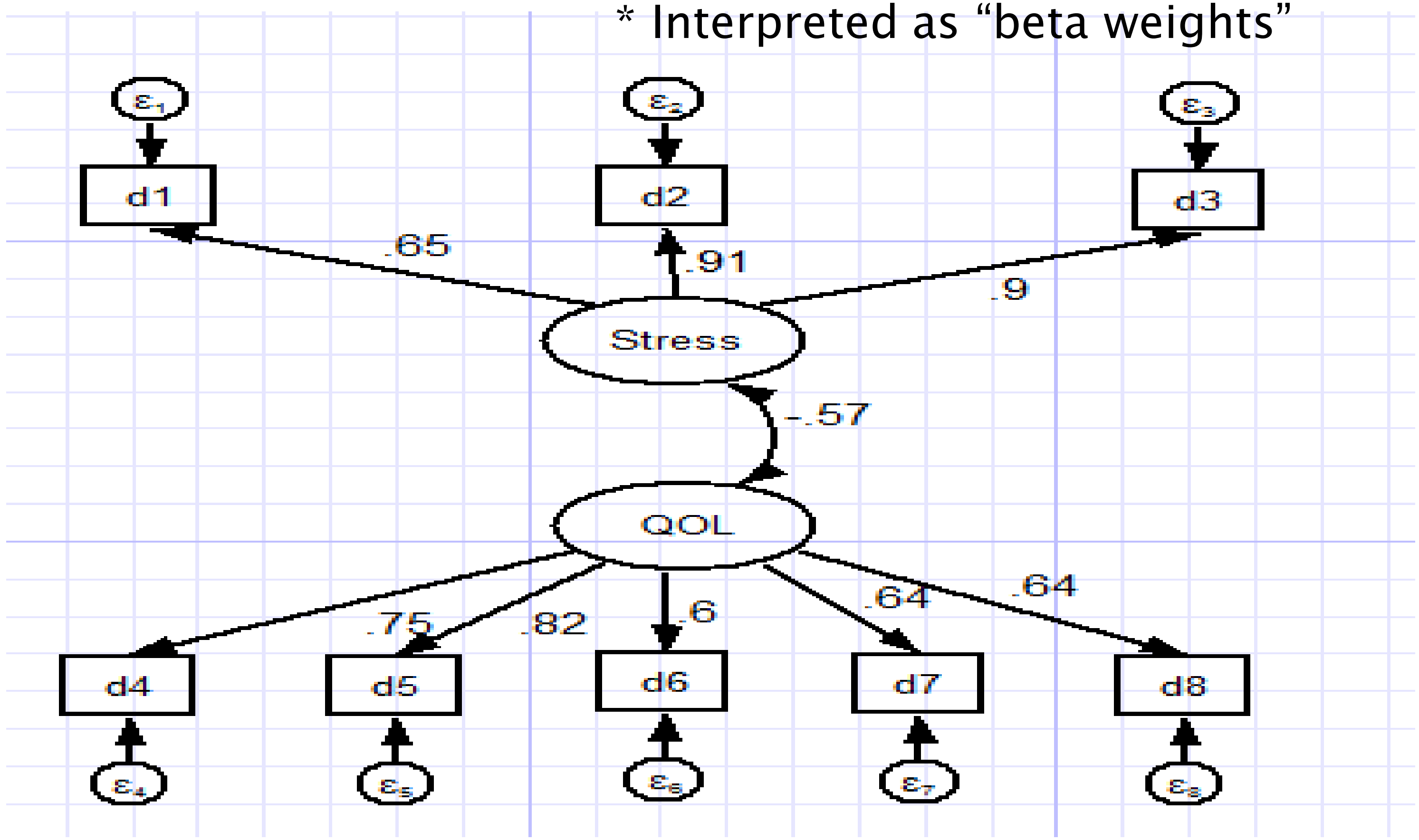
Output

	Coef.	OIM Std. Err.	z	P>|z|	[95% Conf.	Interval]
Measurement						
d1 <-						
Stress	1	(constrained)				
_cons	2.963149	.0867014	34.18	0.000	2.793217	3.133081
d2 <-						
Stress	1.866597	.1053646	17.72	0.000	1.660086	2.073108
_cons	4.396985	.1153336	38.12	0.000	4.170935	4.623035
d3 <-						
Stress	2.312083	.1317451	17.55	0.000	2.053867	2.570299
_cons	4.574539	.143643	31.85	0.000	4.293004	4.856074
d4 <-						
QOL	1	(constrained)				
_cons	14.47236	.1117993	129.45	0.000	14.25324	14.69148
d5 <-						
QOL	.9633735	.0538073	17.90	0.000	.8579131	1.068834
_cons	14.2846	.099069	144.19	0.000	14.09043	14.47877
d6 <-						
QOL	.82803	.0607798	13.62	0.000	.7089038	.9471563
_cons	13.75366	.1150456	119.55	0.000	13.52817	13.97915
d7 <-						
QOL	1.012757	.0692444	14.63	0.000	.877041	1.148474
_cons	4.64992	.1322641	35.16	0.000	4.390687	4.909153
d8 <-						
QOL	.7061149	.0479517	14.73	0.000	.6121313	.8000985
_cons	11.93786	.0920354	129.71	0.000	11.75747	12.11825
var(e.d1)	2.611638	.1629751			2.310975	2.951419
var(e.d2)	1.404555	.2112667			1.045935	1.886134
var(e.d3)	2.289031	.3281658			1.728301	3.031682
var(e.d4)	3.250656	.2499976			2.795811	3.779499
var(e.d5)	1.950899	.1805487			1.627268	2.338893
var(e.d6)	5.014172	.3244824			4.416877	5.692238
var(e.d7)	6.124352	.4092317			5.372575	6.981324
var(e.d8)	2.957146	.1965824			2.595897	3.368668
var(Stress)	1.87609	.2198389			1.491112	2.360461
var(QOL)	4.211301	.4221773			3.460066	5.12564
cov(Stress,QOL)	-1.613276	.1782733	-9.05	0.000	-1.962685	-1.263866

LR test of model vs. saturated: chi2(19) = 108.07, Prob > chi2 = 0.0000

Standardized results*

* Interpreted as "beta weights"




```
. sem (Stress -> d1, ) (Stress -> d2, ) (Stress -> d3, ) (QOL -> d4, ) (QOL ->
d5, ) (QOL -> d6, ) (QOL -> d7, ) (QOL -> d8, ), covstruct(_lexogenous, diagonal)
vce(oim) standardized latent(Stress QOL ) cov( Stress*QOL) nocapslatent
```

Output

Standardized	Coef.	OIM Std. Err.	z	P>|z|	[95% Conf.	Interval]
Measurement						
d1 <-						
Stress	.646567	.026028	24.84	0.000	.595553	.697581
_cons	1.39875	.0575639	24.30	0.000	1.285927	1.511574
d2 <-						
Stress	.9072655	.0153471	59.12	0.000	.8771858	.9373451
_cons	1.560314	.0609429	25.60	0.000	1.440868	1.67976
d3 <-						
Stress	.9023155	.0155001	58.21	0.000	.8719359	.9326952
_cons	1.303394	.0556583	23.42	0.000	1.194305	1.412482
d4 <-						
QOL	.7512453	.0230969	32.53	0.000	.7059762	.7965144
_cons	5.298013	.1586922	33.39	0.000	4.986982	5.609044
d5 <-						
QOL	.8167286	.0200589	40.72	0.000	.777414	.8560433
_cons	5.901237	.175617	33.60	0.000	5.557034	6.24544
d6 <-						
QOL	.6045013	.0300883	20.09	0.000	.5455293	.6634732
_cons	4.892843	.1473947	33.20	0.000	4.603955	5.181731
d7 <-						
QOL	.6431089	.028421	22.63	0.000	.5874048	.698813
_cons	1.438853	.0583863	24.64	0.000	1.324418	1.553288
d8 <-						
QOL	.6443793	.0281589	22.88	0.000	.589189	.6995697
_cons	5.308656	.1589902	33.39	0.000	4.997041	5.620271
var(e.d1)	.5819511	.0336577			.5195848	.6518034
var(e.d2)	.1768694	.0278477			.1299071	.2408089
var(e.d3)	.1858267	.0279719			.13835	.2495956
var(e.d4)	.4356305	.0347029			.372658	.5092442
var(e.d5)	.3329543	.0327653			.274549	.4037843
var(e.d6)	.6345782	.0363768			.5671404	.710035
var(e.d7)	.5864109	.0365556			.5189674	.6626193
var(e.d8)	.5847753	.03629			.5178037	.6604088
var(Stress)	1	.			.	.
var(QOL)	1	.			.	.
cov(Stress,QOL)	-.5739494	.0338289	-16.97	0.000	-.6402528	-.5076461

LR test of model vs. saturated: chi2(19) = 108.07, Prob > chi2 = 0.0000

.

Fit parameters

- **Chi-square test**: null hypothesis = accept the model (covariances between the matrix and the predicted model do not differ). There is no difference between the model and a saturated model. Check p-values and dfs;
- **RMSEA** :Steiger-Lind Root Mean Square Error of Approximation;
- **CFI** :Bentler Comparative Fit Index;
- **SRMR** :Standardized Root Mean Square Residual.

GOF tests

"Ideal" values
Chi2 >0.05
RMSEA <0.05 Upper <0.10
CFI>=0.95
SRMR <= 0.10

```
. estat gof, stats(all)

----------------------------------------------------------------------------
Fit statistic        |      Value   Description
---------------------+------------------------------------------------------
Likelihood ratio     |
         chi2_ms(19) |    108.071   model vs. saturated
            p > chi2 |      0.000
         chi2_bs(28) |   2190.168   baseline vs. saturated
            p > chi2 |      0.000
---------------------+------------------------------------------------------
Population error     |
               RMSEA |      0.089   Root mean squared error of approximation
 90% CI, lower bound |      0.073
         upper bound |      0.105
              pclose |      0.000   Probability RMSEA <= 0.05
---------------------+------------------------------------------------------
Information criteria |
                 AIC |  21011.622   Akaike's information criterion
                 BIC |  21121.420   Bayesian information criterion
---------------------+------------------------------------------------------
Baseline comparison  |
                 CFI |      0.959   Comparative fit index
                 TLI |      0.939   Tucker-Lewis index
---------------------+------------------------------------------------------
Size of residuals    |
                SRMR |      0.048   Standardized root mean squared residual
                  CD |      0.980   Coefficient of determination
```

Improving it: modification indices

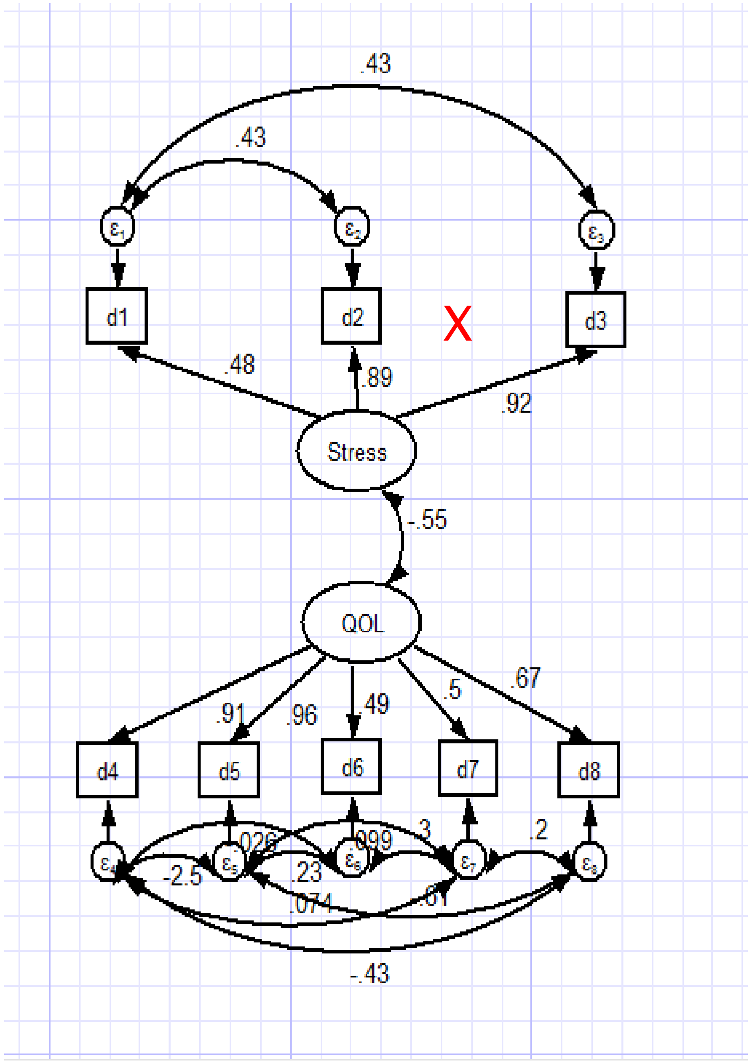
Modification indices

```
. estat mindices
```

	MI	df	P>MI	EPC	Standard EPC
Measurement					
d1 <- QOL	14.122	1	0.00	.1784482	.172865
d4 <- Stress	13.027	1	0.00	-.3172042	-.1590521
d5 <- Stress	5.053	1	0.02	-.1723631	-.0975318
d6 <- Stress	13.264	1	0.00	.3566978	.1738082
d7 <- Stress	19.616	1	0.00	.487967	.2068179
cov(e.d1,e.d5)	6.921	1	0.01	.3005715	.1331601
cov(e.d2,e.d3)	14.123	1	0.00	-2.529193	-1.410545
cov(e.d2,e.d4)	13.624	1	0.00	-.4787569	-.2240579
cov(e.d2,e.d5)	8.795	1	0.00	.3232729	.1952912
cov(e.d3,e.d5)	32.709	1	0.00	-.7787271	-.3685039
cov(e.d3,e.d7)	5.104	1	0.02	.473602	.1264904
cov(e.d5,e.d6)	5.640	1	0.02	.4469742	.142911
cov(e.d6,e.d7)	15.423	1	0.00	1.036673	.1870734
cov(e.d6,e.d8)	8.050	1	0.00	-.5208074	-.1352512
cov(e.d7,e.d8)	6.094	1	0.01	.5118722	.1202805

EPC = expected parameter change

```
. sem (Stress -> d1, ) (Stress -> d2, ) (Stress -> d3, ) (QOL -> d4, ) (QOL -> d5, ) (QOL -> d6, ) (QOL -> d7, ) (QOL ->
d8, ), covstruct(_lexogenous, diagonal) vce(oim) standardized latent(Stress QOL ) cov( Stress*QOL e.d1*e.d2 e.d1*e.d3
e.d4*e.d5 e.d4*e.d6 e.d4*e.d7 e.d4*e.d8 e.d5*e.d6 e.d5*e.d7 e.d5*e.d8 e.d6*e.d7 e.d7*e.d8) nocapslatent
```




```
.
. estat gof, stats(all)
```

Fit statistic	Value
Likelihood ratio	
chi2_ms(8)	28.947
p > chi2	0.000
chi2_bs(28)	2190.168
p > chi2	0.000
Population error	
RMSEA	0.066
90% CI, lower bound	0.041
upper bound	0.093
pclose	0.132
Information criteria	
AIC	20954.498
BIC	21112.607
Baseline comparison	
CFI	0.990
TLI	0.966
Size of residuals	
SRMR	0.014
CD	1.015

Checking where to improve it

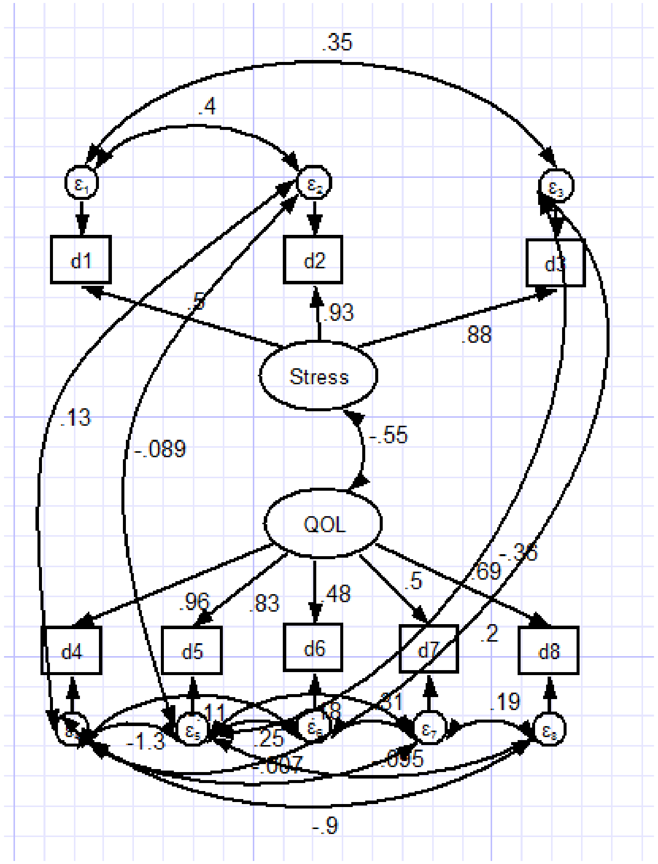
```
. estat mindices

Modification indices
```

	MI	df	P>MI	EPC	Standard EPC
cov(e.d2,e.d4)	8.460	1	0.00	-.4355507	-.3042226
cov(e.d2,e.d5)	17.523	1	0.00	.535461	.612668
cov(e.d3,e.d4)	11.403	1	0.00	.6573815	.4281969
cov(e.d3,e.d5)	22.835	1	0.00	-.7946751	-.8479304

EPC = expected parameter change

“Good fit” model

```
.sem (Stress -> d1, ) (Stress -> d2, ) (Stress -> d3, )
(QOL -> d4, ) (QOL -> d5, ) (QOL -> d6, ) (QOL -> d7, )
(QOL -> d8, ), covstruct(_lexogenous, diagonal) vce(oim)
standardized latent(Stress QOL ) cov( Stress*QOL
e.d1*e.d2 e.d1*e.d3 e.d2*e.d4 e.d2*e.d5 e.d3*e.d4
e.d3*e.d5 e.d4*e.d5 e.d4*e.d6 e.d4*e.d7 e.d4*e.d8
e.d5*e.d6 e.d5*e.d7 e.d5*e.d8 e.d6*e.d7 e.d7*e.d8)
nocapslatent
```




```
. estat mindices
(no modification indices to report, all MI values
less than 3.841458820694123)
```

“Ideal” values
Chi2 >0.05
RMSEA <0.05 Upper <0.10
CFI>=0.95
SRMR <= 0.10

```
estat gof, stats(all)
```

Fit statistic	Value
Likelihood ratio	
chi2_ms(4)	4.845
p > chi2	0.304
chi2_bs(28)	2190.168
p > chi2	0.000
Population error	
RMSEA	0.019
90% CI, lower bound	0.000
upper bound	0.067
pclose	0.818
Information criteria	
AIC	20938.396
BIC	21114.073
Baseline comparison	
CFI	1.000
TLI	0.997
Size of residuals	
SRMR	0.009
CD	1.006

Checking coeffs*, CIs and p-values

Standardized	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Measurement						
d1 <-						
Stress	.5039137	.0794151	6.35	0.000	.3482629	.6595644
_cons	1.398748	.0575641	24.30	0.000	1.285924	1.511571
d2 <-						
Stress	.9270574	.0295647	31.36	0.000	.8691116	.9850031
_cons	1.560313	.0609429	25.60	0.000	1.440867	1.679758
d3 <-						
Stress	.8797725	.0290332	30.30	0.000	.8228684	.9366766
_cons	1.303392	.0556582	23.42	0.000	1.194304	1.41248
d4 <-						
QOL	.9645465	.1600651	6.03	0.000	.6508247	1.278268
_cons	5.297027	.158697	33.38	0.000	4.985987	5.608068
d5 <-						
QOL	.8255588	.1325769	6.23	0.000	.5657129	1.085405
_cons	5.907942	.1757151	33.62	0.000	5.563547	6.252338
d6 <-						
QOL	.4798404	.0502703	9.55	0.000	.3813125	.5783683
_cons	4.892846	.1473947	33.20	0.000	4.603958	5.181734
d7 <-						
QOL	.5020461	.0663271	7.57	0.000	.3720473	.6320448
_cons	1.438854	.0583863	24.64	0.000	1.324419	1.553289
d8 <-						
QOL	.6852245	.0600071	11.42	0.000	.5676127	.8028363
_cons	5.308663	.1589902	33.39	0.000	4.997048	5.620278

* Loadings > 0.40; $p < 0.05$

From global GOF to equation level

```
. estat eqgof

Equation-level goodness of fit
```

depvars	Variance fitted	predicted	residual	R-squared	mc	mc2
observed						
d1	4.487744	1.139568	3.348176	.253929	.5039137	.253929
d2	7.94121	6.824957	1.116253	.8594353	.9270574	.8594353
d3	12.31812	9.534221	2.7839	.7739996	.8797725	.7739996
d4	7.464733	6.944814	.5199189	.93035	.9645465	.93035
d5	5.846065	3.98437	1.861695	.6815473	.8255588	.6815473
d6	7.901572	1.819312	6.082261	.2302468	.4798404	.2302468
d7	10.44378	2.632357	7.811422	.2520502	.5020461	.2520502
d8	5.056879	2.37437	2.68251	.4695326	.6852245	.4695326
overall				1.006482		

```
mc  = correlation between depvar and its prediction
mc2 = mc^2 is the Bentler-Raykov squared multiple correlation coefficient
```

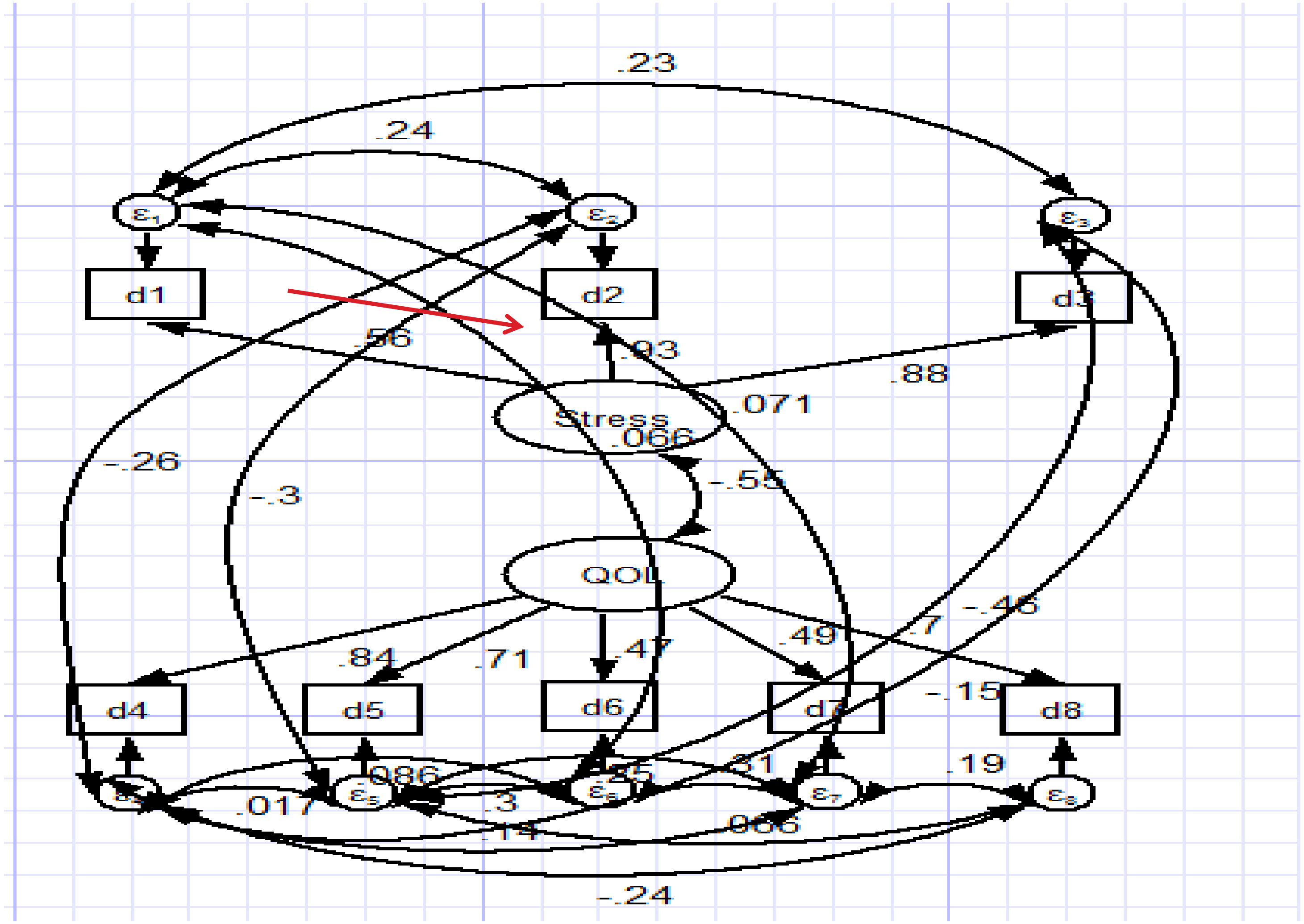
Checking residuals

```
. estat residuals

Residuals of observed variables

  Mean residuals
```

	d1	d2	d3	d4	d5	d6	d7	d8
raw	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

```
  Covariance residuals
```

	d1	d2	d3	d4	d5	d6	d7	d8
d1	0.000							
d2	0.000	0.000						
d3	0.000	0.000	0.000					
d4	0.053	0.013	0.004	-0.003				
d5	0.052	0.017	-0.005	0.003	0.013			
d6	0.200	0.096	-0.083	-0.010	0.042	0.000		
d7	0.258	0.029	0.086	-0.004	0.006	0.000	0.000	
d8	-0.115	-0.038	0.014	0.004	-0.015	0.000	0.000	0.000




```
sem (Stress -> d1, ) (Stress -> d2, ) (Stress -> d3, ) (QOL -> d4, ) (QOL -> d5, ) (QOL -> d6, ) (QOL -
> d7, ) (QOL -> d8, ), covstruct(_lexogenous, diagonal) vce(oim) standardized latent(Stress QOL ) cov(
Stress*QOL e.d1*e.d2 e.d1*e.d3 e.d1*e.d6 e.d1*e.d7 e.d2*e.d4 e.d2*e.d5 e.d3*e.d4 e.d3*e.d5 e.d4*e.d5
e.d4*e.d6 e.d4*e.d7 e.d4*e.d8 e.d5*e.d6 e.d5*e.d7 e.d5*e.d8 e.d6*e.d7 e.d7*e.d8) nocapslatent
```

“Near-equivalent” model

```
. estat gof, stats(all)
```

Fit statistic	Value
Likelihood ratio	
chi2_ms(2)	1.249
p > chi2	0.536
chi2_bs(28)	2190.168
p > chi2	0.000
Population error	
RMSEA	0.000
90% CI, lower bound	0.000
upper bound	0.071
pclose	0.845
Information criteria	
AIC	20938.800
BIC	21123.260
Baseline comparison	
CFI	1.000
TLI	1.005
Size of residuals	
SRMR	0.003
CD	0.988

(no modification indices to report, all MI values less than 3.841458820694123)

```
. estat residuals
```

Residuals of observed variables

Mean residuals

	d1	d2	d3	d4	d5	d6	d7	d8
raw	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Covariance residuals

	d1	d2	d3	d4	d5	d6	d7	d8
d1	-0.000							
d2	0.002	0.000						
d3	-0.003	0.000	0.000					
d4	-0.001	0.004	-0.008	-0.001				
d5	-0.000	0.008	-0.017	0.001	0.008			
d6	-0.008	0.060	-0.127	-0.005	0.031	-0.000		
d7	0.002	-0.015	0.031	0.001	-0.008	-0.000	-0.000	
d8	0.002	-0.017	0.037	0.002	-0.009	-0.000	0.000	0.000

```
. estat eqgof
```

Equation-level goodness of fit

depvars	fitted	Variance predicted	residual	R-squared	mc	mc2
observed						
d1	4.488142	1.427476	3.060666	.3180551	.5639637	.3180551
d2	7.941086	6.830184	1.110902	.8601071	.9274196	.8601071
d3	12.31794	9.526567	2.791372	.7733897	.8794258	.7733897
d4	7.462654	5.219994	2.24266	.6994823	.8363506	.6994823
d5	5.85112	2.932123	2.918997	.5011217	.7078995	.5011217
d6	7.901586	1.768483	6.133102	.2238137	.4730896	.2238137
d7	10.44379	2.55716	7.886634	.2448497	.4948229	.2448497
d8	5.056889	2.442651	2.614238	.4830343	.6950067	.4830343
overall				.9877224		

Comparing models

- Full model (OIM):

```
. sem (Stress -> d1, ) (Stress -> d2, ) (Stress -> d3,) (QOL -> d4, ) (QOL -> d5, ) (QOL -> d6, ) (QOL -> d7, ) (QOL -> d8, ), covstruct(_lexogenous, diagonal) latent(Stress QOL ) cov( Stress*QOL e.d1*e.d2 e.d1*e.d3 e.d2*e.d4 e.d2*e.d5 e.d3*e.d4 e.d3*e.d5 e.d4*e.d5 e.d4*e.d6 e.d4*e.d7 e.d4*e.d8 e.d5*e.d6 e.d5*e.d7 e.d5*e.d8 e.d6*e.d7 e.d7*e.d8) nocapslatent
. estat gof, stats(all)
```

- Models n (random) = 400, 300, 200, 100,75:

```
. set seed 12345
. sample 400, count
(...)
. estat gof, stats(all)
```

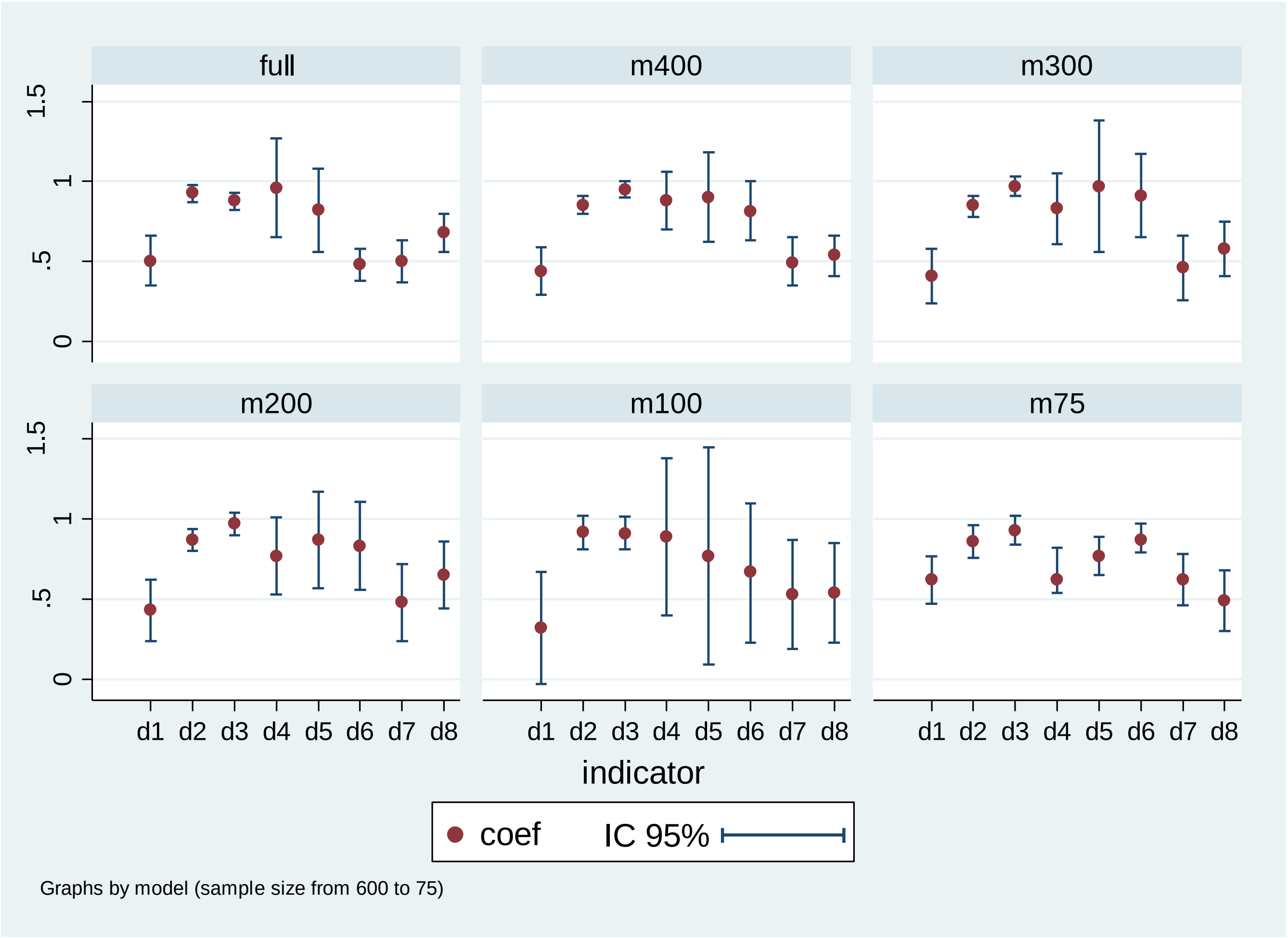
✓ Note: when the model fails to converge, start from a simpler model.

GOF tests – decreasing sample

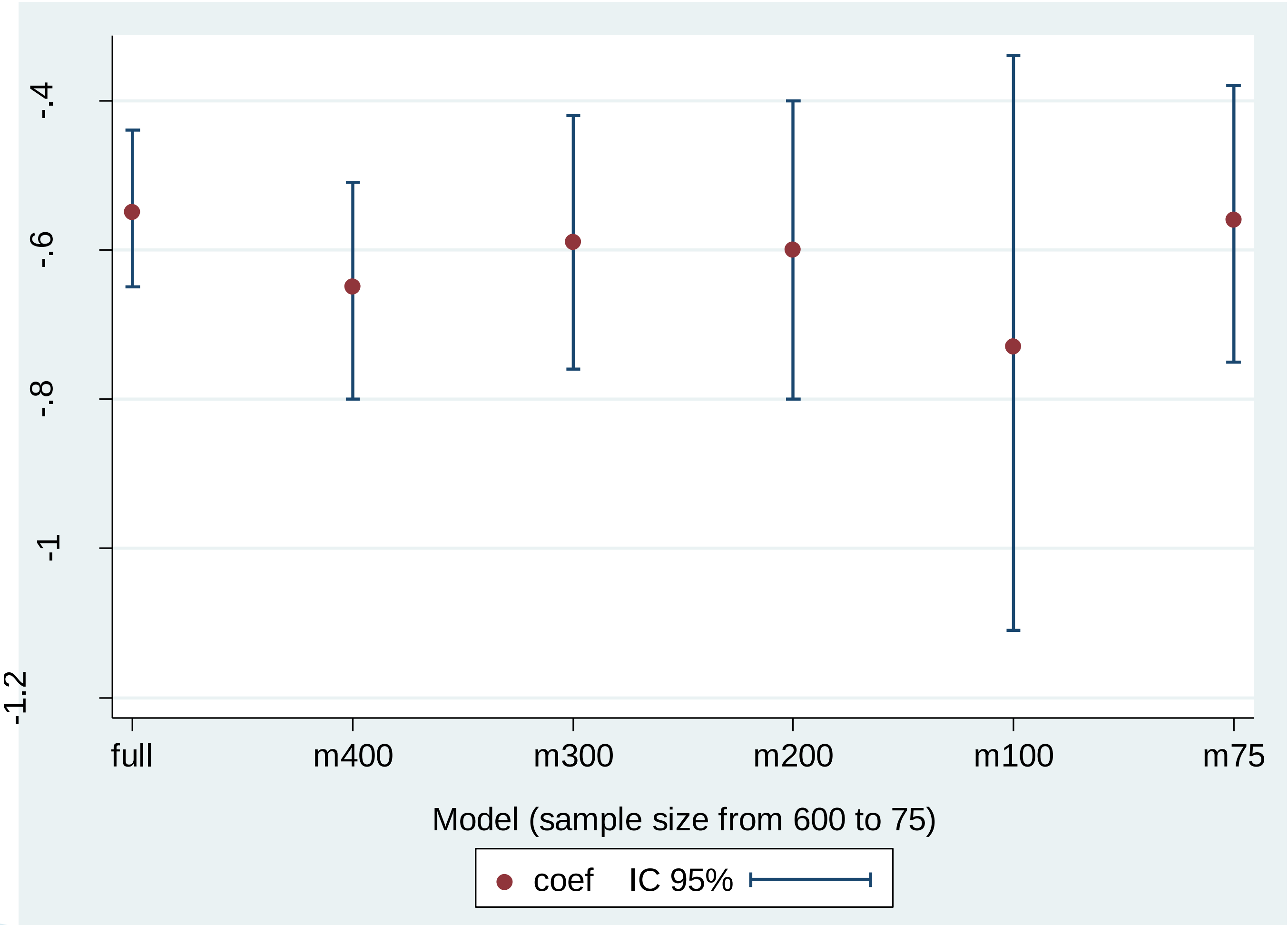
Test/ n	600	400*	300*	200**	100 **	75****
z: p > 0.05 loadings	–	–	–	–	Stress-d1 (0.074)	–
p for Chi2	0.304	0.193	0.129	0.642	0.336	0.280
RMSEA	0.019	0.037	0.052	<0.001***	0.038	0.047
Upper	0.067	0.091	0.112	0.081	0.162	0.116
CFI	1.000	0.999	0.997	1.000	0.999	0.988
SRMR	0.003	0.010	0.003	0.008	0.022	0.060
Stress- QOL	– 0.55	– 0.65	–0.59	–0.60	–0.73	– 0.56
IC 95%	–0.65 –0.44	–0.80 –0.50	–0.76 –0.42	–0.80 –0.41	–1.11 –0.34	–0.75 –0.38

*Lower n leads to higher RMSEA
**Simplified : covariance between d2-d5 excluded due to failure to converge.
***Increase in df leads to lower RMSEA.
****Basic model (slide 13).

Comparing models: loadings (coefficients)



Comparing models: covariance between latent variables



A few caveats – I

- Be aware the GOF tests are "global fit" tests;
- Maximum likelihood – ml – estimation works well under non-severe departs from normality of distribution and provides the widest array of GOF tests and postestimations;
- Under Likert scales, the option vce(robust) shall be taken into consideration;
- With an important fraction of missing values, the option – mlmv – is suggested so as to avoid listwise deletion and decrease of power;

A few caveats – II

- Evaluating p–values from a chi–square test assumes there is an overidentified model (df >0) to “improve”;
- The “best” set of GOF parameters as well as the “ideal” values of each one of the GOF statistics, let alone the relevance, are topics under debate;
- Respecification (or overparameterization) of a model shall be fundamentally based on the rationale, rather than on residuals or GOF tests;
- In this case study, some differences between models may be due to the random sampling.

Closing remarks – I

- Complex and combined questionnaries can be parceled and analysed under SEM models;
- Most GOF tests were somewhat "stable" in spite of a decrease in the sample size;
- Researchers are supposed to present the results under unstandardized and standardized ways;
- Do not be "selective" when presenting GOF tests;
- RMSEA and its upper bound "signalled" earlier a potential lack of fit due to small sample size (but "N" is part of the denominators in the formula).

Closing remarks – II

- Point estimates (for example, those related to the covariance between the latent variables) tended to keep a reasonal level of “stability” when decreasing the sample size;
- Confidence intervals increased, accordingly;
- Under small sample sizes, a more simplified model performed better (and loadings were more similar to the “full” model) than a model with a slighly larger sample size, yet still “complex” in terms of the number of covariances;
- This can be one of the strategies to tackle nonconvergence under short sample size.

References

- The WHOQOL Group. World Health Organization. WHOQOL: measuring quality of life. Geneva: WHO; 1997 (MAS/MNH/PSF/97.4). Also in: http://www.who.int/substance_abuse/research_tools/whoqolbref/en/

- Lipp, M. E. N. & Guevara, A. J. H. 1994. Text in portuguese. Validação empírica do Inventário de Sintomas de Stress. Estudos de Psicologia, 11(3), 43-49.

-

- Acock, Alan C. 2013. Discovering Structural Equation Modeling Using Stata. Revised edition. StataPress.

- Kline, Rex B. 2016. Principles and Practice of Structural Equation Modeling. Fourth edition. Guilford.

- StataCorp. Structural Equation Modeling Reference Manual. Downloadable at: http://www.stata.com/bookstore/structural-equation-modeling-reference-manual/

Thank you!

- Contact:
- *Marcos Almeida, MD PhD*
- Email: virtual.596@gmail.com