

Integrating wastewater and randomised prevalence survey data for national COVID surveillance

Guangquan Li, Peter Diggle and Marta Blangiardo

Supplementary Material

Contents

A Different nowcast models	1
B The directed acyclic graph for the data integration model	3
C Comparing the interval widths between the full model and the simplified model	4

A Different nowcast models

Table 1 summarises the key specifications of the different models for nowcasting local prevalence considered in the paper. Models 1-3 present the different versions of the data integration framework where Model 1 allows the wastewater-prevalence relationship to vary over both space and time while Model 2, a simplified version of Model 1, assumes the relation to vary only spatially but not temporally. Model 3 extends Model 2 by incorporating two LTLA-level covariates, the English Index of Multiple Deprivation (IMD) and the percentage of Black, Asian, and minority ethnic (BAME) people in the population. Model 4 excludes the use of the wastewater viral concentration estimates and thus its nowcast of local prevalence is a function of the national estimates $\alpha + B_t$ plus the estimated local variation U_i . Model 5 produces local prevalence nowcast using the wastewater viral concentrations only, with no anchoring from the national level prevalence.

In the Results Section in the main paper, we considered two scenarios where the spatially coarse prevalence estimates were available at the regional level. For those scenarios, we fitted Model 2 to each of the 9 English regions separately.

Table 1: Key specifications of the different models for nowcasting local disease prevalence

Model	The LTLA-level sub-model with the likelihood $p_{it} \sim \text{Normal}(\mu_{it}, \sigma_{it}^2)$	The national-level sub-model with the likelihood $P_t \sim \text{Normal}(\mu_t, \sigma_t^2)$
1. The full model	$\mu_{it} = (\alpha + U_i) + (B_t + V_{it}) + (d + D_t + M_i) \cdot \tilde{w}_{it}$	$\mu_t = \alpha + B_t + (d + D_t) \cdot \tilde{W}_t$
2. The simplified model	$\mu_{it} = (\alpha + U_i) + (B_t + V_{it}) + (d + M_i) \cdot \tilde{w}_{it}$	$\mu_t = \alpha + B_t + d \cdot \tilde{W}_t$
3. The simplified model + covariates (IMD and BAME) ^a	$\mu_{it} = (\alpha + U_i + \beta_1 \cdot \text{IMD}_i + \beta_2 \cdot \text{BAME}_i) + (B_t + V_{it}) + (d + M_i) \cdot \tilde{w}_{it}$	$\mu_t = \alpha + B_t + d \cdot \tilde{W}_t$
4. Without wastewater	$\mu_{it} = (\alpha + U_i) + (B_t + V_{it})$	$\mu_t = \alpha + B_t$
5. Without national prevalence	$\mu_{it} = (\alpha + U_i) + (B_t + V_{it}) + (d + M_i) \cdot \tilde{w}_{it}$	No national sub-model in this version

^aA weakly informative prior $\text{Normal}(0, 10^6)$ was independently assigned to β_1 and β_2 .

B The directed acyclic graph for the data integration model

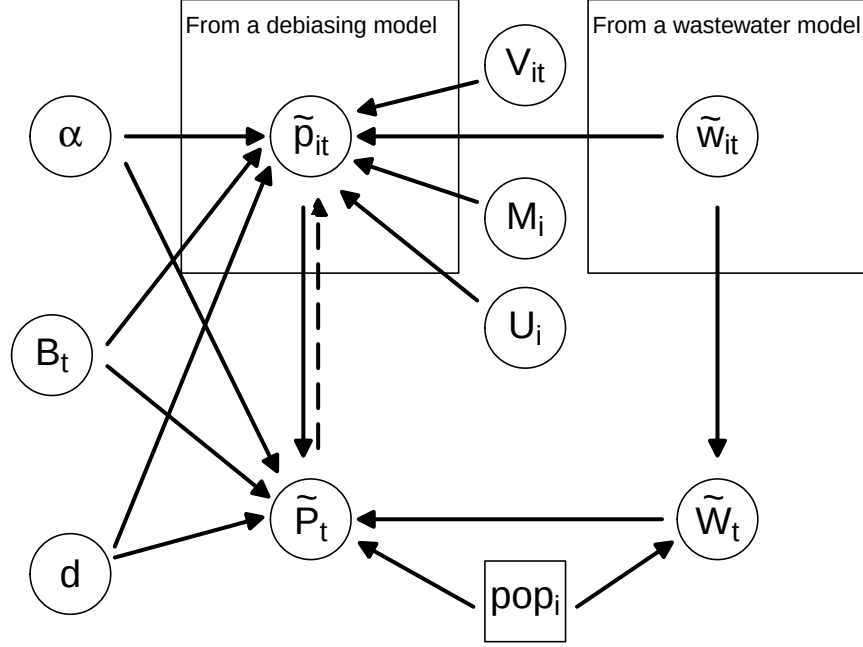


Figure 1: The directed acyclic graph for the data integration model with spatially varying relation between wastewater viral concentration and prevalence, i.e. the simplified model specified in the Methods Section in the main paper. The dashed arrow going from $\tilde{P}_t$ to $\tilde{p}_{it}$ emphasises the contribution of the national prevalence estimates when nowcasting local prevalence.

C Comparing the interval widths between the full model and the simplified model

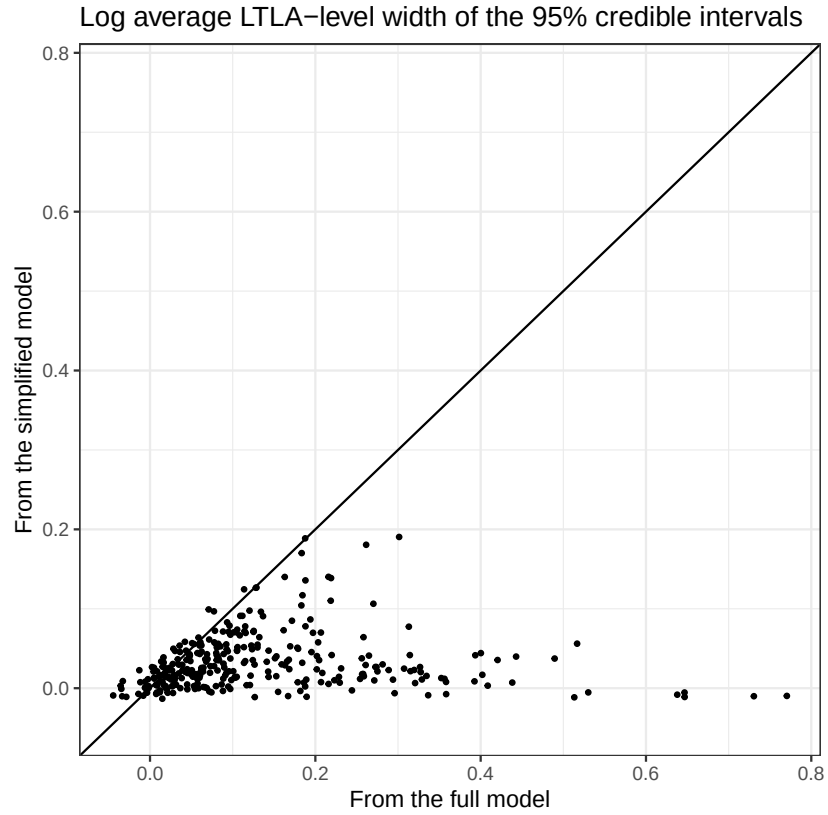


Figure 2: Comparing the log-transformed width of the 95% credible intervals from the simplified model against those from the full model. The interval width of each LTLA is averaged over the 20 week of validation period. There are 261 (out of 307) LTLAs where the average interval lengths from the simplified model are shorter than those from the full model. For 46 LTLAs, the interval lengths from the simplified model are longer but the differences in interval width between the two models are almost negligible.