Analysis of household effects on longitudinal health outcomes using a joint mean-correlation multilevel model with grouped random effects

Analysis of household effects on longitudinal health outcomes using a joint mean-correlation multilevel model with grouped random effects

Fiona Steele*

Department of Statistics, London School of Economics & Political Science, UK

Siliang Zhang*†

School of Statistics, East China Normal University, China

Paul Clarke

Institute for Social and Economic Research, University of Essex, UK

Abstract

Previous cross-sectional research has found correlation in the health outcomes of coresident adults. However, the study of household effects in longitudinal data is challenging due to the complex association structure arising from changes in household membership over time. We propose a 'grouped' multilevel model where the groups (called 'superhouseholds') are specified to capture changes in household structure. Correlated household random effects are used to capture correlations between households sharing an individual(s), and correlations between household pairs can depend on covariates that describe their relationship. We develop a constrained MCMC procedure for model estimation that ensures the group-specific correlation matrices (where dimensions can vary across groups) are positive definite, and implement it as an R package. The performance and robustness of our models is evaluated in a simulation study and then applied in analyses of household and area effects on self-rated physical and mental health in the UK using data from a national household panel survey.

Keywords: Correlated random effects; Correlation model; Joint mean-covariance model.

1 Introduction

There is considerable interest among health researchers in the degree of correlation among coresidents in health-related attitudes, behaviours and outcomes. Much of this research has focused on couples, finding correlation in behaviours such as physical activity, diet and substance use (e.g. Brazeau & Lewis, 2021), clinical measures of physical health such as adiposity and associated biomarkers, and self-rated general health (e.g. Davillas & Pudney, 2017). Other research has gone beyond couples to consider within-household correlation in health among all adult coresidents. Household effects have been found in self-rated measures of general health, physical health functioning and limiting illness (Chandola et al., 2005; Sacker et al., 2006) and of mental health and wellbeing (Weich et al., 2003;

*These authors contributed equally.

†Corresponding author. Email: slzhang@fem.ecnu.edu.cn

Griffith & Jones, 2020). Within-household correlation has also been found across a range of non-health outcomes including social and political attitudes (Brynin et al., 2008; Steele et al., 2019) and voting in political elections (Johnston et al., 2005). Previous research, mainly on couples, suggests that the mechanisms for such household effects include homophily, contagion and shared environmental factors. Homophily, or homogamy in the special case of couples, is the tendency for individuals to choose partners and other coresidents who are similar to themselves with respect to socio-demographic characteristics, values or behaviours (McPherson et al., 2001), while contagion refers to a causal effect of one person's attitudes or behaviours on another's, leading to convergence in their outcomes over time (Davillas & Pudney, 2017). Environmental factors such as household economic circumstances, nutrition, housing quality and area deprivation have been cited in the geography and epidemiology literature as potential explanations for household effects on mental and physical health (Chandola et al., 2005; Griffith & Jones, 2020). Some authors have argued that apparent geographical clustering of health outcomes may be due to omitted household effects (Chandola et al., 2005).

Household panel surveys such as the UK Household Longitudinal Study (UKHLS) track individuals and their coresidents over time, providing rich social and economic data on individuals and their households. Such studies have been conducted in many countries and can be used to study within-household correlations in a range of individual health-related behaviours and outcomes, as well as social inequalities in health and wellbeing. Population registers linked to electronic health records are another source of longitudinal health data on coresident individuals. In spite of the availability of repeated measures data on individuals clustered in households, those studies that have considered household effects have largely been cross-sectional (typically restricting analysis to a single wave of a panel study) because the estimation of household effects with longitudinal data is challenging.

The problem is that, while a household can be defined cross-sectionally as a group of individuals who share accommodation, it is difficult to define households in a longitudinal sense because of changes in household composition over time. For example, an adult child may leave the parental home or a couple may split up; in each case, these demographic events lead to the formation of two new households, but any analysis of household effects must recognise that they are linked to the original household because they share individuals. Perhaps as a result of this complexity, most longitudinal analyses have ignored household effects and considered only individual effects. This not only misses a potentially important component of variation that is of substantive interest, but omitting household effects in longitudinal analysis may also affect inferences in two ways. First, ignoring household clustering will typically lead to misleading inferences regarding the contribution of omitted variables at the individual and higher levels of aggregation (e.g. areas). Second, it is well known that failure to allow for clustering leads to underestimation of standard errors of the estimated coefficients of covariates defined at the cluster level (e.g. Rabe-Hesketh & Skrondal, 2012, p.167).

This paper is motivated by the study of the relative contributions of unmeasured individual, household and area characteristics to repeated measures of individuals' self-rated mental and physical health when individuals change households, and the nature of correlations between linked households, using data from the UKHLS. We propose a longitudinal random effects model that distinguishes individual, household and area effects, allowing for change in household composition. This 'grouped' random effects model has separate household effects for each unique set of coresidents, where the usual independence assumption of standard multilevel models is relaxed to allow for correlation among random effects for households that share individuals over time. Moreover, the correlations between pairs of linked households are modelled as a linear function of covariates to describe the characteristics of and relationship between each pair, for example, the proportion of individuals in common or whether an individual in one household is the parent or the ex-partner of an individual in the other. There is a

long history of joint mean-covariance models that allow covariances among observations to depend on covariates. Much of this research has focused on longitudinal data where within-individual covariance (or correlation) matrices are modeled as functions of time (e.g. Pourahmadi, 1999). Outside the longitudinal setting, models have been proposed for multivariate data in which correlations are modelled as dependent on some measure of distance/similarity (Zou et al., 2017, 2022).

Our key methodological contribution is a novel Markov chain Monte Carlo (MCMC) estimation procedure for positive-definite and covariate-dependent correlation matrices that, unlike Pourahmadi (1999), does not rely on a Cholesky-type decomposition, and allows between-unit relationships to be characterised more generally in terms of pairwise covariates rather than (potentially arbitrary) distances/similarities. This paper instead builds on the work of Zhang et al. (2024), who developed a novel procedure for estimating covariate-dependent correlation matrices that ensures these matrices are positive-definite. However, while their work focused on a single, fixed-dimension correlation matrix for continuous latent variables in a model for cross-sectional multivariate outcomes, we extend this approach to the more general case of longitudinal non-hierarchical multilevel models involving random effects correlation matrices of varying dimensions, which presents a more complex estimation challenge. Our MCMC algorithm is implemented in an R package.

The remainder of the paper is organised as follows. We begin in Section 2 with a review of previous attempts to allow for household effects in longitudinal studies. We then describe our proposed grouped random effects model in Section 3, including an overview of previous research on modelling correlation matrices and an overview of our novel MCMC estimation procedure. Section 4 contains a description of the UKHLS, the dependence structures of which is used in a small-scale simulation study in Section 5 to evaluate our approach and to explore the impact of ignoring or misspecifying household effects. Then in Section 6 we illustrate the application and interpretation of the grouped random effects model in our analysis of annual health data from the UKHLS for 2009–2020. Finally, Section 7 includes a summary of the key contributions of our work, suggestions for further research, and an outline of potential applications of the proposed approach to longitudinal or multivariate data on grouped individuals where covariate-dependent within-group correlations are of interest.

2 Previous research on household effects in longitudinal analyses

Among the studies to have considered household effects, the most common approach has been to sidestep completely the problem of changing household membership by cross-sectionally analysing data from a single wave. The few genuinely longitudinal analyses have taken one of three approaches. We review all three of these approaches below to explain the weaknesses of each and show how our new methodology extends the family of multilevel models to address these issues.

Using matrix notation where all vectors are column vectors unless otherwise stated, and M' indicates the transpose of matrix (or vector) M, for traditional panel designs with n individuals observed at T equally spaced time points, the general form of the linear panel model is

$$\mathbf{Y} = X\boldsymbol{\beta} + \mathbf{r}, \tag{1}$$

where $\mathbf{Y} = (\mathbf{Y}_1', \ldots, \mathbf{Y}_n')'$, $X = (X_1', \ldots, X_n')'$, $\boldsymbol{\beta}$ is a vector of regression coefficients, $\mathbf{r} = (\mathbf{r}_1', \ldots, \mathbf{r}_n')'$; and $\mathbf{r}_i = (r_{1i}, \ldots, r_{Ti})'$ is the vector of zero-mean model residuals. Further, $\mathbf{Y}_i = (Y_{1i}, \ldots, Y_{Ti})'$ is the random vector of outcomes for sample individual $i = 1, \ldots, n$ at each panel wave $t = 1, \ldots, T$; and the covariates for individual i at wave t, $\mathbf{x}_{ti}$ (including a constant), can be similarly combined into matrix X_i with row t given by $\mathbf{x}_{ti}'$. Note that, in practice, it is straightforward to accommodate wave

nonresponse under a missing at random assumption, and that area effects are not introduced until Section 3 to focus on differences in how each method handles household structure. The key difference between approaches is in the residual variance-covariance matrix implied by the model, that is,

$$\text{cov}(\mathbf{Y}|X) = \text{cov}(\mathbf{r}), \tag{2}$$

a large $nT \times nT$ matrix the parameters of which we take to be at least as important to the analyst as $\boldsymbol{\beta}$. Let $h(ti)$ represent the household to which individual i belongs at wave t and $\mathbf{h}_{ti}$ the vector comprising individual i's household membership indicators at wave t for all J households, where each component is given by the indicator function $\text{I}\{h(ti) = j\}$ for $j = 1, \ldots, J$.

Hierarchical linear models. The standard hierarchical multilevel model can be used longitudinally for those who remain in the same household throughout such that $\mathbf{h}_{1i} = \cdots = \mathbf{h}_{Ti} = \mathbf{h}_i$. The hierarchical structure of the resulting model has (at least) three levels with waves nested within individuals within households and leads to

$$\text{cov}(\mathbf{r}) = \sigma_v^2 HH' + \sigma_u^2 CC' + \sigma_e^2 I, \tag{3}$$

where $H = (\mathbf{1}'_T \otimes \mathbf{h}_1, \ldots, \mathbf{1}'_T \otimes \mathbf{h}_n)'$ is the household membership matrix, $C = I_n \otimes \mathbf{1}_T$, I_n is the $n \times n$ identity matrix, $\mathbf{1}_T$ a vector of T ones, and $\otimes$ is the Kronecker product; $CC' = I_n \otimes J_T$, where J_T is a $T \times T$ matrix of ones, captures within-individual autocorrelation. The variance of the wave-specific errors and individual-level random effects are σ_e^2 and σ_u^2, respectively, and σ_v^2 is the variance of the household-level random effects. The usual assumption required for likelihood-based estimation is that these are normally distributed, but estimator performance has been shown to be robust to non-normality (Maas & Hox, 2004).

Three-level random effects models have been proposed for the analysis of repeated measures dyadic data on couples or other family pairs (Atkins, 2005). An alternative but closely related approach suitable for couples and other pairwise data is a bivariate two-level model (Raudenbush et al., 1995) which has been used in studies of spousal similarity of mental health outcomes (Gerstorf et al., 2013) and health behaviours (Brazeau & Lewis, 2021). These approaches have been applied widely in couple research, with analysis samples restricted to individuals who remain with the same partner throughout the observation period, but not in longitudinal analyses of household effects involving multiple household members where the exclusion of households with changing composition may lead to a severe reduction in sample size and selection bias. Conversely, if the sample is extended to include household composition changes such that $\mathbf{h}_{ti} \neq \mathbf{h}_{t'i}$ for at least one pair $t \neq t'$ then individuals are no longer nested in households so that the multilevel model is cross-classified rather than hierarchical (Goldstein, 2010, chapter 12) with $H = (\mathbf{h}_{11}, \ldots, \mathbf{h}_{T1}, \mathbf{h}_{12}, \ldots, \mathbf{h}_{Tn})'$. The principal drawback here is the modelling assumption that random effects for wholly new households, or for households where the composition changes between waves, are assumed to be independent and thus uncorrelated with any past households even those with individuals in common. This assumption may be realistic in certain applications but, we argue, not generally.

Multiple-membership multilevel models. The only existing random effects approach explicitly allowing for household change is the multiple membership multilevel model (Goldstein et al., 2000). Under this model, the household-level random effect for person i is $v^*_{ti} = \sum_{h=1}^{J} q_{h,ti} v_h$, where household random effects $v_h \sim N(0, \sigma_v^2)$ and $q_{h,ti}$ is a weight equal to zero only if individual i was never resident in household h (h indexes the set of J discrete households formed at any point during the observation period). The nonzero weights are not estimated but specified by the analyst depending

on whether i was a member of h for at least one wave. In other words, the household effect v^*_{ti} for individual i at wave t is formed as a weighted sum of effects contributed by the individual's current and previous households. The resulting covariance matrix is

$$\text{cov}(\mathbf{r}) = \sigma^2_v QQ' + \sigma^2_u CC' + \sigma^2_e I, \tag{4}$$

where $Q = (q_{h,ti})$ is the $nT \times H$ matrix of analyst-specified weights. The conditional covariances among Y_{ti} are implied by the choice of the weights $q_{h,ti}$. However, it is not obvious how to choose weights to ensure a substantively meaningful covariance structure, and simple choices can lead to implausible structures. Steele et al. (2019) concluded that there is no way of defining a multiple membership household effect that does not generally induce counterintuitive patterns of association.

Marginal models. Instead, Steele et al. (2019) proposed an alternative marginal modelling approach to estimate household effects in panel data where the mean of each outcome is modelled jointly with the pairwise correlations between the wave-specific outcomes of different individuals (or of the same individual at different waves). The motivation for their approach comes from noting that household membership arises from a social network such that $H = H(\mathcal{E})$, where $\mathcal{E}$ is an $nT \times nT$ matrix of network links, so that the objective is to model $p(\mathbf{y}|X, \mathcal{E})$ rather than $p(\mathbf{y}|X, H)$ as above. Specifically, the marginal model comprises models for the marginal mean $E(\mathbf{y}|X)$ and covariance $\text{cov}(\mathbf{r}) = \text{cov}(\mathbf{r}|\text{X}, \mathcal{E})$.

Because variance-covariance $\text{cov}(\mathbf{r})$ depends on $\mathcal{E}$, it does not generally lead to simple structures like (3) and (4) but to block-diagonal structures $\text{cov}(\mathbf{r}) = \text{diag}(\Omega_1, \ldots, \Omega_{\text{K}})$, where Ω_s is the variance-covariance matrix for the observations in 'superhousehold' $s \in \{1, \ldots, K\}$ and the dimensions of the Ω matrices generally differ. A superhousehold is a cluster constructed to contain individuals who are connected through coresidence, either directly or indirectly through a mutual coresident, in $\mathcal{E}$. Estimation is based on an extension of the generalized estimating equations (GEE) approach of Liang & Zeger (1986) to the simultaneous estimation of models for the marginal mean and for $\text{cov}(\mathbf{r})$ commonly referred to as second-order GEE (GEE2) (Yan & Fine, 2004). The model for the matrix of covariances comprises one (possibly covariate-dependent) model for the scale parameter and another allowing the correlation between each pair of person-wave observations to depend on covariates characterising the relationship between them; for example, indicators which distinguish between observations on the same person at different times and observations on (past, present or future) co-residents.

3 Grouped random effects model with correlated household effects

While the marginal model provides substantively interesting information about the within and between individual association structure, researchers are often interested in estimating the relative contributions of omitted individual and household characteristics. Moreover, its flexibility does not extend to allowing for area effects, or other forms of higher-level clustering, unless superhouseholds are nested within areas to form larger non-overlapping clusters; this is not the case for geographical clustering where an individual's area of residence may be time-varying and thus observations in the same superhousehold may come from multiple areas. A further limitation is that GEE2 estimation may lead to implied correlation matrices for a superhousehold that are not positive definite.

In contrast, random effects models are widely used for the analysis of longitudinal data because, by partitioning variation into different components, the relative contributions of time-varying and time-invariant omitted variables can be estimated and these models are readily extended to accommodate higher-level effects. In health research, for example, there is substantial interest in area effects (e.g. Griffith & Jones, 2020) and whether apparent geographical clustering of health outcomes may be

explained by omitted household effects (Chandola et al., 2005).

We therefore revisit the multilevel modelling framework and propose a random effects model (to be specified in detail below) that allows for changes in household composition over time but without the restrictive assumptions on the covariance structure of the multiple membership model; the new model also allows for time-varying area effects. To do this, we use superhouseholds to structure $\text{cov}(\mathbf{r}) = (\Omega_1, \ldots, \Omega_\text{K})$, but improve on previous random effects models by using correlated *household*-level random effects to more accurately capture the structure of Ω_s than any corresponding submatrix of the form $\sigma_v^2 HH'$ in (3) or $\sigma_v^2 QQ'$ in (4).

3.1 Definition of households and superhouseholds

Households

Household panel datasets routinely include a cross-sectional household identifier which can be used to determine coresident sample members at a given wave t. Several researchers have discussed the problems with the concept of 'longitudinal households'. For instance, Murphy (1996) asks whether the departure of one partner from a couple-with-children household leads to the creation of two new households and, if so, how they should be linked to the original household. We avoid arbitrary rules about which changes define a new household and assign a new household identifier at wave t to individuals following *any* change in their household's membership since the previous observed wave, on the basis that the departure or arrival of any individual may result in a material change in unmeasured household influences on an individual outcome such as health. This household identifier is constant across any waves at which a household's membership is unchanged; these will usually be consecutive waves, but a household may revert to a previous composition after the arrival or departure of an individual, for example following the return of an adult child to the parental home. As the household random effect is defined by the household identifier, it varies with any change in household composition. As discussed below, we allow for links between households, such as in the example from Murphy (1996) above, through between-household correlations which may depend on covariates characterising their link.

Changes in household composition, and the corresponding household identifier, are illustrated in Figure 1A. At the first wave, we begin with household 1 containing a couple (B,C) and their adult daughter (A). By wave 2, A has left the parental home to form new household 2 with a friend D, and the household with (B,C) has been assigned a new identifier of 3. A has returned to the parental home by wave 3, leading their household identifier to revert to 1, and D has formed new household 4 with their partner E. From this example, it is clear that individuals and households are non-nested because an individual can belong to multiple households over time (e.g. A is a member of households 1 and 2) and each household can contain more than one individual.

Superhouseholds

The definition of a 'superhousehold', as originally proposed by Steele et al. (2019), is based on the concept of changes in individuals' coresidents over time as an "evolving network" (Murphy, 1996). The coresidence connections between individuals can be represented by pathways or edges in a network graph. Figure 1B illustrates the coresidence network at wave 3 for the five individuals in Figure 1A. Each pair of individuals is connected either directly or indirectly, with direct links between pairs who have lived together at any of the three waves (e.g. B and C) and indirect links between pairs who have not lived together but have a mutual coresident (e.g. D and C were never coresidents but lived with

A at different waves). The superhousehold evolves over time, starting with (A,B,C) at wave 1 and ending with all five individuals at wave 3. In the random effects model described below, household effects may be correlated within superhousehold clusters defined at the final wave T. If there is no pathway between a pair of individuals by T, they must be in different superhouseholds. Individuals and households (as defined above) are each strictly nested within superhouseholds.

3.2 Model specification

We propose a cross-classified multilevel model where person-wave observations are nested within individuals and within superhouseholds, but individuals and households are non-nested because an individual may belong to more than one household during the observation period and a household may contain more than one individual. Moving from a vector-wise to element-wise description, this model has the form

$$Y_{ti} = \mathbf{x}'_{ti}\boldsymbol{\beta} + w_{a(ti)} + v_{h(ti)} + u_i + e_{ti}, \tag{5}$$

where $w_{a(ti)} \sim N(0, \sigma^2_w)$ is an area effect associated with the area of residence of individual i at wave t, $u_i \sim N(0, \sigma^2_u)$ is an individual effect and $e_{ti} \sim N(0, \sigma^2_e)$ is a time-varying residual. The household effect is $v_{h(ti)}$, where $h(ti)$ denotes the household of individual i at wave t, and thus the effect of household changes whenever there is a change in an individual's coresidents. To allow for possible overlap in the membership of pairs of households, we allow for correlation among $v_{h(ti)}$. Note that none of these classifications are nested in areas because individuals and households can move between areas over time.

The between-household correlations are confined to households in the same superhousehold cluster because only households in the same cluster can have members in common. Suppose there are J households and K superhouseholds and denote by $s(i)$ and $s(h)$ the superhousehold of individual i and household h respectively. Let m_s be the number of households in superhousehold $s \in \{1, \ldots, K\}$ and $\mathbf{v}_s = (v_{s_1}, \ldots, v_{s_{m_s}})'$ be a vector of household effects for households in s. We assume $\mathbf{v}_s \sim N(\mathbf{0}, \boldsymbol{\Omega}_{v_s})$ where $\boldsymbol{\Omega}_{v_s} = \sigma^2_v \mathbf{R}_{v_s}$, $\text{var}(v_h) = \sigma^2_v$ for $h = s_1, \ldots, s_{m_s}$ and $\mathbf{R}_{v_s}$ is a correlation matrix, where $\text{cor}(v_h, v_{h'}) = 0$ when $s(h) \neq s(h')$.

Under model (5), the elements of the combined residual $\mathbf{r}$ defined in (1) are $r_{ti} = w_{a(ti)} + v_{h(ti)} + u_i + e_{ti}$, and the residual covariances are

$$\begin{aligned} \text{cov}(r_{ti}, r_{t'i'}) = {} & \text{I}\{i = i'\}\,\sigma^2_u + \text{I}\{h(ti) = h(t'i')\}\,\sigma^2_v \\ & + \text{I}\{h(ti) \neq h(t'i')\}\,\text{I}\{s(i) = s(i')\}\,\sigma^2_v \rho_{h(ti),h'(t'i')} \\ & + \text{I}\{a(ti) = a(t'i')\}\,\sigma^2_w \end{aligned} \tag{6}$$

where $\rho_{h(ti),h'(t'i')}$ is an element of *correlation* matrix $\mathbf{R}_{\mathbf{v_s}}$. Model (7), allowing the correlations to depend on covariates $\mathbf{z}_{h,h'}$ characterising the households of the person-wave pair, is specified further on. Thus, the implied within-individual autocovariance $\text{cov}(r_{ti}, r_{t'i})$ for $t \neq t'$ is σ^2_u, plus σ^2_v for an individual who was in the same household at t and t' or $\rho_{h(ti),h'(t'i')}$ if there has been a change in household, with a further addition of σ^2_w if the individual has remained in the same area. The between-individual within-household covariance $\text{cov}(r_{ti}, r_{t'i'})$ for $i \neq i'$ and $h(ti) = h(t'i')$ is σ^2_v, plus σ^2_w if the household is in the same area at t and t'. The between-individual within-superhousehold covariance between individuals i and i' who are in the same superhousehold, $s(i) = s(i')$, but not the same household is $\rho_{h(ti),h'(t'i')}$, plus σ^2_w if the individuals are in the same area at t and t' respectively. Finally, the within-area covariance for observations from individuals in different superhouseholds is σ^2_w.

An important difference between the specification of the household effect in model (5) and the multiple membership model is the way in which the influence of an individual's coresidents prior to t on their outcomes at t is handled. The multiple membership model assumes a direct effect of prior coresidents on Y_{ti} which operates through the weighted sum of the effects of current and previous households. In contrast, the proposed model allows for an association between the outcomes of current and past (and possibly future) coresidents through the random effect correlation matrix $\mathbf{R}_{v_s}$. The advantage of this new approach is that it avoids the use of arbitrary weighting schemes which, as noted earlier, can lead to a restrictive and counterintuitive implied covariance structure. Another consideration is the left-truncation of individuals' coresidence histories: unless an individual participated in the study when living in their first household (typically the parental home), their coresidents prior to joining the study are unknown. This is a particular issue for the multiple membership model because it implicitly assigns a zero weight to the effects of households prior to an individual's first observation.

3.3 Correlation matrix modelling and MCMC estimation

In the context of applied work, where a flexible correlation model with straightforwardly interpretable parameters is paramount, we choose to extend the approach developed by Zhang et al. (2024) because it allows the estimation of covariate-dependent correlations while ensuring the estimated correlation matrix is positive-definite. This is achieved without requiring, as do Zou et al. (2017, 2022), that covariates in the correlation model are similarities/distances between observations.

Zhang et al. (2024) do not use a link function to ensure in-range correlations, which is anyway insufficient to ensure the resulting correlation matrix is positive-definite. They instead follow Zou et al. (2017, 2022) by assuming that the correlation between each pair follows a linear model. The linear model yields 'local' estimates which minimise the Kullback-Leibler divergence between the likelihood maximised under the working model and the likelihood under the true model for $\text{cov}(\mathbf{r})$; it also forms the basis for a novel MCMC sampling algorithm that includes a Metropolis-Hastings step in which parameters of the correlation model are drawn only from a range of values commensurate with an implied covariance matrix that is positive-definite at (almost) every point of support of the covariates. However, Zhang et al. (2024) focus on the special case of a four-by-four correlation matrix for continuous latent variables in a model for cross-sectional multivariate binary items that is not suitable here. We hence extend their approach to a more general class of non-hierarchical multilevel models, where the correlation matrices for the household random effects within superhouseholds are of unequal dimensions, as follows.

Define $\rho_{h,h'}(\boldsymbol{z}_{h,h'}) = \text{cor}(v_h, v_{h'})$ as the correlation coefficient between households h and h', which depends on covariates $\boldsymbol{z}_{h,h'}$. A first-order Taylor expansion gives $\rho_{h,h'}(\boldsymbol{z}) \approx \gamma_0 + \boldsymbol{\gamma}_1'\boldsymbol{z}$ so that

$$\text{cor}(v_h, v_{h'}) = \boldsymbol{\gamma}'\mathbf{z}_{h,h'} \quad \text{for } h, h' = s_1, \ldots, s_{m_s};\ h \neq h';\ s(h) = s(h'), \tag{7}$$

where $\mathbf{z}_{h,h'}$ is a vector with leading element one, followed by components characterizing the relationship between households h and h' (e.g., the number of individuals (if any) in both h and h'), and $\boldsymbol{\gamma}$ is a vector of coefficients.

The approximation in (7) is exact when the model is saturated (for example, if all predictors are nominal and all possible interactions are included). A notable special case of (7) arises when $\boldsymbol{\gamma}'\mathbf{z}_{h,h'} = \gamma$, implying equal correlation between household effect pairs within a superhousehold. This simplification is equivalent to decomposing $v_{h(ti)}$ in (5) into the sum of an independent household-level effect and a superhousehold-level random effect.

This local linear modelling approach has several key advantages. First, it provides a simple and interpretable framework for quantifying covariate effects on household correlations. Second, it allows for flexible specification of correlation structures. Third, it enables efficient estimation while maintaining positive definiteness constraints on correlation matrices. Further discussion of limitations and potential extensions of this approach is provided in Section 7.

We now present an equivalent reformulation of the proposed linear correlation structure. This formulation involves ordering pairs of households h, h' within superhousehold s from 1 to $m_s(1-m_s)/2$, recalling that m_s is the number of households within superhousehold s, where each pair is characterised by one of q covariates. Then consider the following proposition:

Proposition 1. *Let $Z_s = (\boldsymbol{z}_{s1}, \ldots, \boldsymbol{z}_{sq})$ denote a $m_s(m_s-1)/2 \times q$ matrix containing covariates for all pairs of random effects in $\boldsymbol{v}_s$, where each row corresponds to a pair (h, h') with $h \neq h'$ and $s(h) = s(h') = s$. The correlation model in (7) can be equivalently expressed as*

$$\mathbf{R}_{v_s} = \mathbf{R}_{v_s}(\boldsymbol{\gamma}, Z_s) = \sum_{l=1}^{q} \gamma_l \mathbf{G}_{sl} + (1 - \sum_{l=1}^{q} \gamma_l) I, \tag{8}$$

where $\mathbf{G}_{sl}$ are symmetric matrices with unit diagonal elements, and $\boldsymbol{z}_{sl}$ contains the lower triangular elements of $\mathbf{G}_{sl}$ arranged as a vector, for $l = 1, \ldots, q$.

Remark 1. *Our formulation extends the local linear correlation model of Zhang et al. (2024) from a single correlation matrix to multiple matrices. While the structure in Proposition 1 shares similarities with frameworks by Anderson (1973) and Zou et al. (2017, 2022), those approaches focus on modelling covariance matrices with similarity matrices $\mathbf{G}_{sl}$. In contrast, our goal is to directly model correlation matrices, which requires different methodological considerations. We address this by developing an efficient Bayesian estimation procedure that ensures the positive definiteness of the correlation matrices.*

A key consideration for MCMC estimation is maintaining positive definiteness of the correlation matrices. Specifically, when sampling $\boldsymbol{\gamma}$ from its posterior distribution, we must ensure that $\mathbf{R}_{v_s}(\boldsymbol{\gamma}, Z_s)$ remains positive definite for all $s = 1, \ldots, K$. The following proposition characterizes the feasible set of $\boldsymbol{\gamma}$ that satisfies these constraints.

Proposition 2. *Let $C_{\gamma,Z}$ denote the set of all $\boldsymbol{\gamma}$ that ensure positive definiteness of the correlation matrices $\mathbf{R}_{v_s}(\boldsymbol{\gamma}, Z_s)$ for $s = 1, \ldots, K$. Then*

$$C_{\gamma,Z} = \left\{ \boldsymbol{\gamma} \in \mathbb{R}^q \mid \sum_{l=1}^{q} \gamma_l \mathbf{G}_{sl} + (1 - \sum_{l=1}^{q} \gamma_l) I \succ 0, \text{ for } s = 1, \ldots, K \right\} \tag{9}$$

is a convex set with $\mathbf{0} \in C_{\gamma,Z}$, where $R \succ 0$ denotes that matrix R is positive definite.

To ensure positive definiteness of all correlation matrices $\mathbf{R}_{v_s}(\boldsymbol{\gamma}, Z_s)$, we specify a constrained prior $p(\boldsymbol{\gamma}) \propto \mathrm{I}\{\boldsymbol{\gamma} \in C_{\gamma,Z}\}$, where $C_{\gamma,Z}$ is defined in Proposition 2. Posterior samples of $\boldsymbol{\gamma}$ are drawn using a random walk Metropolis-Hastings algorithm detailed in Algorithm S1 of the supplementary material. An important property of this specification is that any convex combination of samples from the posterior distribution, such as the posterior mean, necessarily lies within the feasible set $C_{\gamma,Z}$.

Verifying that $\boldsymbol{\gamma}$ is within the set $C_{\gamma,Z}$ requires checking the positive definiteness of K symmetric matrices. While computationally feasible for moderate-sized problems, this verification becomes challenging as the number of superhouseholds K and group size increase. We propose an efficient alternative that replaces the computationally intensive positive definiteness checks (typically implemented via

Cholesky decomposition) with a convex hull verification formulated as a linear programming problem. This procedure incrementally expands the vertices set of $C_{\gamma,Z}$. As the MCMC sampling progresses, the set stabilizes, substantially reducing the frequency of matrix positive definiteness checks.

Full technical details are provided in the supplementary materials: the MCMC algorithm is described in detail in Section S1; proofs of Propositions 1 and 2 are given in Sections S2 and S3, respectively; the procedure for checking semi-positive definite correlation matrices is described in Section S4; and the conditions under which the resulting posterior distribution is concentrated around the true parameter and uniformly converges in distribution to normality are given in Section S5. Our procedure has been implemented in an R package (`mvregrp`) available from `https://github.com/slzhang-fd/mvregrp` along with documentation and simulated data.

4 The UK Household Longitudinal Study

We next describe the UK Household Longitudinal Study (UKHLS), also known as *Understanding Society* (University of Essex, ISER, 2020). In the simulation study presented in Section 5, the data are generated from the same household structure distribution as the UKHLS. The UKHLS data are then analysed in Section 6.

The UKHLS main survey began in 2009 with a sample of approximately 40,000 households. All household members aged 16 and over are invited to complete the adult survey. The survey fieldwork period at each wave is 24 months, but individual sample members are interviewed at approximately 12-month intervals. Interviews are conducted face-to-face in respondents' homes or through a self-completion online survey. We use annual data from waves 1–11 for the period 2009–2020. This sample comprises individuals from three component samples: the General Population Sample; the British Household Panel Survey Sample; and the Ethnic Minority Boost Sample (Lynn, 2009). All three samples were drawn using complex sampling designs involving stratification, clustering and unequal selection probabilities (e.g. the over-sampling of certain ethnic minority groups). The primary sampling units (PSUs) are postal area sectors and secondary sampling units (SSUs) are postal addresses from which all resident individuals are selected (the rare exception being when there are more than three households at an address).

Valid inference generally requires estimators which incorporate survey weights and sample design information on stratification and clustering (all available as UKHLS variables). We avoid the need to weight our analysis by using a 'disaggregated' approach (Chambers, 2003) in which household and area clustering are captured in our models by random effects and ethnic minority membership is included as a covariate along with other predictors correlated with the survey weights. The SSUs effectively correspond to the household level in our model; these dominate the effects of the clustering structure through having the highest intra-cluster correlation. The PSUs correspond to area where, in our models, we include random effects at the Lower Super Output Area (LSOA, defined in the next section) rather than postal area, because LSOAs have higher intra-cluster correlations than postal areas. Capturing the effects of complex survey designs in random effects models is generally a complex issue requiring careful assessment by the analyst that is discussed again in Section 7.

Longitudinally, all members of the wave 1 households constitute the core sample (original sample members, OSMs) who are followed wherever they move within the UK. People who form households with OSMs after wave 1 are referred to as temporary sample members (TSMs), unless they have children with OSMs in which case they become permanent sample members (PSMs); children of OSMs also become PSMs after turning 16. Like OSMs, PSMs are then followed regardless of whether they remain coresident with an OSM. These follow-up rules allow tracking of changes in the coresidents

of OSMs and PSMs over time.

4.1 Superhousehold, household and area identifiers

The household and superhousehold identifiers described in Section 3.1 are both constructed using the cross-wave individual and cross-sectional household identifiers available in UKHLS. Both identifiers use information on *all* members of participating households at a given wave, including children and non-responding adult sample members. A complete enumeration of a household, including basic information such as age and sex of household members and their relationships to one another, is available from the household grid questionnaire, administered before the individual interviews. Further details on the construction of the household and superhousehold identifiers is given in the supplementary materials (Section S6).

We define an area as a Lower Super Output Area (LSOA) based on the 2011 census. LSOAs contain between 400 and 1200 households and an average of 1500 individuals. Larger area classifications were also considered (Primary Sampling Units, Medium Super Output Areas and Local Authority Districts), but in exploratory analysis the proportion of variance explained by area decreased as area size increased, a finding that is consistent with previous research using administrative boundaries (Boyle & Willms, 1999).

4.2 Response variables and covariates for the mean model

The response variables for our analysis are based on the SF-12 instrument which is widely used to measure quality of life. The SF-12 was administered at each wave in UKHLS and comprises seven items about the self-reported impact of an individual's health on their everyday physical functioning and five items about their mental functioning, coded to produce two continuous scales referred to as the Physical and Mental Component Summaries (PCS and MCS). Details of the scoring methods are given in Ware et al. (2002). The PCS and MCS scores have a theoretical range of 0 (low functioning) to 100 (high functioning). In our analysis sample, the mean and standard deviation are 49.4 and 11.2 for the PCS scores and 49.2 and 10.2 for the MCS scores. As the distribution of both variables is strongly negatively skewed, we analyse the log reflected scores which are then standardised to have means of 0 and standard deviations of 1.

The analysis sample for the mean model (5) contains 387,238 person-wave observations from 76,053 adult individuals (aged 16 or older), 63,400 distinct households (as defined above) and 24,335 areas. The number of individuals per household in this sample ranges from 1 to 12 with a median of 2. Over the 11-year observation period, 32.1% of individuals experienced a change in coresidents and were therefore observed in more than one household. Individuals and households are nested in 37,867 superhousehold clusters. The number of individuals per superhousehold ranges from 1 to 13: 39.1% contain one individual, 38.7% have two individuals, 11.4% have three individuals, and the remaining 10.8% have three or more individuals. The number of households per superhousehold ranges from 1 to 8: 65.6% contain only one household, 17.8% have two households, and the remaining 16.6% have three or more households. The median number of person-waves observations, individuals and households per area is 11, 3 and 2 respectively, and 18.2% of individuals were observed in more than one area over time. Single-person households are retained in the analysis; they contribute to estimation of the coefficients of the mean model and, if observed at more than one wave, the sum of the variances of the individual and household random effects. Separation of the individual and household variances is based on the subset of individuals with repeated measurements and coresident respondents for at least one wave.

Informed by previous studies of health in the UK (e.g. Chandola et al., 2003, 2005; Griffith & Jones, 2020), the model for the mean transformed PCS and MCS scores includes the following covariates: age in years (centred at 50), sex, ethnicity (Asian or Asian British, Black or Black British, White, other), lives with a partner, number of children (including biological, adopted and step) and age of the youngest child ($<$ 2, 2-4, 5-10, 11-16, $>$16 years), highest educational qualifications (below university level, university degree), employment status (employed, unemployed, economically inactive), household income (equivalised for household size, adjusted for inflation using the annual Consumer Price Index for the year of interview, and log transformed), and housing tenure (own outright or with mortgage, social housing, private rental). All except income and housing tenure are measured at the individual level, and all except sex and ethnicity are time-varying. The analysis sample size given above is after excluding 12,200 (3%) observations with missing data on at least one covariate (in most cases, one or more of ethnicity, education, household income and housing tenure). Descriptive statistics for the covariates are given in Table S2 in the supplementary materials (Section S8).

4.3 Covariates for the between-household correlation model

The analysis sample for the correlation model (7) contains 49,167 pairs of households from the 13,045 superhouseholds with more than one household. Superhouseholds with only one household, comprising individuals who remain with the same coresidents (or who live alone) throughout the observation period, contribute only to estimation of the mean model. The covariates $\mathbf{z}_{h,h'}$ in (7) describe the connection between the members of households h and h'. We expect that correlations between household effects on health would be highest for households that share closely related individuals. We therefore focus on indicators of whether households share partners (past, current or future) and whether a member of one household is the parent of a member of the other. Such partnership and parent-child links account for 91% of all pairs, with the remainder contributed mainly by unrelated sharers. In addition we include the proportion of the total number of individuals across the two households who are in both households. As for the households and superhousehold identifiers, these variables are derived from the household grid questionnaire which provides information on the age and between-person relationships of all individuals in a household, not only eligible respondents. Table 1 shows the distribution of the covariates considered for pairs of households (h_1, h_2) formed at waves (t_1, t_2) where $t_1 \leqslant t_2$. A and B denote individuals who were coresident partners during the observation period, while P and C denote a parent and child. While these individuals are highlighted in Table 1, h_1 and h_2 typically include additional individuals. After preliminary analysis, some indicators for parent-child links where the composition of h_1 and h_2 are reversed have been combined to avoid small sample sizes; for example we do not distinguish between the formation of a new household after a child leaves the parental home ([P,C,...] [C,...] C16+) and after a child reenters the parental home ([C,...] [P,C,...] C16+).

5 Simulation study

5.1 Design

We carry out a simulation study to assess the impact of sample size and the extent of household clustering on the performance of our proposed approach, and investigate sensitivity to household structure misspecification and non-normality using two alternative models with omitted or misspecified household effects. We begin by setting the data generating model (DGM) to be a special case of the model given by (5) and (7), where area effects are excluded from the mean model (5) and the correlation model of (7) is simplified to $\text{cor}(v_h, v_{h'}) = \gamma$ for all pairs of households h and h' in the

Table 1: Descriptive statistics for the covariates in the correlation model (n=49,167 household pairs). $[h_1]$ $[h_2]$ denotes the members of households h_1 and h_2 formed at waves t_1 and t_2, $t_1 \leqslant t_2$. Individuals A and B were observed as coresident partners and P and C as coresident parent and child during the observation period. h_1 and h_2 may have both partnership and parent-child links and may share other individuals.

$[h_1]$ $[h_2]$	Description	n	Percent
Partnership links[a]			
[A,B,...] [A,B,...]	Same couple	15,371	31.3
[A,B,...] [A,...]	Couple in h_1, 1 partner after split in h_2	6857	13.9
[A,...] [B,...]	Ex-partner hhs	1146	2.3
[A,...] [A,B,...]	Future partner in h_1, couple in h_2	7584	15.4
Other	No partnership link	18,209	37.0
Parent-child links[b]			
[P,C,...] [C,...] C⩾16 *or* [C,...] [P,C,...] C⩾16	Child hh, parent-child hh (child ⩾ 16)	7787	15.8
[P,C,...] [P,...] C⩾16 *or* [P,...] [P,C,...] C⩾16	Parent hh, parent-child hh (child ⩾ 16)	11,583	23.6
[P,...] [P,C,...] C<16	Parent hh, parent-child hh (child < 16)	8347	17.0
[P,...] [C,...] C⩾16 *or* [C,...] [P,...] C⩾16	Separate parent and child hhs (child ⩾ 16)	6408	13.0
Other	No parent-child link	16,422	33.4
Proportion in both h_1 and h_2[c]	Median = 0.50, IQ range = (0.20, 0.67)		

[a]Categories are mutually exclusive; [b]Mainly mutually exclusive with > 1 type of parent-child link in 2.8% of household pairs; [c]Proportion of individuals in intersection of h_1 and h_2, calculated as $|h_1 \cap h_2|/|h_1 \cup h_2|$.

same superhousehold. We then generate data from a model with a heterogeneous between-household correlation structure as in (7).

The correlation matrices are generated from the UKHLS data in two stages. First, superhouseholds are sampled with replacement from the set of 37,867 superhouseholds in the full UKHLS analysis file (described in Section 4). Second, realisations of the response variable are then generated from one of the two DGMs described above, treating the time-varying household structure of the selected superhouseholds as fixed and using the observed values of the covariates. The DGM for the mean includes a constant (x_0), two time-varying individual variables (age centred at 50 years and divided by 10 and a binary indicator for education, x_1 and x_2), and two time-varying household-level variables (indicators for social and private rented housing tenure, x_3 and x_4), with corresponding coefficients $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$. For the heterogeneous between-household correlation case, the DGM for the household random effect correlations includes a constant (z_0), three indicators for the type of connection between households h and h' (parent-child connection but no partnership connection, partnership connection but no parent-child connection, and neither a partnership or parent-child connection, z_1, z_2 and z_3, taking the presence of both types of connection as the reference category) and a continuous variable for the proportion of individuals in the pair who are common to both (centred at 0.5, z_4), with coefficients $(\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4)$.

Data with an equal between-household correlation structure within 20k superhouseholds are drawn as per above. Estimates from the following three models can then be compared: (i) a standard panel model with only individual effects (M1), (ii) an extension of M1 with independent household effects (M2), and (iii) a model with individual effects and correlated household effects (M3) - the correctly specified model in this case. The comparison of models M1 and M3, using superhouseholds drawn from UKHLS, allows us to assess the impact of ignoring household effects when there are realistically complex within and between-household correlation structures and changes in household composition over time. The results are displayed in Table 2 and discussed below.

We then generate data from a model with a heterogeneous between-household correlation structure as in (7) and fit the correlated household effects model (M3). For this design, we consider superhousehold sizes of 20k, 30k and 40k. These are the sample sizes for estimation of the mean model parameters. The corresponding analysis samples for estimation of the correlation model parameters are considerably smaller, containing approximately 6890, 10,335 and 13,780 superhouseholds, and 26k, 39k and 52k household pair observations (based on the proportion of superhouseholds with more than one household and the number of pairwise observations per superhousehold in the full UKHLS dataset described in Section 4.2). For both designs, we consider two values for the household variance partitioning coefficient (VPC_{hh}), calculated as $\sigma_v^2/(\sigma_v^2 + \sigma_u^2 + \sigma_e^2)$: 0.1 (low) and 0.3 (high). Variation in VPC_{hh} is achieved by varying the relative contributions of the individual and household effects, for fixed σ_e^2 and $\sigma_v^2 + \sigma_u^2 + \sigma_e^2 = 1$ for both settings. The results for each sample size and value of VPC_{hh} are based on 200 replicate datasets. The results are displayed in Table 3 and discussed below.

Finally, for the simulation conditions closest to our application (40k superhouseholds and VPC_{hh} = 0.3), we additionally consider the impact of departures from normality for the household random effects. We focus on the household level to assess whether non-normality leads to bias or imprecision in estimates of the household-level variance and coefficients in the model for the between-household correlations. Three non-normal distributions are considered: a t-distribution on 5 degrees of freedom (with substantial kurtosis of 6) and positively and negatively skewed-normal distributions (with a moderate skewness of ± 0.78). Details of data generation using copulas are given in the supplementary materials (Section S7).

5.2 Results

Table 2 shows the results from fitting the first DGM that assumes equal correlation between the household random effects. (Note that M1 and M2 were fitted using the `lme4` package (Bates et al., 2015) which does not give standard error estimates for variance parameters.) For all models the estimates of the parameters of the mean model $(\beta_0, \ldots, \beta_4)$ have negligible bias, although the standard errors of the coefficients for the two household variables (β_3, β_4) are underestimated when the household effects are ignored (M1), and the bias increases with VPC_{hh}.

The between-individual variance σ_u^2 from M1 is overestimated with relative biases of 17.4% and 87.1% for low and high VPC_{hh} respectively (calculated as $100 \times (\bar{\hat{\theta}} - \theta_{\text{true}})/\theta_{\text{true}}$ for a parameter θ). When household effects are omitted (as in M1), the household variance is split between the residual and individual variances; it is not reallocated wholly to the individual level because individuals are not nested within households. When household effects are included but assumed independent (M2), the individual variance is again overestimated but to a much smaller degree at 7.8% and 20.3% for low and high VPC_{hh}, with downward biases of 43.5% and 30.2% in the household variance. M3, which has the same structure as the DGM, produces unbiased estimates of all parameters with similar mean and empirical standard errors and coverage. For most parameters of M3, the estimated interval coverage lies in the 95% range $(0.92, 0.98)$ expected for intervals with nominal 0.95 coverage based on 200 replications. There is slight undercoverage for the individual and household random effect variances (σ_u^2, σ_v^2) and the between-household correlation (γ) when VPC_{hh}=0.1 and $\sigma_v^2 = 0.1$ is close to the zero boundary. For this setting, σ_v^2, its separation from σ_u^2, and the correlation among the household random effects are weakly identified. This leads to asymmetric posterior distributions bounded by zero, which is reflected in the mean posterior SE slightly underestimating the true sampling variability (empirical SD). When σ_v^2 is larger (VPC_{hh}=0.3), the mean SEs align more closely with the empirical SDs, and coverage rates generally improve; increasing from 20k to 40k superhouseholds also leads to

an improvement (see Table 3).

We next assess the performance of M3 when the between-household correlations depend on covariates. Table 3 shows the results for the variance and the between-household correlation parameters. We focus on M3 and the variance and correlation parameters because the estimates of $(\beta_0, \ldots, \beta_4)$ and their standard errors are unbiased and the results for M1 and M2 show a similar pattern to the equal correlation case. Starting with the results for 20k superhouseholds and VPC_{hh}=0.1, the estimator bias of the variance parameters remains negligible, but there is non-negligible bias for one of the correlation parameters (36.0% for γ_3). This bias is reduced to 6.3% respectively when VPC_{hh} is increased to 0.3. The increase in VPC also leads to an improvement in the estimated coverage of the credible intervals (closer to the nominal 0.95) even based on 200 replicates. Increasing the sample size to 40k superhouseholds also leads to bias reduction even when VPC_{hh} is low. (The results for 30k superhouseholds are not shown, but are consistent with a pattern of a decrease in the bias with increasing sample size.)

Table 2: Simulation results for equal between-household correlation design with $r = 200$ replicates of $M = 20,000$ superhouseholds* selected with replacement from the UKHLS data

	β_0	β_1	β_2	β_3	β_4	σ_e^2	σ_u^2	σ_v^2	γ
VPC_{hh}=0.1									
True	0.1	0.25	−0.2	0.3	0.1	0.4	0.5	0.1	0.5
M1: Individual effects only									
Mean	0.1005	0.2501	−0.2004	0.2997	0.0999	0.4131	0.5869	–	–
Mean SE	0.0056	0.0021	0.0077	0.0091	0.0079	–	–	–	–
SD	0.0059	0.0019	0.0079	0.0105	0.0087	0.0014	0.0050	–	–
95% coverage	0.940	0.975	0.950	0.910	0.915	–	–	–	–
M2: Individual and independent household effects									
Mean	0.1004	0.2501	−0.2003	0.2997	0.1001	0.3994	0.5390	0.0565	–
Mean SE	0.0058	0.0021	0.0077	0.0094	0.0084	–	–	–	–
SD	0.0058	0.0019	0.0078	0.0105	0.0087	0.0014	0.0053	0.0022	–
95% coverage	0.950	0.975	0.960	0.915	0.945	–	–	–	–
M3: Individual and correlated household effects									
Mean	0.1004	0.2500	−0.2003	0.2997	0.1001	0.4001	0.5002	0.0997	0.4970
Mean SE	0.0059	0.0021	0.0078	0.0096	0.0084	0.0014	0.0058	0.0044	0.0266
SD	0.0058	0.0019	0.0078	0.0105	0.0087	0.0014	0.0063	0.0046	0.0294
95% coverage	0.950	0.970	0.950	0.910	0.945	0.945	0.900	0.915	0.905
VPC_{hh}=0.3									
True	0.1	0.25	−0.2	0.3	0.1	0.4	0.3	0.3	0.5
M1: Individual effects only									
Mean	0.1007	0.2500	−0.2008	0.2993	0.1002	0.4395	0.5612	–	–
Mean SE	0.0055	0.0021	0.0077	0.0091	0.0081	–	–	–	–
SD	0.0067	0.0021	0.0083	0.0114	0.0096	0.0017	0.0054	–	–
95% coverage	0.890	0.950	0.920	0.880	0.905	–	–	–	–
M2: Individual and independent household effects									
Mean	0.1005	0.2500	−0.2006	0.2995	0.1006	0.3983	0.3610	0.2093	–
Mean SE	0.0059	0.0021	0.0075	0.0098	0.0089	–	–	–	–
SD	0.0065	0.0018	0.0077	0.0112	0.0094	0.0013	0.0049	0.0041	–
95% coverage	0.910	0.975	0.945	0.915	0.925	–	–	–	–
M3: Individual and correlated household effects									
Mean	0.1005	0.2500	−0.2006	0.2997	0.1006	0.4001	0.2998	0.3004	0.4996
Mean SE	0.0063	0.0021	0.0075	0.0103	0.0091	0.0014	0.0041	0.0052	0.0123
SD	0.0064	0.0018	0.0077	0.0111	0.0093	0.0013	0.0044	0.0053	0.0132
95% coverage	0.930	0.980	0.920	0.935	0.915	0.940	0.915	0.935	0.940

*Estimates of γ are based on approximately 6890 superhouseholds containing more than 1 household.

The results for the non-normal household random effects are given in the supplementary materials (Section S7). The parameter estimates, and most standard errors, remain unbiased for the t and skewed normal distributions considered. The exceptions are downward bias in the standard error of the intercept in the correlation model (γ_0) and, for the t-distribution only, the household variance.

Table 3: Simulation results for heterogeneous between-household correlation design with $r = 200$ replicates of $M = 20,000$ and $M = 40,000$ superhouseholds* selected with replacement from the UKHLS data. Results are given only for the variance and between-household correlation parameters of Model M3.

	σ_e^2	σ_u^2	σ_v^2	γ_0	γ_1	γ_2	γ_3	γ_4
$M = 20,000$ **with VPC$_{\text{hh}}$=0.1**								
True	0.4	0.5	0.1	0.6	−0.1	−0.1	0.1	0.4
Mean	0.4001	0.5007	0.0990	0.5928	−0.0977	−0.0957	0.0640	0.4009
Mean SE	0.0014	0.0058	0.0046	0.0270	0.0367	0.0419	0.0631	0.0730
SD	0.0014	0.0061	0.0044	0.0294	0.0367	0.0402	0.0575	0.0764
95% coverage	0.940	0.915	0.945	0.920	0.955	0.955	0.945	0.940
$M = 20,000$ **with VPC$_{\text{hh}}$=0.3**								
True	0.4	0.3	0.3	0.6	−0.1	−0.1	0.1	0.4
Mean	0.4001	0.3001	0.2998	0.5988	−0.0991	−0.0986	0.0937	0.3988
Mean SE	0.0014	0.0041	0.0053	0.0133	0.0199	0.0228	0.0335	0.0385
SD	0.0013	0.0043	0.0053	0.0143	0.0208	0.0223	0.0342	0.0385
95% coverage	0.945	0.925	0.965	0.930	0.945	0.955	0.950	0.955
$M = 40,000$ **with VPC$_{\text{hh}}$=0.1**								
True	0.4	0.5	0.1	0.6	−0.1	−0.1	0.1	0.4
Mean	0.3999	0.5001	0.1000	0.5968	−0.0965	−0.0901	0.0821	0.4021
Mean SE	0.0010	0.0035	0.0018	0.0226	0.0347	0.0368	0.0479	0.0577
SD	0.0010	0.0037	0.0020	0.0239	0.0337	0.0410	0.0485	0.0633
95% coverage	0.950	0.955	0.950	0.930	0.950	0.910	0.920	0.940
$M = 40,000$ **with VPC$_{\text{hh}}$=0.3**								
True	0.4	0.3	0.3	0.6	−0.1	−0.1	0.1	0.4
Mean	0.3999	0.3001	0.3002	0.5995	−0.0993	−0.0979	0.0984	0.4026
Mean SE	0.0010	0.0027	0.0032	0.0116	0.0175	0.0195	0.0255	0.0306
SD	0.0010	0.0029	0.0034	0.0121	0.0179	0.0196	0.0270	0.0317
95% coverage	0.950	0.940	0.940	0.925	0.945	0.945	0.940	0.960

*Estimates of the γ parameters are based on approximately 6890 ($M = 20k$) and 13,780 ($M = 40k$) superhouseholds containing more than 1 household.

Our findings are in line with previous research on the consequences of departures from normality for two-level hierarchical models for continuous responses. Maas & Hox (2004) conducted a simulation study with chi-squared, uniform and Laplace distributions for a range of level 1 and 2 sample sizes and VPCs. For all distributions and settings, they found little impact of non-normality on the parameter estimates and the standard errors of the regression coefficients; however, there was some inaccuracy in the standard errors of the random effect variance for distributions with heavier or lighter tails than the normal.

6 Application: Analysis of physical and mental health in the UK

6.1 Preliminary analysis and model checking

The simulation results indicate that the grouped random effects estimator performs well for data structures similar to the UKHLS with a large sample size (number of household pairs across groups) for the correlation model. We therefore proceed to analyse the UKHLS data, with separate models for the physical and mental health functioning outcomes. The variables described in Section 4.2 were included in the models for mean health given by (5). As the effect of each variable was significant at the 5% level for at least one outcome, all were retained as predictors in both models for ease of comparison. For computational reasons, preliminary models fitted to select the covariates for the mean model included only an intercept term in the between-household correlation model of (7). After selection of the mean model, we then focused on selection of covariates for the between-household correlation model. The final specification of the correlation model includes the variables given in Table 1, but specifications with finer categorisations of the different kinds of connection between a

household pair h_1 and h_2 (in the same superhousehold) were also considered. For example, initial models distinguished cases where h_2 is formed from a parent-child household h_1 after an adult child leaves home ($h_1 = [\mathrm{P}, \mathrm{C}, \ldots]$, $h_2 = [\mathrm{C}, \ldots]$) and the reverse scenario where an adult child returns to the parental home (with h_1 and h_2 switched). Although such pairs differ in the temporal ordering of the parent-child and child households, they have the same proportion of individuals in common and their between-household correlations were not statistically different.

We assess whether the addition of household effects to standard random effects panel models improves model fit using a version of the deviance information criterion, DIC_L, proposed for the comparison of latent variable models (Li et al., 2020). We compare three model specifications of increasing complexity, all with individual and area effects: model A without household effects, B with equally-correlated household effects, and C with covariate-dependent correlations among household effects. For both health outcomes, the value of DIC_L decreases as we move from A to C, with the largest decrease found when household effects are introduced (B vs A). For physical health, DIC_L is -378089 for A, -380750 for B and -380991 for C, while the estimates for mental health are respectively -303530, -306585 and -306772. We therefore conclude that models with heterogeneous between-household correlations are preferred, and focus on the interpretation of these models below.

As for any statistical model, it is important to assess model assumptions, and residual diagnostics for single-level regression have been extended to multilevel models (e.g. Goldstein, 2010; Snijders & Berkhof, 2007). As noted in Section 4.2, the two health outcomes were transformed to reduce the amount of negative skew, but distributions of the empirical Bayes estimates of the random effects were examined to check for major departures from normality. As in the simulation study, we focus on the household random effects. For both outcomes, kurtosis (3.61 for physical and 3.21 for mental health) and skewness (0.55 for physical and 0.42 for mental health) are within acceptable ranges and lower than the values used in our simulations. (See Figure S3 of the supplementary materials for histograms of the distributions.)

The results presented for the selected models are the posterior means and 95% credible intervals for four chains of 20,000 MCMC iterations, each using a different set of starting values, after discarding a burn-in sample of 5000. Convergence was assessed using a range of graphical diagnostics and the potential scale reduction factor (PSRF) (Gelman et al., 2004). Visual inspection of trace plots of each parameter for the multiple chains suggested adequate mixing. Final PSRF estimates were close to 1 for all parameters. Furthermore, increasing the chain length led to little change in the running means of the posterior estimates.

6.2 Relative contributions of individual, household and area effects

We first consider the relative magnitude of wave-specific (residual), individual, household and area effects, and the impact of excluding household effects. Figure 2 shows the variance partition coefficients (VPCs) from models before and after including (correlated) household effects for each outcome before adding covariates to the mean or between-household correlations models. The VPCs are calculated from the random effect variance estimates (shown with 95% credible intervals in Table S3 of the supplementary materials). We begin with an interpretation of the results from the full model with all four components of variation. Before including covariates, household effects account for 25.4% and 16.6% of the total variance in physical and mental health respectively. Coresidents, especially families, may share characteristics such as socieconomic circumstances and lifestyle and genetic factors which have stronger effects on physical than mental health. However, the individual and household covariates explain a higher proportion of the household variance in physical health than in mental

health (Figure S1 in the supplementary materials), resulting in conditional VPCs of 11.5% and 15.5% respectively. Consistent with previous research (Chandola et al., 2005; Sacker et al., 2006; Griffith & Jones, 2020), area effects are small for both outcomes.

To illustrate the importance of modelling the random part correctly, we next consider the impact of excluding household effects. For both outcomes, omitted household variation is picked up primarily by the individual effects. Area effects are also inflated when household effects are ignored, with the contribution of area increasing from 2.2% to 4.0% of the total variance for physical health and 2.7% to 5.0% for mental health. Similar patterns are observed after including covariates in the mean model (see Figure S1). As found in previous cross-sectional studies that consider both household and area effects on health outcomes (e.g. Sacker et al., 2006), we conclude that area effects can be partly explained by household effects.

Table 4: Coefficient estimates from models of mean standardised self-rated physical and mental health functioning with 95% credible intervals. Higher scores indicate lower functioning. Coefficients represent effects in standard deviation units.

	Physical health		Mental health	
	Est.	95% CI	Est.	95% CI
Constant	-0.117	(-0.152, -0.081)	-0.025	(-0.068,0.018)
Age[a]	0.200*	(0.0194, 0.206)	-0.141*	(-0.147,-0.134)
Age squared	0.016*	(0.014, 0.017)	-0.027*	(-0.028,-0.025)
Age cubed	0.004*	(0.004, 0.005)	0.012*	(0.012, 0.013)
Female (vs. Male)	0.052*	(0.042, 0.063)	0.187*	(0.176, 0.197)
Ethnicity (vs. White)				
Asian/ Asian British	0.232*	(0.212, 0.252)	-0.000	(-0.022, 0.023)
Black/ Black British	0.027*	(0.001, 0.054)	-0.165*	(-0.194, -0.136)
Other	0.083*	(0.051, 0.116)	0.080*	(0.045, 0.116)
Partnered (vs. Single)	-0.005	(-0.015,0.004)	-0.175*	(-0.186, -0.163)
Number of children (vs. 1)				
None[b]	0.053*	(0.037, 0.047)	0.109*	(0.090, 0.128)
2[c]	-0.044*	(-0.056, -0.032)	0.026*	(0.012, 0.040)
3+[c]	-0.044*	(-0.061, -0.026)	0.058*	(0.038, 0.078)
Age of youngest child (vs. < 2 years)[d]				
2–4 years	0.011	(-0.001, 0.023)	0.051*	(0.036, 0.066)
5–10 years	0.032*	(0.019, 0.045)	0.076*	(0.060, 0.092)
11–16 years	0.063*	(0.048, 0.058)	0.122*	(0.103, 0.140)
>16 years	0.092*	(0.075, 0.108)	0.160*	(0.139, 0.180)
Highest education degree level (vs. Lower)	-0.145*	(-0.155, -0.134)	-0.006	(-0.017, 0.006)
Employment status (vs. Employed)				
Unemployed	0.055*	(0.043, 0.067)	0.211*	(0.197, 0.226)
Economically inactive	0.172*	(0.164, 0.180)	0.119*	(0.110, 0.128)
Log annual household income (GBP)	-0.004*	(-0.007, -0.001)	-0.009*	(-0.013, -0.006)
Housing tenure (vs. Owner-occupier)				
Social rent	0.274*	(0.261, 0.287)	0.185*	(0.169, 0.200)
Private rent	0.106*	(0.095, 0.118)	0.077*	(0.064, 0.090)
Random effect variances				
Area σ_w^2	0.016	(0.014, 0.018)	0.021	(0.018, 0.024)
Household σ_v^2	0.088	(0.082, 0.095)	0.150	(0.143, 0.157)
Individual σ_u^2	0.355	(0.349, 0.362)	0.316	(0.309, 0.322)
Wave σ_e^2	0.311	(0.310, 0.313)	0.480	(0.477, 0.482)

*95% credible interval excludes zero; [a]Age in years centred at 50 and divided by 10; [b]Contrast of no children versus 1 child aged < 2 years; [c]Contrast of 2 or 3+ children vs 1 child where youngest is aged < 2 years; [d]Effects of age of youngest child among respondents with children.

6.3 Covariate effects on mean health outcomes

Parameter estimates for the models of mean physical and mental health functioning are shown in Table 4. Age effects are represented by a cubic polynomial and plotted in Figure S2 of the supplementary

materials. Physical health declines almost linearly up to the early 60s and then accelerates. There is a much weaker and more nonlinear association between mental health and age, with poorer health in the 30s followed by an almost linear improvement with age until a small decline from age 80. Turning to the other covariates, the following individual and household characteristics are associated with poorer physical health: female sex, Asian or Asian British ethnicity, no children or only older (age $\geqslant$ 11) children, less than degree-level education, out of employment, lower household income, and social or private rented housing. The effects on physical health are strongest for ethnicity, education, economic activity and housing tenure. The following characteristics are associated with poorer mental health: female sex, an ethnicity other than Black or Black British, no coresident partner, no or more than one older child (age $\geqslant$ 11), out of employment, lower household income, and social or private rented housing. In general, the effects are weaker for mental health, with the exception of sex, partnership status, unemployment and household income. There is little impact of omitting household effects on the coefficient estimates (results not shown). As expected, the credible intervals for the household characteristics (income and tenure) are narrower when household effects are ignored, but our conclusions are unaffected because the 95% credible intervals associated with these variables (Table 4) all exclude zero.

6.4 Covariate effects on between-household random effect correlations

Parameter estimates from the models of the between-household (within-superhousehold) random effect correlations for physical and mental health functioning are shown in Table 5. There is a strong positive effect of the proportion of individuals in both households on the correlation, especially for mental health. For example, an increase from 0.20 to 0.67 in the proportion of overlap (corresponding to the interquartile range, see Table 1) is associated with an increase in the between-household correlation of 0.11 for physical and 0.21 for mental health (adjusting for the effects of any partnership or parent-child link). For both health outcomes, the correlation is highest for household pairs that share the same couple. Other partnership links are associated with a reduction in the correlation relative to 'same couple' pairs, with more pronounced variation among different configurations of past, current and future partners for physical health. Among parent-child links, the correlation is lowest for pairs formed when a young child (usually a newborn) joins an adult household and, for mental health, highest for separate parent and child households. However, the effects of parent-child and partnership links should be considered jointly as they are strongly associated. For example, most household pairs with a parent-child connection also share the same couple (the parents).

To assess the joint effects of the covariates on the between-household correlation, the parameter estimates in Table 5 were used to calculate predicted correlations for each pair of households. The mean predictions were then computed for each possible combination of the indicators of partnership and parent-child links. Table 6 shows the mean predicted correlations for the nine most frequent combinations (out of a total of 35), which account for 85.5% of all household pairs. For each combination, there is variation in the predictions because the correlation for a given pair also depends on the proportion of overlap in their household members; the mean overlaps are also shown in Table 6. The predicted correlations provide insights into the demographic events leading to change in household membership that are associated with the largest changes in unmeasured household influences on health: a low predicted correlation between a pair of households in the same superhousehold suggests that the two households have largely distinct unmeasured characteristics. We find that all correlations are moderate to high, indicating that households within a superhousehold share characteristics that affect the health of individual household members.

Table 5: Coefficient estimates from models of between-household random effect correlations for standardised self-rated physical and mental health functioning. Correlations for pair of households h_1 and h_2 formed at waves t_1 and t_2, $t_1 \leqslant t_2$.

	Physical health		Mental health	
	Est.	95% CI	Est.	95% CI
Constant[a]	0.716*	(0.664, 0.763)	0.622*	(0.571, 0.665)
Partnership links (vs. Same couple)				
Couple in h_1, 1 partner after split in h_2	-0.345*	(-0.404, -0.284)	-0.280*	(-0.334,-0.218)
Ex-partner hhs	-0.522*	(-0.669, -0.347)	-0.176	(-0.350, 0.005)
Future partner in h_1, couple in h_2	-0.245*	(-0.314, -0.172)	-0.184*	(-0.253, -0.107)
No partnership link	-0.246*	(-0.311, -0.180)	-0.062	(-0.124, 0.006)
Parent-child links (vs. No link)[b]				
Child hh, parent-child hh (child $\geqslant$ 16)	-0.018	(-0.076, 0.036)	-0.051	(-0.115, 0.008)
Parent hh, parent-child hh (child $\geqslant$ 16)	-0.015	(-0.061, 0.031)	0.008	(-0.034, 0.051)
Parent hh, parent-child hh (child $<$ 16)	-0.126*	(-0.178, -0.075)	-0.109*	(-0.159, -0.058)
Separate parent and child hhs (child $\geqslant$ 16)	0.083	(-0.031, 0.195)	0.105*	(0.019, 0.188)
Proportion in both h_1 and h_2[c]	0.230*	(0.088, 0.362)	0.453*	(0.341, 0.565)

*95% credible interval excludes zero; [a]Represents correlation when $[h_1]$ and $[h_2]$ share the same couple with no parent-child link and proportion in both households is 0.5; [b]Categories are almost mutually exclusive ($>$ 1 type of parent-child link in 2.8% of pairs); [c]Centred at 0.5.

Table 6: Mean predicted between-household random effect correlations and proportion overlap by type of connection between households h_1 and h_2 formed at waves t_1 and t_2, $t_1 \leqslant t_2$. Results for 9 most frequent patterns on the indicators of partnership and parent-child links in decreasing order of prevalence.

Connection between h_1 and h_2	Cor(v_{h_1}, v_{h_2})[a]		Overlap[b]	%	Cum. %[c]
	Physical	Mental			
1. Same couple; parent and parent-child households (child $\geqslant$ 16), i.e. adult child leaves or enters parental household	0.737	0.700	0.653	14.2	14.2
2. No partnership link; separate parent and adult child households	0.439	0.439	0	12.6	26.8
3. Same couple; parent and parent-child households (child $<$ 16), i.e. couple household before and after birth of a child	0.633	0.598	0.688	11.8	38.6
4. Couple in h_1, 1 partner after split in h_2; no parent-child link	0.366	0.331	0.477	11.2	49.8
5. No partnership or parent-child link, e.g. adult sharers where $\geqslant$ 1 person moves in or out	0.413	0.447	0.249	8.9	58.7
6. Future partner in h_1, couple in h_2; no parent-child link	0.460	0.416	0.452	8.7	67.4
7. No partnership link; child and parent-child households (child $\geqslant$ 16)	0.407	0.418	0.300	6.8	74.2
8. No partnership link; parent and parent-child households (child $\geqslant$ 16), single parent	0.459	0.575	0.515	6.0	80.2
9. Future partner in h_1, couple in h_2; child and parent-child households (child $\geqslant$ 16), i.e. child leaves home to live with partner	0.387	0.257	0.212	5.3	85.5

[a]Mean predicted correlation between random effects for households h_1 and h_1; [b]Mean proportion of individuals in both h_1 and h_2; [c]The cumulative percentage of household pairs in this and all previous (larger) covariate pattern categories.

The most common type of household pair results from an adult child either leaving or entering their parental household (type 1 in Table 6). Pairs of this type have both a partnership link (as they share a couple) and a parent-child link, and on average 65% of their members are in both households. For both physical and mental health, the correlation is highest for these pairs, and higher than the correlation for pairs comprising a couple before and after the birth of a(nother) child (type 3). A comparison of the correlations for types 1 and 3 suggests that, conditional on the covariates in the mean model, a birth leads to a greater shift in the unmeasured household influences on the health of parents and any other adult residents than the departure or arrival of an adult child. Type 8 is similar to type 1, but for a single parent; this leads to a lower mean overlap between households which partly explains the lower correlation relative to the type 1 couple pairs. Pairs of type 7 and 9 are child and parent-child households where, in most cases, the child household is formed after the child leaves the parental household. For type 9 pairs, there is also a partnership link as the child moves in with a partner, while for type 7 the child does not have a coresident partner. For both outcomes, but especially for mental health, the correlation is lower for type 9 than for type 7, which may again be due to a lower mean overlap in household members. Even when the households have no one in common, as in the case of separate parent and child households (type 2), there is a moderate correlation between the household random effects, possibly due to shared lifestyle or genetic factors that persist after coresidence ends.

Pairs of type 4 and 6 share a partnership but not a parent-child link. Following a partnership breakdown (type 4), there is a moderate correlation between the household effects for the couple household and the household containing one of the ex-partners (and possibly a new partner or other coresidents). Type 6 pairs result from partnership formation, where the first household contains one of the future partners and the second contains the new couple. Again there is a moderate correlation between these households. The majority of the remaining household pairs (type 5) contain unrelated adults; the two households have neither a partnership nor parent-child link, but nevertheless have individuals in common and therefore a positive correlation.

Table 7: Conditional intra-class correlations (ICC) $\mathrm{Cor}(r_{ti}, r_{t'i'})$ implied by models for mean standardised self-rated physical and mental health functioning and between-household random effect correlations with covariates. In each case, we assume observations ti and $t'i'$ are from the same area, i.e. $a(ti) = a(t'i')$.

Type of ICC	Physical	Mental
ICC_{ind}: Same hh $[h(ti) = h(t'i')]$, same individual $[i = i']$, different waves $[t \neq t']$	0.596	0.504
ICC_{hh}: Same hh $[h(ti) = h(t'i')]$, different individuals $[i \neq i']$	0.136	0.177
ICC_{shh}: Same super-hh $[s(i) = s(i')]$, different hh $[h(ti) \neq h(t'i')]$	0.077 (0.018)[a]	0.096 (0.024)[a]
ICC_{area}: Same area $[a(ti) = a(t'i')]$, different super-hh $[s(i) \neq s(i')]$	0.021	0.022

[a] ICC_{shh} varies between household pairs due to dependence of between-household correlation on covariates; estimates shown are the mean and standard deviation.

Finally, Table 7 shows estimates of the conditional intra-class correlations (ICCs) for individual, household, superhousehold and area implied by the random effect variance estimates in Table 4 and the between-household correlation coefficients in Table 5. The corresponding estimates before accounting for covariates in the mean model are given in Table S4 of the supplementary materials. The ICCs are based on estimates of the covariances given by (6). For both outcomes, there is a relatively high within-person correlation (ICC_{ind}) over the observation period, especially for physical health. After accounting for covariates in the mean model, the within-household correlation (ICC_{hh}) is higher for mental than for physical health. There is also a weak correlation between household pairs in the same superhousehold (ICC_{shh}), ranging from 0.03 to 0.11 and from 0.02 to 0.15 for physical and mental health respectively, according to the type of connection between the pair. As noted earlier,

area effects are negligible (ICC_{area}). Comparing the conditional and unconditional ICCs, we find that the individual and household covariates in Table 4 explain a substantially higher proportion of the within-individual and within-household correlations for physical health than for mental health. The covariates explain only 3% of the total variance in mental health compared to 23% for physical health (as both outcomes are standardised, the overall R^2 is estimated as 1 minus the sum of the variance component estimates in Table 4).

7 Discussion

Household panel data provide the opportunity to disentangle individual and household effects on health outcomes, but changes in household membership over time present an analytical challenge. Most previous research on household effects has sidestepped the problem by carrying out cross-sectional analysis of data from a single wave, which is wasteful and severely limits the research questions that can be explored. This paper presents a flexible random effects model that uses data from all waves and allows estimation of the relative contributions of individual and household effects, together with effects due to higher levels of clustering such as geographical area. We incorporate a correlation model from which we can infer the types of demographic event leading to changes in household composition that are related to changes in the unmeasured household influences on health. We also propose a novel MCMC estimation procedure to estimate the parameters of this joint model while ensuring the covariance matrix is positive definite.

Our analysis of self-rated health illustrates the sorts of substantive insight that can result from using the model we propose. This analysis reveals strong household effects, especially for physical health functioning, although covariates explain a larger proportion of household variance for physical than for mental health. Chandola et al. (2005) suggest some of the mechanisms that could explain clustering of health outcomes within households, including shared nutritional and hygiene behaviours, physical characteristics of the building (such as the presence of damp) and causal effects of one person's health or economic circumstances on the health of others in the household. As in previous research across a range of health outcomes, we find far greater heterogeneity within areas than between areas. One possible explanation for the lack of area effects is the arbitrariness of census boundaries; however, like other authors, we find consistently small area effects across different area definitions. Another limitation of our analysis is that information on area of residence is available only for the study period, so we cannot allow for duration effects or contributions from areas lived in prior to our observation window. It is also possible that area effects are stronger for some population subgroups, for example elderly and other vulnerable groups who may be more susceptible to between-area differences in access to high-quality healthcare services.

Another feature of our application is the way in which the stratified multi-stage sampling design used by UKHLS is incorporated into the model. Within a disaggregated modelling approach targeting parameters of the superpopulation model that generated the data, random effects are included to handle the effects of clustering induced by the design, and a rich set of predictor variables correlated with the survey weights adjust for unequal selection probabilities (including non-response). Our application has the advantage that the clustering levels in the analysis model, namely area and household, correspond to the stages of the multi-stage sampling scheme. However, as far as clustering is concerned, it should be noted that the clustering induced by the sampling design plays a relatively minor role in a (short) time series analysis of 11 waves such as this: individuals change household and area of residence following sample inclusion, and the purpose of our approach is to allow these changes to be incorporated into the model.

The other issue regarding complex sampling designs arises when the selection probabilities for PSUs, SSUs, etc. are unequal, even if the overall design weights are equal. Methods for dealing with this problem have been proposed but only for simple two- and three-level models for cross-sectional data (e.g. Rabe-Hesketh & Skrondal, 2006). As far as the impact of this is concerned, we have implicitly assumed throughout that the predictor variables in our model adequately capture the correlation between the model residuals and all of the multilevel selection probabilities (Pfeffermann et al., 1998, p. 24). To support our belief that this strategy works for UKHLS, we ran a simple analysis of wave 1 data using a two-level model with individuals nested in households for which existing methods for handling unequal selection probabilities at levels one and two can be used: if the predictor variables we have included are inadequate then the results from fitting a simple two-level model will differ substantively from those obtained using the unequal weighting adjustment. The most complete implementation of multilevel models for complex survey designs is available in Stata (namely, the `svy` version of the `meglm` command) and the results from using it are given in the supplementary materials (Tables S5 and S6). The results from the unadjusted and adjusted multilevel models are very similar, despite wave 1 being where we would expect to see the largest design effects. More generally, aiming for robustness guarantees, extending our method to allow for unequal selection probabilities at different stages in multi-stage clustered designs, and extending it to incorporate design effects when the levels of the analysis model do not correspond to those used in the sampling design, would be interesting and challenging directions for further work.

The simulation study explored the performance of the MCMC procedure when applied to complex correlation structures such as those found in our application, and more generally dependence on the intra-household correlation and the sample size for estimation of the correlation model. Some bias in the correlation parameters was found for the scenario where the intra-household correlation is low (0.1) and the number of household pairs is small (26k observations from 20k superhousehold clusters). However, inference for the correlation model was not only shown to improve for a larger sample size, as expected, but also for an increase in the intra-household correlation from 0.1 to a modest 0.3.

The correlation matrix modelling approach presented in Section 3.3 employs a local linear approximation, where the correlation coefficient is expressed as a first-order Taylor expansion in terms of the covariates. While this approach offers simplicity and interpretability, it has the limitation of being a 'local' model only valid within a small neighbourhood of the true correlations. Consequently, the feasible parameter set $C_{\gamma,Z}$ ensuring positive definiteness is an estimate that depends on the observed covariates and so cannot be guaranteed to induce a positive-definite correlation matrix when applied to new samples. In terms of residual non-normality, our simulation results showed inference to be fairly robust to local model misspecification larger than that found in our application, but further work on developing likelihood adjustments to minimise the impact of likelihood misspecification (e.g. developing Royall & Tsou (2003) to ensure the misspecified likelihood satisfies the second Bartlett identity) would strengthen the applicability of our procedure.

The proposed model can be extended in a number of ways to handle different types of multilevel structure and to investigate further research questions. Our grouped random effects formulation provides a flexible framework that accommodates various correlation matrix specifications beyond the local linear approximation used here, such as the Cholesky decomposition. While we consider a continuous response, it is straightforward to generalise the MCMC algorithm to probit models for binary and ordered or unordered categorical responses, following the approach of Albert & Chib (1993). Another generalisation is the addition of random coefficients at the individual level, for example to allow individual-specific health trajectories with age as in a growth curve model (Laird & Ware, 1982), or autocorrelated residuals (Goldstein et al., 1994). Random coefficients may also be added at the

household or area levels; an area-specific coefficient for age, for example, would allow the between-area variance in health to depend on age.

Although motivated by the study of household effects with longitudinal data, the proposed model can be applied more generally. One potential application is to longitudinal data on individuals who are clustered, for example, in families, schools or workplaces which are fixed over time. We can specify a two-level model with individual-specific random effects where the within-cluster correlation between a pair of individual random effects depends on covariates that characterise the pair, for example their respective roles within the cluster (parent-child, manager-employee), sex or age difference. Such data are traditionally analysed using a three-level model with cluster random effects, but this assumes equal correlation between individuals within a cluster. Another application is to cross-sectional multivariate data on clustered individuals. We can specify a latent variable model where one or more latent variables represent individual traits measured by the observed multivariate responses, such as ability, personality traits or attitudes. Using our approach, traditional latent trait models can be extended to allow within-cluster correlation among latent variables to depend on pair-specific covariates.

Acknowledgements

The authors thank the reviewers for their valuable suggestions which have led to a much improved paper.

Funding

S.Z. gratefully acknowledges research support from the National Key R&D Program of China (grants 2023YFA1008700 and 2023YFA1008703) and the National Natural Science Foundation of China (grants 12301373, 12271166 and 12371289). P.C. gratefully acknowledges funding from the UK Economic and Social Research Council (ESRC) through grants ES/T002611/1 (Understanding Society: The UK Household Longitudinal Study, Waves 13–15) and ES/S012486/1 (MiSoC: The ESRC Research Centre on Micro-Social Change). Understanding Society is an initiative funded by the ESRC and various government departments, with scientific leadership by the Institute for Social and Economic Research, University of Essex, and survey delivery by NatCen Social Research and Kantar Public. The research data are distributed by the UK Data Service.

Data availability

Understanding Society data are publicly available to registered users from the UK Data Service at `https://datacatalogue.ukdataservice.ac.uk/series/series/2000053#access-data`. The relevant study numbers are: SN 6614 for the main datasets and, under Special License Access, SN 7248 for the Lower Layer Super Output Area identifiers.

The MCMC algorithm is implemented in an R package (`mvregrp`) which is available, with documentation and a simulated dataset, from `https://github.com/slzhang-fd/mvregrp`.

Supplementary material

Supplementary material is available online at Journal of the Royal Statistical Society: Series A.

Author contributions

F.S. and S.Z. contributed equally and are listed alphabetically.

References

Albert, J. H., & Chib, A. (1993). Bayesian analysis of binary and polychotomous response data. *Biometrika*, *85*(2), 347-361.

Anderson, T. W. (1973). Asymptotically efficient estimation of covariance matrices with linear structure. *The Annals of Statistics*, *1*(1), 135–141.

Atkins, D. C. (2005). Using multilevel models to analyze couple and family treatment data: basic and advanced issues. *Journal of Family Psychology*, *19*(1), 98-110.

Bates, B., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Boyle, M., & Willms, J. (1999). Place effects for areas defined by administrative boundaries. *American Journal of Epidemiology*, *149*(6), 577-585.

Brazeau, H., & Lewis, N. A. (2021). Within-couple health behavior trajectories: the role of spousal support and strain. *Health Psychology*, *40*(2), 125-134.

Brynin, M., Longhi, S., & Martínez Pérez, Á. (2008). The social significance of homogamy. In M. Brynin & J. Ermisch (Eds.), *Changing Relationships* (p. 73-90). New York: Routledge.

Chambers, R. (2003). Introduction to Part A. In R. Chambers & C. Skinner (Eds.), *Analysis of Survey Data* (p. 11-28). Chichester: Wiley.

Chandola, T., Bartley, M., Wiggins, R., & Schofield, P. (2003). Social inequalities in health by individual and household measures of social position in a cohort of healthy people. *Journal of Epidemiology and Community Health*, *57*, 56-62.

Chandola, T., Clarke, P., Wiggins, R., & Bartley, M. (2005). Who you live with and where you live: setting the context for health using multiple membership multilevel models. *Journal of Epidemiology and Community Health*, *59*(2), 170-175.

Davillas, A., & Pudney, S. (2017). Concordance of health states in couples: analysis of self-reported, nurse administered and blood-based biomarker data in the UK Understanding Society panel. *Journal of Health Economics*, *56*, 87-102.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian Data Analysis* (2nd ed.). Boca Raton, FL: Chapman and Hall/CRC.

Gerstorf, D., Windsor, T. D., Hoppmann, C. A., & Butterworth, C. (2013). Longitudinal change in spousal similarities in mental health: between-couple and within-couple perspectives. *Psychology and Aging*, *28*(2), 540-554.

Goldstein, H. (2010). *Multilevel Statistical Models* (4th ed.). London: Wiley.

Goldstein, H., Healy, M. J. R., & Rasbash, J. (1994). Multilevel time series models with applications to repeated measures data. *Statistics in Medicine*, *13*(16), 1643-1655.

Goldstein, H., Rasbash, J., Browne, W. J., Woodhouse, G., & Poulain, M. (2000). Multilevel models in the study of dynamic household structures. *European Journal of Population*, *16*, 373-387.

Griffith, G. J., & Jones, K. (2020). When does geography matter most? Age-specific geographical effects in the patterning of, and relationship between, mental wellbeing and mental illness. *Health and Place*, *64*, 102401.

Johnston, R., Jones, K., Propper, C., Sarker, R., Burgess, S., & Bolster, A. (2005). A missing level in the analyses of British voting behaviour: the household as context as shown by analyses of a 1992–1997 longitudinal survey. *Electoral Studies*, *24*, 201-225.

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, *38*(4), 963-974.

Li, Y., Yu, J., & Zeng, T. (2020). Deviance information criterion for latent variable models and misspecified models. *Journal of Econometrics*, *216*(2), 450-493.

Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*(1), 13-22.

Lynn, P. (2009). Sample design for Understanding Society. *Understanding Society Working Paper Series*(01), 1–16.

Maas, C., & Hox, J. (2004). The influence of violations of assumptions on multilevel parameter estimates and their standard errors. *Computational Statistics & Data Analysis*, *46*(3), 427–440.

McPherson, M., Smith-Lovin, L., & Cook, J. (2001). Birds of a feather: homophily in social networks. *Annual Review of Sociology*, *27*(1), 415-444.

Murphy, M. (1996). The dynamic household as a logical concept and its use in demography. *European Journal of Population*, *12*(4), 363-381.

Pfeffermann, D., Skinner, C. J., Holmes, D. J., Goldstein, H., & Rasbash, J. (1998). Weighting for unequal selection probabilities in multilevel models (with discussion). *Journal of the Royal Statistical Society B*, *60*(2), 23–56.

Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: unconstrained parameterisation. *Biometrika*, *86*(3), 677–690.

Rabe-Hesketh, S., & Skrondal, A. (2006). Multilevel modelling of complex survey data. *Journal of the Royal Statistical Society A*, *169*(4), 805–827.

Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and Longitudinal Modeling Using Stata, Volume I* (3rd ed.). College Station, TX: Stata Press.

Raudenbush, S. W., Brennan, R. T., & Barnett, R. C. (1995). A multivariate hierarchical model for studying psychological change within married couples. *Journal of Family Psychology*, *9*(2), 161-174.

Royall, R., & Tsou, T.-S. (2003). Interpreting statistical evidence by using imperfect models: robust adjusted likelihood functions. *Journal of the Royal Statistical Society B*, *64*(2), 391-404.

Sacker, A., Wiggins, R. D., & Bartley, M. (2006). Time and place: putting individual health into context. A multilevel analysis of the British household panel survey, 1991–2001. *Health and Place*, *12*(3), 279-290.

Snijders, T. A. B., & Berkhof, J. (2007). Diagnostic checks for multilevel models. In J. de Leeuw & I. Kreft (Eds.), *Handbook of Multilevel Analysis* (p. 141-175). New York: Springer.

Steele, F., Clarke, P., & Kuha, J. (2019). Modeling within-household associations in household panel studies. *Annals of Applied Statistics*, *13*(1), 367-392.

University of Essex, ISER. (2020). Understanding Society: Waves 1-11, 2009-2020 and Harmonised BHPS: Waves 1-18, 1991-2009. [data collection] (16th ed.). University of Essex, Institute for Social and Economic Research. UK Data Service. SN: 6614, `http://doi.org/10.5255/UKDA-SN-6614-16`.

Ware, J. E., Kosinski, D. M., M. amd Turner-Bowker, & Gandek, B. (2002). *How to score version 2 of the SF-12® health survey (with a supplement documenting version 1).* Lincoln, RI: QualityMetric Incorporated.

Weich, S., Holt, G., Twigg, L., Jones, K., & Lewis, G. (2003). Geographic variation in the prevalence of common mental disorders in Britain: a multilevel investigation. *American Journal of Epidemiology*, *157*(8), 730-737.

Yan, J., & Fine, J. (2004). Estimating equations for association structures. *Statistics in Medicine*, *23*(6), 859–874.

Zhang, S., Kuha, J., & Steele, F. (2024). Modelling correlation matrices in multivariate data, with application to reciprocity and complementarity of child-parent exchanges of support. *Annals of Applied Statistics*, *18*, 3024–3049.

Zou, T., Lan, W., Li, R., & Tsai, C.-L. (2022). Inference on covariance-mean regression. *Journal of Econometrics*, *230*(2), 318–338.

Zou, T., Lan, W., Wang, H., & Tsai, C.-L. (2017). Covariance regression analysis. *Journal of the American Statistical Association*, *112*(517), 266–281.

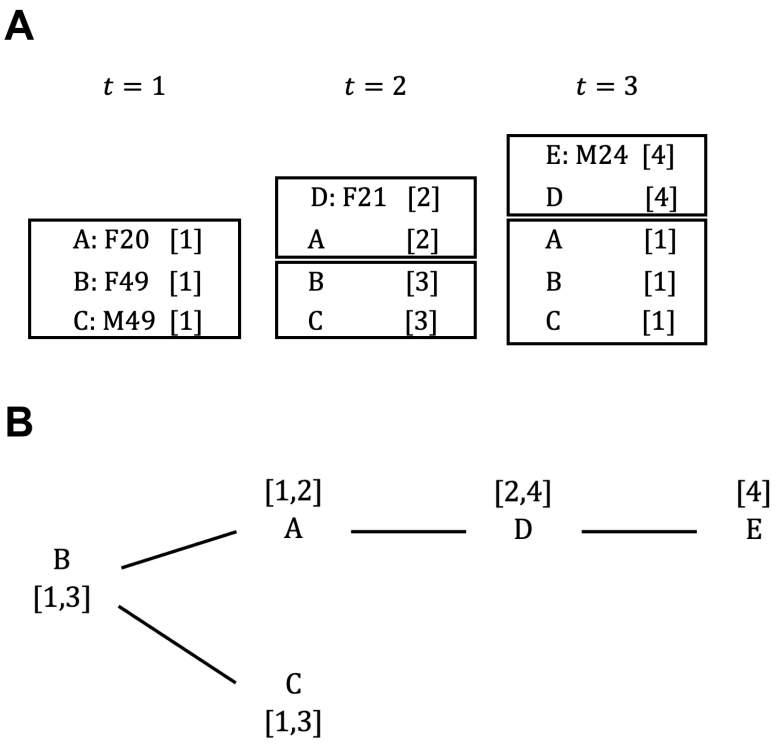

Figure 1: Illustration of the evolution of a household network over three waves. (A) Household membership at each wave with sex and age of each individual at entry to the panel. Coresidents are grouped together and household identifiers $h(ti)$ are in square brackets. (B) Network members by $t = 3$ with ever coresidence indicated by paths. Household identifiers $h(ti)$ over all three waves are in square brackets.

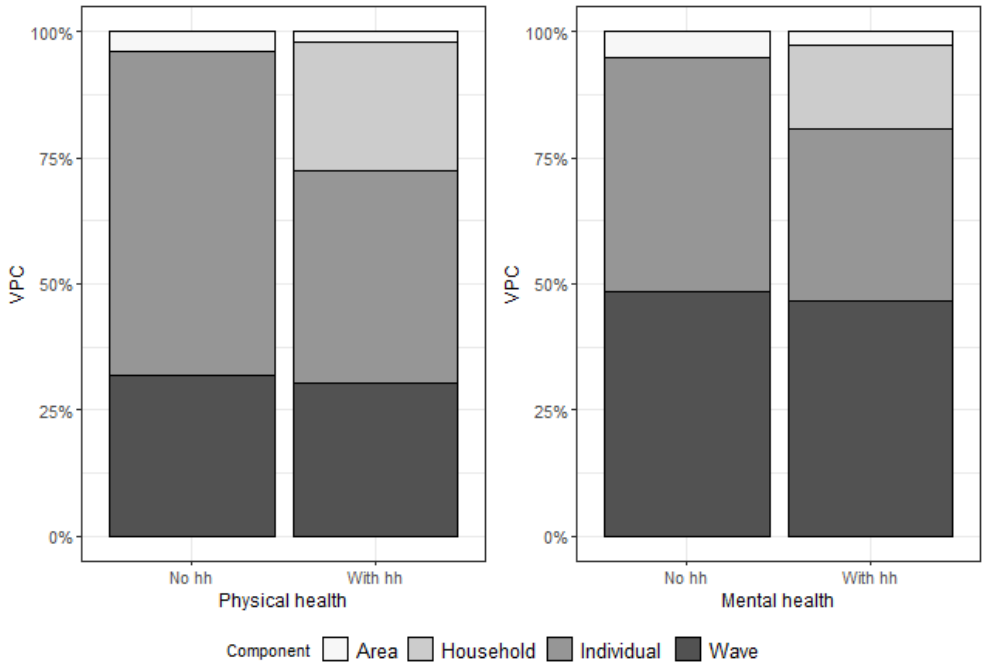

Figure 2: Unconditional variance partition coefficients from models without and with household effects. Intercept only in mean model and between-household correlation model.