Supplementary Material S1. Symptom diary

Supplementary Material S2. Case report form

2011 04 08 TARGET CRF Bristol centre version 1.0

Supplementary material S3. LLCA models – model fit and comparisons.

LL= log-likelihood, ΔLL= change in log-likelihood, aBIC = Sample-size adjusted BIC;

BLRT = Bootstrap Likelihood Ratio Test; LMR = Lo-Mendell Rubin test

Model fit statistics considered

aBIC - sample-size adjusted Bayesian Information Criterion.

The Bayesian Information Criterion (BIC: Schwartz, G., 1978) is the most commonly-used fit statistic for comparing mixture models. A function of both the likelihood and the number of estimated parameters, the BIC penalises model complexity. We opted for the sample-size adjusted version which incorporates the samplesize as an additional term. BIC will typically decrease and then increase following the incremental additional of classes. Using this statistic the model with the lowest BIC (or other models with BIC values in the vicinity) would be deemed satisfactory however in some instances this statistic did not reach a minimum within the range of models considered..

Bootstrap tests for nested models.

The Bootstrap Likelihood Ratio Test (BLRT) and the Lo-Mendell-Rubin (LMR) test statistics (Nylund et al., 2007) both assess change in model fit when adding an additional class. Here a high p-value for a k-class model indicates no substantial improvement in fit compared to the k-1 class solution. Unlike the LMR, The BLRT makes no distributional assumptions and simulation work has so far shown this measure to be superior (Nylund et al., 2007) however in our experience the BLRT can be extremely conservative and may reject all the models considered.

Entropy.

Mixture modelling output consists primarily of class-assignment probabilities which describe the confidence with which each participant can be assigned to each latent class. Entropy, also referred to as classification accuracy, summarises this information as a single measure which can take values form zero to one, with one indicating no assignment uncertainty. Entropy is of little use in determining the optimal model (Tein et al, 2013) and can be poor in simulation studies even when the correct model is estimated (Heron et al., 2015). Whilst LCA has been promoted as a method to facilitate targeted interventions (Lanza and Rhoades, 2013) we propose that such a strategy is dependent on clearly defined and well-separated groups of individuals. Consequently, we regard entropy as an indicator of model *utility* since if entropy is low and individuals can only be poorly classified then the resulting classification is of little use as a targeting tool.

In addition, entropy has been shown to be important when it comes to the level of bias resulting from a standard three-step analysis (Vermunt, 2010; Bakk, Tekle, & Vermunt, 2013; Bakk, Oberski, & Vermunt, 2014). Whilst not directly related to the issue of model utility, entropy will influence the analytical approach employed when assessing covariate and outcome associations.

Smallest class size.

As a latent class analysis is usually the initial stage of a project with the intention of deriving a number of groups for further research, analysts often place a limit on the size of the results classes. This pragmatic decision is to facilitate the planned further study since there is little one could reasonably do with a class of ten participants other than drop them from the sample. Here we only considered models where all classes contained at least 5% of the participants.

Bivariate residuals

In addition to the figures shown in the previous table we also examined the bivariate residuals from each model. Mixture modelling, as with continuous trait modelling, is based on the assumption of conditional independence, namely that the class indicators should be independent conditional on the latent variable. The pattern and magnitude of these residuals is examined over the following few pages.

Bakk, Z., Tekle, F. T., & Vermunt, J. K. (2013). Estimating the association between latent class membership and external variables using bias-adjusted three-step approaches. *Sociological Methodology*, 43, 272-311.

Bakk, Z., Oberski, D. L., & Vermunt, J. K. (2014). Relating Latent Class Assignments to External Variables: Standard Errors for Correct Inference. *Political Analysis*, 22, 520-540.

Heron, J. E., Croudace, T. J., Barker, E. D., & Tilling, K. (2015). A comparison of approaches for assessing covariate effects in latent class analysis. *Longitudinal and Life Course Studies*; Vol 6, No 4.

Lanza ST and Rhoades BL. Latent Class Analysis: An Alternative Perspective on Subgroup Analysis in Prevention and Treatment. *Prevention Science*. 2013, Volume 14, Issue 2, pp 157-168

Nylund KL, Asparouhov T, Muthen BO. Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modelling: A Monte Carlo Simulation Study. *Structural Equation Modelling: A Multidisciplinary Journal* 2007;14(4):535-569.

Schwarz G. Estimating the dimension of a model. *Annals of Statistics* 1978;6:461–464

Tein J-Y, Coxe S , Cham H. Statistical Power to Detect the Correct Number of Classes in Latent Profile Analysis. *Structural Equation Modeling: A Multidisciplinary Journal*. Vol. 20, Iss. 4, 2013.

Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, 18, 450-469.

Bivariate residuals from models of 15 consecutive days' data on COUGH.

Graphs over next two pages show residuals for 1 class, 2 class, ... 6 class models. As each class indicator has three categories, there are nine residuals for each pair of measures (3x3 cells in the contingency table). Red indicates residuals between adjacent time points, green indicates residuals between measures two days apart and blue the remaining residuals. It's apparent that these models are failing to capture two aspects of the data (i) the strong association between measures taken very close together and (ii) the association between measures taken towards the start and end of the two-week period when the majority of children exhibit little change from day to day.

7

Not surprisingly, we see marked improvement in the magnitude of the residuals when adjacent measurements are dropped from the model. These figures show pairwise residuals from 1 through 6-class models for the cough data using alternate measures. Once again, but to a lesser extent, we see that the model is less able to model the data at the two extremes of the two-week measurement window.

4 class 5 class 6 class

 $\frac{1}{5}$

Distribution of bivariate residuals with red-lines indicating +/- 1.96 (Skip 0: All 15 measures)

Distribution of bivariate residuals with red-lines indicating +/- 1.96 (Skip 1: Alternate measures)

Distribution of bivariate residuals with red-lines indicating +/- 1.96 (Skip 2: Every third measure)

Estimated trajectories

Modelling three-category ordinal measures poses a slight problem when it comes to plotting the results as all three categories describe a trajectory through time meaning two plots are necessary. To avoid this we have collapsed the within-class profiles by turning the probabilities into a single measure of severity (severity = 0*P(Category 0) + 1*P(Category 1) + 2*P(Category 2)).

3-class models

4-class models

5-class models

† 'Improved' and 'worse' is defined by change of ≥1 point on a 7 level Likert scale.

\$ Number of children less than 1,408 due to missing data.

^a Pearson's chi².

b In the 24 hours prior to consultation.