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

Supplementary materials

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

S1 Details of the block-wise Gibbs sampling procedure

S1.1 Sampling the random effects

Sampling individual random effects For individual $i = 1, \ldots, n$, the individual-level random effect u_i is drawn from the posterior distribution

$$\begin{aligned} p(u_i \mid \boldsymbol{y}_i, \boldsymbol{v}_{h(i)}, \boldsymbol{w}_{a(i)}) &\propto \prod_{t=1}^{T} p(y_{ti} \mid u_i, v_{h(ti)}, w_{a(ti)}) p(u_i) \\ &\propto \exp\left[-\frac{1}{2\sigma_e^2} \sum_{t=1}^{T} \left(y_{ti} - \boldsymbol{\beta}' \mathbf{x}_{ti} - u_i - v_{h(ti)} - w_{a(ti)} \right)^2 - \frac{1}{2\sigma_u^2} u_i^2 \right], \end{aligned} \tag{S1}$$

where $\boldsymbol{y}_i = (y_{1i}, \ldots, y_{Ti})'$ are observed responses for individual i, and $\boldsymbol{v}_{h(i)} = (v_{h(1i)}, \ldots, v_{h(Ti)})'$, $\boldsymbol{w}_{a(i)} = (w_{a(1i)}, \ldots, w_{a(Ti)})'$ are vectors of their household and area effects. Specifically, this is a normal distribution with variance and mean

$$\sigma^2_{u,\text{pos}} = \left(\frac{T}{\sigma_e^2} + \frac{1}{\sigma_u^2} \right)^{-1}, \quad \mu_{u,\text{pos}} = \frac{\sigma^2_{u,\text{pos}}}{\sigma_e^2} \sum_{t=1}^{T} \left(y_{ti} - \boldsymbol{\beta}' \mathbf{x}_{ti} - v_{h(it)} - w_{a(ti)} \right).$$

*These authors contributed equally.

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

Sampling the grouped household random effects For household random effect v_h, suppose household h is in superhousehold s, and let A be the total number of areas. The posterior distribution of v_h is

$$p(v_h \mid \boldsymbol{y}, \boldsymbol{u}, \boldsymbol{w}) \propto \prod_{(i,t):h(ti)=h} p(y_{ti} \mid u_i, v_h, w_{a(ti)}) p(v_h \mid \boldsymbol{v}_{s_{-h}}), \tag{S2}$$

where $\boldsymbol{y} = (\boldsymbol{y}_1', \ldots, \boldsymbol{y}_n')'$, $\boldsymbol{u} = (u_1, \ldots, u_n)'$, $\boldsymbol{w} = (w_1, \ldots, w_A)'$, and (i,t) denotes pairs of i and t such that $h(ti) = h$. $\boldsymbol{v}_{s_{-h}}$ represents the vector $\boldsymbol{v}_s$ excluding v_h. In fact, instead of sampling each v_h in $\boldsymbol{v}_s = (v_{s_1}, \ldots, v_{s_{m_s}})'$ sequentially, we sample the whole vector $\boldsymbol{v}_s$ together to improve efficiency and mixing performance. The posterior distribution for $\boldsymbol{v}_s$ is

$$p(\boldsymbol{v}_s \mid \boldsymbol{y}, \boldsymbol{u}, \boldsymbol{w}) \propto \exp\left[-\frac{1}{2}(\boldsymbol{v}_s - \boldsymbol{t}_s)' \Omega_s^{-1} (\boldsymbol{v}_s - \boldsymbol{t}_s) - \frac{1}{2} \boldsymbol{v}_s' \Omega_{v_s}^{-1} \boldsymbol{v}_s \right] \tag{S3}$$

where

$$\begin{aligned} \boldsymbol{t}_s &= (t_{s_1}, \ldots, t_{s_{m_s}})', \\ t_h &= \frac{1}{n_h^{(v)}} \sum_{(i,t):h(ti)=h} \left(y_{ti} - \boldsymbol{\beta}' \mathbf{x}_{ti} - u_i - w_{a(ti)} \right), \\ \Omega_s^{-1} &= \operatorname{diag}\left(\frac{n_{s_1}^{(v)}}{\sigma_e^2}, \ldots, \frac{n_{s_{m_s}}^{(v)}}{\sigma_e^2} \right), \end{aligned}$$

and $n_h^{(v)} = \sum_{i=1}^{n} \sum_{t=1}^{T} \mathrm{I}\{h(ti) = h\}$. Specifically, the posterior distribution of $\boldsymbol{v}_s$ is a multivariate normal distribution with covariance matrix and mean

$$\Omega_{v_s,\mathrm{pos}} = \left(\Omega_s^{-1} + \Omega_{v_s}^{-1} \right)^{-1}, \quad \boldsymbol{\mu}_{v_s,\mathrm{pos}} = \Omega_{v_s,\mathrm{pos}} \Omega_s^{-1} \boldsymbol{t}_s.$$

In particular, if $m_s = 1$ (i.e. superhousehold s contains only one household), the posterior for $\boldsymbol{v}_s = (v_h)$ is

$$\sigma_{v_s,\mathrm{pos}}^2 = \left(\frac{n_h^{(v)}}{\sigma_e^2} + \frac{1}{\sigma_v^2} \right)^{-1}, \quad \mu_{v_s,\mathrm{pos}} = \frac{\sigma_{v_s,\mathrm{pos}}^2}{\sigma_e^2} n_h^{(v)} t_h.$$

Sampling the area random effects For area $a = 1, \ldots, A$, the random effect $w_a, a = a(ti)$ is drawn from the posterior distribution

$$\begin{aligned} p(w_a \mid \boldsymbol{y}, \boldsymbol{v}, \boldsymbol{u}) &\propto \prod_{(i,t):a(ti)=a} p(y_{ti} \mid u_i, v_{h(ti)}, w_a) p(w_a) \\ &\propto \exp\left[-\frac{1}{2\sigma_e^2} \sum_{(i,t):a(ti)=a} \left(y_{ti} - \boldsymbol{\beta}' \mathbf{x}_{ti} - u_i - v_{h(ti)} - w_a \right)^2 - \frac{1}{2\sigma_w^2} w_a^2 \right], \end{aligned} \tag{S4}$$

which is a normal distribution with variance and mean

$$\sigma_{w,\mathrm{pos}}^2 = \left(\frac{n_a^{(w)}}{\sigma_e^2} + \frac{1}{\sigma_w^2} \right)^{-1}, \quad \mu_{w,\mathrm{pos}} = \frac{\sigma_{w,\mathrm{pos}}^2}{\sigma_e^2} \sum_{(i,t):a(ti)=a} \left(y_{ti} - \boldsymbol{\beta}' \mathbf{x}_{ti} - u_i - v_{h(it)} \right),$$

where $n_a^{(w)} = \sum_{i=1}^{n} \sum_{t=1}^{T} \mathrm{I}\{a(ti) = a\}$.

S1.2 Sampling the parameters

The parameters in the model are $\boldsymbol{\beta}, \sigma_u^2, \sigma_v^2, \sigma_w^2, \sigma_e^2$ and $\boldsymbol{\gamma}$. We set out their sampling procedures below.

Sampling $\boldsymbol{\beta}$ By assuming the multivariate normal prior $N(\mathbf{0}, \sigma_0^2\mathbf{I})$ for $\boldsymbol{\beta}$, we have the posterior

$$\boldsymbol{\beta} \mid \boldsymbol{y}, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \sim N(\boldsymbol{\mu}_\beta, \Omega_\beta),$$

where $\boldsymbol{v} = (\boldsymbol{v}_1', \ldots, \boldsymbol{v}_K')'$ and

$$\begin{aligned}\Omega_\beta &= \left((1/\sigma_0^2)I + \Omega_e^{-1} \otimes (X'X)\right)^{-1} \text{ and} \\ \boldsymbol{\mu}_\beta &= \Omega_\beta(\Omega_e^{-1} \otimes X')\boldsymbol{r}^\beta,\end{aligned}$$

in which $\Omega_e = \sigma_e^2 I$, $\boldsymbol{r}^\beta = (r_{11}^\beta, \ldots, r_{T1}^\beta, r_{12}^\beta, \ldots, r_{Tn}^\beta)'$, and $r_{ti}^\beta = y_{ti} - u_i - v_{h(ti)} - w_{a(ti)}$.

Sampling σ_u^2, σ_v^2, σ_w^2, and σ_e^2 For sampling σ_u^2, we specify an inverse gamma prior $\mathrm{IG}(\alpha, \beta)$ where for weakly informative priors (α, β) are specified as small positive values, e.g. 0.01. Thus given individual random effects $u_i, i = 1, \ldots, n$, the posterior distribution of σ_u^2 is

$$\sigma_u^2 \mid \boldsymbol{u} \sim \mathrm{IG}\left(\alpha + \frac{n}{2}, \beta + \frac{\boldsymbol{u}'\boldsymbol{u}}{2}\right).$$

The posterior distribution of σ_v^2 is

$$\sigma_v^2 \mid \boldsymbol{v} \sim \mathrm{IG}\left(\alpha + \frac{\sum_s m_s}{2}, \beta + \frac{\sum_s \boldsymbol{v}_s'\mathbf{R}_{v_s}^{-1}\boldsymbol{v}_s}{2}\right).$$

Similarly, the posterior distribution of σ_w^2 is

$$\sigma_w^2 \mid \boldsymbol{w} \sim \mathrm{IG}\left(\alpha + \frac{A}{2}, \beta + \frac{\boldsymbol{w}'\boldsymbol{w}}{2}\right).$$

Finally the posterior of σ_e^2 is

$$\sigma_e^2 \mid \boldsymbol{e} \sim \mathrm{IG}\left(\alpha + \frac{nT}{2}, \beta + \frac{\sum_i \sum_t e_{it}^2}{2}\right),$$

where $e_{it} = y_{it} - \boldsymbol{\beta}'\mathbf{x}_{ti} - u_i - v_{h(ti)} - w_{a(ti)}$.

Sampling $\boldsymbol{\gamma}$ For each superhousehold that contains more than one household, there is a m_s-variate random effect vector

$$\boldsymbol{v}_s = (v_{s_1}, \ldots, v_{s_{m_s}})', \quad \boldsymbol{v}_s \sim N(0, \sigma_v^2\mathbf{R}_{v_s}),$$

where the elements of $\mathbf{R}_{v_s}$ (i.e. correlation coefficients) are further modelled as linear functions of covariates $\boldsymbol{z}$. In particular, we have

$$\mathrm{cor}(v_h, v_{h'}) = \boldsymbol{\gamma}'\boldsymbol{z}_{h,h'}, \quad \text{for } h, h' = 1, \ldots, m_s; h \neq h'; s(h) = s(h'). \tag{S5}$$

Let $Z_s = (\boldsymbol{z}_{s1}, \ldots, \boldsymbol{z}_{sq})$ be a $m_s(m_s-1)/2 \times q$ matrix containing covariates for each pair of random effects in $\boldsymbol{v}_s$ (i.e., the row vector of Z_s is $\boldsymbol{z}_{h,h'}, h \neq h', s(h) = s(h') = s$). Then we have

$$\mathbf{R}_{v_s} = \mathbf{R}_{v_s}(\boldsymbol{\gamma}, Z_s) = \gamma_1\mathbf{G}_{s1} + \cdots + \gamma_q\mathbf{G}_{sq} + (1 - \sum_{l=1}^{q}\gamma_l)I, \tag{S6}$$

where $\mathbf{G}_{sl}, l = 1, \ldots, q$ are symmetric matrices with off-diagonal elements $\boldsymbol{z}_{sl}$ and diagonal elements equal to one.

During the sampling, we require that all correlation matrices $\mathbf{R}_{v_s}, s = 1, \ldots, K$, are positive definite. Thus we specify the prior for $\boldsymbol{\gamma}$ as $p(\boldsymbol{\gamma}) = \mathrm{I}\{\boldsymbol{\gamma} \in C_{\gamma,Z}\}$, where

$$C_{\gamma,Z} = \left\{ \boldsymbol{\gamma} \in \mathbb{R}^q \mid \mathbf{R}_{v_s} = \sum_{l=1}^{q} \gamma_l \mathbf{G}_{sl} + (1 - \sum_{l=1}^{q} \gamma_l) I \succ 0, \text{ for } s = 1, \ldots, K \right\}, \tag{S7}$$

and a matrix $R \succ 0$ indicates positive definiteness. It is easy to verify that $C_{\gamma,Z}$ is a convex set for $\boldsymbol{\gamma}$ due to the fact that the set of all positive definite matrices is convex. By using the prior $C_{\gamma,Z}$, any convex combination (e.g. the posterior mean) of posterior samples from Algorithm S1 is still in $C_{\gamma,Z}$. For each element in $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)'$, we perform the following:

Algorithm S1 (Random walk Metropolis-Hastings for sampling $\boldsymbol{\gamma}$)**.** *For $l = 1, \ldots, q$:*

1. *Generate proposal*

$$\boldsymbol{\gamma}^* = \boldsymbol{\gamma} + \epsilon \boldsymbol{e}_l,$$

   where $\boldsymbol{e}_l$ is a vector with the l-th element sampled from the standard normal distribution and the remaining elements set to 0, and ϵ is the step size for the MH update which should be adjusted to achieve an acceptance rate of approximately 25%.

2. *Accept $\boldsymbol{\gamma}^*$ with the probability*

$$\alpha(\boldsymbol{\gamma} \to \boldsymbol{\gamma}^*) = \min \left\{ 1, \frac{\prod_s p(\boldsymbol{v}_s | \sigma_v^2, \boldsymbol{\gamma}^*) p(\boldsymbol{\gamma}^*)}{\prod_s p(\boldsymbol{v}_s | \sigma_v^2, \boldsymbol{\gamma}) p(\boldsymbol{\gamma})} \right\},$$

   and stay at $\boldsymbol{\gamma}$ with probability $1 - \alpha(\boldsymbol{\gamma} \to \boldsymbol{\gamma}^*)$.

The MCMC estimation procedure has been implemented in an R package which will be made available via a `Github` repository.

S2 Proof of Proposition 1

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 (9) of the main paper can be equivalently expressed as*

$$\mathbf{R}_{v_s} = \sum_{l=1}^{q} \gamma_l \mathbf{G}_{sl} + (1 - \sum_{l=1}^{q} \gamma_l) I, \tag{S8}$$

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 $l = 1, \ldots, q$.

Proof. The proof is straightforward. Let

$$\mathbf{G}_{sl} = \begin{bmatrix} 1 & & & \\ z_{12}^{(sl)} & 1 & & \\ \vdots & \vdots & \ddots & \\ z_{1,m_s}^{(sl)} & z_{2,m_s}^{(sl)} & \cdots & 1 \end{bmatrix}, \text{and} \quad \boldsymbol{z}_{sl} = \begin{bmatrix} z_{12}^{(sl)} \\ \vdots \\ z_{m_s-1,m_s}^{(sl)} \end{bmatrix}_{m_s(m_s-1)/2 \times 1}. \tag{S9}$$

Then the equation $\mathbf{R}_{v_s} = \sum_{l=1}^{q} \gamma_l \mathbf{G}_{sl} + (1 - \sum_{l=1}^{q} \gamma_l) I$ can be written as:

$$\begin{bmatrix} 1 & & & \\ \rho_{12} & 1 & & \\ \vdots & \vdots & \ddots & \\ \rho_{1,m_s} & \rho_{2,m_s} & \cdots & 1 \end{bmatrix} = \sum_{l=1}^{q} \gamma_l \begin{bmatrix} 1 & & & \\ z_{12}^{(sl)} & 1 & & \\ \vdots & \vdots & \ddots & \\ z_{1,m_s}^{(sl)} & z_{2,m_s}^{(sl)} & \cdots & 1 \end{bmatrix} + (1 - \sum_{l=1}^{q} \gamma_l) I$$

$$= \begin{bmatrix} 1 & & & \\ \sum_{l=1}^{q} \gamma_l z_{12}^{(sl)} & 1 & & \\ \vdots & \vdots & \ddots & \\ \sum_{l=1}^{q} \gamma_l z_{1,m_s}^{(sl)} & \sum_{l=1}^{q} \gamma_l z_{2,m_s}^{(sl)} & \cdots & 1 \end{bmatrix} = \begin{bmatrix} 1 & & & \\ \boldsymbol{\gamma}' \boldsymbol{z}_{12}^{(s)} & 1 & & \\ \vdots & \vdots & \ddots & \\ \boldsymbol{\gamma}' \boldsymbol{z}_{1,m_s}^{(s)} & \boldsymbol{\gamma}' \boldsymbol{z}_{2,m_s}^{(s)} & \cdots & 1 \end{bmatrix},$$

where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)'$ and $\boldsymbol{z}_{h,h'}^{(s)} = (z_{h,h'}^{(s1)}, \ldots, z_{h,h'}^{(sq)})'$. Therefore we have

$$\rho_{h,h'} = \boldsymbol{\gamma}' \boldsymbol{z}_{h,h'}^{(s)}.$$

This is equivalent to the correlation model in (9) of the main paper, which completes the proof. □

S3 Proof of Proposition 2

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

$$C_{\gamma,Z} = \left\{ \boldsymbol{\gamma} \in \mathbb{R}^q \mid \sum_{l=1}^{q} \gamma_l \mathbf{G}_{sl} + (1 - \sum_{l=1}^{q} \gamma_l) I \succ 0, \text{ for } s = 1, \ldots, K \right\} \tag{S10}$$

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

Proof. First, note that $C_{\gamma,Z} \subseteq \mathbb{R}^q$ and $\mathbf{0} \in C_{\gamma,Z}$. This is because, for $\boldsymbol{\gamma} = (0, \ldots, 0)'$, we have

$$\sum_{l=1}^{q} \gamma_l \mathbf{G}_{sl} + (1 - \sum_{l=1}^{q} \gamma_l) I = I, \quad \text{for } s = 1, \ldots, K,$$

which is positive definite.

Next, we show that $C_{\gamma,Z}$ is convex. Let $\boldsymbol{\gamma}_1 = (\gamma_{1,1}, \ldots, \gamma_{1,q})', \boldsymbol{\gamma}_2 = (\gamma_{2,1}, \ldots, \gamma_{2,q})' \in C_{\gamma,Z}$ and $\lambda \in [0, 1]$. Then for any $s = 1, \ldots, K$, we have

$$\mathbf{R}_{v_s,1} = \sum_{l=1}^{q} \gamma_{1,l} \mathbf{G}_{sl} + (1 - \sum_{l=1}^{q} \gamma_{1,l}) I \succ 0, \text{ and } \mathbf{R}_{v_s,2} = \sum_{l=1}^{q} \gamma_{2,l} \mathbf{G}_{sl} + (1 - \sum_{l=1}^{q} \gamma_{2,l}) I \succ 0.$$

Furthermore, let $\boldsymbol{\gamma}_\lambda = \lambda \boldsymbol{\gamma}_1 + (1 - \lambda) \boldsymbol{\gamma}_2$. Then we have

$$\begin{aligned} \mathbf{R}_{v_s,\lambda} &= \sum_{l=1}^{q} \gamma_{\lambda,l} \mathbf{G}_{sl} + (1 - \sum_{l=1}^{q} \gamma_{\lambda,l}) I \\ &= \sum_{l=1}^{q} \left[\lambda \gamma_{1,l} + (1 - \lambda) \gamma_{2,l} \right] \mathbf{G}_{sl} + (1 - \sum_{l=1}^{q} \left[\lambda \gamma_{1,l} + (1 - \lambda) \gamma_{2,l} \right]) I \\ &= \lambda \left[\sum_{l=1}^{q} \gamma_{1,l} \mathbf{G}_{sl} + (1 - \sum_{l=1}^{q} \gamma_{1,l}) I \right] + (1 - \lambda) \left[\sum_{l=1}^{q} \gamma_{2,l} \mathbf{G}_{sl} + (1 - \sum_{l=1}^{q} \gamma_{2,l}) I \right] \\ &= \lambda \mathbf{R}_{v_s,1} + (1 - \lambda) \mathbf{R}_{v_s,2} \succ 0, \end{aligned}$$

which implies $\boldsymbol{\gamma}_\lambda \in C_{\gamma,Z}$. Therefore, $C_{\gamma,Z}$ is convex. □

S4 Efficient SPD checking through Linear Programming

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 superhousehold 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. We elaborate on this in the following.

First, we maintain a minimal set of $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_R$ from previous iterations whose convex hull is a subset of $C_{\gamma,Z}$. Note that $C_{\gamma,Z}$ is convex and $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_R$ are its vertices. Let $\Gamma = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_R) \in \mathbb{R}^{q \times R}$. Next, for a new proposal $\boldsymbol{\gamma}^*$, we check if it can be expressed as a convex combination of the points (column vectors) in Γ. This involves pursuing a solution of

$$\begin{bmatrix} \Gamma \\ \mathbf{1}' \end{bmatrix} \boldsymbol{x} = \begin{bmatrix} \boldsymbol{\gamma}^* \\ 1 \end{bmatrix}, \quad \text{s.t. } x_r \geqslant 0, r = 1, \ldots, R, \tag{S11}$$

which is known as the feasible point problem and can be solved efficiently using linear programming. Specifically, we further reformulate the problem (S11) as the following linear programming problem

$$\begin{array}{ll} \max_{\boldsymbol{x},s} & s \\ \text{s.t.} & x_r \geqslant s, \ r = 1, \ldots, R \\ & A\boldsymbol{x} = \boldsymbol{b}, \end{array}$$

which can be efficiently solved using standard linear programming packages. Importantly, a high precision solution is not required, and the initial value can be set as the solution from the previous iteration. Once we find $(\boldsymbol{x}, s)$ that satisfies $s \geqslant 0$, we have $\boldsymbol{\gamma}^* \in C_{\gamma,Z}$. Otherwise, we check the positive definiteness of K matrices. If $\boldsymbol{\gamma}^*$ is in $C_{\gamma,Z}$, we set $p(\boldsymbol{\gamma}^*) = 1$ and append $\boldsymbol{\gamma}^*$ at the last column of Γ; if not, set $p(\boldsymbol{\gamma}^*) = 0$. In summary, 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. The estimation algorithm has been implemented in an R package which will be made available on publication, with documentation and simulated data.

S5 Theoretical properties of the proposed estimators

S5.1 Formal setup and assumptions

We establish the asymptotic properties of the posterior distribution for the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\sigma}^2) \in \Theta$, where $\boldsymbol{\sigma}^2 = (\sigma_v^2, \sigma_u^2, \sigma_w^2, \sigma_e^2)$. The data consist of observations from K independent superhouseholds, indexed by $s = 1, \ldots, K$. The data for superhousehold s is denoted by $\mathbf{Y}_s$, a vector of length n_s. The model is specified by the likelihood

$$p(\mathbf{Y}_1, \ldots, \mathbf{Y}_K|\boldsymbol{\theta}) = \prod_{s=1}^{K} p(\mathbf{Y}_s|\boldsymbol{\theta}),$$

where $\mathbf{Y}_s \sim N(\mathbf{X}_s\boldsymbol{\beta}, \boldsymbol{\Omega}_s(\boldsymbol{\gamma}, \boldsymbol{\sigma}^2))$. The $n_s \times n_s$ covariance matrix for superhousehold s has the structure

$$\boldsymbol{\Omega}_s(\boldsymbol{\gamma}, \boldsymbol{\sigma}^2) = \sigma_e^2\mathbf{I}_{n_s} + \sigma_u^2\mathbf{C}_s\mathbf{C}_s' + \sigma_w^2\mathbf{A}_s\mathbf{A}_s' + \sigma_v^2\mathbf{H}_s\mathbf{R}_{vs}(\boldsymbol{\gamma})\mathbf{H}_s',$$

where $\mathbf{C}_s, \mathbf{A}_s, \mathbf{H}_s$ are known design matrices for individual, area, and household effects respectively, and $\mathbf{R}_{vs}(\boldsymbol{\gamma})$ is the $m_s \times m_s$ household-level correlation matrix, which is a linear function of $\boldsymbol{\gamma}$ as defined in Equation (10) of the main text. Let $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \boldsymbol{\sigma}_0^2)$ denote the true value of the parameter.

Our theoretical results rely on the following set of regularity conditions.

Assumption 1 (Proper prior)**.** *The prior distribution $\pi(\boldsymbol{\theta})$ on the parameter space $\Theta \subset \mathbb{R}^d$ is proper, i.e., $\int_\Theta \pi(\boldsymbol{\theta})d\boldsymbol{\theta} = 1$. The prior density $\pi(\cdot)$ is assumed to be continuous and positive in a neighborhood of the true parameter $\boldsymbol{\theta}_0$, which is assumed to be an interior point of Θ.*

Assumption 2 (Asymptotic regime and superhousehold structure)**.** *The asymptotic regime is characterized by an increasing sample size drawn from the population, observed over a fixed, finite time horizon T. We assume that the underlying process of household formation and dissolution, over this fixed period T, results in a dependency structure that satisfies the following conditions:*

(i) The number of independent superhouseholds K tends to infinity ($K \to \infty$).

(ii) The size of the superhouseholds is uniformly bounded. That is, there exists a constant $M < \infty$ (which may depend on T) such that the number of observations n_s and the number of households m_s within each superhousehold s satisfy $\sup_s n_s \leqslant M$ and $\sup_s m_s \leqslant M$.

Assumption 3 (Covariates and design)**.** *The entries of the covariate matrices $\mathbf{X}_s$ and $\mathbf{Z}_s$ (used to construct $\mathbf{R}_{vs}(\boldsymbol{\gamma})$) are uniformly bounded. Furthermore, the Fisher information matrix with respect to $\boldsymbol{\theta}$ for a single superhousehold is well-behaved, and its average over all superhouseholds converges to a positive definite matrix $\bar{\mathbf{I}}(\boldsymbol{\theta}_0)$ as $K \to \infty$.*

Assumption 4 (Identifiability)**.** *The parameter vector $\boldsymbol{\theta}$ is identifiable from the likelihood. That is, for any $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$ in Θ, the set $\{\mathbf{Y} : p(\mathbf{Y}; \boldsymbol{\theta}_1) \neq p(\mathbf{Y}; \boldsymbol{\theta}_2)\}$ has positive measure. This implies that the Kullback-Leibler (KL) divergence $KL(p_{\boldsymbol{\theta}_0} \| p_{\boldsymbol{\theta}}) > 0$ for all $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$.*

Assumption 5 (Stability of the parameter space)**.** *The feasible set for $\boldsymbol{\gamma}$, $C_{\boldsymbol{\gamma},\mathbf{Z}_K} = \{\boldsymbol{\gamma} : \mathbf{R}_{vs}(\boldsymbol{\gamma})$ is positive definite for all $s = 1, \ldots, K\}$, is data-dependent. We assume that the collection of covariates $\{\mathbf{Z}_1, \ldots, \mathbf{Z}_K\}$ is such that $C_{\boldsymbol{\gamma},\mathbf{Z}_K}$ converges to a fixed, compact, convex limiting set $C_{\boldsymbol{\gamma}}^*$ as $K \to \infty$. We assume the true parameter $\boldsymbol{\gamma}_0$ lies in the interior of $C_{\boldsymbol{\gamma}}^*$.*

Remark 1 (On the validity of Assumption 2)**.** *In longitudinal studies, the dynamic process of household formation and dissolution could merge previously independent clusters (superhouseholds). Unconstrained, this process could lead to the emergence of a "giant component" – a single massive cluster encompassing most of the sample, thereby violating the requirements of $K \to \infty$ and bounded superhousehold size. Restricting the analysis to a fixed, finite time horizon T ensures the dependency structure remains sufficiently sparse. This "short panel" constraint limits recombination, allowing K to grow proportionally with the total sample size and justifying the application of this framework to large-scale, short-duration surveys rather than scenarios requiring long-term network asymptotics.*

S5.2 Main theoretical results: posterior concentration

Under the stated assumptions, we prove that the posterior distribution concentrates all of its mass in any arbitrarily small, fixed neighborhood of the true parameter value $\boldsymbol{\theta}_0$ as the sample size K grows to infinity.

Theorem 1 (Posterior concentration)**.** *Let the posterior distribution be denoted by* $\Pi(\cdot|\mathbf{Y}_1, \ldots, \mathbf{Y}_K)$*. Under Assumptions 1–5, for any* $\varepsilon > 0$*,*

$$\Pi\left(\boldsymbol{\theta} \in \Theta : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_2 > \varepsilon \mid \mathbf{Y}_1, \ldots, \mathbf{Y}_K\right) \to 0,$$

in $\mathbb{P}_{\boldsymbol{\theta}_0}$*-probability as* $K \to \infty$*, where* $\|\cdot\|_2$ *is the distance measured as the square root of the sum of squares of the differences between the components of the parameter vector.*

A direct consequence of posterior concentration is the weak consistency of the posterior mean as a point estimator for $\boldsymbol{\theta}_0$.

Corollary 1 (Consistency of the posterior mean)**.** *Let* $\hat{\boldsymbol{\theta}}_K = \mathrm{E}[\boldsymbol{\theta} \mid \mathbf{Y}_1, \ldots, \mathbf{Y}_K]$ *be the posterior mean of* $\boldsymbol{\theta}$*. Under the conditions of Theorem 1, the posterior mean is a (weakly) consistent estimator of* $\boldsymbol{\theta}_0$*, i.e.,*

$$\hat{\boldsymbol{\theta}}_K \to \boldsymbol{\theta}_0 \quad \textit{in } \mathbb{P}_{\boldsymbol{\theta}_0}\textit{-probability.}$$

The results above provide a rigorous theoretical justification for the proposed estimation procedure. They ensure that with a sufficiently large number of superhouseholds, our Bayesian method will recover the true underlying parameters with high precision.

Proof of Theorem 1 and Corollary 1. The proof is a direct application of Schwartz's Theorem for posterior consistency (see, e.g., Ghosal & Van der Vaart, 2017). The theorem requires two main conditions to hold: (i) the prior distribution must assign positive mass to any Kullback-Leibler (KL) neighborhood of the true data-generating density, and (ii) there must exist exponentially consistent tests for testing the true parameter $\boldsymbol{\theta}_0$ against any alternative outside a given neighborhood. We show that our assumptions are sufficient to verify these two conditions.

We first verify the Kullback-Leibler prior support condition. For a parametric model, this condition requires that for any $\delta > 0$, the prior assigns positive mass to the set $\{\boldsymbol{\theta} : KL(p_{\boldsymbol{\theta}_0} \| p_{\boldsymbol{\theta}}) < \delta\}$. The KL divergence between the densities corresponding to $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}$ is given by the limit of the average KL divergences:

$$KL(p_{\boldsymbol{\theta}_0} \| p_{\boldsymbol{\theta}}) = \lim_{K \to \infty} \frac{1}{K} \sum_{s=1}^{K} \mathrm{E}_{\boldsymbol{\theta}_0}[\log(p(\mathbf{Y}_s|\boldsymbol{\theta}_0)/p(\mathbf{Y}_s|\boldsymbol{\theta}))].$$

Under the regularity of our model (differentiability of the log-likelihood, bounded covariates), the function $\boldsymbol{\theta} \mapsto KL(p_{\boldsymbol{\theta}_0} \| p_{\boldsymbol{\theta}})$ is continuous in a neighborhood of $\boldsymbol{\theta}_0$, with

$KL(p_{\boldsymbol{\theta}_0} \| p_{\boldsymbol{\theta}_0}) = 0$. By Assumption 1, the prior density $\pi(\cdot)$ is continuous and strictly positive at $\boldsymbol{\theta}_0$. Therefore, any open ball $B(\boldsymbol{\theta}_0, \eta) = \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| < \eta\}$ in the parameter space has positive prior mass. Due to the continuity of the KL divergence, for any $\delta > 0$, we can find a sufficiently small η such that this ball $B(\boldsymbol{\theta}_0, \eta)$ is a subset of the KL neighborhood $\{\boldsymbol{\theta} : KL(p_{\boldsymbol{\theta}_0} \| p_{\boldsymbol{\theta}}) < \delta\}$. Since the ball has positive prior mass, the KL neighborhood must also have positive prior mass. Thus, the condition is satisfied.

Next, we verify the exponentially consistent tests condition. Let U be any open neighborhood of $\boldsymbol{\theta}_0$. We must show that there exists a sequence of tests ϕ_K for the hypothesis $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against the alternative $H_1 : \boldsymbol{\theta} \in U^c$ (the complement of U) such that the error probabilities decay to zero exponentially fast.

Consider the likelihood ratio test statistic. For any alternative $\boldsymbol{\theta}_1 \in U^c$, by the Law of Large Numbers for independent, non-identically distributed random variables:

$$\frac{1}{K} \sum_{s=1}^{K} \log \frac{p(\mathbf{Y}_s|\boldsymbol{\theta}_1)}{p(\mathbf{Y}_s|\boldsymbol{\theta}_0)} \xrightarrow{\mathbb{P}_{\boldsymbol{\theta}_0}} -KL(p_{\boldsymbol{\theta}_0} \| p_{\boldsymbol{\theta}_1}).$$

By the identifiability condition (Assumption 4), $KL(p_{\boldsymbol{\theta}_0} \| p_{\boldsymbol{\theta}_1}) > 0$. This means the average log-likelihood ratio converges in probability to a strictly negative number. This separation between the null and the alternative allows for the construction of a powerful test. For instance, a test that rejects H_0 if the average log-likelihood ratio is greater than $-c$ (for some small $c > 0$) will have Type I and Type II error probabilities that can be shown to decay exponentially to zero. Since the parameter space Θ is finite-dimensional, we can cover the compact set $\Theta \backslash U$ with a finite number of such alternatives and construct a uniformly exponentially consistent test.

Since both the KL prior support condition and the existence of exponentially consistent tests are satisfied under our assumptions, Schwartz's Theorem applies and the posterior distribution is consistent, as stated in Theorem 1. The consistency of the posterior mean in Corollary 1 follows from the fact that the posterior distribution concentrates its mass in an arbitrarily small neighborhood of $\boldsymbol{\theta}_0$. □

S5.3 Asymptotic validity of credible intervals

An important tool for inference in this paper is the 95% credible interval. While posterior consistency guarantees that uncertainty vanishes asymptotically, the Bernstein-von Mises (BvM) theorem provides a much stronger, finite-sample approximation that justifies the accuracy and interpretation of these intervals. We state the theorem and its direct consequences for our model, which hold under the regularity conditions specified in Assumptions A1-A5.

Theorem 2 (Bernstein-von Mises)**.** *Under Assumptions A1-A5, the posterior distribution of $\boldsymbol{\theta}$ is asymptotically normal. Let $\hat{\boldsymbol{\theta}}_K$ be a consistent and efficient estimator, such as the posterior mean. Then, the posterior distribution of the centered and scaled parameter vector converges in total variation distance to a Normal distribution:*

$$\sup_{B \in \mathcal{B}(\mathbb{R}^d)} \left| \Pi\left(\sqrt{K}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_K) \in B \mid \mathbf{Y}_1, \ldots, \mathbf{Y}_K\right) - N(0, \bar{\mathbf{I}}(\boldsymbol{\theta}_0)^{-1})(B) \right| \to 0,$$

in $\mathbb{P}_{\boldsymbol{\theta}_0}$-probability as $K \to \infty$. Here, $\mathcal{B}(\mathbb{R}^d)$ denotes the Borel σ-algebra (i.e., the smallest σ-algebra that contains all open sets) on $\mathbb{R}^d$, and $\bar{\mathbf{I}}(\boldsymbol{\theta}_0)^{-1}$ is the inverse of the asymptotic Fisher information matrix.

Remark 2. *The Bernstein-von Mises (BvM) theorem above provides essential reassurance regarding the practical validity of the Bayesian inference for our model and the reliability of our uncertainty estimates.*

It guarantees that with a large number of superhouseholds (K), the posterior distribution for our parameters $\boldsymbol{\theta}$ (including regression effects, correlation parameters, and variance components) becomes approximately Normal centered near the true values. This has two critical implications for practice: first, it ensures that our Bayesian credible intervals are well-calibrated in a frequentist sense; for example, a 95% credible interval will cover the true parameter value in approximately 95% of repeated experiments. Second, our estimators are statistically efficient, achieving the best possible precision.

In essence, the BvM theorem confirms that our reported credible intervals are not only meaningful summaries of posterior belief but are also accurate and efficient measures of uncertainty for the real-world quantities we aim to estimate.

Proof of Theorem 2. The proof consists of demonstrating that our model, under Assumptions A1-A5, satisfies the standard regularity conditions required for a parametric Bernstein-von Mises theorem to hold (see, e.g., Chapter 10, Van der Vaart, 2000). The key conditions are the existence of a suitable local approximation of the likelihood function (specifically, the Local Asymptotic Normality property) and a well-behaved prior distribution.

The BvM theorem requires the prior distribution $\pi(\boldsymbol{\theta})$ to have a density that is continuous and strictly positive in a neighborhood of the true parameter $\boldsymbol{\theta}_0$. This is explicitly stated in our Assumption 1.

Next, we show that the log-likelihood function is sufficiently regular to permit a local quadratic approximation around the true parameter $\boldsymbol{\theta}_0$. 1). The log-likelihood function, $\log p(\mathbf{Y}_s|\boldsymbol{\theta})$, is at least twice continuously differentiable with respect to all components of $\boldsymbol{\theta}$. This is guaranteed because the covariance matrix $\boldsymbol{\Omega}_s(\boldsymbol{\theta})$ is a smooth function of $\boldsymbol{\theta}$, and the matrix logarithm and inverse are smooth operations on the space of positive definite matrices. Assumption 5 ensures that $\boldsymbol{\theta}_0$ is an interior point of the parameter space, so these derivatives are well-defined. 2). The existence and convergence of the average Fisher information matrix to a positive definite limit, $\bar{\mathbf{I}}(\boldsymbol{\theta}_0)$, is guaranteed by Assumption 3. This matrix forms the curvature of the limiting quadratic approximation. The boundedness of covariates and superhousehold sizes (Assumptions 2 and 3) ensures that the derivatives of the log-likelihood have finite moments. 3). The posterior distribution is consistent, which was established as a consequence of Assumptions A1-A5.

Since our model satisfies the standard conditions for a parametric model for the likelihood, the conclusion of the Bernstein-von Mises theorem applies directly. This establishes that the posterior distribution, when properly centered and scaled, converges to a $N(0, \bar{\mathbf{I}}(\boldsymbol{\theta}_0)^{-1})$ distribution in total variation distance.

□

S6 Construction of household and superhousehold identifiers

The first step is to construct the wave 11 superhousehold clusters. The superhousehold identifier is initialised at the wave 1 cross-sectional household identifier, and extended at each subsequent wave $t = 2, \ldots, 11$ to include any individuals who have joined a wave $t-1$ household within a wave $t-1$ superhousehold. Households joining the study after wave 1 are assigned a new superhousehold identifier at the time of entry. In practice, the construction of superhouseholds is challenging because household panel studies have complex designs. In particular, any algorithm must account for new entrants (from single individuals to entire households) at each wave, individuals who rejoin previous coresidents (e.g. children returning to the parental home), and wave non-response of households and individuals within households. Further details are provided in the supplementary material to Steele et al. (2019).

After creating the superhousehold clusters, the cross-wave household identifiers defining the household random effects are constructed. These are nested within superhouseholds and are initialised at the cross-sectional household identifier for the wave at which their superhousehold is first observed. Thereafter, an individual's coresidents at t are compared with their coresidents at $t-1$. If there is no change, their cross-wave household identifier (and that of their coresidents) remains constant. For individuals whose coresidents changed between $t-1$ and t, their coresidents prior to $t-1$ are also scanned: if there is an exact match between their coresidents at t and those at $t' < t-1$, their household identifier at t reverts to that at t'; otherwise a new household identifier is assigned at t.

S7 Additional simulation results: non-normal household random effect distributions

To assess the robustness of our method against violations of the normality assumption, we simulated the household random effects from two non-normal distributions using a Gaussian copula. This approach preserves the target correlation matrix $\mathbf{R}_h$ while allowing for flexible marginal distributions. Specifically, for each household h, we first drew $\mathbf{Z}_h \sim \text{MVN}(\mathbf{0}, \mathbf{R}_h)$ and transformed its components to uniform variables, $U_{hj} = \Phi(Z_{hj})$. The final random effects v_{hj} were obtained by applying the appropriate inverse CDF, F^{-1} and then standardizing to ensure mean zero and variance σ_v^2:

$$v_{hj} = \left(\frac{F^{-1}(U_{hj}) - \mathbb{E}[F^{-1}(U_{ij})]}{\sqrt{\text{Var}(F^{-1}(U_{hj}))}} \right) \sigma_v. \tag{S12}$$

We define three scenarios by setting F^{-1} to be the quantile function of either a t-distribution with 5 degrees of freedom (for a heavy-tailed case) or a skew-normal distribution with shape parameter $\alpha = \pm 4$ (moderate negative and positive skew).

Table S1: Simulation results for heterogeneous between-household correlation design with $r = 200$ replicates of $M = 40,000$ superhouseholds* selected with replacement from the UKHLS data and VPC_{hh}=0.3. Results are given for Model M3 with non-normal distributions for the household random effects; the individual and area random effects are assumed to be normally distributed.

t-distribution on 5 d.f.

	β_0	β_1	β_2	β_3	β_4			
True	0.1	0.25	−0.2	0.3	0.1			
Mean	0.0994	0.2501	−0.1999	0.3006	0.1005			
Mean SE	0.0046	0.0015	0.0053	0.0073	0.0065			
SD	0.0043	0.0015	0.0050	0.0073	0.0066			
95% coverage	0.955	0.940	0.965	0.945	0.955			
	σ^2_e	σ^2_u	σ^2_v	γ_0	γ_1	γ_2	γ_3	γ_4
True	0.4	0.3	0.3	0.6	−0.1	−0.1	0.1	0.4
Mean	0.4000	0.2998	0.3001	0.5870	−0.1006	−0.0982	0.0995	0.4019
Mean SE	0.0010	0.0027	0.0032	0.0117	0.0178	0.0197	0.0263	0.0308
SD	0.0010	0.0028	0.0044	0.0123	0.0184	0.0201	0.0265	0.0342
95% coverage	0.965	0.935	0.810	0.790	0.940	0.950	0.965	0.930

Skewed normal ($\alpha = 4$, skewness=0.78)

	β_0	β_1	β_2	β_3	β_4			
True	0.1	0.25	−0.2	0.3	0.1			
Mean	0.1000	0.2502	−0.2001	0.3006	0.1009			
Mean SE	0.0046	0.0015	0.0053	0.0073	0.0065			
SD	0.0045	0.0013	0.0051	0.0069	0.0063			
95% coverage	0.970	0.960	0.955	0.975	0.965			
	σ^2_e	σ^2_u	σ^2_v	γ_0	γ_1	γ_2	γ_3	γ_4
True	0.4	0.5	0.1	0.6	−0.1	−0.1	0.1	0.4
Mean	0.3999	0.2997	0.3005	0.5911	−0.1014	−0.1014	0.0969	0.3978
Mean SE	0.0010	0.0027	0.0032	0.0117	0.0177	0.0196	0.0261	0.0305
SD	0.0010	0.0028	0.0036	0.0129	0.0187	0.0199	0.0271	0.0319
95% coverage	0.960	0.940	0.940	0.865	0.945	0.945	0.950	0.940

Skewed normal ($\alpha = -4$, skewness=−0.78)

	β_0	β_1	β_2	β_3	β_4			
True	0.1	0.25	−0.2	0.3	0.1			
Mean	0.0997	0.2503	−0.2001	0.3003	0.1006			
Mean SE	0.0046	0.0015	0.0053	0.0073	0.0065			
SD	0.0045	0.0014	0.0051	0.0073	0.0064			
95% coverage	0.945	0.970	0.955	0.950	0.945			
	σ^2_e	σ^2_u	σ^2_v	γ_0	γ_1	γ_2	γ_3	γ_4
True	0.4	0.5	0.1	0.6	−0.1	−0.1	0.1	0.4
Mean	0.3999	0.2997	0.3006	0.5908	−0.1012	−0.1016	0.0985	0.3997
Mean SE	0.0010	0.0027	0.0032	0.0117	0.0177	0.0196	0.0257	0.0309
SD	0.0010	0.0027	0.0035	0.0126	0.0188	0.0201	0.0270	0.0310
95% coverage	0.955	0.955	0.920	0.850	0.925	0.945	0.925	0.955

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

S8 Additional results from analysis of physical and mental health

Table S2: Descriptive statistics for the covariates in the mean model (n=387,238 person-wave observations).

Variable	n	Percent
Respondent characteristics		
Age (years)	Mean=48.4	SD=18.3
Sex		
Female	215,634	55.7
Male	171,604	44.3
Ethnicity		
Asian / Asian British	33,127	8.6
Black / Black British	14,694	3.8
White	330,068	85.2
Other	9349	2.4
Partnership status		
Partnered	245,219	63.3
Single	142,019	36.7
Number of children		
None	231,970	59.9
1	62,635	16.2
2	63,110	16.3
3+	29,523	7.6
Age of youngest child (among those with children)		
$0-1$ years	19,295	12.4
$2-4$ years	24,202	15.6
$5-10$ years	35,094	22.6
$11-16$ years	30,896	19.9
>16 years	45,781	29.5
Highest educational qualification		
Below degree level	242,390	62.6
Degree	144,848	37.4
Employment status		
Employed	215,060	55.5
Unemployed	17,622	4.6
Economically inactive	154,556	39.9
Household characteristics		
Annual equivalised household income (GBP)	Mean=21,636	SD=28,888
Housing tenure		
Owner-occupier	280,324	72.4
Social rent	60,730	15.7
Private rent	46,184	11.9

Table S3: Unconditional random effect variances from variance components models of mean standardised self-rated physical and mental health functioning.

	Physical health				Mental health			
	No hh effects		hh effects		No hh effects		hh effects	
	Est.	95% CI	Est.	95% CI	Est.	95% CI	Est.	95% CI
Area σ^2_w	0.042	(0.039, 0.045)	0.023	(0.020, 0.025)	0.051	(0.048, 0.055)	0.027	(0.024, 0.030)
Household σ^2_v	–	–	0.267	(0.256, 0.277)	–	–	0.171	(0.164, 0.178)
Individual σ^2_u	0.657	(0.649, 0.666)	0.443	(0.435, 0.451)	0.475	(0.468, 0.482)	0.350	(0.343, 0.357)
Wave σ^2_e	0.325	(0.324, 0.327)	0.317	(0.315, 0.319)	0.494	(0.492, 0.497)	0.480	(0.477, 0.482)
Household cor. γ	–	–	0.855	(0.843, 0.865)	–	–	0.605	(0.579, 0.630)

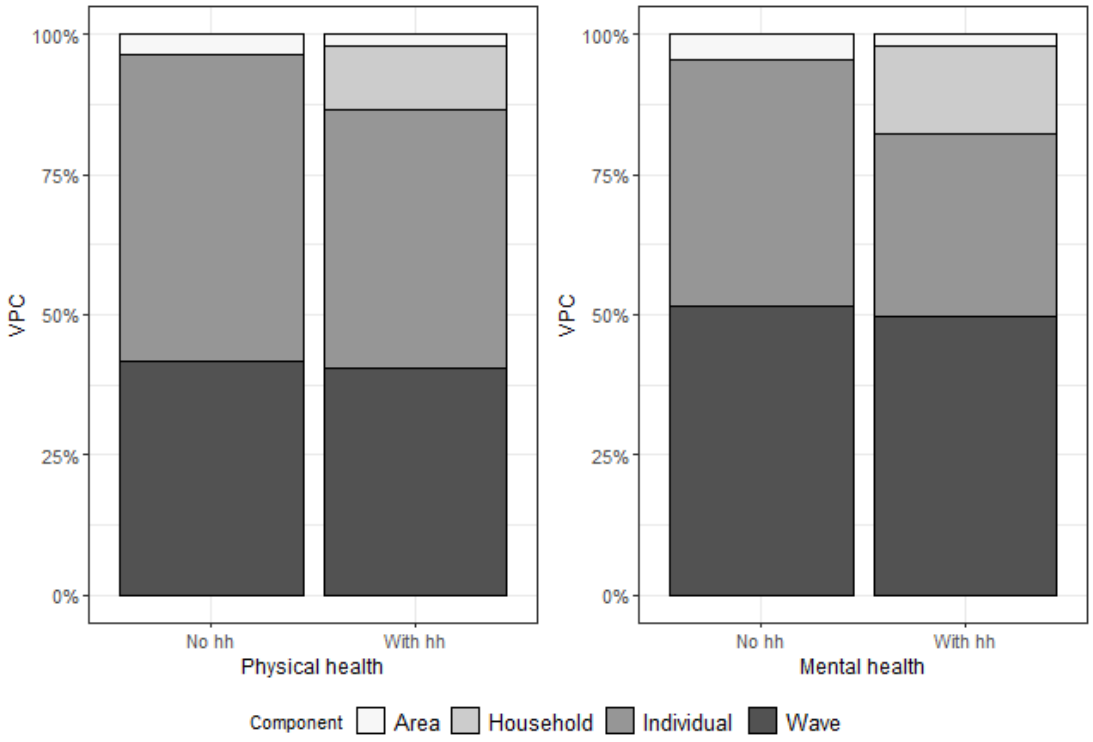

Figure S1: Conditional variance partition coefficients from models without and with household effects. The mean model includes the full set of covariates while the between-household correlation model includes only an intercept term.

Table S4: Unconditional intra-class correlations (ICC) $\text{Cor}(r_{ti}, r_{t'i'})$ implied by models for mean standardised self-rated physical and mental health functioning and between-household random effect correlations with covariates. In each case, observations ti and $t'i'$ are assumed to be from the same area, i.e. $a(ti) = a(t'i')$.

Type of ICC	Definition	Physical	Mental
ICC_{ind}: Same ind., diff. waves	$t \neq t'; i = i'; h(ti) = h(t'i)$	0.699	0.535
ICC_{hh}: Diff. ind., same hh	$i \neq i'; h(ti) = h(t'i')$	0.278	0.198
ICC_{shh}: Diff. hhs, same super-hh	$i \neq i'; h(ti) \neq h(t'i'); s(i) = s(i')$	0.230 (0.016)[a]	0.166 (0.011)[a]
ICC_{area}: Diff. super-hhs, same area	$s(i) \neq s(i')$	0.021	0.025

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

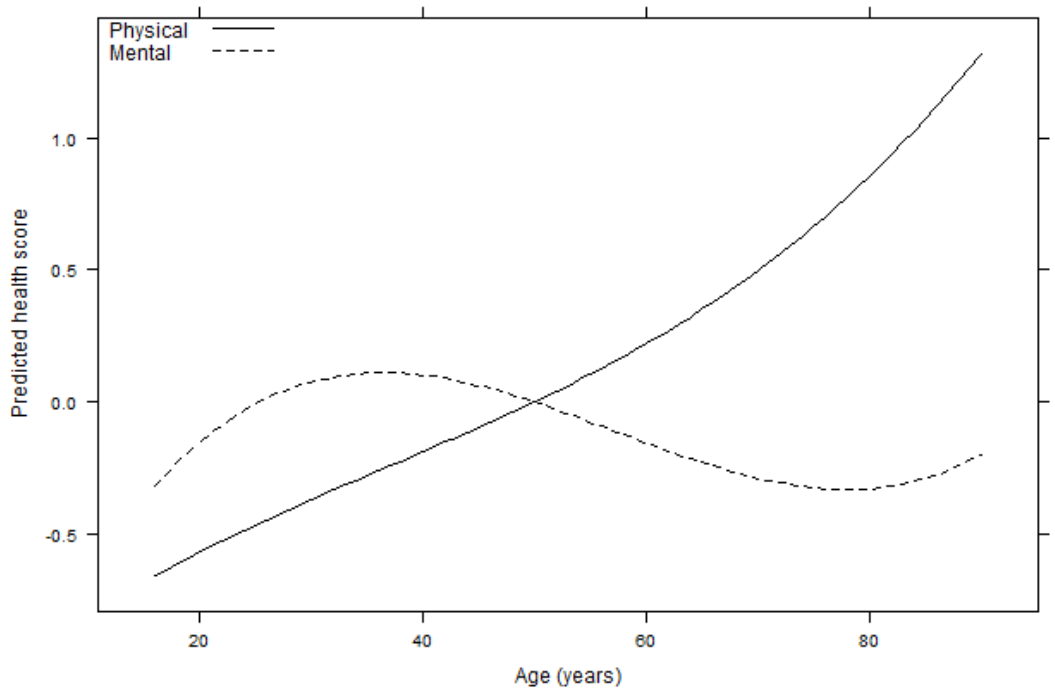

Figure S2: Predicted standardised self-rated physical and mental health functioning by age. Higher scores indicate lower functioning.

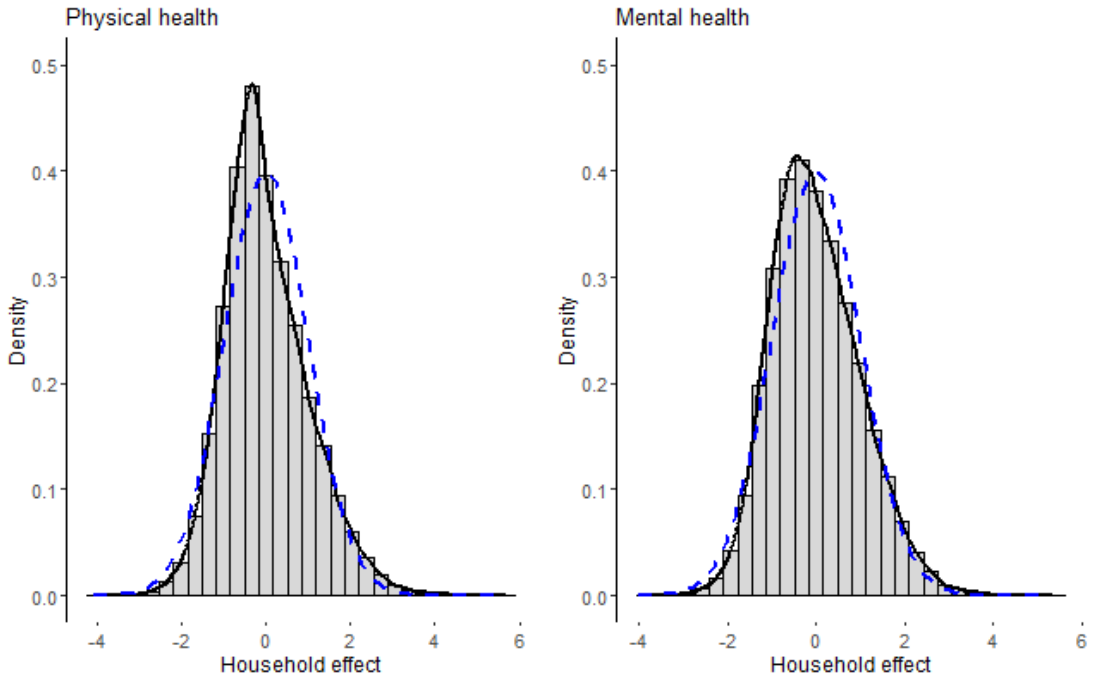

Figure S3: Histograms of standardised empirical Bayes estimates of the household random effects. The normal distribution is given by the blue dashed curve.

Table S5: Results from multilevel models of standardised self-rated **physical** health functioning at Wave 1 with household random effects, before and after adjusting for survey design weights and strata. Adjusted estimates account for household strata and weights at the household and individual levels using a combination of the `svy` and `meglm` commands in Stata. Higher scores indicate lower functioning and coefficients represent effects in standard deviation units.

	Unadjusted		Adjusted	
	Est.	(|Est./s.e.|)	Est.	(|Est./s.e.|)
Constant	-0.174	(3.789)	-0.134	(2.572)
Age[a]	0.118	(17.949)	0.097	(13.351)
Age squared	-0.003	(1.832)	-0.002	(1.107)
Age cubed	0.009	(13.490)	0.010	(13.882)
Female (vs. Male)	0.012	(1.512)	-0.0001	(0.017)
Ethnicity (vs. White)				
Asian/ Asian British	0.114	(7.860)	0.078	(5.048)
Black/ Black British	-0.076	(3.971)	-0.115	(5.629)
Other	-0.013	(0.531)	-0.039	(1.286)
Partnered (vs. Single)	-0.007	(0.658)	0.002	(0.155)
Number of children (vs. 1)				
None[b]	0.099	(4.781)	0.116	(5.978)
2[c]	-0.055	(3.540)	-0.072	(4.622)
3+[c]	-0.050	(2.577)	-0.086	(4.286)
Age of youngest child (vs. < 2 years)[d]				
2–4 years	0.016	(0.719)	0.015	(0.715)
5–10 years	0.034	(1.524)	0.038	(1.801)
11–16 years	0.103	(4.448)	0.108	(4.841)
>16 years	0.163	(6.960)	0.174	(7.284)
Highest education degree level (vs. Lower)	-0.162	(17.127)	-0.166	(16.823)
Employment status (vs. Employed)				
Unemployed	0.094	(5.257)	0.060	(2.884)
Economically inactive	0.432	(38.230)	0.439	(32.087)
Log annual household income (GBP)	-0.005	(1.296)	-0.011	(2.302)
Housing tenure (vs. Owner-occupier)				
Social rent	0.361	(29.192)	0.393	(26.229)
Private rent	0.083	(6.389)	0.087	(6.520)
Random effect variances				
Household	0.093	(14.578)	0.081	(10.385)
Individual	0.685	(93.781)	0.702	(76.273)

[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.

Table S6: Results from multilevel models of standardised self-rated **mental** health functioning at Wave 1 with household random effects, before and after adjusting for survey design weights and strata. Adjusted estimates account for household strata and weights at the household and individual levels using a combination of the `svy` and `meglm` commands in Stata. Higher scores indicate lower functioning and coefficients represent effects in standard deviation units.

	Unadjusted		Adjusted	
	Est.	(|Est./s.e.|)	Est.	(|Est./s.e.|)
Constant	0.056	(1.085)	0.093	(1.531)
Age[a]	-0.161	(22.209)	-0.172	(21.383)
Age squared	-0.048	(28.767)	-0.048	(24.475)
Age cubed	0.014	(18.750)	0.015	(17.181)
Female (vs. Male)	0.162	(18.841)	0.166	(18.438)
Ethnicity (vs. White)				
Asian/ Asian British	0.005	(0.317)	-0.013	(0.660)
Black/ Black British	-0.150	(6.979)	-0.154	(5.838)
Other	0.067	(2.422)	0.024	(0.716)
Partnered (vs. Single)	-0.203	(17.908)	-0.194	(15.153)
Number of children (vs. 1)				
None[b]	0.004	(0.192)	0.010	(0.420)
2[c]	-0.002	(0.140)	-0.012	(0.671)
3+[c]	0.026	(1.211)	0.016	(0.662)
Age of youngest child (vs. < 2 years)[d]				
2–4 years	-0.002	(0.081)	-0.007	(0.256)
5–10 years	-0.009	(0.354)	-0.029	(1.154)
11–16 years	0.053	(2.029)	0.056	(2.063)
>16 years	0.085	(3.227)	0.086	(3.078)
Highest education degree level (vs. Lower)	-0.069	(6.684)	-0.061	(5.894)
Employment status (vs. Employed)				
Unemployed	0.295	(15.202)	0.292	(11.964)
Economically inactive	0.241	(19.753)	0.245	(17.301)
Log annual household income (GBP)	-0.020	(4.313)	-0.025	(4.548)
Housing tenure (vs. Owner-occupier)				
Social rent	0.255	(18.080)	0.272	(15.499)
Private rent	0.099	(6.665)	0.128	(7.832)
Random effect variances				
Household	0.260	(33.269)	0.252	(26.840)
Individual	0.678	(91.676)	0.670	(72.086)

[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.

References

Ghosal, S., & Van der Vaart, A. W. (2017). *Fundamentals of nonparametric bayesian inference* (Vol. 44). Cambridge University Press.

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

Van der Vaart, A. W. (2000). *Asymptotic statistics* (Vol. 3). Cambridge university press.