BJPsych The British Journal of Psychiatry (2011) 198, 17–23. doi: 10.1192/bjp.bp.110.080861

Online supplement

Derivation of the measurement model

The measurement model was derived using as much data as possible – ignoring the availability of other covariates. The model was fitted initially using those individuals with a complete Short Mood and Feelings Questionnaire (SMFQ) data-set ($n = 2501$) and the results compared with those obtained from a sample with at least one SMFQ measure (partial SMFQ response data-set: $n = 3887$). Incorporation of timing of menarche into these models reduced the sample to 2868 (partial SMFQ) and 2063 (complete SMFQ) and the addition of confounders resulted in further reductions to 2184 and 1957 respectively. As shown in Fig. 1 in the article, we refer to these four samples as N_1–N_4.

The fit of a range of unconditional confirmatory factor analysis (CFA) models with different levels of measurement invariance (constraints applied to parameter values) was compared. Model fit was assessed using root mean square error of approximation (RMSEA, <0.05 desirable) and comparative fit index/Tucker–Lewis index (CFI/TLI, >0.9 desirable). The following parameters were assigned to be variant/invariant across time.

(a) Factor loadings – the factor variances were fixed equal to unity; consequently, 12 loadings could be freely estimated at each time point. Loadings were allowed to vary between items, but fixed to be equal for the same item at different time points.

(b) Item thresholds – individual items were three-level categorical variables, hence there were two thresholds to measure for each item. Thresholds were constrained to be constant through time, similar to the loading constraint.

(c) Correlation of residual variances – residual variances were permitted to be correlated between the same items measured at different times. These correlations were invariant through time, but could vary between items.

(d) Scale factors – using the default Delta parameterisation for categorical variables within Mplus, scale factors for continuous latent response variables of observed categorical dependent variables are allowed to be parameters in the model, but residual variances for continuous latent response variables are not. In the constrained model, scale factors T_1 (10.5 years) were all fixed to unity. All other scale factors were constrained to be equal, but freely estimated.

(e) Factor means and variances – it was possible to relax the constraint on the factor means and variances to permit the level and spread of depressive symptoms in the population to change over time. The mean and variance for the first factor (10.5 years) was fixed at 0 and 1 respectively, whereas the mean/variance for T_2 and T_3 (13 and 14 years) were freely estimated.

The successions of CFA models along with the parameter constraints applied are shown in Table DS1. Models were fitted initially for the sample with at least one of the three sets of SMFQ items ($n = 3887$) and repeated for the smaller sample with complete SMFQ data ($n = 2501$). Model fit statistics shown in Table DS1 show that for the larger sample, all models were of acceptable fit (CFI/TLI >0.95, RMSEA <0.05). Fit statistics changed slightly with the reduced sample of 2501, but changes were minor.

Whereas models 4 and 5 were more parsimonious, model 7 was chosen on the basis that the freed factor means and variance captured an additional aspect, i.e. the change in symptoms across the time period modelled. Factor means and variances for model 7 ($n = 3887$) were as follows: means 0, 0.13, 0.48; variances 1.00, 1.20, 1.22 for each of the three time-points (10.5, 13 and 14 years) respectively. As a consequence of these freed parameters, the somewhat arbitrary scaling of our latent trait is now all in the metric of the first time point. It can be seen that the population mean increases by approximately 0.5 standard deviations between 10.5 and 14 years, with most of this increase occurring during the final year.

Table DS2 shows the factor loadings obtained under the chosen measurement model. Although all items would be deemed to load significantly on the depressive symptoms factor, there is clearly a wide variability in the estimated loadings with 'I felt so tired I just sat around and did nothing' (0.429) being relatively unimportant compared with 'I felt I was no good any more' (0.824), 'I hated myself' (0.833) and 'I thought nobody really loved me' (0.795). The assumption of equal loadings implied by the use of a simple sum-score would miss the large discrepancy between the items. We observe very similar results when comparing loadings across the larger and smaller samples. For the current analysis the item 'I was very restless' was omitted from the 13-item SMFQ because preliminary work suggested that some individuals in the study sample were uncertain about the meaning of this item.

Finally, the correlations between the latent depressive symptom factors at the different time points were: correlation (T_1, T_2) 0.466 (s.e. = 0.02), correlation (T_1, T_3) 0.315 (s.e. = 0.02), correlation (T_2, T_3) 0.613 (s.e = 0.02) in the larger sample ($n = 3887$); and correlation (T_1, T_2) 0.460 (s.e. = 0.02), correlation (T_1, T_3) 0.303 (s.e. = 0.02), correlation (T_2, T_3) 0.607 (s.e. = 0.02) in the reduced sample ($n = 2501$).

Table DS1 Measurement model comparison

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Constraint							
Factor loadings	Unequal	Equal	Equal	Equal	Equal	Equal	Equal
Thresholds	Unequal	Unequal	Equal	Equal	Equal	Equal	Equal
Scale factors	All = 1	All = 1	All = 1	All = 1	Partly freed[a]	Partly freed[a]	Partly freed[a]
Error variances	Uncorrelated	Uncorrelated	Uncorrelated	Correlated	Correlated	Correlated	Correlated
Factor variances	Equal (unity)	Equal (unity)	Equal (unity)	Equal (unity)	Equal (unity)	Freed	Freed
Factor means	Equal (zero)	Equal (zero)	Equal (zero)	Equal (zero)	Equal (zero)	Freed	Freed
Free parameters	111	87	39	51	52	54	56
Partial SMFQ data-set, $n = 3887$							
Comparative fit index	0.972	0.973	0.944	0.951	0.950	0.961	0.980
Tucker–Lewis index	0.989	0.986	0.973	0.976	0.977	0.984	0.992
Root mean square error of approximation	0.028	0.031	0.043	0.040	0.040	0.033	0.024
Complete SMFQ data-set, $n = 2501$							
Comparative fit index	0.967	0.968	0.938	0.947	0.945	0.958	0.977
Tucker–Lewis index	0.987	0.985	0.972	0.975	0.976	0.983	0.991
Root mean square error of approximation	0.033	0.036	0.049	0.046	0.045	0.038	0.028

SMFQ, Short Mood and Feelings Questionnaire.
a. All equal to 1 at 10.5 years, free but constrained equal at 13 and 14 years.

Table DS2 Factor loadings for the chosen measurement model (model 7)

	Factor loading (s.e.)	
Item	$n = 3887$	$n = 2501$
I felt miserable or unhappy	0.648 (0.010)	0.650 (0.012)
I didn't enjoy anything at all	0.510 (0.015)	0.509 (0.017)
I felt so tired I just sat around and did nothing	0.429 (0.013)	0.423 (0.015)
I felt I was no good any more	0.824 (0.009)	0.821 (0.010)
I cried a lot	0.691 (0.010)	0.690 (0.012)
I found it hard to think properly or concentrate	0.558 (0.011)	0.557 (0.013)
I hated myself	0.833 (0.009)	0.833 (0.011)
I was a bad person	0.613 (0.013)	0.619 (0.015)
I felt lonely	0.741 (0.009)	0.747 (0.011)
I thought nobody really loved me	0.795 (0.009)	0.790 (0.011)
I thought I could never been as good as other kids	0.732 (0.009)	0.721 (0.011)
I did everything wrong	0.726 (0.010)	0.721 (0.012)