Statistical Methods in Medical Research

http://smm.sagepub.com

Latent mixture models for multivariate and longitudinal outcomes

Andrew Pickles and Tim Croudace

Stat Methods Med Res 2010; 19; 271 originally published online Jul 16, 2009;
DOI: 10.1177/0962280209105016

The online version of this article can be found at:
http://smm.sagepub.com/cgi/content/abstract/19/3/271

Published by:

http://www.sagepublications.com

Additional services and information for *Statistical Methods in Medical Research* can be found at:

Email Alerts: http://smm.sagepub.com/cgi/alerts

Subscriptions: http://smm.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.co.uk/journalsPermissions.nav

Citations http://smm.sagepub.com/cgi/content/refs/19/3/271

Downloaded from http://smm.sagepub.com at UCLA on July 6, 2010

Statistical Methods in Medical Research 2010; **19**: 271–289

Latent mixture models for multivariate and longitudinal outcomes

Andrew Pickles Biostatistics, Health Methodology Research Group, University of Manchester, University Place, Oxford Road, Manchester, M13 9PL, UK and **Tim Croudace** Department of Psychiatry, University of Cambridge, Addenbrooke's Hospital, Hills Road, Cambridge, CB2 2QQ, UK

Repeated measures and multivariate outcomes are an increasingly common feature of trials. Their joint analysis by means of random effects and latent variable models is appealing but patterns of heterogeneity in outcome profile may not conform to standard multivariate normal assumptions. In addition, there is much interest in both allowing for and identifying sub-groups of patients who vary in treatment responsiveness. We review methods based on discrete random effects distributions and mixture models for application in this field.

1 Introduction

In a variety of therapeutic contexts the outcome targeted by an intervention is the time path of one or more variables. In many cases growth curve models, multilevel models in which the intercept and various functions of time may be considered as random effects, lend themselves to the analysis of such data, allowing treatment effects to be defined by a small number of growth parameters capturing change for individuals or groups, that are often characterised simply as just a linear trend. The timescale over which parametric, or more flexible 'response to treatment' trajectories are defined, may be the natural units of follow-up or some linearising transformation. Where a discontinuous trend is expected, more complex, piecewise forms can be considered. These models commonly assume *individual* trajectories as being smoothly distributed around *group* (average) or covariate-adjusted trends, with typical applications involving univariate, bivariate or trivariate Gaussian normal random effects distributions.[1]

However, in some fields, particularly psychological and behavioural development, consideration of both data and theory has suggested that a typological representation might be more appropriate.[2–5] By typological, we mean the identification of a small number of groups, in which the time course for evolution of trends across the repeated measures defines type/group.[6] This perspective, often referred to as *group-based trajectory modelling*, by one of its originators,[7,8] or as latent-class growth analysis LCGA by

Address for correspondence: Andrew Pickles, Biostatistics, Health Methodology Research Group, University of Manchester, University Place, Oxford Road, Manchester, M13 9PL, UK.
E-mail: andrew.pickles@manchester.ac.uk

© The Author(s), 2010. Reprints and permissions
http://www.sagepub.co.uk/journalsPermissions.nav

10.1177/0962280209105016

Downloaded from http://smm.sagepub.com at UCLA on July 6, 2010

another[9,10] is based on the assumption that the data arise from observing, subject to measurement error, a mixture of individuals from a limited number of latent classes, each with their own developmental time path (trajectory type or group).[11]

Levine and Rabinowitz[12] studied response to antipsychotic medications in an RCT of patients with schizophrenia, arguing that little is known about the extent of heterogeneity of symptomatology in treated early-onset psychosis, using data from a clinical trial of two active treatments (haloperidol and risperidone). Trajectories on two types of psychotic symptoms were studied and related to demographic (age, sex) diagnostic groups, social and cognitive measures. They found that a number of variables, including the age of onset of the disorder, the type of schizophrenia diagnosis, level of cognitive and premorbid functioning had prognostic value in predicting treatment response trajectories (class memberships).

Of course this perspective can be motivated in many therapeutic contexts in which disease remission or course progression is studied, but has been particularly prominent in the context of psychological therapy research where there are acute challenges in identifying the principles and processes of change. Trajectory groups may correspond to client groups who are either 'recovering' or 'non-responsive' or any other variety of more subtle, potentially graduated patterns.[13] This context has also pioneered the conceptualisation and study of such latent trajectory groups, observed under routine practice conditions, outside of the controlled context of clinical trial intervention studies.[14] These authors explicitly set out to measure client progress under the expectation that they might follow highly variable temporal courses, using growth mixture modelling (GMM) of six repeated measures recorded on clients treated as psychotherapy outpatients. GMM can be considered as an extension of multilevel growth models that allow for latent classes, or an extension of latent class growth analyses that allow for random effects. Their flexibility and also their complexity arises from the occurrence of correlated random effects/growth factors within latent classes, and hence combines notions of quantitative and qualitative heterogeneity. Stultz *et al.*[14] found five client groups with different profiles on the Clinical Outcomes in Routine Evaluation – Outcome Measure described as: (a) high initial impairment, (b) low initial impairment, (c) early improvement, (d) medium impairment with continuous treatment progress or (e) medium impairment with discontinuous treatment progress. Relationships with demographic and psychopathology measures were identified.

Kreuter and Muthén[15] have advanced this type of methodology in criminological research where outcomes over time are often repeated event counts, following pioneering applications of mixture models to cross-sectional data.[16,17] Also, it is in this context that many of the most vociferous debates about what the groups represent or capture have been had.[18] The methods have also been used in addiction treatment studies.[19] Sequential process models (one growth model linked in time to another) have also been formulated, often relating trajectory groups observed across periods of development and in school settings where preventive interventions are tested.[20] When trajectory models span a pre-intervention period, then prevention-oriented behavioural science hypotheses are examined. In these and other settings it is also not unusual for there to be more than one variable whose time path is to be targeted and several may be studied in parallel simultaneously. In developmental studies it is often prudent, or necessary, to study the joint evolution of key variables and in intervention studies there will often be interest, or

Downloaded from http://smm.sagepub.com at UCLA on July 6, 2010

additional value, in studying the parallel and/or divergent course of secondary outcomes. Adaptation of the generalised growth curve mixture model[21] for such a circumstance can consider correlated sets of random effects, for two or more processes (polynomial growth or piecewise growth models for each measure), or potentially more parsimoniously, factor analytic growth curve models, that harness the information for multiple responses into a single latent variable outcome whose single trajectory is then examined. These more general latent variable approaches to potential multivariate trajectory indicators offer a framework within which the effects of treatment can be partitioned into common and specific effects for each target variable. In the mixture model approach the latent typology allows classes with distinctively different combinations of trajectories over each target outcome whose differential association with treatment can be of interest. When applied to pre-intervention longitudinal data these offer groupings akin to principal strata.

Another generalisation is to use latent class mixtures for missing data, as a means of tackling non-ignorable differences between sub-groups of patients with different patterns of available data.[22] This approach generates the latent class equivalent of missing-data pattern mixture models.[23,24] These are valuable where there are numerous sparse patterns of missingness that in the standard pattern-mixture approach would lead to too numerous and poorly estimated parameters or hard to justify restrictions on parameters. Random coefficient expansions have been entertained.

In this framework, the intention of treatment may be to raise the probability of an individual belonging to a recovering group or increasing the probability of switching from a pathological or non-responsive group to a more benign or recovering group. But these models can also be used as a tool to assist in accounting for selective receipt of treatment, for example as an alternative to or as the first step in a propensity score adjustment approach.[25] Treatment compliance can also be approached along the pattern-mixture lines of the previous paragraph (especially since missingness and non-compliance often co-occur), an approach that is especially appealing where latent compliance classes can be operationalised to overcome the problem of being unable to observe compliance to a counter-factual treatment.

In this necessarily selective overview and introduction we illustrate some of these methods, describe their key features, and discuss some of the less familiar and more distinctive statistical issues. We report preliminary findings from applying some simple variants of these methods using data from the SoCRATES trial (also used elsewhere in this issue) analysed in Mplus and Stata `gllamm`.

2 SoCRATES trial description

The SoCRATES trial[26,27] was a three-centre prospective, rater-blind, randomised, controlled trial of three treatments for psychosis: CBT and treatment as usual (TAU), supportive counselling (SC) and TAU, or TAU alone, with 18 months follow-up. In summary, 101 participants were allocated to CBT+TAU, 106 to SC+TAU and 102 to TAU alone. Of these, 257 (83.20%) were first admission patients. A total of 225 participants (75% of those randomised) were interviewed at 18 months follow-up, 75 in the CBT+TAU arm, 79 in the SC+TAU arm and 71 after receiving TAU alone.

Downloaded from http://smm.sagepub.com at UCLA on July 6, 2010

The remaining participants died during the follow-up period (7), withdrew consent (4) or were lost (73).

In this article, we consider the CBT and SC treatments as a single treatment group. Both consisted of a preliminary 5-week intensive phase followed by a small number of booster sessions until 3 months. Treatment was delivered by a small number of therapists, implying that observations may exhibit some degree of within-centre clustering.

We examine the primary outcome, The Positive and Negative Syndromes Schedule (PANSS),[28] an interview-based scale for rating 30 psychotic and non-psychotic symptoms administered blind to condition allocation. The PANSS was administered at baseline, once a week over the first 6 weeks and then at 3 months, 9 months and 18 months. In analyses we log transformed the timescale, and considered linear and quadratic terms, interacted with group (CBT vs. (SC+TAU and TAU)), collapsing the two non-CBT arms for simplicity.

3 Single outcome growth and trajectory models

In the traditional linear growth curve model observations $\{y_{it}\}$ for individual i made at times $t = 0, \ldots, T$ are assumed to follow a growth trajectory such that

$$y_{it} = \alpha + \beta t + a_i + b_i t + \varepsilon_{it}$$

where a_i and b_i are random coefficients (intercept and slope, respectively) and ε_{it} for $t = 0, \ldots, T$ are commonly assumed to have a mean zero Gaussian distribution with a diagonal covariance matrix and uncorrelated with the freely correlated bivariate normal distribution of the random coefficients.

Under maximum likelihood, observations need not be available for all individuals. Thus, the measurement regime does not have to be the same provided that each regime can be assumed to correspond to a missing-at-random missing data mechanism. With more than two values of t the model equation is easily extended to include further functions of t, such as polynomials, with fixed and random or just fixed coefficients as required. With a measurement regime of fixed time intervals growth curve models are easily set up within the confirmatory factor analysis framework given by

$$y_{it} = \eta_{1i} + \lambda_t \eta_{2i} + \varepsilon_{it}$$

where the factor η_1 corresponds to the random intercept and the factor η_2 the random slope that are jointly bivariate normal, and the factor loadings are temporal basis functions (linear or higher order functions of time). As commonly applied, the models differ from the random coefficient model in that the flexibility of possible factor loadings lends itself to fitting a piecewise linear growth (equivalent to a saturated polynomial in time in the random coefficient setup) and in a more obviously explicit choice as to the constancy of var(ε_{it}) with t. Figure 1(a) and (b) illustrates these two models.

In both multilevel model and factor analysis model set-ups, Empirical Bayes' (EB) estimates of individual growth trajectories can be extracted, in the former case as functions of the level-2 residuals and in the latter as functions of the factor scores.

Downloaded from http://smm.sagepub.com at UCLA on July 6, 2010

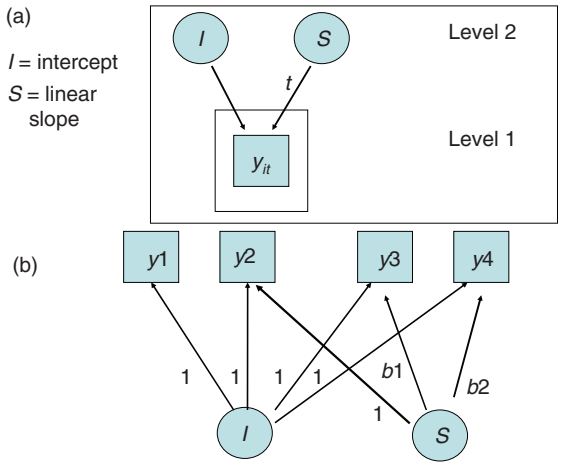

Figure 1 Graphical representation of linear growth random effects model (a) with an unspecified number of measurement occasions and a random coefficient for t, and a structural equation growth model (b) with factor loadings that vary over four fixed occasions of measurement.

Latent class trajectory models i.e. the group-based or LCGM approach (often abbreviated to LCGA), replaces the bivariate Gaussian assumption for the random coefficients or factors by a discrete distribution of classes in two dimensions, with M masses of size p_j at locations $d1_j$ and $d2_j$, $j = 1, \ldots, M$, where $d1$ corresponds to the intercept dimension and $d2$ the slope. Here too, EB estimates of the individual growth trajectories can be obtained. These are given as the weighted sum of the growth trajectories of each class, the weights being the individual posterior probabilities of class membership. The maximum *a posteriori* probability (MAP) can also be used to assign individuals to their most likely class, with models in which such allocations can be made with high confidence, being indicative of the quality of the resulting classification and thus typology. Second stage or simultaneous analyses can then relate groups to covariates.

Application of these mixture models seems to adopt three different positions on determining the number of classes that should be fit.

The first position sees the models more as a data reduction device intended to capture parsimoniously the bulk of the variation in the data, perhaps mapping the identified classes to some usually fairly coarse articulation of theory. The second position attempts to identify the true number of classes by adding classes until no significant improvement in fit occurs. Standard likelihood ratio test statistics cannot be compared to the usual chi-square distribution as setting class probabilities to zero leaves other parameters unidentified and it is unclear how many parameters are being tested. Comparisons of Akaike and Bayesian Information Criterion (AIC $= -2 \log \mathrm{L} + 2p$ and BIC $= -2 \log \mathrm{L} + p \log(n)$ where p is the number of model parameters and n the number of participants) can be used.[29] Strict BIC superiority is often used as the selection criterion but Raferty[30] suggests a more demanding improvement before selecting the more complex model. Other proposed tests include the Vuong inspired LMR[31] likelihood ration (LR) test of a K class model versus a K-1 class model in which the test statistic is

Downloaded from http://smm.sagepub.com at UCLA on July 6, 2010

compared against an approximation to a mixture of likelihoods, and a parametric bootstrap likelihood ratio test.[32] Nylund *et al.*[33] undertook a range of simulations to examine the validity of these tests and criteria. The LMR test was a substantial improvement over the standard LR test that consistently over-estimated the number of classes, but the parametric bootstrap LR test generally outperformed the LMR test in both power and Type-1 error rate. The standard BIC outperformed both the AIC and the sample-size adjusted BIC. The parametric bootstrap LR test marginally outperformed the BIC, but has the disadvantages of assuming that the K-1 class model is correctly specified, being much more computationally intensive, and being difficult to apply in small sample sizes where not all bootstrap samples may deliver valid solutions. Lukociene and Vermunt[34] have reported work on the appropriate choice of n in these formula for clustered or repeated measures data settings, where a value reflecting either individuals or groups could be chosen (they argued for the n of clusters).

Whether the classes are theoretically or empirically determined, both these perspectives essentially regard the classes as 'real' (see Pickles and Angold,[35] Nagin[8] and Kreuter and Muthén[15] for further discussion) or sufficiently so to be used to guide theory development, policy formulation or clinical decision-making.[18,36]

By contrast, the third perspective regards the discrete masses merely as a way of representing what may be a non-normal distribution of the random coefficients or factors. In a limited number of simple cases it can be shown theoretically that the non-parametric maximum likelihood (NPML) estimator of a random effect or latent variable is a finite set of discrete masses.[37,38] Adding more and more classes does not give a progressive increase in likelihood but instead reaches a maximum, after which additional masses are assigned zero-mass or locate at the same position as an existing mass. Even where the underlying distribution is known and continuous, this non-parametric representation gives a better fit to the data than does the correct distribution itself. While the range of models for which it has been shown to apply theoretically is small and the models are simple, empirically it has been shown to apply more generally (e.g. Davies and Pickles[39] and Rabe-Hesketh *et al.*[40]). In these more complex settings, though the number of classes required to achieve a NPML estimator is surprisingly small, commonly less than 10 and often much fewer if the response variables are categorical, the number of classes identified is still often larger than most theory would support as distinct types and some classes may be considered too small to be of clinical significance. Thus, from this position the classes are not seen as representing an actual typology but are instead merely a non-parametric representation of some unknown, probably continuous, distribution (an idea pursued further by Kreuter and Muthén[15]).

As finite mixture modellers are only too aware, the likelihood surfaces are often multimodal.[41] It is necessary therefore for careful (or automated and exhaustive) checks to be made on the optimality of the reported solution, and its stability. For example, both Mplus and Latent Gold software provide estimation from multiple starting points to enable the user to check that a global optimum has been achieved, while `gllamm` implements a different procedure that performs a grid search for the best location for adding a K-th class to the $K-1$ class solution.

In the study of Levene and Rabinowitz[12] (cited above), this discrete random effects approach is applied to the therapeutic context of treatment for early schizophrenia. Essentially what they describe is the use of latent class trajectory models of in-treatment

Downloaded from http://smm.sagepub.com at UCLA on July 6, 2010

response trajectory for repeated measures of severity of symptoms (up to 24 months). They used an implementation of the model (which they referred to as *mixed-mode latent class regression*, following the user-written R library that was used for parameter estimation) as a model based clustering tool to assign individuals to 'recovery' classes. The association of MAP class assignments to a range of pre-treatment participant characteristics was then examined using ANOVA where MAP class was a simple between-subjects factor. Of course this association of baseline covariates with trajectory class can also be estimated jointly inside the trajectory model. Usually a multinomial model is used to link covariates to class probabilities such that

$$\pi_i(c = k) = \exp(X_i\beta_k) \Big/ \sum_{k=1\ldots K} \exp(X_i\beta_k)$$

Such models allow association with treatment to be estimated, with or without further covariate adjustment in the fixed-part of the model. However, the treatment effect may be distributed across several classes and may include association to trajectory groups that are not unequivocally beneficial, for example showing early response followed by later relapse. An overall treatment effect may be obtained by associating trajectories with some ultimate outcome as part of the model or by defining some class specific summary statistic, such as area-under-the-curve, that allows the relative benefits of one trajectory over another to be compared and a weighted sum over classes to be calculated.

4 Group-based latent growth (LCGA) model applied to SoCRATES

For illustration we fitted a two-class linear growth model to the positive symptom profiles on the log time scale using `gllamm`. Assigning participants according to their MAP, Figure 2 shows their raw symptom profiles by treatment group and class. Class 1, 81% of the sample, evidently improves more rapidly than Class 2. A simple chi-square association test of treatment allocation group and participant MAP class gave a marginally significant Pearson chi-square (df = 1) of 3.05. Estimating the association simultaneously with the trajectory model gave a log-odds coefficient for the association of treatment 1.103 (SE adjusted for therapist clustering 0.528).

We have no theoretical justification for restricting the model to two classes. Table 1 gives various criteria (from Mplus[42]) for assessing models from the same family but with an increasing number of classes. Here the minimum BIC might suggest four classes whereas the parametric bootstrapping would suggest five. Further additional classes were also small in size. Figure 3 shows the groups means for the five-class solution. The great majority of participants (85%) showed variable but substantial initial reduction in symptoms that was maintained. But both the four- and five-class solutions identified a group whose symptoms tended to return to post-treatment levels during follow-up, while the five-class solution identified a further group (4.7%) who had also shown little improvement in the treatment phase. Both these latter groups would be of interest to clinical researchers concerned either to better target treatment on patients most likely to be responsive or to explore what modifications to the treatment protocol might allow a better outcome.

Downloaded from http://smm.sagepub.com at UCLA on July 6, 2010

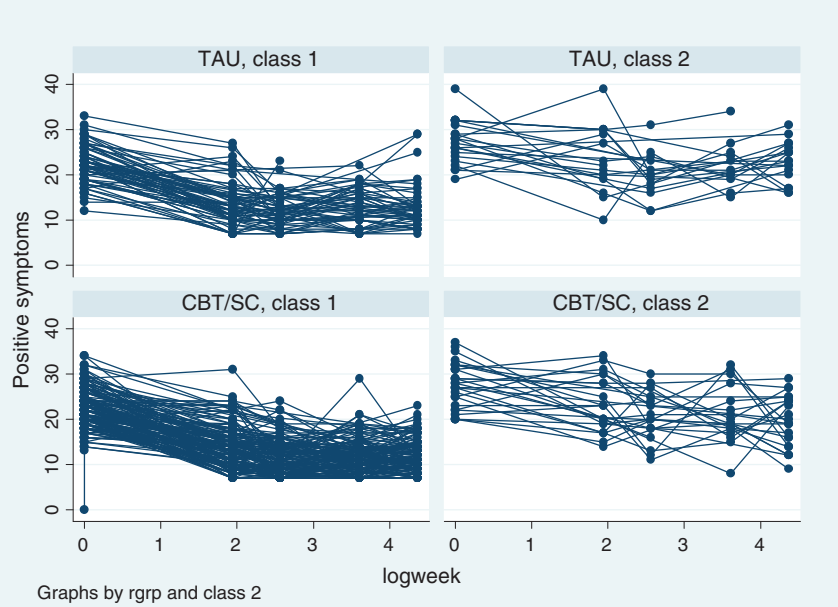

Figure 2 Participant positive symptom profiles by treatment and maximum *a posteriori* probability (MAP) class (SoCRATES).

Table 1 Fit and test criteria for group-based (LCGM) with increasing classes for the positive symptoms (SoCRATES)

Model: all in log (time + 1) metric 3 non-parametric random effects	Likelihood	Number of free params	AIC	BIC	ssaBIC	BLRT p value for $k-1$ latent classes	Vuong-Lo-Mendell-Rubin likelihood ratio test for k versus $k-1$ classes
Latent classes							
2	−3481.23	11	6984.45	7025.52	6990.63	<0.001	<0.001
3	−3462.02	15	6954.03	7010.03	6962.46	<0.001	<0.001
4	−3441.81	19	6921.62	6992.55	6932.29	<0.001	0.180
5	−3431.21	23	6908.40	6994.28	6921.34	0.0128	0.347

Table 2 gives mean classification probabilities after allocation for this five-class model. This model gave an entropy of 0.727 and with the exception of the intermediate response group 5, gave mean individual MAPs of 0.84 and above.

5 Multivariate outcome

For trajectories in more than one measure there are two distinct approaches to generalisation. The first is to consider a set of single measure growth models (Figure 4), one

Downloaded from http://smm.sagepub.com at UCLA on July 6, 2010

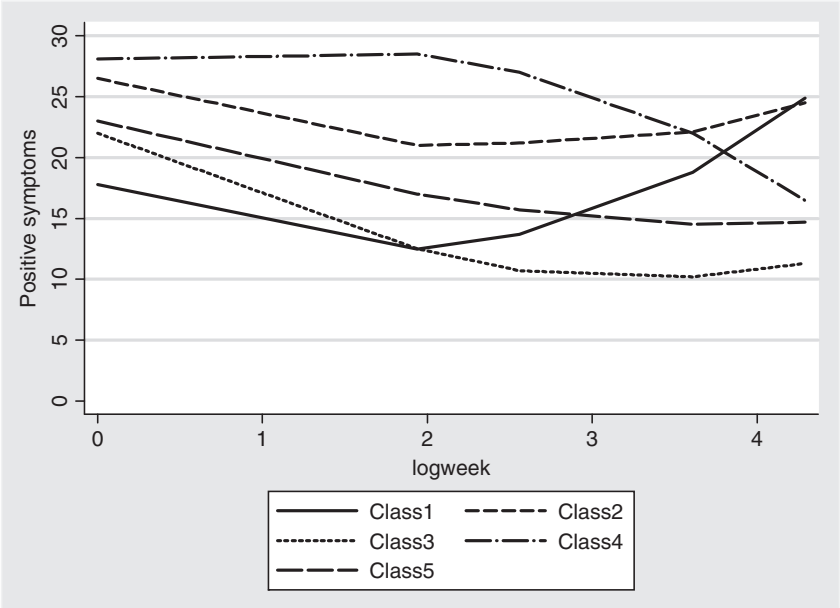

Figure 3 Five group latent growth (LCGA) model mean positive symptom profiles for each class by logweek (SoCRATES).

Table 2 Classification quality and class allocations based on MAPs (Mplus Version 5.1)

Latent class	$n\hat{}$	%	Classification accuracy				
			1	2	3	4	5
1	7	2.265	*0.838*	0.018	0.095	0.000	0.049
2	18	5.825	0.007	*0.841*	0.000	0.044	0.108
3	165	53.398	0.019	0.005	*0.855*	0.002	0.118
4	11	3.560	0.000	0.075	0.000	*0.907*	0.018
5	108	34.951	0.009	0.052	0.160	0.031	*0.748*

Classification of individuals based on most likely class membership (highest posterior probability). Entropy estimated by Mplus: 0.727. Entries in italics are average latent class probabilities for class membership, by class for which probability is highest (MAP).

for each measure k, $k = 1, \ldots, K$ (where K is now used for outcome, not number of classes) with observations $\{y_{ikt}\}$ such that

$$y_{ikt} = \alpha_k + \beta_k t + a_{ik} + b_{ik} t + \varepsilon_{ikt}$$

Assumptions as to the correlation structure of the $2 \times K$ random coefficient vector $(a_{i1}, b_{i1}, \ldots, a_{iK}, b_{iK})$ must be made. For discrete trajectory classes the equivalent of an unstructured covariance matrix is obtained by allowing the class probability masses to range freely over the $2 \times K$ dimensions. Alternatively, constraints can be imposed such

Downloaded from http://smm.sagepub.com at UCLA on July 6, 2010

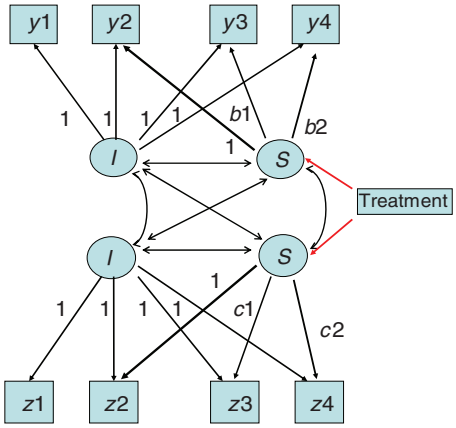

Figure 4 Correlated growth models for a bivariate flexible ($b1$, $b2$, $c1$, $c2$) response profile.

that the multidimensional classes are considered as arising from a cross-classification of classes defined on each of K dimensions. In either formulation, the occasion-specific measurement errors may also be correlated across k, adding to the complexity of estimation. Treating the observations as clustered and using robust parameter covariance matrix estimators may be adequate for some purposes, but will preclude the use of almost all the criteria suggested for empirically determining the optimum number of classes.

The second approach, more suited to the circumstance where it is expected that the profiles of development will be similar across measures, is to assume a factor model for the k measures at each time t, η_{it}, implying that the measures while being on different scales, measure the same underlying construct (Figure 5). A change over time in the underlying construct is described by a single growth curve model as before.

$$y_{ikt} = \alpha_k + \lambda_k \eta_{it} + \varepsilon_{it}$$
$$\eta_{it} = \eta^*_{1i} + \lambda_t \eta^*_{2i} + d_{it}$$

For trajectory class models it is the distribution over higher order factors or random coefficients η^* that is specified as discrete rather than the more usual Gaussian.

6 Bivariate group-based latent growth (LCGA) model applied to SoCRATES

Applied to the joint positive and negative symptom profiles of the SoCRATES dataset, four classes were defined by the cross-classification of being 'high' or 'low' on each of P and N symptom profiles. A class was thus defined by six locations, with intercept, linear and quadratic slopes for each of P and N, with constraints on class locations to ensure that 'high' on P in the high-P/High-N class is the same as that in the high-P/Low-N class, etc. Table 3 compares the results from two models estimated in Mplus,[42]

Downloaded from http://smm.sagepub.com at UCLA on July 6, 2010

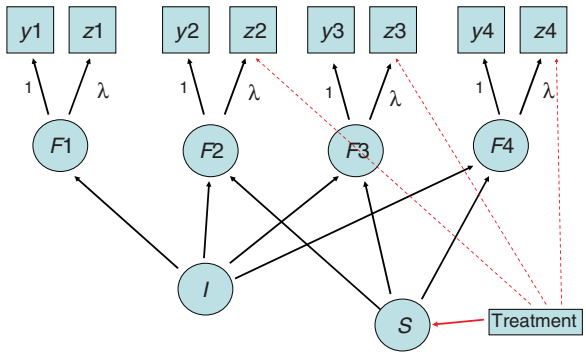

Figure 5 Latent factor growth curve model for a bivariate response profile.

Table 3 Fit criteria with and without correlated errors for a bivariate latent growth/LCGA model with four classes located on a rectangle in the 2 × 2 positive by negative symptoms space

Bivariate model	Class sizes-after modal allocation	AIC	BIC	ssaBIC	Covariance Estimation (S.E.)	P classes Response Variable	N classes Response Variable
2 classes *P* by 2 classes *N* Log likelihood −7159.933	*P*1/*N*1 $n = 26$ *P*1/*N*2 $n = 37$ *P*2/*N*1 $n = 47$ *P*2/*N*2 $n = 199$	14357.87	14428.80	14368.54	None	18.081 (1.055)	21.125 (1.345)
2 classes *P* by 2 classes *N* (*P* with *N* covariance) Log likelihood −7088.839	*P*1/*N*1 $n = 27$ *P*1/*N*2 $n = 31$ *P*2/*N*1 $n = 49$ *P*2/*N*2 $n = 202$	14217.68	14292.34	14228.91	7.474 (0.832)	18.376 (1.072)	21.679 (1.528)

one that assumes the occasion-specific errors for positive and negative symptoms to be uncorrelated, and the second model, which clearly fits much better, allows them to be correlated within occasion. Although the class sizes differ relatively little between the models, it is likely that the tests and criteria used to assess the number of classes required could be affected substantially by the appropriate choice of error covariance.

7 Allowing the classes to differ in more than means: mixtures of growth curve models

In all the models described above the discrete distribution corresponding to potential latent classes has been used to replace a continuous distribution for the random coefficients of a more traditional growth curve model. Except in the case of NPML representation, the discrete model is not a generalisation of the continuous model in the sense of the continuous model being nested within it. However, such a nesting is possible if instead of the latent class representing a group of participants with a common mean

Downloaded from http://smm.sagepub.com at UCLA on July 6, 2010

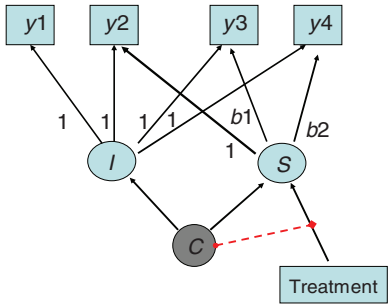

Figure 6 GMM for a single response profile and between class treatment effect heterogeneity.

and residual variance structure, the latent classes should represent participants with a common mean and overall covariance structure i.e. we allow each latent class to have growth structure variability within it. Following Muthén[10] we refer to this as a GMM. In practice it rarely seems possible to allow unrestricted differences in the covariance structure, but even allowing for a common within class random growth structure this approach is more general and does not assume that the only within class variation is in time-specific errors. However, the implied classes are now heterogeneous, rather more overlapping and perhaps less easily considered typologically (Figure 6).

Many applied studies e.g. Stulz *et al.*[14] cited earlier, use this more general random effects mixture model,[43] which in software tutorials by applied journals is addressed alongside the group-based latent growth model (LCGA).[11] Current practice in this area largely follows guidance offered by Muthén and Muthén.[9,10] Typical reports use the full range of BIC, AIC, entropy and LMR LR tests to choose the optimal number of classes and dimensions of between class variability in applications described as GMMs. Muthén and Asparouhov[44] extending the original exposition in Muthén *et al.*[1] apply this to a small set of pharmaceutical trial data on antidepressant treatment. Separate analysis of the placebo group finds evidence of a placebo response trajectory class with a strong initial improvement, followed by a later worsening. A separate analysis of the medication group shows two types of responder classes, one with an initial improvement only and one with a sustained improvement. A joint analysis of the placebo and medication groups makes it possible to estimate medication effects in the presence of placebo-response effects and shows benefits of medication.

8 Latent classes for compliance and drop-out

In addition to data defining outcome trajectory, compliance data may also be included in the analysis. This allows the estimation of the Complier-Average-Causal-Effect (CACE). Often this coincides or is strongly associated with missing data and thus overlaps with approaches to missing data such as pattern-mixture models.[22]

Consider the growth curve mixtures model in the previous section. We may consider the outcome trajectory data as being a mixture of data derived from a set of growth models, where each component is formed by a class of participants that differ on their level of compliance. In the treated group, compliance data is assumed to be available,

Downloaded from http://smm.sagepub.com at UCLA on July 6, 2010

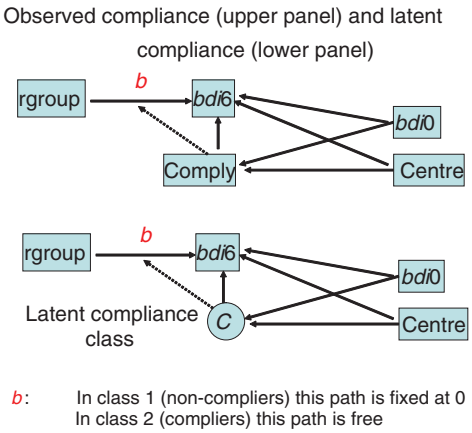

Figure 7 Path diagram for complier average causal effect for a single outcome (*bdi*6), with baseline covariates (*bdi*0 and centre) when compliance is unobserved for participants assigned to the non-active treatment randomisation group.

and hence is informative as regards compliance class membership. In the untreated participants compliance class is commonly not observed, but due to randomisation, we might be willing to assume that compliance class prevalences are equal to those in the treated group (Figure 7).

In the simplest case, compliance is considered to be the binary compliant ($c = 1$) or non-compliant ($c = 0$). In the treated group ($r = 1$), we observe the two conditional distributions $f(y|\ r = 1,\ c = 1)$ or $f(y|\ r = 1,\ c = 0)$ while in the control group ($r = 0$) we can observe only the marginal $\pi f(y|\ r = 0,\ c^* = 1) + (1 - \pi)f(y|r = 0,\ c^* = 0)$ where π is the proportion of $c = 1$ in the treated group and c^* is latent (unobserved) compliance. Estimates are then obtained for the effect of treatment within strata defined by the partially observed compliance groups. Skrondal and Rabe-Hesketh (pp. 427–32)[45] provide an example in relation to the benefits of a social intervention (job-training) on depression.

Morgan-Lopez and Fals-Stewart[46] consider the use of this kind of model for estimation of treatment effects in trials in which therapy is delivered by open-enrollment groups. Over and above the fact that the trials are clustered, with therapy being delivered within groups,[47] these have the additional complexity that the groups are not formed at a single point in time and then closed to new participants, but instead have a membership that evolves as participants are recruited into the groups and others drop out. Morgan-Lopez and Fals-Stewart consider a model of similar form to that illustrated in Figure 8. In this model data on participation $d1$–$d4$ at each of four occasions is used to form latent participation classes, say $c = k$ for $k = 1, \ldots, K$. Class membership may also be correlated with the week of entry into the trial (start week) and might be influenced by treatment assignment (dotted arrow). The dashed lines indicate that the effect of treatment may vary by participation class (though because of randomisation the average treatment over latent classes on the intercept should be zero).

Such a model delivers treatment effects by latent participation class, with an average effect being obtained as the mean of these estimates weighted by class size.

Downloaded from http://smm.sagepub.com at UCLA on July 6, 2010

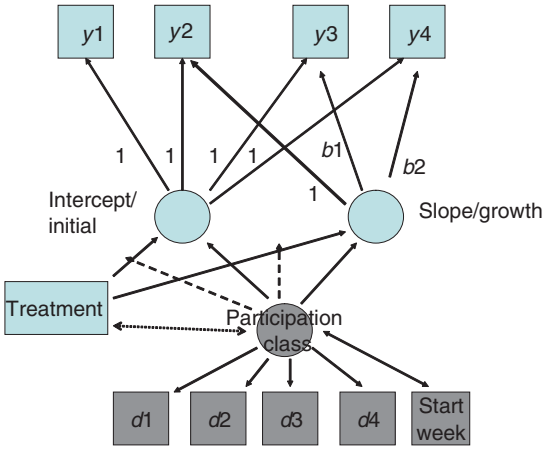

Figure 8 Participation data pattern-mixture latent class model for a growth curve with intercept, slope and treatment effect differences by class.

Their simulation results suggest that the average treatment effect estimated from this mixture model is less biased than the standard simple growth curve model estimated under missing-at-random assumptions even when the number of latent classes is smaller than that used to simulate the data. Beunckens *et al.*[48] give a latent-class mixture treatment in the more standard incomplete data case.

As the diagram makes clear the association between participation class and response is via the growth intercept and slope. This rules out the more immediate or contemporaneous association of non-participation and response that you might expect when examining, say, a drug treatment on a causally proximal biomarker response.

9 Models for mediation and moderation

In a similar fashion to the formation of latent compliance groups, Emsley *et al.*[49] in this issue make use of discrete latent classes to operationalise the principal stratification approach to examine moderation and mediation. In that case, the inability to observe the value of a binary mediating (or moderating) variable under the counterfactual treatment regime means that the expected counterfactual outcome is derived from a consideration of the outcomes under the two latent states of the mediating (moderating) variable.

10 Instrumental variables

Models for causal effect estimation based on the principle of instrumental variables (IV) have a long history. Their use and equivalence to other approaches to effect estimation, such as nested marginal structural means models are examined elsewhere in this issue. An IV model structure for a growth mixture application is shown in Figure 9, in which the instrument, usually random group assignment and interactions of this with baseline characteristics are allowed to influence the growth mixture parameters only through treatment received.

Downloaded from http://smm.sagepub.com at UCLA on July 6, 2010

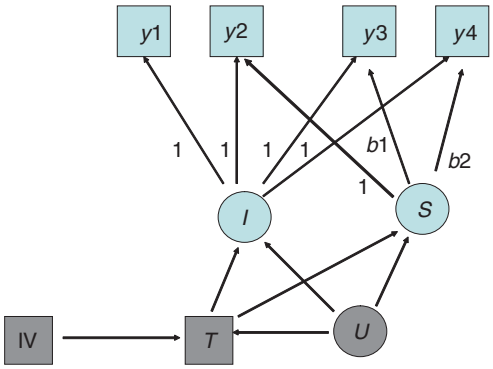

Figure 9 Growth curve model within which causal effects on a response profile of partly non-randomly assigned treatment T with unobserved confounder U is identified by an IV.

These variables are taken as IVs on the basis of one of a number of alternative restrictions that broadly rule out direct effects of the IVs on the profile of responses (and that all indirect effects occur through difference in treatment exposure). Such assumptions identify the model in which treatment exposure and outcome response are correlated due to selection and the presence of confounders.

As yet there has been little work to develop the use of IVs in the latent class/growth mixture setting. It is important to distinguish the settings where the treatment allocations may be two or more treatment regimes,[50] requiring binomial or multinomial treatment selection/exposure models, from those where there are two or more trajectory classes, requiring binomial or multinomial class membership models. Further work is also required to clarify the interpretation to be given to the class-specific treatment effect estimates when within class treatment effect heterogeneity (essential heterogeneity[51]) is being postulated.

While simultaneous estimation of a joint model of treatment exposure and growth curve response can be undertaken, there may be scope for simpler two-stage methods that begin by estimating predicted treatment based on covariates and IVs, and then replacing treatment by predicted treatment in the LCGA or GMM.

11 Using latent classes for pre-treatment trajectories

Havilland and Nagin[25] and Haviland *et al.*[52] consider the case where there is an extended pre-treatment record for the response variable and this history is likely to be correlated with treatment receipt. They examine use of trajectory classes formed from these pre-baseline histories as a means to achieve a data reduction prior to propensity score construction or as an alternative to propensity scores. They considered assigning participants to their pre-treatment response history class according to their MAP and treatment effects estimated by class or pooled across classes weighted by prevalence. Alternatively, participants can be considered as potentially belonging to all classes and class-specific treatment effects/pooled effects estimated by an average across all participants weighted

Downloaded from http://smm.sagepub.com at UCLA on July 6, 2010

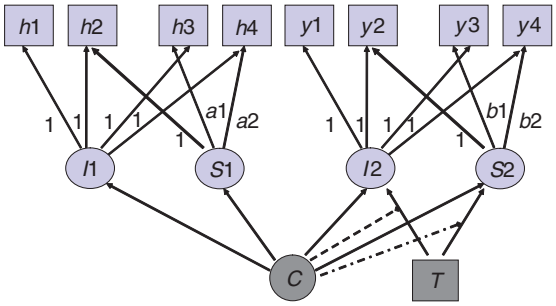

Figure 10 GMM pre-treatment growth history class to post-treatment growth response class transition model.

by their class specific posterior probabilities. If the treatment effects and history trajectory classes were estimated simultaneously the model would be structured similarly to that of Morgan-Lopez and Fals-Stewart shown in Figure 8, where d1–d4 are replaced by pre-baseline measures of *y*.

As in the case considered earlier of forming trajectory classes over response profiles for more than one measure, so too can classes be formed over both pre-treatment and post-treatment response histories. This formulation allows the effect of treatment to be considered explicitly as influencing the probability of switching between trajectory latent classes. Figure 10 shows such a model adapted from the trial setting of antidepressant medication treatment with a repeat as described by Muthén *et al*.[53]

12 Discussion

There can be little doubt regarding the potential value of the group-based approaches for exploratory analysis of differential response to therapy. The combination of being able to use the model specification to define the space of variability to be considered, combined with the relatively non-parametric nature of the class identification allows scope for both clinical insight to guide search and yet for counter-intuitive findings driven by the data to show through. The approach responds to the need to investigate for whom a treatment is effective by allowing for different treatment effects in different trajectory classes. Caution is nonetheless required in the interpretation of the classes. GMMs, in allowing within group variability in growth trajectories, are likely to lead to less distinctive classifications and a wider area of the response profile space where class membership is more equivocal. They may, however, be more realistic and better fit the data than the group-based latent growth (LCGA) models in which within-group homogeneity of trajectory is assumed.

The models may also be viewed as data-reduction tools, collapsing variability across a number of dimensions into a typology, whether that is informed primarily by patterns of outcome response, patterns of pre-treatment response history, patterns of treatment participation or patterns of missingness.

Downloaded from http://smm.sagepub.com at UCLA on July 6, 2010

Further work is required in the formalisation of these methods into both the traditional framework of pre-specified analysis plans and into the more formal frameworks of causal inference.

Acknowledgements

AP's contribution was supported by MRC grant G0600555 to Dunn, Davies, Green, Pickles, Roberts and Windjmeier. TC is funded by a Career Scientist Award from the UK Department of Health and his contribution was facilitated by the NIHR CLAHRC for Cambridgeshire & Peterborough. Thanks to investigators and participants of the SoCRATES trial for use of data for illustration and the UK Mental Health Research Network (MHRN) for facilitating collaboration.

References

1 Muthén B, Brown CH, Masyn K *et al*. General growth mixture modeling for randomized preventive interventions. *Biostatistics* 2002; **3**: 459–75.
2 Moffitt TE. Adolescence-limited and life-course persistent antisocial behavior: a developmental taxonomy. *Psychological Review* 1993; **100**: 674–701.
3 Moffitt TE, Caspi A, Dickson N, Silva P, Stanton W. Childhood-onset versus adolescent-onset antisocial conduct problems in males: natural history from ages 3 to 18 years. *Development and Psychopathology* 1996; **8**: 399–424.
4 D'Unger A, Land K, McCall P, Nagin D. How many latent classes of delinquent/criminal careers? Results from mixed Poisson regression analyses of the London, Philadelphia, and Racine cohort studies. *American Journal of Sociology* 1998; **103**: 1593–630.
5 Broidy LM, Nagin DS, Tremblay RE *et al*. Developmental trajectories of childhood disruptive behaviors and adolescent delinquency: a six-site cross-national study. *Developmental Psychology* 2003; **39**(2): 222–45.
6 Croudace TJ, Jarvelin MR, Wadsworth ME, Jones PB. Developmental typology of trajectories to nighttime bladder control: epidemiologic application of longitudinal latent class analysis. *American Journal of Epidemiology* 2003; **157**(9): 834–42.
7 Nagin DS. Analyzing developmental trajectories: a semi-parametric, group-based approach. *Psychological Methods* 1999; **4**: 139–77.
8 Nagin DS. *Group-based modeling of development*. Harvard University Press, Cambridge, MA; 2005.
9 Muthén B, Muthén L. Integrating person-centered and variable-centered analyses: growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research* 2000; **24**: 882–91.
10 Muthén B. Second-generation structural equation modeling with a combination of categorical and continuous latent variables: new opportunities for latent class-latent growth modeling. In Collins LM, Sayers AG, eds. *New methods for the analysis of change*. American Psychological Association, Washington, DC; 2001: pp. 291–322.
11 Jung T, Wickrama KAS. An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass* 2008; **2/1**: 302–17.
12 Levine SZ, Rabinowitz J. Trajectories and antecedents of treatment response over time in early-episode psychosis. *Schizophrenia Bulletin*, epub 2008.
13 Laurenceau JP, Hayes AM, Feldman GC. Some methodological and statistical issues in the study of change processes in psychotherapy. *Clinical Psychology Review* 2007; **27**(6): 682–95.
14 Stulz N, Lutz W, Leach C, Lucock M, Barkham M. Shapes of early change in

Downloaded from http://smm.sagepub.com at UCLA on July 6, 2010

psychotherapy under routine outpatient conditions. *Journal of Consulting and Clinical Psychology* 2007; **75**: 864–74.
15 Kreuter F, Muthén B. Longitudinal modeling of population heterogeneity: methodological challenges to the analysis of empirically derived criminal trajectory profiles. In Hancock GR, Samuelsen KM, eds. *Advances in latent variable mixture models*. Information Age Publishing, Charlotte, NC; 2007; 53–75.
16 Nagin DS, Land KC. Age, criminal careers, and population heterogeneity: specification and estimation of a nonparametric, mixed Poisson model. *Criminology*, 1993; **31**: 327–62.
17 Roeder K, Lynch K, Nagin DS. Modeling uncertainty in latent class membership: a case study in criminology. *Journal of the American Statistical Association,* 1999; **94**: 766–76.
18 Nagin DS, Tremblay, RE. Developmental trajectory groups: fact or a useful statistical fiction? *Criminology* 2005; **43**: 873–903.
19 Gueorguieva R, Wu R, Pittman B *et al*. New insights into the efficacy of naltrexone based on trajectory-based reanalyses of two negative clinical trials. *Biological Psychiatry* 2007; **61**(11): 1290–5.
20 Wang CP, Brown CH, Bandeen-Roche K. Residual diagnostics for growth mixture models: examining the impact of a preventive intervention on multiple trajectories of aggressive behavior. *Journal of the American Statistical Association* 2005; **100**: 1054–76.
21 Muthén, B, Shedden K. Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics* 1999; **55**: 463–9.
22 Yau LHY, Little RJ. Inference for the complier-average causal effect from longitudinal data subject to noncompliance and missing data, with application to a job training assessment for the unemployed. *Journal of the American Statistical Association* 2001; **96**: 1232–44.
23 Roy J. Modeling longitudinal data with nonignorable dropouts using a latent dropout class model. *Biometrics* 2003; **59**(4): 829–36.
24 Roy J. Latent class models and their application to missing-data patterns in longitudinal studies. *Statistical Methods in Medical Research* 2007; **16**(*5*):441–56.
25 Haviland A, Nagin DS. Causal inference with group-based trajectory models. *Psychometrika* 2005; **70**: 1–22.
26 Lewis S, Tarrier N, Haddock G *et al*. Randomised, controlled trial of cognitive-behavioural therapy in early schizophrenia: acute-phase outcomes. *British Journal of Psychiatry* 2002; **181**(suppl. 43): s91–s97.
27 Tarrier N, Lewis S, Haddock G *et al*. Cognitive-behavioural therapy in early schizophrenia: 18 month follow up of a randomised, controlled trial. *British Journal of Psychiatry* 2004; **184**: 231–9.
28 Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophrenia Bulletin* 1987; **13**: 261–76.
29 Schwartz G. Estimating the dimension of a model. *The Annals of Statistics* 1978; **6**: 461–4.
30 Raferty AE. Bayesian model selection in social research. *Sociological Methodology* 1995; **25**: 111–96.
31 Lo Y, Mendell NR, Rubin DB. Testing the number of components in a normal mixture. *Biometrika* 2001; **88**: 767–78.
32 McLachlan GJ, Peel D. *Finite mixture models*. John Wiley and Sons, New York; 2000.
33 Nylund KL, Asparouhov A, Muthén B. Deciding on the number of classes in latent class analysis and growth mixture modelling. A Monte Carlo study. *Structural Equation Modelling Journal* 2007; **14**(4): 535–69.
34 Lukociene O, Vermunt JK. Determining the number of components in mixture models for hierarchical data. In *Studies in classification, data analysis, and knowledge organization*. Springer, Berlin-Heidelberg; 2008.
35 Pickles A, Angold A. Natural categories or fundamental dimensions: On carving nature at the joints and the re-articulation of psychopathology. *Development and Psychopathology* 2003; **15**: 529–51.
36 Bauer DJ, Curran PJ. Distributional assumptions of growth mixture models: Implications for over-extraction of latent trajectory classes. *Psychological Methods* 2003; **8**: 338–63.
37 Laird N. Nonparametric maximum likelihood estimation of a mixture distribution. *Journal of the American Statistical Association* 1978; **73**: 805–11.
38 Lindsay B, Clogg C, Grego J. Semiparametric estimation in the Rasch model and related exponential response models, including a simple latent class model for item analysis.

Downloaded from http://smm.sagepub.com at UCLA on July 6, 2010

Journal of the American Statistical Association 1991; **86**: 96–107.
39 Davies RB, Pickles AR. A joint trip-timing/store-type choice model for grocery shopping including feedback, inventory effects and non-parametric control for omitted variables. *Transportation Research* 1987; **21**: 345–61.
40 Rabe-Hesketh S, Pickles A, Skrondal A. Correcting for measurement error in logistic regression using non-parametric maximum likelihood estimation. *Statistical Modelling* 2003; **3**: 215–32.
41 Hipp JR, Bauer DJ. Local solutions in the estimation of growth mixture models. *Psychological Methods*, 2006; **11**: 36–53.
42 Muthén LK, Muthén B. *Mplus user's guide* (4th edn). Muthén & Muthén, Los Angeles, CA; 1998–2007.
43 Verbeke G, Lesaffre E. A linear mixed-effects model with heterogeneity in the random effects population. *Journal of the American Statistical Association* 1996; **91**: 217–21.
44 Muthén B, Asparouhov T. Growth mixture modeling: analysis with non-Gaussian random effects. In Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G, eds. *Advances in longitudinal data analysis*. Chapman & Hall/CRC Press, London; 2007; 143–165.
45 Skrondal A, Rabe-Hesketh S. *Generalized latent variable modelling*. Chapman & Hall/CRC Press, London; 2004.
46 Morgan-Lopez AA, Fals-Stewart W. Consequences of misspecifying the number of latent treatment attendance classes in modeling group membership turnover within ecologically valid behavioral treatment trials. *Journal of Substance Abuse Treatment* 2008; **35**: 396–409
47 Roberts C, Roberts SA. Design and analysis of clinical trials with clustering effects due to treatment. *Clinical Trials* 2005; **2**(2):152–62.
48 Beunckens C, Molenberghs G, Verbeke G, Mallinckrodt C. A latent-class mixture model for incomplete longitudinal Gaussian data. *Biometrics* 2008; **64**(1): 96–105.
49 Emsley R, White I, Dunn G. Mediation and moderation of treatment effects in randomised controlled trials of complex interventions. *Statistical Methods in Medical Research* 2010; **19**(3): 237–70.
50 Heckman JJ, Urzua S, Vytlacil E. Understanding instrumental variables in models with essential heterogeneity. *Review of Economics & Statistics*, 2006; **88**: 389–422.
51 Heckman JJ, Urzua S, Vytlacil E. Instrumental variables in models with multiple outcomes: the general unordered case. Discussion Paper 3565, IZA Institute for the Study of Labor, P.O. Box 7240, 53072 Bonn, Germany, 2008.
52 Haviland A, Nagin DS, Rosenbaum PR. Combining propensity score matching and group-based trajectory analysis in an observational study. *Psychological Methods* 2007; **12**(3): 247–67.
53 Muthén B, Brown H, Leuchter A, Hunter A. General approaches to analysis of course: applying growth mixture modeling to randomized trials of depression medication. In Shrout PE, ed. *Causality and psychopathology: finding determinants of disorders and their cures*. American Psychiatric Association, Washington, (In Press).

Downloaded from http://smm.sagepub.com at UCLA on July 6, 2010