Last updated January 4, 2012

This document provides a description of Mplus code for implementing mixture factor analysis with four latent class components with and without covariates described in the following paper:

Wall MM, Guo J, Amemiya Y. Mixture factor analysis for approximating a non-normally distributed continuous latent factor with continuous and dichotomous observed variables. To appear in *Multivariate Behavioral Research.*

Implementation in Mplus

The Mplus programs for fitting a mixture factor analysis with four latent class components with and without covariates are shown below in the Appendix and described in detail here where capital letters are used to denote Mplus command names and lowercase letters are user specified names (Muthen & Muthen, 1998-2010). For the model without covariates, i.e. Figure 2c, in the VARIABLE section of the Mplus code, the USEVARIABLES ARE command specifies that y1 y2 y3 and u will be included in the model. CATEGORICAL ARE u indicates that the outcome variable u, here a dichotomous measure of non-adherence to treatment, is to be taken as ordered categorical. The CLASSES = c(4) command tells Mplus that in the mixture model there should be 4 latent class components and that the name of the latent class variable should be c. In the ANALYSIS section of the code, the TYPE=MIXTURE command specifies that a mixture model will be fit and the ALGORITHM = INTEGRAL option is necessary because of the need for numerical integration (here Gaussian quadrature) as described above. Robust standard errors associated with the maximum likelihood estimates are output because of the inclusion of the ESTIMATOR=MLR command. Maximum likelihood optimization is done in two stages. The STARTS 100 20; and STITERATIONS = 10; commands tell Mplus to generate in the first stage 100 different random starting values for the parameters and to do 10 iterations of the maximization for all of them. Then in the second stage, it takes the parameter estimates associated with the best 20 likelihood values obtained from those partial optimizations in the first stage and uses them as starting values for an optimization that continues until default convergence settings are satisfied. The STSEED = 1234; simply sets the seed of the random generation to ensure that if the program is re-run with the same seed, the exact same results will be obtained.

The MODEL section of the code includes %OVERALL% and class specific, e.g., %c#1% commands. Model commands placed under the %OVERALL% statement indicate relationships (and associated model parameters) that are fixed to be the same across all latent classes. Here the latent factor f is measured by the 3 continuous observed variables through the command: f BY y1* y2 y3@1; and this measurement (i.e. the loadings and measurement errors) are fixed to be the same across all latent classes (i.e. measurement invariance). The * notation and the @0 notation following y1 and y3 respectively in the first command tell Mplus to allow the factor loading for y1 to be freely estimated (by default the first variable has the loading fixed to 1) and instead fix the loading of y3 to be 1. The [y3@0]; command indicates to Mplus to fix the intercept of y3 to zero and is necessary for identifiability because the mean of the latent factor (through the mixture part of the model) will be freely estimated. The dichotomous outcome is predicted by the latent factor through the command: u ON f; and the addition of the

command [u$1](1); is necessary in order to fix the intercept of the relationship to be the same across latent classes. The notation (1) following the intercept [u$1] gives this threshold parameter a name, ``1", that is then fixed across classes. If it was not included, a different intercept for the relationship between u and f would be estimated for each class.

The command [f*] under each of the class specific statements indicates to Mplus that a separate mean should be fit for the latent factor within each of the 4 classes. By default without specifying otherwise, a common variance is assumed for f within each latent class. For the model that also consider covariates, Figure 2D, besides adding these covariates to the USEVARIABLES command, we include u ON f age sex; f ON age sex; and c ON age sex; to specify that covariates can influence the outcome u, the latent factor f and the probability of being in any one of the 4 latent classes c. This command is added under the %OVERALL% statement because we assume these relationships are invariant across class.

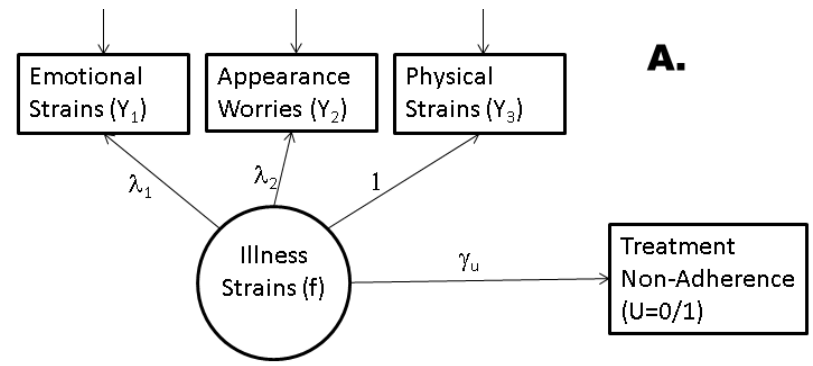
Finally, a useful feature that allows multiple processors to be utilized (assuming they are available on the local computer) is available in Mplus 6. The command ``PROCESS = 4 (STARTS);" could be added to the code described above in the ANALYSIS section and would split the model fitting of the multiple random starting values across 4 processors (if the computer has 4 or more processors) and hence run approximately four times faster.

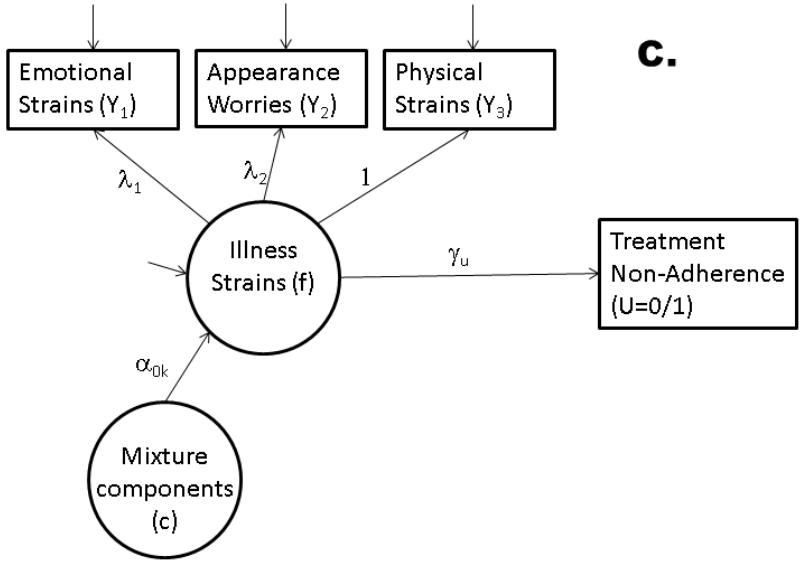
Appendix

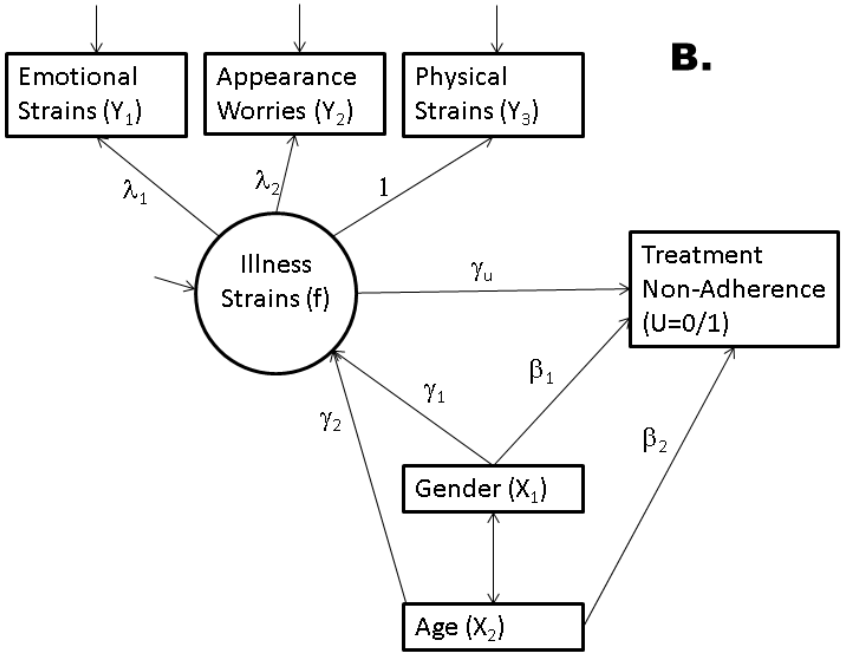
Line-by-line explanation of the Mplus code is given above

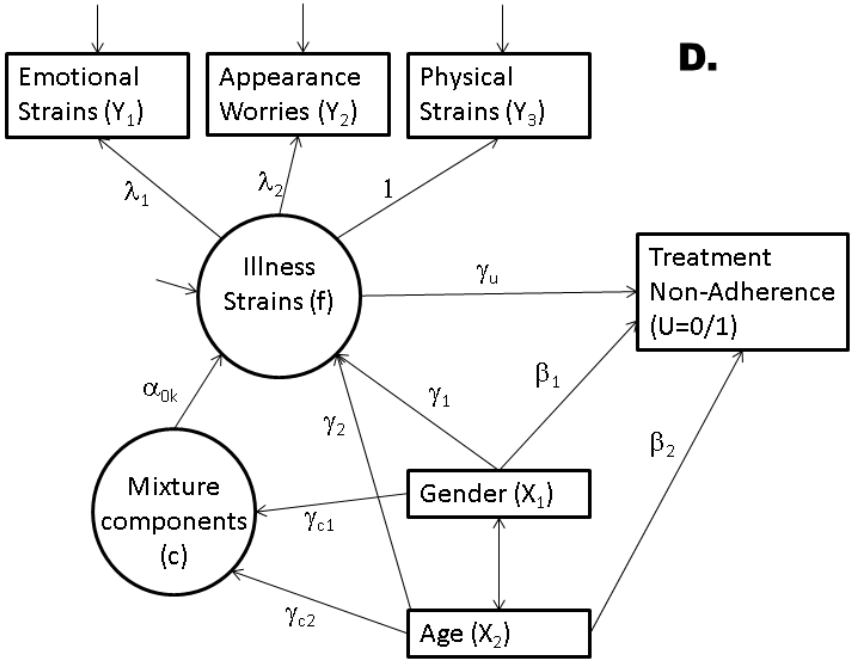
```
 **** Example of data frame (cysticfibrosis.dat);
      y1          y2          y3        u   age sex
 -0.591712369 -1.273512337 -0.813941913    0    2   0
  0.597256375  0.489545613  0.484037295    0    1   1
 -0.163289883  0.182509503  0.039055639    1    3   1
 -0.277251534 -0.598088431 -1.086112851    1   -2   1
                   ... ...
TITLE:  Model for Figure 2C with a mixture of 4 components;
DATA:       FILE IS cysticfibrosis.dat;
VARIABLE:   NAMES ARE y1 y2 y3 u age sex;
         USEVARIABLES ARE y1 y2 y3 u;
         CATEGORICAL ARE u;
         CLASSES = c(4);
ANALYSIS:   TYPE = MIXTURE; ALGORITHM=INTEGRATION; ESTIMATOR=MLR;
         STARTS 100 20; STITERATIONS 10; STSEED 1234;
MODEL:      %OVERALL%
         f BY y1* y2 y3@1; [y3@0];
         u ON f; [u$1](1);
         %c#1%
         [f*];
         %c#2%
         [f*];
         %c#3%
         [f*];
         %c#4%
         [f*];
```

```
TITLE:  Model for Figure 2D with a mixture of 4 components;
DATA:       FILE IS cysticfibrosis.dat;
VARIABLE:   NAMES ARE y1 y2 y3 u age sex;
        USEVARIABLES ARE y1 y2 y3 u age sex;
        CATEGORICAL ARE u;
        CLASSES = c(4);
ANALYSIS:   TYPE = MIXTURE; ALGORITHM=INTEGRATION; ESTIMATOR=MLR;
        STARTS 100 20; STITERATIONS 10; STSEED 1234;
MODEL:      %OVERALL%
        f BY y1* y2 y3@1; [y3@0];
        u ON f age sex; [u$1](1);
        f ON age sex;
        c ON age sex;
        %c#1%
        [f*];
        %c#2%
        [f*];
        %c#3%
        [f*];
        %c#4%
        [f*];
```


A.
Emotional Strains (Y_1)
Appearance Worries (Y_2)
Physical Strains (Y_3)
λ_1
λ_2
1
Illness Strains (f)
γ_u
Treatment Non-Adherence (U=0/1)


C.
Emotional Strains (Y_1)
Appearance Worries (Y_2)
Physical Strains (Y_3)
λ_1
λ_2
1
Illness Strains (f)
γ_u
Treatment Non-Adherence (U=0/1)
α_{0k}
Mixture components (c)


B.
Emotional Strains (Y_1)
Appearance Worries (Y_2)
Physical Strains (Y_3)
λ_1
λ_2
1
Illness Strains (f)
γ_u
Treatment Non-Adherence (U=0/1)
γ_1
β_1
γ_2
β_2
Gender (X_1)
Age (X_2)


D.
Emotional Strains (Y_1)
Appearance Worries (Y_2)
Physical Strains (Y_3)
λ_1
λ_2
1
Illness Strains (f)
γ_u
Treatment Non-Adherence (U=0/1)
α_{0k}
γ_1
β_1
γ_2
β_2
Mixture components (c)
γ_{c1}
Gender (X_1)
γ_{c2}
Age (X_2)