Supplementary Material

Please refer to website https://github.com/MarcelaCespedes/Brain_wombling for R code (including code for plots) used in analysis outlined in the manuscript *A Bayesian hierarchical approach to jointly model structural biomarkers and covariance networks.*

1 Simulation study

The simulation study described in Section 3.1 of the manuscript described a thorough assessment of the Bayesian brain wombling algorithm. The four scenarios in the simulation study were; contiguous balanced (each person had an equal number of replicates) and unbalanced (the number of replicated varied per person), and a structured balanced and unbalanced designs. The results for fixed effect parameters $\beta$ and residual variance $\sigma^2$ are shown in Table ???. While the results for the structured configuration show a slightly lower recovery of fixed effect parameters, they do no represent a potential biological configuration. Hence performance of the wombling algorithm was better assessed on the contiguous configuration, whose performance of the recovery of the parameters is approximately 95%.

As mentioned in Section 3.2, spatial scale variance $\sigma_s^2$ is a biased estimate and was not recovered in our simulation study.

|          | Contig Unbal | Contig Bal | Struc Unbal | Struc Bal |
|----------|--------------|------------|-------------|-----------|
| $\beta_0$ | 50           | 49         | 46          | 46        |
| $\beta_1$ | 50           | 50         | 46          | 45        |
| $\sigma^2$ | 47           | 49         | 47          | 48        |

Table S1: Fixed effect and residual variance recovery for four scenarios, each with 50 simulations.

Figure ?? shows the histograms on the percentage of the recovered random effects for each scenario. As the simulation study comprised of 50 independently simulated data sets for each scenario, each data set consisted of $I = 100$ simulated participants, each with $K = 35$ ROI resulting in 3,500 random effects per simulated data set to estimate. The recovery of the random effects are summarised in Figure ??, whose histograms are summarised over the 50 simulated data sets. Overall we can see that there is approximately 95% recovery of the random effects for each scenario.

As $\rho$ in Model (1) is a fixed value, we investigated the effect recovering the parameters in the structured scenario for $\rho$ values [0.85, 0.9, 0.95, 0.99]. The table below shows the parameter values for this simulation study.

|          | 0.85       | 0.9        | 0.95       | 0.99       |
|----------|------------|------------|------------|------------|
| $\beta_0$ | 3.1 (2.9, 3.2) | 3 (2.8, 3.2) | 3.0 (3.2, 3.5) | 2.8 (2.2, 3.4) |
| $\beta_1$ | 0.3 (0.1, 0.5) | 0.3 (-0.01, 0.5) | 0.1 (-0.3, 0.5) | 0.4 (-0.3, 0.4) |
| $\sigma^2$ | 0.5 (0.5, 0.7) | 0.5 (0.5, 0.5) | 0.5 (0.5, 0.5) | 0.6 (0.6, 0.7) |

Table S2: Parameter values for $\rho$ set to 0.85, 0.9, 0.95, 0.99 values.
Figure S1: Each scenario (structured balanced and unbalanced, contiguous balanced and unbalanced) had 100 simulated participants and each participant had 35 ROI (3,500 random effects in total for each simulation). As each scenario comprised of 50 simulations, there are \( 50 \times 3,500 \) random effects to assess. Each histogram denoted the percentage of the number of random effects recovered (that is random effects whose solution within the 95% credible interval).

2 Posterior diagnostic checks for AIBL data set

Posterior predictive plots for each AIBL group analysed were used to assess goodness-of-fit for each wombling model. The plots in Figures ??, ?? and ?? show the expected mean of the data was recovered well, however there is a slight overestimation of the variance, as the proportion of predicted values inside the 95% credible intervals.
was slightly over 0.95. However these results show our models adequately estimated the uncertainty in the data.

Contact author for trace, density and additional diagnostics plots. Table ?? shows the Gelman-Rubin diagnostic, upper 95% confidence interval for convergence checks of the four chains for $\beta_0$, $\beta_1$, $\sigma^2$ and $\sigma^2_s$.

| Gelman-Rubin diagnostic | $\beta_0$ | $\beta_1$ | $\sigma^2$ | $\sigma^2_s$ |
|--------------------------|-----------|-----------|------------|--------------|
| HC                       | 1.15      | 1.16      | 1          | 1            |
| AD                       | 1.09      | 1.16      | 1          | 1            |
| Age A                    | 1.14      | 1         | 1          | 1            |
| Age B                    | 1.04      | 1.06      | 1          | 1.01         |
| Age C                    | 1.07      | 1         | 1          | 1            |
| APOE carrier             | 1.02      | 1.01      | 1          | 1            |
| APOE non-carrier         | 1.03      | 1.02      | 1          | 1            |

Table S3: Gelman-Rubin diagnostic upper confidence limit values for each group in AIBL study. As the combinations of four chains for each group had values close to one, we are confident the MCMC algorithm for each group has reached convergence.

Figure S2: Posterior predictive plots for healthy control (HC, left) and Alzheimer’s disease (AD, right) models presented in the manuscript. The proportion of response values inside the predictive 95% credible intervals (in red) is 0.981 and 0.979 for HC and AD models respectively.
Figure S3: Posterior predictive plots for age groups; A (59-69), B (69-79) and C (79 - 93). The proportion of response values inside the predictive 95% credible intervals (in red) are 0.986, 0.982 and 0.981 for age groups A, B and C.
Figure S4: Posterior predictive plots for APOE ε4 non-carriers (negative, left) and carriers (positive, right). The proportion of response values inside the predictive 95% credible intervals (in red) are 0.979 for both APOE groups.

3 Wombling cortical thickness estimates at the ROI level

As discussed in Section 4.2.1 of the manuscript, we investigated an adaptation to the wombling model to account for ROI means via fixed effect parameters. The extended model is of the form

\[ y_{irk} | b_{ik}, \beta, \sigma^2 \sim N(\beta_0 + \beta_1 R_2 + \beta_2 R_3 + \ldots + \beta_{34} R_{35} + b_{ik}, \sigma^2) \]

\[ b_i \sim MVN(0, \sigma^2 Q) \]

\[ Q^{-1} = \rho(D_w - W) + (1 - \rho)I. \]  \hspace{1cm} (1)

Where the response \( y_{irk} \), spatial random effects \( b_{ik} \), residual \( \sigma^2 \) and spatial scale variance \( \sigma_s^2 \) terms are the same as those presented in Section 2.2 of the manuscript. The precentral gyrus is the baseline ROI whose cortical thickness (in mm) is estimated by \( \beta_0 \). The fixed effect parameter \( \beta_{k-1} \) estimates the deviation of ROI \( k \) away from \( \beta_0 \) when the binary indicator variable \( R_k \) is equal to one. Estimation of \( \beta \) is attained by the same conditional distribution described in Section 2.2 of the manuscript, with minor modifications to account for the design matrix \( R \) rather than \( X \). Figure ?? shows the posterior means of \( W \) for HC (top), MCI (middle) and AD (bottom) groups. While the posterior mean for the HC group is similar that in Figure 4 of the manuscript, with the same 36 links present in both networks and 468 absent connections in common, the matrices for MCI and AD group show the probability of each link is close to 0.5. We believe that the reason for this is because the HC group has a substantially larger sample size (120 individuals) compared to the MCI and AD groups (with 21 and 26 individuals respectively). Hence, the more complex model in Expression (??) requires data with larger sample sizes, compared to the original wombling model, in order to derive meaningful \( W \) estimates.
Figure S5: Left column: Posterior means for $W$ for HC (top), MCI (middle) and AD (bottom). Right column: Binarised matrices at posterior probability cut-off values of 0.7 for HC and 0.6 for MCI and AD.
Figure ?? shows the marginal posterior densities for the ROI means for 35 regions. These results resemble the independent Bayesian LME ROI estimates in Figure ??, particularly for ROIs associated with early onset of AD such as the inferior, middle and superior temporal gyrus, posterior cingulate gyrus.

Figure S6: Marginal posterior distributions for 35 ROI means, for HC (green), MCI (blue) and AD (red) groups.
4  Wombling cortical thickness estimates at the participant level

As described in Sections 2.4.4 and discussed in Section 3.2.4, the wombling model derived participant specific estimates on all ROIs. Figures ?? and ?? shows the posterior means and 95% credible intervals (as error bars) for each participant.

Figure S7: Participant specific cortical thickness estimates for all 35 ROI for HC and AD group comparison, derived by the wombling algorithm.
Figure S8: Participant specific cortical thickness estimates for all 35 ROI for age groups A, B and C group comparison, derived by the wombling algorithm.
5 APOE wombling results

Carriers of the Apolipoprotein (APOE) ε4 gene have been associated to be at higher risk of developing AD, hence in neuroimaging studies, it is a key biomarker to investigate. We applied the wombling model on AIBL data divided into APOE ε4 carrier and non-carrier groups. Figures ??, ?? and ?? show the cortical networks, global estimates across all ROI and participant specific rankings for key AD regions as described in Section 3.2.3 of the manuscript.

There were no strong differences APOE ε4 carrier and non-carrier groups in any of the ROI. We believe the reason for this, is because the carrier and non-carrier groups consist of participants across the entire diagnosis spectrum (HC, MCI and AD), age groups and many other AD related biomarkers, making it difficult to assess the deterioration differences due to APOE ε4 gene. Unfortunately due to our low sample size, we did not have sufficient data to investigate more meaningful biomarker groups such as APOE ε4 carrier and non-carrier groups that were clinically diagnosed as HC or AD.
Figure S9: Left: Posterior mean of $W$ heat map for APOE $\varepsilon 4$ carriers (top) and non-carriers (bottom). Right: Cortical network from binarised heatmap with threshold $\tau = 0.7$ for the respective APOE $\varepsilon 4$ carrier groups. Node size reflects the number of edges on each vertex. Total number of edges for each network (top and bottom) are 152 and 150 for APOE $\varepsilon 4$ non-carriers and carriers groups.
6 Bayesian linear mixed effect models on each ROI

As described in Section 2.3 and discussed in Results Sections 4.1, 4.2 and 4.3, Bayesian linear mixed effect models were independently applied to each ROI on groups; diagnosis levels HC, MCI and AD, age groups A, B and C and APOE ε4 allele carriers and non-carriers. All models were of the form

\[ y_{ij} | \sigma^2, \mu_{0i} \sim N(\mu_{0i} + \beta_1 x_i, \sigma^2) \]
\[ \mu_{0i} \sim N(\mu_0, \sigma^2_0). \]  

In order to make the models comparable with the wombling approach, covariate \( x_i \) is gender as described in Section 4.2 in the manuscript, with \( x_i = 1 \) for male and 0 otherwise. The residual variance prior for \( \sigma^2 \) and the random effects prior, \( \sigma^2_0 \), is the same as discussed in Section 3.1 of the manuscript. Similarly, the prior for the intercept effect \( \mu_0 \) is also relatively vague with a \( N(0, 10) \) distribution.
Figure S11: Posterior mean and 95% credible interval on participants estimated cortical thickness (random effects $b_i$), for nine key regions associated with AD cortical signature. Colour coded as follows; (red) APOE $\varepsilon 4$ carriers and (green) APOE $\varepsilon 4$ non-carriers.

Figures ?? to ?? show the marginal posterior mean population distributions and participants ranked according to posterior means with 95% credible interval.
7 WAIC results

As described in Section 2.4.4 of the manuscript, we applied the WAIC criterion on the wombled and independent Bayesian LME models to assess model choice. Table ?? shows the results of the WAIC for the wombling model applied to each group, and the combined WAIC criterion for the independent Bayesian LME analyses for each region.

| Group | WAIC wombled model | WAIC LME models |
|-------|--------------------|-----------------|
| HC    | -33 598.26         | -12 424.94      |
| MCI   | -4 355.94          | -1 589.94       |
| AD    | -3 045.20          | -689.97         |
| Age A | -6 029.15          | -2 870.38       |
| Age B | -13 065.38         | -6 051.11       |
| Age C | -10 321.63         | -4 587.98       |

Table S4: WAIC values for diagnosis groups. Smaller WAIC values denotes a more parsimonious model compared to the alternative, here the wombled model is preferred to the independent Bayesian LME models.
Figure S13: Marginal posterior distributions of total cortical thickness estimates for each ROI. AD, MCI and HC diagnosis are denoted by red, yellow and green posterior densities.

8 Pearson correlation networks for each group

Cortical networks derived by Pearson’s pairwise correlation networks for each group are shown in Figures ?? to ?? . As Pearson’s pairwise networks does not accommodate the repeated measure structure of the data, we derived networks at both baseline (independent and identically distributed (IID) observations) as well as on the whole data, with repeated measures treated as IID.
Figure S14: Participant specific estimates of each ROI ranked according to posterior mean with 90% credible interval error bar. Colour coded according to the diagnosis levels HC, MCI and AD (green, yellow and red respectively) derived from the Bayesian LME models.
Figure S15: Marginal posterior distributions of total cortical thickness estimates for each ROI. Age groups A, B and C denoted by green, yellow and green posterior densities derived from the Bayesian LME models.
Figure S16: Participant specific estimates of each ROI ranked according to posterior mean with 90% credible interval error bar. Colour coded according to age groups A, B and C (colours green, yellow and red respectively) derived from the Bayesian LME models.
Figure S17: Marginal posterior distributions of total cortical thickness estimates for each ROI. APOE ε4 carrier (red) and non-carrier (green) groups.
Figure S18: Participant specific estimates of each ROI ranked according to posterior mean with 90% credible interval error bar. Colour coded according to APOE ε4 carrier (red) and non-carrier (green) groups derived from the Bayesian LME models.
Figure S19: Pearson pairwise correlation plots for baseline (left top and bottom) and repeated measures (right top and bottom) on HC diagnosis. Top: networks binarised according to threshold of $\tau = 0.7$ applied on the absolute value of each element on matrices below.
Figure S20: Pearson pairwise correlation plots for baseline (left top and bottom) and repeated measures (right top and bottom) on MCI diagnosis. Top: networks binarised according to threshold of $\tau = 0.7$ applied on the absolute value of each element on matrices below.
Figure S21: Pearson pairwise correlation plots for baseline (left top and bottom) and repeated measures (right top and bottom) on AD diagnosis. Top: networks binarised according to threshold of $\tau = 0.7$ applied on the absolute value of each element on matrices below.
Figure S22: Pearson pairwise correlation plots for baseline (left top and bottom) and repeated measures (right top and bottom) on age group A. Top: networks binarised according to threshold of $\tau = 0.7$ applied on the absolute value of each element on matrices below.
Figure S23: Pearson pairwise correlation plots for baseline (left top and bottom) and repeated measures (right top and bottom) on age group B. Top: networks binarised according to threshold of $\tau = 0.7$ applied on the absolute value of each element on matrices below.
Figure S24: Pearson pairwise correlation plots for baseline (left top and bottom) and repeated measures (right top and bottom) on age group C. Top: networks binarised according to threshold of $\tau = 0.7$ applied on the absolute value of each element on matrices below.
Figure S25: Pearson pairwise correlation plots for baseline (left top and bottom) and repeated measures (right top and bottom) on APOE ε4 carriers. Top: networks binarised according to threshold of $\tau = 0.7$ applied on the absolute value of each element on matrices below.
Figure S26: Pearson pairwise correlation plots for baseline (left top and bottom) and repeated measures (right top and bottom) on APOE ε4 non-carriers. Top: networks binarised according to threshold of $\tau = 0.7$ applied on the absolute value of each element on matrices below.