Investigating the UK measles outbreak between 1944 to 1962 using the hh4 model
Kajal Dodhia\textsuperscript{1}, Jordan J Hood\textsuperscript{2}

University of Glasgow\textsuperscript{1}, STOR-i, Lancaster University\textsuperscript{2}

Background

- SARS-CoV-2 (Covid-19) virus has had a significant impact on our lives since January 2020. Vaccines do not provide complete immunity, hence SARS-CoV-2 will remain a part of our lives until we reach a resolution.
- World events and societies changing behaviours greatly influence what global infections look like. However, such a scenario can be determined from other infectious disease such as Influenza which exhibits local endemic and epidemic phases.
- The pre-vaccine UK measles data set from 1944-62 is analysed using the hh4 model, starting with a basic model, adding covariates and taking into account random effects in order to explore the best fitting model and predicting future outbreaks.

Basic Model

An endemic-epidemic multivariate time-series model for infectious disease counts $Y_{it}$ from units $i = 1, \ldots, 60$ during periods $t = 1, \ldots, 493$ where $i$ denotes city and $t$ denotes fortnightly time. The hh4 model assumes that, $Y_{it} | F_{i} \sim NB(\mu_{it}, \psi_{i})$ where
\[
\mu_{it} = \exp(\bar{\alpha}_{i}^{(v)} + \beta_{i} \log(E_{it}) + \gamma \sin(\omega t) + \delta \cos(\omega t)) \nonumber
\]

and overdispersion parameter $\psi_{i} > 0$ such that the conditional variance of $Y_{it} | F_i$ is $\mu_{it}(1 + \psi_{i}\mu_{it})$. The link function is $\log(\mu_{i})$.

Final Model

The final model with independent random effects in all three components:

$\alpha_{i}^{(v)} \sim N(\bar{\alpha}_{i}^{(v)}, \sigma_{\alpha}^{2})$, $\alpha_{i}^{(\lambda)} \sim N(\bar{\alpha}_{i}^{(\lambda)}, \sigma_{\alpha}^{2})$ and $\alpha_{i}^{(\phi)} \sim N(\bar{\alpha}_{i}^{(\phi)}, \sigma_{\alpha}^{2})$,

$\mu_{it} = \exp(\nu_{i}^{(v)} + \lambda Y_{it-1} + \phi \sum_{j \neq i} w_{ij} Y_{j, t-1})$,

$\log(v_{i}) = \alpha^{(v)} + \beta_{i} \log(E_{it}) + \gamma \sin(\omega t) + \delta \cos(\omega t)$

Model parameters explained below:

1. Endemic log-linear predictor $\nu_{i}$:
   - Temporal variation of disease incidence incorporates an overall trend and a sinusoidal wave of frequency $\omega = \frac{2\pi}{60}$.
   - Population fraction as multiplicative offset $\nu$.

2. Epidemic Component:
   - Autoregressive parameter: $\lambda = \exp(\alpha^{(\lambda)})$.
   - Spatio-temporal parameter: $\phi = \exp(\alpha^{(\phi)})$.
   - These are assumed homogeneous across cities and constant over time and in this model the epidemic can only arrive from directly adjacent districts.
   - The Negative Binomial which accounts for overdispersion is a better fit compared to the Poisson model as it has a lower Akaike information criterion (AIC) value of 218770.4.

Adding Covariates Step by Step

1. Scaled infections against average employment domain score ($E$) and average domain income score ($I$) had a positive correlation $0.5854129$ and $0.4786521$ respectively. By performing AIC-based model selection, the lowest AIC for both employment and income shows they can be added to the endemic predictor as a covariate.

2. Random effects are added to the model as cities display heterogeneous incidence levels not explained by observed covariates. This could be caused by under reporting (GLMM).

Conclusions

The final model which incorporates the covariates for average of employment and income domain score and random effects is the best fit for the measles data set. However, in predicting, it does not capture the large number of cases for the year 1962. The largest portion of the fitted mean results from within-city autoregressive component, a very small spatio-temporal and almost negligible endemic component to the data.

References


Figure 1: Fitted components in the initial model for the cities with more than 80,000 total infections. Data are drawn for positive weekly counts. The basic model shows that the largest proportion of the fitted mean results from the within-city autoregressive component to the data and captures seasonality.

Figure 2: Fitted components in the final model for Birmingham.

Figure 3: Simulation-based forecast of 1962 starting from the second last week in 1961 (vertical bar on the left), showing the counts aggregated over all districts. The fortnightly mean of the fitted components in Figure 1 but there is a slight increase in the proportion of fitted mean captured by the spatio-temporal component for Birmingham, Figure 2. The lack of transmission between cities could be due to less ways and needs of travelling between cities, hence less physical contact to pass the disease amongst cities compared to today.