Multilevel structural equation models for longitudinal data where predictors are measured more frequently than outcomes: An application to the effects of stress on the cognitive function of nurses

Fiona Steele,[1] Paul Clarke,[2] George Leckie,[3] Julia Allan,[4] Derek Johnston[4]

[1]London School of Economics, [2]University of Essex, [3]University of Bristol, [4]University of Aberdeen

Summary. Ecological momentary assessment (EMA) is used to measure subjects' mood and behaviour repeatedly over time, leading to intensive longitudinal data. Variability in EMA assessment schedules creates an analytical challenge because predictors are measured more frequently than responses. We consider this problem in a study of the effect of stress on the cognitive function of telephone helpline nurses, where stress is measured for each call and cognitive outcomes are measured at the end of a shift. We propose a flexible structural equation model which can handle multiple levels of clustering, measurement error, time trends, and mixed variable types.

Keywords: ecological momentary assessment; real-time assessment; high-frequency data; intensive longitudinal data; multilevel latent variable model; simultaneous equation model; occupational stress

1. Introduction

Increasingly, social and medical researchers are interested in determining how people's thoughts, feelings and actions in everyday life influence their behaviour and their likelihood of health-related events. For example, how exposure to stress can lead to relapses in smoking; how the experience of pain over time determines patient's treatment decisions; how workplace events affect employee performance; how social interactions influence mental health; and so on. In order to assess the relationships between these real-world exposures and events of interest, researchers must be able to collect and analyse longitudinal data. 'Ecological momentary assessment' (EMA; Stone and Shiffman 1994) is one such methodological approach. EMA relies on the repeated assessment of participants' current states in their natural environment, and as such maximizes the ecological validity of data while avoiding problems traditionally associated with retrospective recall (Tourangeau 2000). EMA assessment schedules vary widely, from collecting infrequent data over an extended time period (e.g. Jamison *et al.* (2001); daily data over 1 year) to extremely frequent data collection over a more compressed period (e.g. Shapiro *et al.* (2002); data collection every 45 minutes for 4 days). The flexibility of the EMA approach allows researchers to examine dynamic temporal associations between variables over time, but the use of different measurement schedules for different variables leads to complex longitudinal data that are often challenging to analyse (Bolger and Laurenceau 2013, Walls and Schafer 2006).

The particular question we consider is how to analyse the relationship between a predictor and an outcome using EMA data which contain frequent measures of the predictors but relatively infrequent measures of the outcomes. This question is motivated by data from an investigation into the effect of stress (predictor) on the cognitive function (outcome) of nurses who work on the Scottish NHS24 healthcare telephone service. For each nurse, EMA data were collected over two working shifts, with stress measured after each call, and cognitive function measured at the start and the end of each shift (Allan *et al.* 2009, Farquharson *et al.* 2012). While these data have a three-level hierarchical structure (with calls at level 1 nested within shifts at level 2 and nurses at level 3), the design is nonstandard because the outcome is defined at level 2 rather than level 1.

In this paper, motivated by the richness of the NHS24 data set and the analytical problems it poses, we develop a family of multilevel structural equation models (SEM) for analysing data with the aforementioned structure. Following the approach of Preacher *et al.* (2010), we specify an SEM which comprises two multilevel models: a measurement model for a latent variable representing the 'true' level of the predictor over the observation period; and a structural model in which the latent variable predicts the outcome. However, there are features of the NHS24 study design which require us to extend the model as follows:

- First, stress is the level 1 predictor in NHS24 and it is measured on an ordinal scale, but previous models for level 1 predictors and higher-level outcomes have treated both the predictor and outcome as continuous variables (Croon and van Veldhoven 2007, Griffin 1997, Preacher, *et al.* 2010). Hence, we propose a model that allows for the ordinal measurement of stress *and* non-continuous outcomes.

- Second, each nurse is recorded over two shifts and so the data have a three-level structure. However, in the case of a level 1 predictor and higher-level outcome, most previous research has focussed on two-level designs where the model for the outcome is a single-level model, which means it is not possible to separate the between and within effects of a level 1 predictor. But it is generally desirable to separate the effect of a level 1 predictor into between-cluster and within-cluster components (e.g. Curran and Bauer 2011). This decomposition is particularly important for NHS24 because both types of effect are of interest: the between-nurse component represents time-invariant influences on stress, such as the nature of the job and the nurse's ability to cope with stressful situations; and the within-nurse component represents influences specific to a particular shift, which make that shift more or less stressful than is usual for a given nurse; for example, the nature of the calls handled during a shift and external influences which vary from day to day. When both the predictor and outcome variables are self-reported, as with NHS24, we argue that the within-subject effect is of greater interest because it represents the effect of the predictor adjusted for unobserved time-invariant reporting tendency (captured by the between-subject effect). Hence, we propose a three-level model that allows these effects to be identified.

- Third, previous research has used a measurement model which ignores all temporal information on the predictors and requires us to assume that the predictor is constant over the observation period. Curran and Bauer (2011) show that, in a 2-level longitudinal design with both predictor and outcome at level 1, the usual between-within decomposition assumes a random intercepts model and fixed measurement occasions. This assumption is less realistic in the NHS24 study, where it is more plausible to hypothesise that stress changes over a shift. Thus, as suggested by Curran and Bauer, we extend the measurement model to a latent growth curve model by including random slopes so that temporal information can be incorporated and trends allowed to vary between individuals.

In terms of the novelty of the second extension, it should be noted that Preacher (2011) does consider three-level models for level-1 predictors and a level-2 outcome in a brief simulation study, but does not apply these models to real data: the main concern of that paper is with three-level models for designs where all variables are measured at level 1. It should also be pointed out that the models we describe can be viewed as special cases of the general multilevel SEM frameworks developed by Skrondal and Rabe-Hesketh (2004) and Muthén and Asparouhov (2009), and can be estimated in a range of statistical software packages.

The remainder of the paper is structured as follows. Section 2 provides background on the central substantive question – the consequences of workplace stress for nurses' ability to process information quickly and accurately – and a description of the NHS24 study. In Section 3 we set out the measurement model for stress and its assumptions. This model forms part of the multilevel SEM for the estimation of the effects of stress on cognitive function which is described in Section 4. Section 5 presents applications of various specifications of the SEM in analyses of the effects of stress on three measures of nurses' cognitive function that were collected in the NHS24 study. Finally we discuss the implications of our results and possible directions for future research in Section 6.

2. Background and overview of the NHS24 Study

2.1 Stress and cognitive function among nurses

Nursing is a stressful profession (Health Care Commission 2011, Jones and Johnston 2000, McVicar 2003), with high stress levels attributed to factors such as the need to deal with difficult patients, frequent contact with death and bereavement, high workload, conflicting demands, and inadequate resources and support (Chang *et al.* 2005). In addition to the consequences that occupational stress has for nurses themselves, stress in nurses is associated with poorer patient outcomes (Aiken *et al.* 2002) , and more frequent mistakes and accidents in the workplace (Johnston and Pollard 1991, Nolting *et al.* 2002). Changes in these performance-related outcomes may stem from stress-related changes in cognitive function. Stress can lead to difficulties in maintaining concentration (Maslach and Schaufeli 2000, Schaufeli and Enzmann 1998), to changes in decision making processes (Keinan 1987) and to more frequent failures of attention and memory (van der Linden *et al.* 2005). As nursing is a dynamic and responsive occupation,

reductions in the speed and accuracy with which information can be processed and evaluated would have clear implications for job performance. The present NHS24 study assessed the relationship between stress and cognitive function in a sample of nurses employed by the Scottish NHS24 service, a telephone helpline that members of the public can access around the clock for advice on symptoms (Allan*, et al.* 2009, Farquharson*, et al.* 2012). Nurses working for NHS24 face unusually high cognitive demands as they must evaluate symptoms and make decisions based solely on the verbal information that they are given (Edwards 1998). Good performance in this context is contingent upon nurses' ability to maintain concentration during calls, to quickly and accurately process information received and to make appropriate decisions. Consequently, the NHS24 study set out to investigate the association between nurses' stress levels and three key cognitive variables – speed of information processing, accuracy of information processing, and frequency of cognitive slips and errors during shifts.

2.2 The NHS24 study and measurement of stress and cognitive function

The analysis presented in this paper is based on 147 NHS24 telephone-advice nurses from four Scottish call centres. This sample represents approximately 35% of the NHS24 workforce. The mean age was 44 years and nine of the nurses were male. Nurses were observed over two working shifts on two separate days during 2008/09. Nurses' incoming and outgoing calls (nurses phoning non-urgent patients back) were date and time stamped. The data can be viewed as a three-level hierarchy with 4913 calls (at level 1) nested within 270 shifts (level 2) within 147 nurses (level 3). Not all nurses provided useable data on both days: 136 did on Day 1, and 134 nurses on Day 2. The mean shift duration was 7 hours and the mean number of calls per shift was 18. However, shift length ranged from 2.9 to 10.8 hours. Similarly, the number of calls per shift ranged from as few as two calls to as many as 39 calls. The average call was 15 minutes long and 29% of calls were outgoing calls.

The stress variable used in the analysis is measured at the end of each call. Nurses reported how stressful they found each call on a five-point scale, with higher scores indicating a more stressful call. The average call was rated 1.5 out of 5. Table 1 shows the full distribution of stress across all calls and for the first and last call in each shift. The probability of rating a call as 'not all stressful' is slightly higher for the last call than for the first, but the probability of a highly stressful call is low at both ends of the shift. Nevertheless, these average probabilities mask substantial heterogeneity between nurses and their shifts.

We considered three cognitive function outcomes which were all measured at the end of each shift: two objective measures from computerised tasks to assess the speed and accuracy of information processing, and the self-reported number of cognitive failures made during the shift. To assess speed and accuracy of information processing, the nurses each completed a computerized version of a classic choice reaction time task (Smith 1968) immediately before and after both of their shifts. Nurses were presented with 100 single and randomly ordered words one after another and were asked to use designated computer keys to categorize each word, as quickly and accurately as possible, as an animate or inanimate object. The words used were

drawn randomly from a pool of items matched for word length and frequency of use (Kučera and Francis 1967) in order to minimize practice and familiarity effects. Reaction time (RT) was defined as the time taken (in milliseconds) to correctly categorize each word. The number of errors made when categorising words (out of 100) was used to measure the accuracy of information processing. The third measure of cognitive function used is the self-reported frequency of lapses in concentration or memory experienced over the shift, which is assessed using the 15-item work-specific cognitive failures questionnaire (Wallace and Chen 2005). Although, as expected, quicker reaction times in the post-shift task were associated with more classification errors (r = -0.17), the three outcomes were not highly correlated.

All models for the post-shift RT and accuracy outcomes conditioned on the corresponding pre-shift measure. In the analysis of the self-reported cognitive outcomes, social desirability bias and dispositional negative and positive affect were considered as covariates to control for variability in reporting style. Social desirability bias was assessed with the 17-item Social Desirability Scale (Stober 2001), which is designed to quantify respondents' tendency to answer questions in a socially desirable manner. Positive and negative affect were assessed using the Positive and Negative Affect Schedule (PANAS; Watson *et al.* 1988), which has been shown to be a valid and reliable self-report measure of dispositional affect (Crawford and Henry 2004).

Table 1 shows summary statistics for all three cognitive function outcomes, call stress and covariates. As with all repeated cognitive tasks, reaction times improve from first to second administration (as participants acquire more practice at the task). However, if stress negatively impacts on cognitive efficiency, there should be a measurable attenuation in this practice effect, that is, stress should lead to slower post-shift reaction times after controlling for pre-shift reaction times.

3. Approaches to estimating effects of level 1 variables on higher-level outcomes

In this section, we review some existing approaches for modelling the effect of a level 1 variable on a level 2 outcome, and highlight some of the implicit assumptions made by these models in their application to our study of the influence of stress on cognitive function. This review is used to motivate our novel approach, which we introduce in Section 4.

Any useful model must address two main issues. The first is that it should reflect our scientific understanding about stress and cognition, and enable us to test hypotheses regarding this relationship. The second is that stress is measured imperfectly and so measurement error must be accounted for to avoid bias.

Modelling assumptions

We begin by considering the first of these issues and assume that we are able to measure stress perfectly and treat it as a continuous variable; in reality, stress is a latent variable which we do not directly observe. We further take it that each nurse is observed only during a single shift before extending to two shifts further on.

Now suppose that the relationship between cognitive function y_k and stress s_k^* for nurse k follows the linear model

$$y_k = \beta + \gamma s_k^* + \epsilon_k \qquad (1)$$

$$\epsilon_k \sim N(0, \sigma_\epsilon^2)$$

where the superscript * is used to indicate that stress is a latent variable. But what is s_k^* given that, as we have already discussed, stress is measured more frequently than cognition, and is therefore defined at a lower level?

Cognitive function is measured only at the start and end of each shift but, thinking conceptually, we can view stress for nurse k as varying continuously over the shift in response to the allocated calls according to function $s_k^*(t)$. While the design only allows us to observe $s_{ik}^* = s_k^*(t_{ik})$, where t_{ik} is the time at which nurse k was measured following call i, this is not problematic if we can assume that between-call stress fluctuations do not influence cognition. However, to use a strategy based on model (1), we must chose s_k^* to be a summary of s_{ik}^* for each nurse that reflects the way in which call stress affects cognition.

The resulting data lead to what is sometimes referred to as a 'micro-macro' design (Snijders and Bosker 2012, Chapter 2). Multilevel models were originally developed for 'macro-micro' designs where the outcome is at level 1 and covariates are a mixture of level 2 and level 1 variables. The usual approach for micro-macro designs is to use aggregated level 1 covariates as level 2 covariates, and estimate a regression model of the level 2 outcome on the aggregated level 2 covariates. The most common method is to specify s_k^* to be $\bar{s}_{.k}^*$, namely, the level 2 mean of s_{ik}^*. In other words, the stress experienced during each call is assumed to influence end-of-shift cognition equally, whether it is the first or final call. Two examples of alternative choices for model (1) are that only the most stressful call affects cognition (i.e. $s_k^* = \max(s_{ik}^*)$), and that only the final call has any effect (i.e. $s_k^* = s_{n_k k}^*$ where n_k is the number of calls handled by nurse k).

Note that it is assumed throughout that the allocation of calls to nurses is made irrespectively of how stressed a nurse is at the time, or any other of the characteristics included as covariates when elaborating model (1). This is a reasonable assumption for the NHS24 study because calls were simply allocated to the first available nurse.

Measurement error

Even if stress were measured on a continuous scale, we would not observe latent s_{ik}^* but only an imperfect measure $s_{ik} = s_{ik}^* + m_{ik}$, where m_{ik} indicates measurement error at call i. Croon and van Veldhoven (2007) considered estimating model (1) using mean stress measured for nurse k $\bar{s}_{.k}$ rather than his/her true average stress $\bar{s}_{.k}^*$. They show that the estimated coefficient of $\bar{s}_{.k}$ is biased for γ, where the extent of the bias depends on the reliability of the sample mean as a measure of the true mean; the bias decreases as both the intra-class correlation in s_{ik} (i.e. the within-shift correlation in stress across calls) and the number of calls per shift increase.

To remedy this problem, they propose a stepwise estimation procedure based on an ordinary least squares (OLS) regression of y_k on an adjusted level 2 mean of s_{ik}, followed by a standard error correction to allow for heteroskedasticity when the number of level 1 units (calls) varies across level 2 units (shifts). Griffin (1997) proposes an alternative approach in which we estimate a multilevel model for s_{ik} with a nurse-level random effect. Empirical Bayes estimates of the nurse-level random effects are then substituted for s_k^* in (1) and the model estimated via OLS as before.

Implicit in both of these approaches is the assumption that latent stress at each call can be decomposed as

$$s_{ik}^* = \alpha + v_k, \qquad (2)$$

where α is the mean stress experienced among the population of nurses, and v_k is a nurse-level effect satisfying $E(v_k) = 0$ that indicates how much greater/less the stress experienced by nurse k is than the average. In other words, it is assumed that a nurse experiences a 'blip' of stress after a call and that this nurse's blip is the same for every call. The full model for observed stress is thus

$$s_{ik} = \alpha + v_k + m_{ik},$$

where it is assumed that $E(m_{ik}|v_k) = 0$, that is, a classical measurement error model in which the measurement errors cancel out over the distribution of calls that could be experienced by a nurse. Curran and Bauer (2011) show that for the 2-level case the standard approach of estimating the between and within components of s_{ik} by $\bar{s}_{.k}$ and $s_{ik} - \bar{s}_{.k}$ assumes an underlying measurement error model of this form. Both of the above approaches can be described as two-stage aggregation methods where the first stage involves deriving an estimate of s_k^* which is then included as a predictor of y_k in a second-stage regression, thus accounting for measurement error.

4. Multilevel structural equation models for stress and cognitive function

Croon and van Veldhoven's method is a form of SEM comprising two submodels: 'structural model' (1) and a 'measurement model' for s_{ik}. Rather than estimating these submodels in two separate stages, it is generally more efficient and convenient to estimate a joint multilevel structural equation model (SEM) for the measurement and structural submodels (e.g. Preacher, *et al.* 2010).

We now generalise this joint modelling approach to allow for (i) the observation of nurses in multiple shifts, leading to a three-level hierarchical structure; and (ii) ordinal measurement of stress, which may be combined with a discrete measure of cognitive function. We note that our approach can be framed as an extension of the two-level latent covariate model described by Lüdtke *et al.* (2008) in which ordinal s_{ik} are treated as indicators of the level 2 latent construct s_k^*. In so doing, we assume that the primary purpose of recording stress at each call is to measure s_k^*, using the between-call within-person variance in s_{ik} to allow for measurement error in s_k^*.

7

The model we present can also be viewed as a generalisation of the multilevel multivariate model for mixed responses at different levels proposed by Goldstein and Kounali (2009), where we allow the random effects from the model for one response to predict another response.

4.1 Measurement model for stress

Suppose that nurses were observed for more than one shift, and denote by s_{ijk}^* the true latent stress of call i in shift j of nurse k. We can extend (2) to decompose s_{ijk}^* into shift and nurse components

$$s_{ijk}^* = \alpha + v_k + u_{jk}, \qquad (3)$$

where α is the intercept or mean stress, v_k is a nurse-level effect representing the time-invariant component of stress that is fixed across shifts, and u_{jk} is a shift-level effect for nurse k capturing factors that vary across shifts for a given nurse. In other words, u_{jk} represents the level of stress experienced during a particular shift as a deviation from a nurse's overall (across-shift) level of stress, and may be interpreted as the degree to which shift j is more or less stressful than usual for nurse k.

Under (3), it is assumed that a nurse experiences the same blip of stress during every call on the same shift, but the size of these blips varies between shifts within the same nurse as well as between nurses.

The measurement model discussed in Section 3 is applicable if observed stress s_{ijk} for call i of shift j of nurse k is measured on a continuous scale. However, in our application, s_{ijk} is measured on an ordinal scale and so we must develop an ordinal model for stress. To define our multilevel ordinal model, we define the following latent variable model

$$s_{ijk}^{**} = \alpha + v_k + u_{jk} + m_{ijk} \qquad (4)$$
$$= s_{ijk}^* + m_{ijk}$$

where s_{ijk}^{**} is a continuous variable underlying the observed ordinal measure s_{ijk} which is assumed to be measured with error.

To define the multilevel model, we assume $v_k \sim N(0, \sigma_v^2)$ and $u_{jk} \sim N(0, \sigma_u^2)$ where σ_v^2 is the between-nurse variance in stress, and σ_u^2 is the between-shift within-nurse variance. In addition we assume that m_{ijk} follows a standard logistic distribution which has mean zero and variance σ_m^2 (the between-call within-shift variance) fixed at $\pi^2/3 = 3.29$. This leads to an ordered logistic model, also known as a proportional odds model.

The level of stress of each call s_{ijk} is reported on a 5-point ordinal scale. This ordinal measurement is related to the underlying latent variable s_{ijk}^{**} via a set of threshold parameters $\tau_1, \dots \tau_4$ such that

$$s_{ijk} = \begin{cases} 1 & \text{if } s_{ijk}^{**} < \tau_1 \\ r & \text{if } \tau_{r-1} \le s_{ijk}^{**} < \tau_r \ (r = 2, 3, 4) \\ 5 & \text{if } s_{ijk}^{**} \ge \tau_4 \end{cases}$$

Denoting the response probabilities by $\pi_{rijk} = \text{Pr}(s_{ijk} = r)$, $r = 1, \ldots, 5$, the proportional odds model is specified in terms of the cumulative response probabilities $\gamma_{rijk} = \text{Pr}(s_{ijk} > r) = \pi_{r+1,ijk} + \cdots + \pi_{5ijk}$. The decomposition of s_{ijk}^{**} in (4), with the identification constraint $\alpha = 0$, then leads to the random effects proportional odds model

$$\log\left(\frac{\gamma_{rijk}}{1 - \gamma_{rijk}}\right) = v_k + u_{jk} - \tau_r, \qquad r = 1, \ldots 4 \qquad (5)$$

$$v_k \sim N(0, \sigma_v^2), \qquad u_{jk} \sim N(0, \sigma_u^2)$$

The measurement model given by (5) can be viewed as a multilevel model, specifically a three-level random intercept model (Skrondal and Rabe-Hesketh 2004, Chapter 10, Snijders and Bosker 2012, Chapter 4). It can also be formulated as a type of multilevel item response theory model where the repeated measures s_{ijk} are multiple indicators of the latent variables (or random effects) v_k and u_{jk} (Liu and Hedeker 2006, Liu *et al.* 2013).

Covariates may be added to (5) to investigate whether observed characteristics of calls and nurses predict stress; for example incoming calls are more stressful than outgoing calls, and older nurses report lower levels of stress. However, our objective is to estimate the total (unadjusted) effects of stress on cognitive function, rather than the effects of the component of stress that is unexplained by call (and possibly shift and nurse) characteristics. After establishing whether there is an association between stress and cognitive function, it may then be of interest to determine whether that association can be explained by covariates. Another situation where covariates may be included is when both stress and the outcome are self-reported, in which case individual-level measures of personality traits can be included to adjust for reporting tendency. We consider this approach in our analysis of the subjective measure of accuracy during the shift.

4.2 Multilevel SEM for effect of stress on cognitive function

Specifying s_{jk}^*, the true latent stress for shift j of nurse k, to be the shift-level mean $\bar{s}_{.jk}^*$, the decomposition of (3) implies that the mean-centred true latent stress for shift j of nurse k is

$$\bar{s}_{.jk}^* - \alpha = v_k + u_{jk}$$

and thus the stress experienced in a given shift can be represented by $v_k + u_{jk}$ in a model for the effect of shift-level stress on a shift-level measure of cognitive function y_{jk}. For continuous y_{jk} we specify the linear two-level random effects model

$$y_{jk} = \boldsymbol{\beta}^T \boldsymbol{x}_{jk} + \lambda(v_k + u_{jk}) + w_k + \epsilon_{jk} \qquad (6)$$

$$w_k \sim N(0, \sigma_w^2), \qquad \epsilon_{jk} \sim N(0, \sigma_\epsilon^2)$$

where $\boldsymbol{x}_{jk}$ is a vector of covariates with coefficients $\boldsymbol{\beta}$, w_k is a nurse-level random effect allowing for dependency in y_{jk} across shifts for the same nurse, and ϵ_{jk} is a shift-specific residual. The covariates $\boldsymbol{x}_{jk}$ may be defined at the shift or nurse level and include a pre-shift measure of y_{jk} for the cognitive test responses (reaction time and number of errors). The combination of $\boldsymbol{x}_{jk}$, w_k and ϵ_{jk} capture all observed and latent influences on y_{jk} other than stress.

The availability of repeated measures on stress over calls in the same shift and for two shifts per nurse allows the effect of stress to be decomposed into between-nurse and within-nurse effects, leading to

$$y_{jk} = \boldsymbol{\beta}^T \boldsymbol{x}_{jk} + \lambda_B v_k + \lambda_W u_{jk} + w_k + \epsilon_{jk} \qquad (7)$$

where λ_B is the between-nurse effect of stress and λ_W is the within-nurse between-shift effect of stress. The separation of between and within effects is common in multilevel models (Curran and Bauer 2011, Neuhaus and Kalbfleisch 1998). Here, the between effect represents the nurse-level relationship between stress and cognitive function (conditional on $\boldsymbol{x}_{jk}$), that is the effect of nurse-level stress on the nurse-level mean of y_{jk}. The within effect is the shift-level relationship between stress and y_{jk}, adjusted for the effect of the nurse-level component of stress. As our measure of stress is self-reported, the separation of the effect of stress into between and within nurse effects is of particular importance when y_{jk} is also self-reported. This is because unmeasured nurse-level characteristics related to reporting tendency will be absorbed into both v_k and w_k, leading to a non-zero correlation between the latent predictor v_k and residual w_k in (7) and thus a biased estimate of λ_B. As u_{jk} varies within but not between individuals, $\text{cov}(u_{jk}, w_k) = 0$ and the estimator of the within effect λ_W will be unbiased. Nevertheless, the effect of the total shift-level stress $v_k + u_{jk}$ is also of interest because v_k captures not only reporting tendency but factors related to the true level of stress experienced by a nurse that are constant across shifts, for example the stressfulness of the job and sensitivity to stress. In the analyses presented in Section 5, we consider both the effect of total stress and the effects of the shift-specific and nurse-level components of stress.

Models (6) and (7) can be replaced by generalised linear multilevel models if y_{jk} is a discrete variable. In our application, for example, the number of failures in the post-shift cognitive test is a count variable which is analysed using a log-linear Poisson multilevel model. The combination of models (5) and (7), or (5) and (6), form a multilevel SEM.

Specifying a measurement model for s_{ijk} and estimating it jointly with a model for y_{jk} offers several advantages over two-stage estimation procedures. First, as described in Section 3, two-stage methods may produce biased estimates of the coefficients and standard errors must be corrected to account for the uncertainty in estimation of v_k and u_{jk}. SEM allows for effects on y_{jk} of latent stress variables which represent 'true' nurse and shift-level stress. Second, the

measurement model can be adapted to reflect the measurement and distribution of s_{ijk}. Third, the SEM is more easily generalised to allow for covariate effects on s_{ijk} or y_{jk}, and for the effects of change in stress over the shift on cognitive function. In Section 5, we demonstrate how covariates can be included in the model for s_{ijk} in an effort to adjust for reporting tendency when y_{jk} is also self-reported.

We next consider how the measurement model (5) can be extended to a latent growth model with individual-specific random intercepts and slopes for call number to allow for change in stress over a shift; these are then included as predictors of y_{jk} in models (6) and (7).

4.3 Multilevel SEM for effect of change in stress on cognitive function

As was discussed in Section 3, the measurement model for stress given by (5) assumes that stress is the same after each call during a shift. This random intercept model is implied by the model for true latent stress s_{ijk}^* of (3) where each nurse has equal blips in stress during every call on a given shift. A more realistic model would include call effects to allow for heterogeneity in these blips across calls. Curran and Bauer (2011) consider, for a 2-level design, a growth curve model for a time-varying predictor and show how estimates of its between and within components can be adjusted to allow for a subject-specific time trend. These adjusted estimates can form part of a two-stage approach, but we follow Curran and Bauer's suggestion of specifying a joint model for the predictor and outcome.

Previous research has found that, rather than remaining constant, stress tends to change over the course of a shift (Johnston *et al.* 2013). We therefore extend the measurement model of (5) to allow stress to change as a linear function of call number c_{ijk} with variation between nurses in both the level of stress at the first call (the intercept) and the rate of change over the shift (the slope of c_{ijk}), as follows:

$$\log\left(\frac{\gamma_{rijk}}{1 - \gamma_{rijk}}\right) = \delta\, c_{ijk} + v_{0k} + u_{0jk} + v_{1k}c_{ijk} + u_{1jk}c_{ijk} - \tau_r, \qquad r = 1, \dots 4 \qquad (8)$$

$$\begin{pmatrix} v_{0k} \\ v_{1k} \end{pmatrix} \sim N(\mathbf{0}, \mathbf{\Omega}_v), \;\; \mathbf{\Omega}_v = \begin{pmatrix} \sigma_{v0}^2 & \\ \sigma_{v01} & \sigma_{v1}^2 \end{pmatrix}$$

$$\begin{pmatrix} u_{0jk} \\ u_{1jk} \end{pmatrix} \sim N(\mathbf{0}, \mathbf{\Omega}_u), \;\; \mathbf{\Omega}_u = \begin{pmatrix} \sigma_{u0}^2 & \\ \sigma_{u01} & \sigma_{u1}^2 \end{pmatrix}$$

Equation (8) is a 3-level random slope growth model. If call number is coded such that $c_{ijk} = 0$ at the first call in a shift, v_{0k} is the time-invariant component of stress at the start of any shift for nurse k and u_{0jk} is the deviation between her stress at the start of shift j and her usual initial stress; σ_{v0}^2 and σ_{u0}^2 are the between-nurse and between-shift within-nurse variances at the first call. The change in stress for each additional call during shift j of nurse k is $\delta + v_{1k} + u_{1jk}$, where v_{1k} and u_{1jk} allow the change in stress over the shift to vary between nurses and between shifts

worked by the same nurse respectively. Figure 1 shows three regression lines depicting the relationship between the latent stress response s_{ijk}^{**} and call number for the first three calls in shift j of nurse k at the population level (with slope δ), the nurse level (slope $\delta + v_{1k}$) and the shift level (slope $\delta + v_{1k} + u_{1jk}$). The relationship between the intercepts and slopes of the between-nurse and within-nurse regressions are captured by the covariances σ_{v01} and σ_{u01}. For example, the combination of a negative overall slope and negative intercept-slope covariance at the nurse level ($\delta < 0$ and $\sigma_{v01} < 0$) would imply that nurses with above-average stress at the start of a shift ($v_{0k} > 0$) tend to report a faster-than-average decline in stress over the course of a shift ($v_{1k} < 0$).

The model given by (6) for cognitive outcome y_{jk} is then extended to allow for effects of initial stress and change in stress over the shift as follows

$$y_{jk} = \boldsymbol{\beta}^T \boldsymbol{x}_{jk} + \lambda_0(v_{0k} + u_{0jk}) + \lambda_1(v_{1k} + u_{1jk}) + w_k + \epsilon_{jk} \qquad (9)$$

$$w_k \sim N(0, \sigma_w^2), \qquad \epsilon_{jk} \sim N(0, \sigma_\epsilon^2)$$

where λ_0 is the effect of stress at call 1 and λ_1 the effect of change in stress.

We can also separate the effects of the nurse and shift level random effects, leading to a generalisation of (7):

$$y_{jk} = \boldsymbol{\beta}^T \boldsymbol{x}_{jk} + \lambda_{B0}v_{0k} + \lambda_{B1}v_{1k} + \lambda_{W0}u_{0jk} + \lambda_{W1}u_{1jk} + w_k + \epsilon_{jk} \qquad (10)$$

We now distinguish two types of between-nurse effect: the effect of a nurse's average stress at call 1 (λ_{B0}) and the effect of the average amount of change in stress over the course of a shift (λ_{B1}) on her average cognitive outcome, where 'average' refers to a typical shift. The between effects describe the nurse-level relationship between y and stress at the start and during a shift. The within-nurse effects λ_{W0} and λ_{W1} are the effects of initial stress and change in stress on y_{jk} for a given shift, adjusted for the effects of the corresponding nurse-level component of stress. They describe the shift-level relationship between y and stress, holding constant time-invariant influences on stress.

4.4 Estimation

The multilevel SEMs described above can be specified as generalisations of a multilevel multivariate response model. In the simplest multivariate model, the responses are all continuous and defined at level 1 and the coefficients of all random effects are constrained to 1. Such a model can be estimated as a standard multilevel model after stacking the responses into a single vector and including binary response indicators as explanatory variables with random coefficients at each level (Snijders and Bosker 2012, Chapter 16). The models considered in Sections 4.2 and 4.3 extend the basic multivariate model in three ways. First, the responses in the application are not all continuous: s_{ijk} is ordinal and y_{jk} is either continuous or a count. Second,

12

s_{ijk} and y_{jk} are defined at different levels in the hierarchical structure. Third, the coefficients of the stress random effects in the model for y_{jk} are unconstrained parameters.

The multilevel SEM for s_{ijk} and y_{jk} combines a multilevel generalised linear model for multivariate mixed response types (e.g. Goldstein *et al.* 2009, Skrondal and Rabe-Hesketh 2004, Chapter 14) with a multilevel factor model (e.g. Goldstein 2010, Chapter 7, Muthén 1991, Steele and Goldstein 2006) and falls within the 'generalised linear latent and mixed models' (GLLAMM) framework of Skrondal and Rabe-Hesketh (2004) and the growth mixture modelling framework of Muthén and Asparouhov (2009). For example, the joint model specified by (5) and (7) can be framed as a multilevel factor model for the multivariate response formed of $n_{1k} + n_{2k}$ measurements of stress for calls over two shifts and two cognitive measurements for nurse k, where the random effects u_{jk}, v_k and w_k are shift and nurse-level factors. The factor loadings (coefficients) of u_{jk} and v_k are constrained to 1 for the stress responses and freely estimated for the cognitive response, while the loadings of the second nurse-level factor w_k are constrained to zero for the stress responses. The model is completed by including i.i.d. residuals m_{ijk} and ϵ_{jk} for the stress responses and cognitive responses respectively.

All models were estimated using maximum likelihood with numerical integration via Guassian quadrature to 'integrate out' the random effects for the discrete responses (stress and number of errors in the computerised task), as implemented in the aML software (Lillard and Panis 1998-2003). Other software options include Mplus, Stata (gllamm and gsem), SAS (proc nlmixed) and, using MCMC, WinBUGS. Details of the required data structure and annotated aML syntax and output files for selected models are provided as online supplementary materials.

5. Application

5.1 Random intercept SEM for overall effect of stress on cognitive function

We begin with an application of the multilevel SEMs defined by (5) and (6), and by (5) and (7). In these random intercept models mean stress is permitted to vary between nurses and between shifts within nurses but, for all nurses, stress is assumed to be constant across a shift apart from random fluctuations represented by m_{ijk} in (4).

Table 2 shows estimates from fitting the measurement model (5). Estimates of the intra-shift and intra-nurse correlations in the underlying continuous latent variable stress s_{ijk}^{**} can be derived by expressing the multilevel ordinal logit model as a linear model for s_{ijk}^{**}, as in (4), where m_{ijk} follows a standard logistic distribution (with variance 3.29) and v_k and u_{jk} are normally distributed as before. The correlation between stress responses for two randomly selected calls in the same shift (for the same nurse) is estimated as $(1.327^2 + 0.615^2)/(1.327^2 + 0.615^2 + 3.29) = 0.39$, which suggests substantial intra-shift variation in reported stress levels across calls. The correlation between the mean shift-level stress for two shifts from the same nurse is

$1.327^2/(1.327^2 + 0.615^2) = 0.82$. Thus, although nurses report similar levels of stress for each shift, there is some within-nurse variation. This is important for distinguishing the between-nurse and within-nurse between-shift effects of stress on cognitive outcomes as in (7).

Estimates of the effects of stress on each cognitive outcome y_{jk} were obtained from multilevel SEMs fitted to each cognitive measure in turn. As the distribution of accuracy of information processing in the computerised task was highly positively skewed with 0-2 categorisation errors made in almost 70% of shifts, this outcome was analysed using a Poisson regression model. The distribution of the self-reported number of errors was also positively skewed, but the log-transformed variable had an approximately normal distribution. Both reaction time and the logarithm of the self-reported number of errors were analysed using linear models of the same form as equations (6) and (7). A different model was fitted to each cognitive outcome due to the weak correlations between the outcomes. In each case the measurement model (5) was estimated jointly with either equation (6) (Model M1) or equation (7) (Model M2). All models for reaction time and number of errors in the post-shift computerised tasks included the corresponding pre-shift score as a covariate to adjust for pre-existing differences between nurses in their cognitive function at the start of the shift. In addition, a dummy variable for shift was included and interacted with pre-shift scores to allow the mean post-shift-pre-shift change in reaction time and accuracy to vary across shifts.

Table 3 shows the estimated effects of stress on the three cognitive outcomes. Model M1 includes the total shift-level latent stress, $v_k + u_{jk}$, which combines the nurse-specific and shift-specific components of stress. The effect of stress is then decomposed into between-nurse and within-nurse between-shift effects in Model M2 by including v_k and u_{jk} as separate predictors. Starting with reaction times in the post-shift computerised tasks (adjusted for pre-shift scores), the effect of total shift-level stress is almost significant at the 5% level with longer reaction times observed among nurses who reported higher levels of stress during the shift. When shift-level stress is decomposed into between and within nurse components, however, we find that this effect is largely due to nurse-specific factors: nurses with higher overall stress levels (across both shifts) tend to have longer reaction times, but there is little evidence to suggest that having a more stressful shift than usual affects reaction times.

Turning to accuracy of information processing, we find no effect of stress on the number of errors made in the post-shift computerised task. In contrast, there is a positive association between the perceived level of stress during a shift and the self-reported number of errors or lapses in concentration during a shift, a subjective measure of accuracy. However, after separating the nurse-level and shift-level components of stress, nurse-specific factors are again found to dominate. As both stress and the number of cognitive failures during the shift are self-reported, the apparent effect of nurse-level stress could be due to differences between nurses in their reporting style rather than differences in the level of stress experienced or sensitivity to stress. In an attempt to adjust for self-report bias, we therefore extended both the measurement model (5) for stress and model (7) for the number of cognitive failures to include predictors of reporting

tendency measured before the first shift as covariates. Three measures were considered: social desirability bias and negative and positive affect (see Section 2.2 for details). To adjust for change in the between-nurse variance in stress (σ_v^2) when nurse-level covariates are added to the model, we compare estimates of the standardised between-nurse effect of stress ($\lambda_B \sigma_v$) before and after inclusion of covariates. The addition of covariates led to a reduction in the standardised between effect from 0.334 to 0.246, but the effect remained strongly significant at the 1% level. As expected, the addition of nurse characteristics had no effect on the standardised within-nurse effect of stress ($\lambda_W \sigma_u$). To the extent that these three measures capture reporting tendency, the adjusted effect of nurse-level stress is more likely to represent the effects of time-invariant sources of stress such as differences between nurses in the characteristics of their job or individual differences in the ability to cope with stressful situations.

5.2 Random slope SEM for effect of baseline stress and change in stress on cognitive function

We next consider the extended measurement model of equation (8). This random slope model allows stress to change over a shift as a linear function of call number, and for the rate of change to vary between nurses and between shifts for the same nurse. It was necessary to fit a restricted version of (8) with a single random slope effect that combined nurse and shift factors because of convergence problems when the slope was decomposed into between-nurse and within-nurse between-shift effects. The fitted model has the form

$$\log\left(\frac{\gamma_{rijk}}{1-\gamma_{rijk}}\right) = \delta\,c_{ijk} + v_{0k} + u_{0jk} + \tilde{u}_{1jk}c_{ijk} - \tau_r, \qquad r = 1, \dots 4 \qquad (8a)$$

$$\tilde{u}_{1jk} = v_{1k} + u_{1jk}$$

$$v_{0k} \sim N(0, \sigma_{v0}^2)$$

$$\begin{pmatrix} u_{0jk} \\ \tilde{u}_{1jk} \end{pmatrix} \sim N(\mathbf{0}, \mathbf{\Omega}_{\tilde{u}}), \quad \mathbf{\Omega}_{\tilde{u}} = \begin{pmatrix} \sigma_{u0}^2 & \\ \sigma_{\tilde{u}01} & \sigma_{\tilde{u}1}^2 \end{pmatrix}$$

The above specification still allows the rate of change in stress with call number to vary across shifts, but does not separate slope variation due to unmeasured factors that are fixed across shifts for the same nurse and unmeasured factors that vary between shifts. Thus the random slope effect $\tilde{u}_{1jk}$ is no longer a within-nurse effect, but combines unobserved factors at the shift and nurse level that affect the rate of change. As in the full specification (8), variation at call 1 (the intercept) is partitioned into nurse-level and shift-specific components.

The estimates for the model of equation (8a) are shown in Table 4. There is a negative linear effect of call number, suggesting that on average stress declines over a shift. Variation in the rate of decline between shifts is captured by $\sigma_{\tilde{u}1}$ and, based on the normality assumption for $\tilde{u}_{1jk}$, we would expect the rate of change to lie in the range -0.027 ± 1.96 $\hat{\sigma}_{\tilde{u}1} = (-0.152, 0.098)$ for the middle 95% of shifts. Thus, although the slope of the average prediction line is negative, there is

substantial variation across shifts in the direction and amount of change in stress over a shift. The negative estimate of -0.324 for the intercept-slope correlation, $\rho_{\tilde{u}01} = \sigma_{\tilde{u}01}/(\sigma_{u0}\sigma_{\tilde{u}1})$, indicates that nurses who report higher-than-average stress at the start of the shift ($u_{0jk} > 0$) tend to have a faster-than-average decline in stress over the shift ($\tilde{u}_{1jk} < 0$). The variation in intercepts and slopes of stress trajectories is illustrated in Figure 2 which shows the predicted log-odds of reporting some stress ($s_{ijk} > 1$) for selected shifts.

The measurement model in equation (8a) was then estimated jointly with models for each post-shift cognitive outcome y_{jk} to investigate the effects on y_{jk} of stress at the start of the shift and change in stress over the shift. Table 5 shows estimates of the effects of stress on two of the cognitive outcomes. (The same model was fitted to the number of errors in the post-shift task but, as in the simpler model of Table 3, no effects of stress were found.) As in the random intercepts analysis of Section 5.1, the model for reaction time in the post-shift computerised task includes pre-shift reaction time as a covariate. The model for self-reported number of errors (and the jointly estimated measurement model) includes nurse-level covariates as an adjustment for reporting tendency. The models for both outcomes are SEMs in which stress and change in stress are represented by latent variables; these latent variables are the nurse and shift random effects from the random slopes measurement model of equation (8a). Both Models M3 and M4 include an effect of change in stress on cognitive outcomes, but they differ in the treatment of stress at the start of the shift. Model M3 includes an overall effect which combines nurse-level and shift-specific influences on stress, as in equation (9). Model M4 separates the effect of v_{0k} (the between-nurse effect of stress in call 1) from the effect of u_{0jk} (the within-nurse between-shift effect of stress in call 1), in a restricted form of equation (10) without the between-within decomposition of the effects of change in stress (i.e. with $\lambda_{B1} = \lambda_{W1} = \lambda_1$).

The positive estimate for λ_0 in Model M3 for speed in the computerised task indicates that higher stress at the start of a shift is associated with a longer reaction time at the end, conditional on reaction time at the start. From the decomposition into between-nurse and within-nurse between-shift effects (Model M4), we find that this is driven by a between-nurse effect (λ_{B0}) rather than a within-nurse effect ($\tilde{u}_{W0}$). Nurses reporting above-average stress at the first call in a shift tend to perform worse in the post-shift task than nurses who start less stressed. This may be because nurses who come onto a shift in a stressed state (i.e. who have experienced high levels of stress outside of the workplace) are more cognitively depleted and may be less able to rapidly process information after a full shift at work. Cognitive resources are finite, and become depleted with use (Schmeichel 2007). There is evidence to suggest that the efficiency of the cognitive processes involved in information processing and decision making reduces as people work, only becoming replenished during breaks from work (Danziger *et al.* 2011). There is little evidence of an effect of rate of change in stress over the shift on post-shift reaction time.

Turning to the subjective measure of cognitive failures during the shift, we again find that higher initial stress is associated with a poorer cognitive outcome (more errors or lapses in concentration). Again, this turns out to be a between-nurse effect (λ_{B0} in Model M4), that is an

effect of the nurse-level component of stress rather than factors specific to a particular shift. The positive estimate of the coefficient for the slope random effect (λ_1) suggests that nurses whose stress declines more rapidly than average over a shift ($\tilde{u}_{1jk} < 0$) report making fewer errors during that shift. In contrast, nurses with fairly constant stress levels ($\tilde{u}_{1jk} > 0$) tend to report more errors. A faster decline in stress over a shift could indicate the speed at which nurses are habituating to work, and those who do this most efficiently process information faster. Although attitudinal and personality measures have been included as covariates, we cannot rule out the possibility that our estimates of the effects of stress are subject to self-report bias. Suppose, for example, that nurses who overstate their level of stress in any call also tend to report making more errors during the shift. If not fully captured by covariates, this reporting tendency will lead to a positive correlation between v_{0k} and the nurse-level residual w_k in the model for y_{jk}, and thus a biased estimate of λ_{B0} the coefficient of v_{0k} in the model for y_{jk} . However, it is less plausible that reporting tendency would affect a nurse's *change* in reported stress over a shift. Under the assumption that any overstatement or understatement of stress is the same for each call, apart from random fluctuations unrelated to call number, the estimate of λ_1 the coefficient of $\tilde{u}_{1jk}$ will be unaffected by self-report bias.

6. Discussion

In this paper, we considered the problem of analysing multilevel data where the predictors are defined at a lower level than the outcome variable. Traditional multilevel models cannot be applied in this situation, but such 'micro-macro' designs are common in longitudinal studies. For instance, trials involving repeated measures of exposure and confounders are now regularly analysed using marginal structural models to allow for time-varying confounding (e.g. Daniels *et al.* 2013). Our motivating example used intensive longitudinal data collected using EMA methods with different assessment schedules for different variables. Another application is to analyses of the effect of childhood physical and cognitive development on adult outcomes (e.g. Sayers *et al.* 2015). Such studies are usually based on birth cohort data where individuals are measured at several time points in childhood, but far less frequently in adulthood. Micro-macro designs also arise in cross-sectional studies where groups of individuals form the higher-level units. For example, Sampson *et al.* (1997) used a multilevel SEM to investigate the effect of individuals' perceptions of their neighbourhood on neighbourhood-level violent crime. In organisational psychology there is interest in the effect of individual attitudes and work practices on team productivity. Other examples can be found in Croon and van Veldhoven (2007).

Previous research has shown that simply aggregating the covariate to the level of the outcome variable leads to biased effects of the covariate and its standard error. This has led to the use of SEMs where, in the longitudinal case, repeated measurements of the covariate are treated as indicators of an individual-level latent variable. We have considered several generalisations of this approach in our analysis of the effect of stress on nurses' cognitive function, which have widespread potential applications in social and health research. The first of these extensions was to allow for additional levels of clustering in either or both the covariate and outcome. In our

application the availability of stress measures for calls over two shifts and post-shift cognitive outcomes leads to a three-level structure for stress and a two-level structure for cognitive outcomes. We show how this enables the effect of stress on cognitive outcomes to be decomposed into between-nurse and within-nurse (between-shift) effects. Three-level designs are common in EMA data where participants are often observed repeatedly over several days, and we have argued that the between-within decomposition is especially useful when both the covariate(s) and outcome are self-reported measures. Our second extension was to allow the level 1 covariates and higher-level outcomes, treated as responses in the SEM approach, to be mixtures of continuous and discrete variables. The final generalisation was to specify the measurement model for 'true' stress as a random slopes growth curve model and to allow the random intercept and slopes from this model to influence cognitive effects.

In our analysis of the effect of stress on cognitive outcomes, there was a suggestion that higher stress levels reported during a shift were associated with longer reaction times in a post-shift task, but a decomposition into between-nurse and within-nurse between-shift effects revealed that this was due to nurse-level factors rather than shift-specific factors. There was strong evidence that higher stress was associated with reporting more errors and lapses in concentration, but this was largely a between-nurse effect which may reflect individual differences in the ability to cope with stress or reporting tendency rather than a true effect of stress exposure per se. Using a random slopes model, we found a significant effect of the rate of change in stress over the shift, with nurses whose stress declined more rapidly than average reporting fewer errors. While we have argued that effects of change in stress are less susceptible to self-report bias than effects of the level of stress, our conclusions are limited because we were unable to estimate the between-within decomposition of the effect of change. The convergence problems for the most complex within-effects random slopes model may indicate that these effects are inestimable for the NHS24 design. Alternatively, it may indicate that these effects are non-identified for designs of this type and so cannot be estimated unless we observe further shifts on each nurse: a referee noted the similarity of the NHS24 design to the more conventional one with two repeated measurements per individual, and that it is possible to estimate the random intercept or the random slope of a two-level model but not both. This requires further work to resolve.

The multilevel SEM framework is highly flexible and the model presented here can be extended in a number of ways. We have shown how the measurement model for stress can be extended to a random slopes model to allow shift and nurse-specific stress trajectories. A further extension is to specify a random slopes model for the structural model to allow the effect of shift-level stress to vary across nurses (see Preacher, *et al.* (2010) for further details in the two-level case). Another extension would be to specify the measurement model as a growth mixture model (Muthén and Asparouhov 2009) where individual stress trajectories over a shift are grouped into latent classes with between-individual variance in intercepts and slopes within classes. It is also possible to fit a joint model for stress and all three cognitive outcomes in a multilevel multivariate SEM. Another generalisation would be to model the variability in nurses' stress levels during a shift and to allow

both the mean level of stress over a shift and the within-subject variance in stress to influence their subsequent cognitive performance. Mixed-effects location scale models have been developed for this purpose and have been applied in analyses of EMA data on adolescent mood variability and smoking (Hedeker *et al.* 2012); see also Leckie *et al.* (2014) for applications in educational research.

Acknowledgements

References

Aiken, L., Clarke, S., Sloane, D., Sochalski, J. and Silber, J. (2002) Hospital nurse staffing and patient mortality, nurse burnout and job dissatisfaction. *Journal of the American Medical Association,* 288, 1987-1993.

Allan, J., Farquharson, B., Choudhary, C., Johnston, D. W., Jones, M. C. and Johnston, M. (2009) Stress in telephone helpline nurses: research protocol for a study of theoretical determinants, physiological aspects and behavioural consequences. *Journal of Advanced Nursing* 65, 2208-2215.

Bolger, N. and Laurenceau, J.-P. (2013) *Intensive Longitudinal Methods: An Introduction to Diary and Experience Sampling Research*, New York: Guilford.

Chang, E. M., Hancock, K. M., Johnson, A., Daly, J. and Jackson, D. (2005) Role stress in nurses: review of related factors and strategies for moving forward. *Nursing and Health Sciences,* 7, 57-65.

Crawford, J. R. and Henry, J. D. (2004) The positive and negative affect schedule (PANAS): construct validity, measurement properties and normative data in a large non-clinical sample. *British Journal of Clinical Psychology,* 43, 245-265.

Croon, M. A. and van Veldhoven, M. J. P. M. (2007) Predicting group-level outcome variables from variables measured at the individual level: a latent variable multilevel model. *Psychological Methods,* 12, 45-57.

Curran, P. J. and Bauer, D. J. (2011) The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology,* 62, 583-619.

Daniels, R. N., Cousens, S. N., De Stavola, B. L., Kenward, M. G. and Sterne, J. A. C. (2013) Methods for dealing with time-dependent confounding. *Statistics in Medicine,* 32, 1584-1618.

Danziger, S., Levav, J. and Avnaim-Pesso, L. (2011) Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences,* 108, 6889-6892.

Edwards, B. (1998) Seeing is believing - picture building: a key component of telephone triage. *Journal of Clinical Nursing,* 7, 51-57.

Farquharson, B., Allan, J., Johnston, D. W., Johnston, M., Choudhary, C. and Jones, M. C. (2012) Stress amongst nurses working in a healthcare telephone-advice service: relationship with job satisfaction, intention to leave, sickness absence, and performance. *Journal of Advanced Nursing,* 68, 1624-1635.

Goldstein, H. (2010) *Multilevel Statistical Models*, London: Wiley.

Goldstein, H., Carpenter, J., Kenward, M. G. and Levin, K. A. (2009) Multilevel models with multivariate mixed response types. *Statistical Modelling,* 9, 173-197.

Goldstein, H. and Kounali, D. (2009) Multilevel multivariate modelling of childhood growth, numbers of growth measurements and adult characteristics. *Journal of the Royal Statistical Society A (Statistics in Society),* 172, 599-613.

Griffin, M. A. (1997) Interaction between individuals and situations: using HLM procedures to estimate reciprocal relationships. *Journal of Management,* 23, 759-773.

Health Care Commission (2011) NHS Staff Survey 2010, London: Department of Health.

Hedeker, D., Mermelstein, R. J. and Demirtas, H. (2012) Modeling between-subject and within-subject variances in ecological momentary assessment data using mixed-effects location scale models. *Statistics in Medicine,* 31, 3328-3336.

Jamison, R. N., Raymond, S. A., Levine, J. G., Slawsby, E. A., Nedeljkovic, S. S. and Katz, N. P. (2001) Electronic diaries for monitoring chronic pain: one-year validation study. *Pain,* 91, 277-285.

Johnston, D. W., Jones, M. C., McCann, C. K. and McKee, L. (2013) Stress in nurses: stress-related affect and its determinants examined over the nursing day. *Annals of Behavioural Medicine,* 45, 348-356.

Johnston, M. and Pollard, B. (1991) Length of nurses' working shift: stress and information processing In *Psychology in Hospitals* (ed D. Biondi), pp. 101-105, Rome: Nuova Editrice Spada.

Jones, M. C. and Johnston, D. W. (2000) A critical review of the relationship between perception of the work environment, coping and mental health in trained nurses, and patient outcomes. *Clinical Effectiveness in Nursing,* 4, 75-85.

Keinan, G. (1987) Decision making under stress: scanning of alternatives under controllable and uncontrollable threats. *Journal of Personality and Social Psychology,* 52, 639-644.

Kučera, H. and Francis, W. N. (1967) *Computational Analysis of Present-day American English*, Providence, R.I.: Brown University Press.

Leckie, G. B., French, R., Charlton, C. and Browne, W. J. (2014) Modeling heterogeneous variance-covariance components in two-level models. *Journal of Educational and Behavioral Statistics,* 39, 307-332.

Lillard, L. A. and Panis, C. W. A. (1998-2003) *aML User's Guide and Reference Manual, Version 2*, Los Angeles: EconWare.

Liu, L. C. and Hedeker, D. (2006) A mixed effects regression model for longitudinal multivariate ordinal data. *Biometrics,* 62, 261-268.

Liu, L. C., Hedeker, D. and Mermelstein, R. J. (2013) Modeling nicotine dependence: an application of a longitudinal IRT model for the analysis of adolescent nicotine dependence syndrome scale. *Nicotine and Tobacco Research,* 15, 326-333.

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T. and Muthén, B. (2008) The multilevel latent covariate model: a new, more reliable approach to group-level effects in contextual studies. *Psychological Methods,* 13, 203-229.

Maslach, C. and Schaufeli, W. (2000) Job burnout. *Annual Review of Psychology,* 52, 397-422.

McVicar, A. (2003) Workplace stress in nursing: a literature review. *Journal of Advanced Nursing,* 44, 633-642.

Muthén, B. (1991) Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement,* 28, 338-354.

Muthén, B. and Asparouhov, T. (2009) Growth mixture modeling: analysis with non-Gaussian random effects In *Longitudinal Data Analysis* eds G. M. Fitzmaurice, M. Davidian, G. Verbeke and G. Molenberghs), Boca Raton, FL: CRC Press.

Neuhaus, J. M. and Kalbfleisch, J. D. (1998) Between-and within-cluster covariate effects in the analysis of clustered data. *Biometrics,* 54, 638-645.

Nolting, H. D., Berger, J., Schiffhorst, G., Genz, H. O. and Kordt, M. (2002) Job strain as a risk factor for occupational accidents among hospital nursing staff. *Gesundheitswesen,* 64, 25-32.

Preacher, K. J. (2011) Multilevel SEM strategies for evaluating mediation in three-level data. *Multivariate Behavioral Research,* 46, 691-731.

Preacher, K. J., Zyphur, M. J. and Zhang, Z. (2010) A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods,* 15, 209-233.

Sampson, R. J., Raudenbush, S. W. and Earls, F. (1997) Neighborhoods and violent crime: a multilevel study of collective efficacy. *Science,* 277, 918-924.

Sayers, A., Heron, J., Smith, A. D. A. C., Macdonald-Wallis, C., Gilthorpe, M. S., Steele, F. and Tilling, K. (2015) Joint modelling compared with two stage methods for analysing longitudinal data and prospective outcomes: A simulation study of childhood growth and BP. *Statistical Methods in Medical Research*.

Schaufeli, W. and Enzmann, D. (1998) *The burnout companion to study and practice: A critical analysis*, London: Taylor & Francis.

Schmeichel, B. J. (2007) Attention control, memory updating and emotion regulation temporarily reduce the capacity for executive control. *Journal of Experimental Psychology: General,* 2, 241-255.

Shapiro, D., Jamner, L. D., Davydov, D. M. and James, P. (2002) Situations and moods associated with smoking in everyday life. *Psychology of Additive Behaviors,* 16, 342-345.

Skrondal, A. and Rabe-Hesketh, S. (2004) *Generalized Latent Variable Modelling: Multilevel, Longitudinal, and Structural Equation Models*, Boca Raton, FL: Chapman & Hall/CRC.

Smith, E. E. (1968) Choice reaction time: an analysis of major theoretical positions. *Psychological Bulletin,* 69, 77-119.

Snijders, T. A. B. and Bosker, R. J. (2012) *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*, London: Sage Publishers.

Steele, F. and Goldstein, H. (2006) A multilevel factor model for mixed binary and ordinal indicators of women's status. *Sociological Methods & Research,* 35, 137-153.

Stober, J. (2001) The social desirability scale-17 (SDS-17): convergent validity, discriminant validity, and relationship with age. *European Journal of Psychological Assessment,* 17, 222-232.

Stone, A. A. and Shiffman, S. (1994) Ecological Momentary Assessment in behavioral medicine. *Annals of Behavioral Medicine,* 16, 199-202.

Tourangeau, R. (2000) Remembering what happened: memory errors and survey reports In *The Science of Self-Report: Implications for Research and Practice* eds A. A. Stone, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman and V. S. Cain), Mahwah, NJ: Erlbaum.

van der Linden, D., Keijsers, G. J. P., Eling, P. and van Schaijk, R. (2005) Work-related stress and attention to action: An initial study on burnout and executive control. *Work and Stress,* 19, 1-14.

Wallace, J. C. and Chen, G. (2005) Development and validation of a work-specific measure of cognitive failure: Implications for occupational safety. *Occupational and Organizational Psychology,* 78, 615-632.

Walls, T. A. and Schafer, J. L. (2006) *Models for Intensive Longitudinal Data*, New York: Oxford University Press.

Watson, D., Clark, L. and Tellegen, A. (1988) Development and validation of a brief measure of positive and negative affect: the PANAS Scale. *Journal of Personality and Social Psychology,* 54, 1063-1070.

Table 1 Descriptive statistics for cognitive function outcomes, call stress and covariates

Cognitive function	Mean	SD	n
Reaction time in post-shift computerized task (ms)	605.0	88.5	270
No. errors in post-shift computerized task (out of 100)	2.1	2.3	270
Self-reported no. errors during shift (max = 75)	25.6	8.4	247[a]

Stressfulness of call	Percentage		
	All calls (n=4913)	First call (n=270)	Last call (n=270)
1 (not at all)	65.0	57.4	63.7
2	24.1	30.7	23.7
3	9.0	10.4	9.3
4	1.6	1.1	3.0
5 (extremely)	0.3	0.4	0.4

Covariates	Mean	SD	n
Reaction time in pre-shift computerized task (ms) [b]	616.9	90.2	270
No. errors in pre-shift computerized task (out of 100) [b]	2.6	5.8	270
Social desirability bias [c]	9.2	2.7	147
Positive affect [c]	32.2	7.2	147
Negative affect [c]	16.0	6.4	147

[a] The self-reported number of errors was missing for 23 nurses; [b] Pre-shift reaction time and number of errors were included as predictors in models for their corresponding post-shift measure; [c] These nurse-level variables are included in the analysis of self-reported number of errors during a shift in an attempt to adjust for differences in nurses' reporting tendencies.

Table 2 Estimates from multilevel ordered logit model of call stress, the variance components measurement model of equation (5)

Parameter	Estimate	(SE)
Thresholds		
τ_1	0.796	(0.126)
τ_2	2.850	(0.135)
τ_3	5.101	(0.173)
τ_4	7.102	(0.301)
Random effect standard deviations		
Between-nurse (σ_v)	1.327	(0.097)
Between-shift within-nurse (σ_u)	0.615	(0.073)

Table 3 Effects of stress on measures of cognitive function from random intercept multilevel structural equation models

Parameter	Estimate	(SE)	p-value	Stan. est.[e]
Reaction time in post-shift task [a]				
M1. Overall between-shift (λ) [c]	0.059	(0.031)	0.053	0.086
M2. Decomposition [d]				
Between-nurse (λ_B)	0.090	(0.052)	0.084	0.119
Within-nurse between-shift (λ_W)	-0.043	(0.091)	0.635	-0.026
Log(no. errors in post-shift task) [b]				
M1. Overall between-shift (λ)	0.044	(0.050)	0.376	0.064
M2. Decomposition				
Between-nurse (λ_B)	0.011	(0.061)	0.857	0.015
Within-nurse between-shift (λ_W)	0.190	(0.168)	0.259	0.117
Log(self-reported no. failures during shift) [a]				
M1. Overall between-shift (λ)	0.251	(0.051)	<0.001	0.367
M2. Decomposition				
Between-nurse (λ_B)	0.252	(0.061)	<0.001	0.334
Within-nurse between-shift (λ_W)	0.251	(0.155)	0.106	0.154

Separate models were estimated for each cognitive outcome. [a] Reaction time and log(self-reported no. failures) were standardised and treated as continuous responses; [b] Number of errors in post-shift task was treated as a count variable and analysed using Poisson regression. [c] In model M1, given by equations (5) and (6), λ is the coefficient of the sum of between-nurse stress and within-nurse between-shift stress $v_k + u_{jk}$; [d] In model M2, given by equations (5) and (7), λ_B is the coefficient of v_k and λ_W is the coefficient of u_{jk}; [e] Standardised coefficients representing the effect of a 1 standard deviation change in the associated stress random effect, calculated using estimates of the random effect standard deviations from the measurement model. For example, a standardised estimate of λ is $\hat{\lambda}^* = \sqrt{\hat{\sigma}_u^2 + \hat{\sigma}_v^2}\,\hat{\lambda}$.

Table 4 Estimates from multilevel ordered logit model of call stress, the latent growth model with random slope for call number of equation (8a)

Parameter	Estimate	(SE)
Thresholds		
τ_1	0.581	(0.133)
τ_2	2.685	(0.142)
τ_3	5.003	(0.181)
τ_4	7.032	(0.307)
Linear effect of call number (γ)	-0.027	(0.007)
Random effect standard deviations		
Between-nurse (σ_v)	1.356	(0.091)
Within-nurse between-shift		
Intercept, call 1 (σ_{u0})	0.573	(0.124)
Between shift		
Slope of call number ($\sigma_{\tilde{u}1}$)	0.064	(0.008)
Intercept-slope correlation ($\rho_{\tilde{u}01}$)	-0.324	(0.160)

Table 5 Effects of stress at start of shift and change in stress during shift on selected measures of cognitive function from multilevel random slope structural equation models

Parameter	Estimate	(SE)	p-value	Stan. est.[d]
Reaction time in post-shift task [a]				
M3. Overall between-shift effects [b]				
Initial (call 1) stress (λ_0)	0.061	(0.034)	0.074	0.090
Change in stress (λ_1)	0.032	(0.814)	0.968	-0.002
M4. Decomposition [c]				
Between-nurse initial (λ_{B0})	0.079	(0.042)	0.062	0.107
Within-nurse between-shift initial (λ_{W0})	-0.050	(0.123)	0.683	-0.029
Change in stress (λ_1)	-0.117	(0.795)	0.882	-0.007
Log(self-reported no. failures during shift) [a]				
M3. Overall between-shift effects [b]				
Initial (call 1) stress (λ_0)	0.179	(0.054)	0.001	0.264
Change in stress (λ_1)	4.581	(1.471)	0.002	0.293
M4. Decomposition [c]				
Between-nurse initial (λ_{B0})	0.182	(0.062)	0.003	0.247
Within-nurse between-shift initial (λ_{W0})	0.125	(0.187)	0.504	0.072
Change in stress (λ_1)	4.383	(1.514)	0.004	0.281

Separate models were estimated for each cognitive outcome. [a] Reaction time and log(self-reported no. failures) were standardised and treated as continuous responses; [b] In model M3, λ_0 is the coefficient of the sum of between-nurse stress and within-nurse between-shift stress intercept effects and λ_1 is the coefficient of the sum of the corresponding slope effects; [c] In model M4, λ_{B0} is the coefficient of v_{0k} and λ_{W0} is the coefficient of u_{0jk}; [d] Standardised coefficients representing the effect of a 1 standard deviation change in the associated stress random effect, calculated using estimates of the random effect standard deviations from the measurement model.
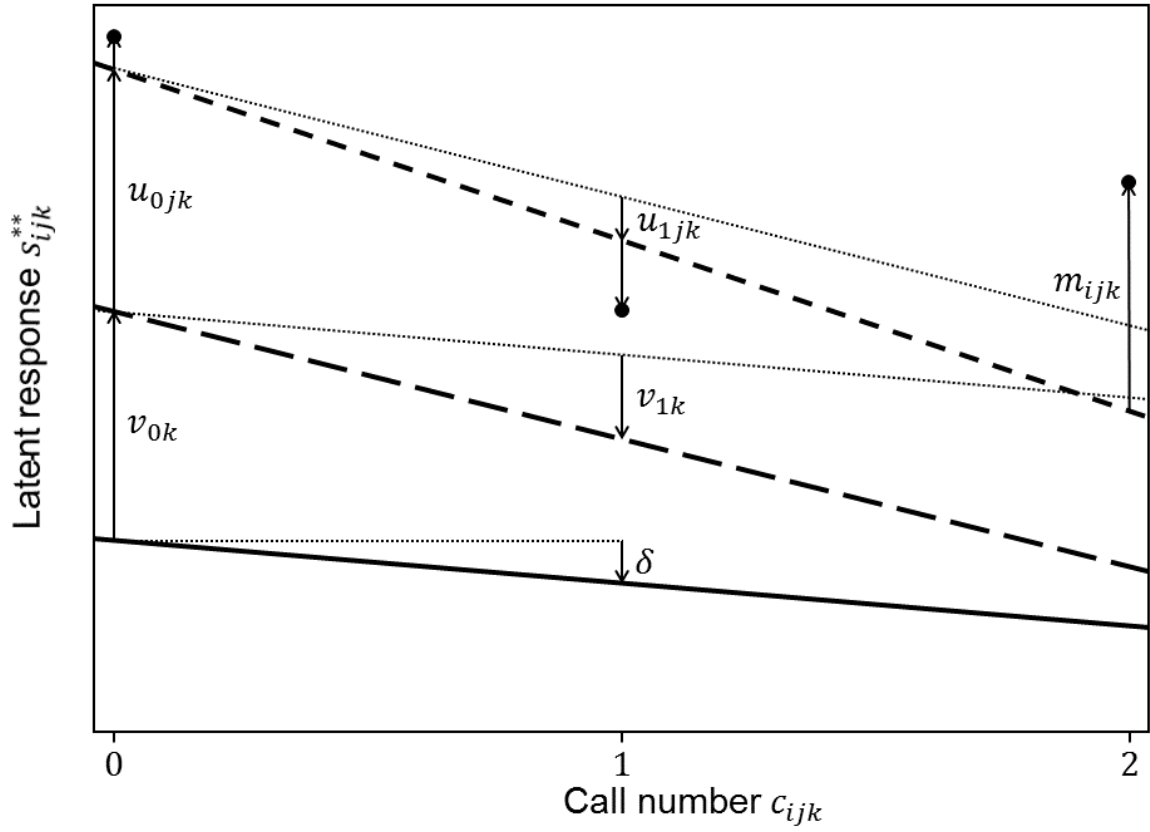
Figure 1 Illustration of a three-level measurement model for latent continuous stress s_{ijk}^{**} underlying the observed ordinal stress measure s_{ijk} with random slopes at the shift and nurse level for call number c_{ijk}. This is the latent response formulation of the proportional odds model of equation (8). The thick solid line denotes the population-averaged relationship between stress and call number for the first 3 calls in a shift; the thick long-dashed line denotes the relationship for nurse k; and the thick short-dashed line denotes the relationship for nurse k on shift j
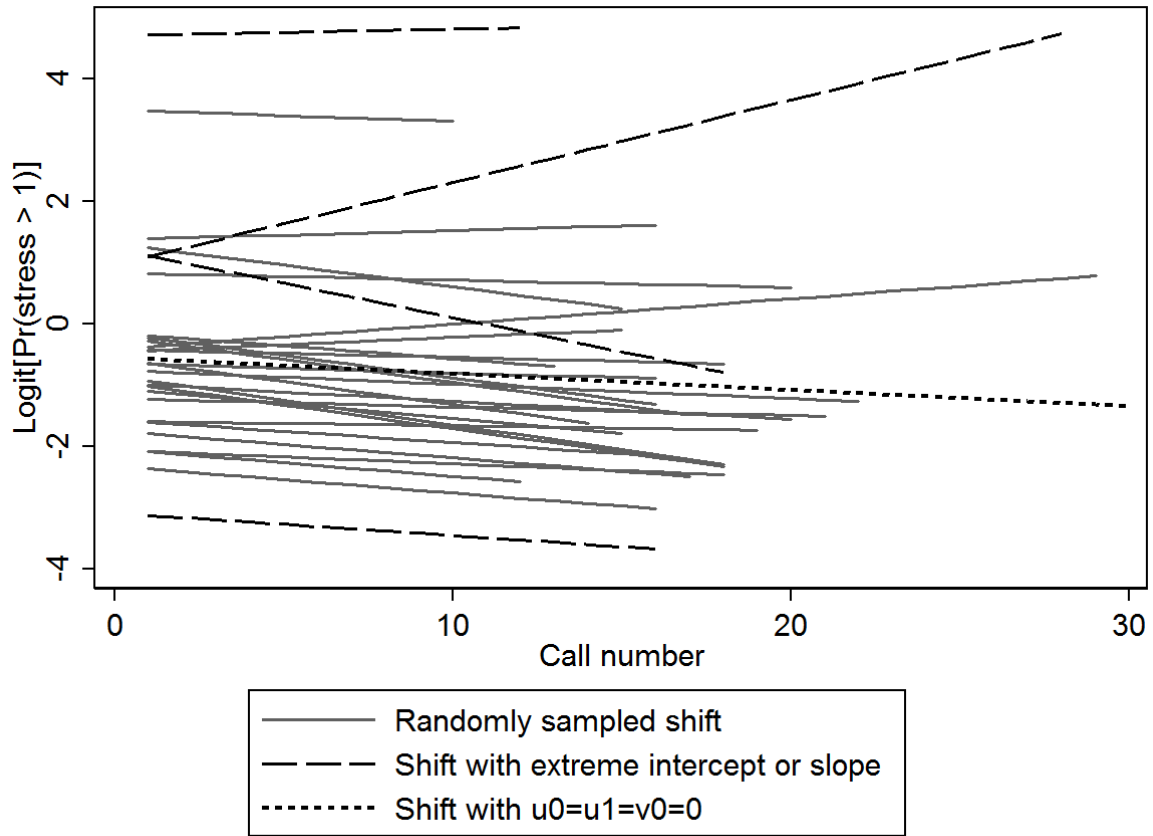
Figure 2 Predicted stress trajectories from the random slopes model of equation (8a). The plot shows the log-odds that stress $s_{ijk} > 1$ (above "not at all") by call number. The long-dashed lines are trajectories for shifts with the largest and smallest estimates for the intercept ($u_{0jk} + v_{0k}$) and slope ($\tilde{u}_{1jk}$). The short-dashed line corresponds to a shift and nurse at the mean of the random effect distributions. The other lines are the trajectories for a randomly selected shift for a random sample of 25 nurses.

Supplementary Materials for "Multilevel structural equation models for longitudinal data where predictors are measured more frequently than outcomes: An application to the effects of stress on the cognitive function of nurses"

F. Steele, P. Clarke, G. Leckie, J. Allan and D. Johnston

Examples of multilevel SEMs for effect of call-level stress on shift-level cognitive outcomes and their estimation in the aML software

1. Introduction

In this appendix we describe how the multilevel simultaneous equations models (SEM) presented in the paper can be fitted using the aML software. We give a brief explanation of the structure of the input file and provide aML syntax, but readers are referred to the aML manual for further details. The software and documentation can be downloaded free from www.applied-ml.com.

aML references

Lillard, L. A., & Panis, C. W. A. (1998-2003). *aML User's Guide and Reference Manual, Version 2*. Los Angeles: EconWare. Download from www.applied-ml.com

Steele, F. (2004). A review of aML (Release 2.0): Centre for Multilevel Modelling, University of Bristol. Download from http://www.bris.ac.uk/cmm/learning/mmsoftware/aml.html

2. Structure of input data file

The input data file has 1 record per call, with the values of shift-level variables replicated across calls from the same shift. aML requires variables to be grouped according to the level at which they are defined, starting with those at the highest level in the hierarchy (nurse in our case). The first variable in the data file must be a numeric identifier for the highest-level units (nurse). There are additional requirements for the format required by aML for more complex structures – see the aML manual for further details.

The aML program `raw2aml` converts an ascii text file into an aML data file (`call_based.dat`), which is the input for the estimation program `aml`.

The following variables are referred to in the aML syntax given below.

Variable name	Description
_id	Nurse identifier. This is the first variable in the aML data file which aML will recognise as the identifier for the highest level in the hierarchy (and is always referred to as _id).
shift	Shift ID, coded 1 and 2 within nurses
shift2	Dummy variable for the 2nd shift (coded 1 for shift=2, and 0 for shift=1)
callc1	Call number (coded 0, 1, 2 …)
stress	Ordinal stress measure (coded 1 to 5)
zpostrt	Post-shift reaction time (standardised)
zprert	Pre-shift reaction time (standardised)

3. Random intercept model with overall between-shift effect of stress and continuous cognitive outcome (Structural model is M1 in Table 1)

We begin with a joint model for stress and reaction time. The measurement model for stress is a 3-level random intercept ordered logit model, as in equation (5) of the paper. The structural model for reaction time is a 2-level model including as predictors the stress and shift level random effects from the stress model. In this first model, we estimate the effect of overall shift-level stress on reaction time, as in equation (6).

The syntax can be broken down as follows:

- The maximum number of iterations is set to 200 and robust standard errors are requested.

- Data set name (dsn) is **call_based.dat**

- Define threshold parameters for the ordinal model of stress (τ_k)

- Define coefficient of stress random effect (λ) in model for reaction time

- Specify explanatory variables for the stress and reaction time equations (called 'regressor sets' in aML) and name them **BetaX** and **AlphaX**. **BetaX** for stress includes only a constant; its coefficient is later constrained to zero for identification as the inclusion of τ_k makes an overall intercept redundant. **AlphaX** for post-shift reaction time includes a constant, pre-shift reaction time, a dummy for shift 2, and their interaction.

- Specify four independent normally distributed random effects for stress (named **vs** and **us**) and reaction time (**vc** and **uc**). The level at which each random effect varies, and the response to which each is attached, is specified in the later model statements.

- Specify an ordered logit model for stress with explanatory variables in **BetaX** and random effects at the nurse and shift level. `intres` specifies a residual (which must be integrated out for a generalised linear model). Specifying `draw=_id` defines **vs** as a nurse-level residual because aML recognises **_id** as the first variable in the dataset which is the identifier for the highest level units. Similarly `draw=shift` defines **us** as a shift-level residual.

- Specify a linear regression model for reaction time (**zpostrt**) with explanatory variables in **AlphaX** and random effects at the nurse and shift level. This model is fitted only to the first call for each shift because **zpostrt** is defined at the shift level. As for the stress model, we have a residual at the nurse level (**vc**). A residual at the shift level (**uc**) is specified using `draw=_iid`; these are the usual level 1 residuals in a linear regression which are assumed to be independent and identically distributed.

- Provide starting values for each parameter in the order in which they have been specified: thresholds for the ordinal model for stress (**Tau**), coefficients of latent variables (**lambda**), coefficients of variables in the regressor set **BetaX**, coefficients of variables in the regressor set **AlphaX**, and finally the residual (random effect) standard deviations and correlations. For random effect parameters, starting values must be listed in the same order as in the `define normal distribution` statements.

  You can name the parameters whatever you like. The 'T' next to each parameter tells aML that you wish it to be estimated. Parameters can be constrained at a particular value using 'F' instead of 'T'. Here, for example, the intercept in the stress model (**cons_st**) is fixed at zero.

  Starting values will usually come from fitting preliminary, simpler models. In general it is advisable to start with simple models, building up to the full model gradually. It is also possible to estimate a model in stages using the aML `update` program so that estimates from one model are used as starting values for another.

2

See the aML User Guide for a full description of all aML commands with plenty of examples and guidance on starting values.

Syntax

```
/* Joint model for stress and post-shift reaction time */
/* Stress: 3-level model with nurse, shift and call residuals */
/* Cognitive RT: 1-level model with nurse residual */
/* No covariates in stress model, pre-shift RT and interaction with shift in model for post-
shift RT*/

/* Include sum of nurse stress and shift stress (within nurse) residuals as predictor of
reaction time */

option numerical standard errors;
option iterations=200;

dsn='call_based.dat';

define vector Taus; dim=4;

define parameter lambda;

define regset BetaX;
  var = 1;

define regset AlphaX;
  var = 1 zprert shift2 zprert*shift2;

define normal distribution; dim=1; number of integration points=16; name=vs;
define normal distribution; dim=1; number of integration points=16; name=us;

define normal distribution; dim=1; number of integration points=16; name=vc;
define normal distribution; dim=1; number of integration points=16; name=uc;

ordered logit model;
  outcomes=stress-1 stress;
  thresholds=Taus;
  model = regset BetaX + intres(draw=_id, ref=vs) + intres(draw=shift, ref=us);

continuous model;
  keep if callno==1;
  outcome=zpostrt;
  model = regset AlphaX
      + par lambda*res(draw=_id, ref=vs)
      + par lambda*res(draw=shift, ref=us)
      + res(draw=_id, ref=vc) + res(draw=_iid, ref=uc);


starting values;
tau1        T    0.8
tau2        T    2.9
tau3        T    5.1
tau4        T    7.1
lambda      T    0
cons_st     F    0
cons_rt     T    0
prert       T    0.6
shift2      T    0.1
preXsh2     T    0
sigvs       T    1.3
sigus       T    0.6
sigvc       T    0.5
siguc       T    0.5
;
```

Output

```
======================================================================
Log Likelihood:   -4116.6183
                                    Based on numerical Hessian
   Parameter     Free?     Estimate        Std Err       T-statistic

   1  Tau1         T      .81032266729     .12255484467       6.6119
   2  Tau2         T     2.8630882888     0.1318548489       21.7139
   3  Tau3         T     5.1130796553      .17079461211      29.9370
   4  Tau4         T     7.1079491765      .29968680188      23.7179
   5  lambda       T      .05902263579     .03052190768       1.9338
   6  cons_st      F             0.0
   7  cons_rt      T     -.06012307481     .05918458097      -1.0159
   8  prert        T      .62596420898     .05177194954      12.0908
   9  shift2       T      .11458476254     .06363620249       1.8006
  10  preXsh2      T      .01489769146     .06199692706       0.2403
  11  sigvs        T     1.3324098359      .09649164755      13.8086
  12  sigus        T      .59385556635     .07480836861       7.9384
  13  sigvc        T      .49038995865     .04471271306      10.9676
  14  siguc        T      .44450711406     .02869942354      15.4884


----------------------------------------------------------------------
```

4. Random intercept model with between-within decomposition of effect of stress and continuous cognitive outcome (Structural model is M2 in Table 3)

In the next model, the structural model for reaction time again includes random effects from the stress model as predictors, but we now estimate different effects at the nurse and shift levels. In other words, we decompose the effect of stress into a between-nurse component and within-nurse between-shift component, as in equation (7). The measurement model for stress is a 3-level random intercepts ordered logit model as before.

Changes to the aML syntax are indicated in red. We now:

- Define coefficients for the nurse-level (λ_B, **lambdav**) and shift-level (λ_W, **lambdau**) random effects for stress in the model for reaction time.

- Specify **lambdav** as the coefficient of random effect **vs** and **lambdau** as the coefficient of random effect **us** in the model for reaction time.

- Specify starting values for **lambdav** and **lambdau**.

Syntax

```
/* Joint model for stress and post-shift reaction time */
/* Stress: 3-level model with nurse, shift and call residuals */
/* Cognitive RT: 1-level model with nurse residual */
/* No covariates in stress model, pre-shift RT and interaction with shift in model for post-
shift RT*/

/* Include nurse stress residual as predictor of reaction time */
/* Include shift stress (within nurse) residual as predictor of reaction time */


option numerical standard errors;
option iterations=100;

dsn='call_based.dat';

define vector Taus; dim=4;

define parameter lambdav;
```

```
define parameter lambdau;

define regset BetaX;
  var = 1;

define regset AlphaX;
  var = 1 zprert shift2 zprert*shift2;

define normal distribution; dim=1; number of integration points=16; name=vs;
define normal distribution; dim=1; number of integration points=16; name=us;

define normal distribution; dim=1; number of integration points=16; name=vc;
define normal distribution; dim=1; number of integration points=16; name=uc;

ordered logit model;
  outcomes=stress-1 stress;
  thresholds=Taus;
  model = regset BetaX + intres(draw=_id, ref=vs) + intres(draw=shift, ref=us);

continuous model;
  keep if callno==1;
  outcome=zpostrt;
  model = regset AlphaX
      + par lambdav*res(draw=_id, ref=vs)
      + par lambdau*res(draw=shift, ref=us)
      + res(draw=_id, ref=vc) + res(draw=_iid, ref=uc);

starting values;
tau1      T    .81
tau2      T    2.86
tau3      T    5.11
tau4      T    7.11
lambdav   T    0
lambdau   T    0
cons_st   F    0
cons_rt   T    -0.06
prert     T    0.63
shift2    T    0.12
preXsh2   T    0.01
sigvs     T    1.33
sigus     T    0.59
sigvc     T    0.49
siguc     T    0.44
;
```

Output

```
========================================================================

Log Likelihood:  -4115.8689
                                    Based on numerical Hessian
  Parameter    Free?     Estimate        Std Err        T-statistic

   1  Tau1        T      .80163860034    .12524739186       6.4004
   2  Tau2        T      2.8550246957    .13432119021      21.2552
   3  Tau3        T      5.1058537495    .17281571467      29.5451
   4  Tau4        T      7.0981232759    .30054948085      23.6172
   5  lambdav     T      .09036978568    .05214263841       1.7331
   6  lambdau     T     -.04309918718    .09098450126      -0.4737
   7  cons_st     F             0.0
   8  cons_rt     T     -.06179075119    .05892744471      -1.0486
   9  prert       T      .62586633647    .05140101551      12.1761
  10  shift2      T      .11805643452    .06335650292       1.8634
  11  preXsh2     T      .01897278826    .06180491085       0.3070
  12  sigvs       T      1.3373997678    .10395574162      12.8651
  13  sigus       T      .60262581173    .07398374418       8.1454
  14  sigvc       T      .47946296956    .04613538454      10.3925
  15  siguc       T      .44321121721    .02872071177      15.4318


------------------------------------------------------------------------
```

5. Random slope model with between-within decomposition of effect of stress and continuous cognitive outcome (Structural model is M4 in Table 5)

In the final model, the measurement model is generalised to a 3-level growth model where call stress is a linear function of call number, as in equation (8a). A random slope is fitted at the shift level; at the nurse level there is only a random intercept. The structural model includes the three random effects from the stress model as predictors, with separate coefficients estimated for each.

As before, changes to the aML syntax compared to the previous model are indicated in red. We now:

- Define coefficients for each of the random effects for stress in the model for reaction time: random intercepts at the nurse-level (λ_{B0}, **lambdav**) and shift-level (λ_{W0}, **lambdau0**) and for the random slope at the shift level (λ_1, **lambdau1**).

- Add call number (**callc1**) to the regressor set **BetaX** for the stress model.

- Define intercept and slope random effects at the shift level (**u0s** and **u1s**) and specify their distribution as bivariate normal.

- Include a shift-level random intercept (**u0s**) and random slope (**u1s**) in the model for stress. The random slope effect is multiplied by **callc1**.

- Specify **lambdav**, **lambdau0** and **lambdau1** as the coefficients of random effects **vs**, **u0s** and **u1s** respectively in the model for reaction time.

- Specify starting values for **lambdav**, **lambdau0** and **lambdau1**, the standard deviations of random effects **u0s** and **u1s**, and their correlation.

Syntax

```
/* Joint model for stress and post-shift reaction time */
/* Stress: 3-level model with nurse, shift and call residuals */
/* Cognitive RT: 2-level model with nurse and shift residuals */
/* No covariates in stress model, pre-shift RT and interaction with shift in model for post-
shift RT*/

/* Linear call effect in stress model */
/* Call number centred at 1st call (callno=1) */
/* Random coefficient on call number at shift level (but not nurse level) in stress model */


/* Include nurse stress residual as predictor of RT */
/* Include shift stress (within nurse) intercept and slope residuals as predictor of RT */

option numerical standard errors;
option iterations=200;

dsn='call_based.dat';

define vector Taus; dim=4;

define parameter lambdav;
define parameter lambdau0;
define parameter lambdau1;

define regset BetaX;
  var = 1 callc1;

define regset AlphaX;
  var = 1 zprert shift2 zprert*shift2;

define normal distribution; dim=1; number of integration points=16; name=vs;
define normal distribution; dim=2; number of integration points=16;
  name=u0s; name=u1s;
```

```
define normal distribution; dim=1; number of integration points=16; name=vc;
define normal distribution; dim=1; number of integration points=16; name=uc;

ordered logit model;
  outcomes=stress-1 stress;
  thresholds=Taus;
  model = regset BetaX + intres(draw=_id, ref=vs)
       + intres(draw=shift, ref=u0s)
       + callc1*intres(draw=shift, ref=u1s);

continuous model;
  keep if callno==1;
  outcome=zpostrt;
  model = regset AlphaX
       + par lambdav*res(draw=_id, ref=vs)
       + par lambdau0*res(draw=shift, ref=u0s)
       + par lambdau1*res(draw=shift, ref=u1s)
       + res(draw=_id, ref=vc) + res(draw=_iid, ref=uc);

starting values;
tau1        T    0.80
tau2        T    2.86
tau3        T    5.11
tau4        T    7.10
lambdav     T    0
lambdau0    T    0
lambdau1    T    0
cons_st     F    0
callc1      T    0
cons_rt     T   -0.06
prert       T    0.63
shift2      T    0.12
preXsh2     T    0.02
sigvs       T    1.34
sigu0s      T    0.60
sigu1s      T    0.01
rhou01s     T    0
sigvc       T    0.48
siguc       T    0.44
;
```

Output

```
=======================================================================

Log Likelihood:  -4088.1014
                                        Based on numerical Hessian
   Parameter    Free?     Estimate       Std Err      T-statistic

   1  Tau1       T     .57794557036    .13451690898      4.2965
   2  Tau2       T     2.6865437681    .14277114574     18.8171
   3  Tau3       T     5.0147367893    .18300067136     27.4028
   4  Tau4       T     7.0531798831    .31064010909     22.7053
   5  lambdav    T     .07857194664    .04204537424      1.8687
   6  lambdau0   T    -.05019279237    .12285385712     -0.4086
   7  lambdau1   T    -.11737679459    .79486016269     -0.1477
   8  cons_st    F            0.0
   9  callc1     T    -.02549762177    .00730854125     -3.4887
  10  cons_rt    T    -.06292847806    .05930572001     -1.0611
  11  prert      T     .62557099119    .05206572827     12.0150
  12  shift2     T     .11899326309    .06366256456      1.8691
  13  preXsh2    T     .01774720603    .06202201183      0.2861
  14  sigvs      T     1.3660422903    .10091316953     13.5368
  15  sigu0s     T      0.573226996    .12487436705      4.5904
  16  sigu1s     T     .06102864752    .00819908411      7.4433
  17  rhou01s    T    -.32865469601   0.1891743091      -1.7373
  18  sigvc      T     .48748931172    .04516558179     10.7934
  19  siguc      T     .44326463758    .02876747528     15.4085
-----------------------------------------------------------------------
```