1 Pseudo-code for hierarchical selection of data sources for influenza forecasting

**S1 Algorithm.** Hierarchical data source selection.

1. **Given** : Set of predictor data sources \( D \) and forecasting target \( G \)
2. \( D_{\text{chosen}} : \{D_1, D_2, D_3, \ldots, D_{n-1}\} \)
3. \( D_{\text{remaining}} : D \setminus D_{\text{chosen}} \)
4. To choose \( D_n \),
5. for \( q \in D_{\text{remaining}} \) do
6. Fit \( \Omega_q : G \sim D_1 + D_2 + \cdots + D_{n-1} + q \) on the full historical time series.
7. \( S \) : historical seasons of \( G \)
8. for \( s_i \in S \) do
9. for \( d \in D_{\text{chosen}} \cup q \) do
10. Make Bayes’ forecast of \( s_i \) for \( d \) from week \( w \), with \( (S \setminus s_i) \) as priors.
11. Apply \( \Omega_q \) to forecast \( G \) for season \( s_i \) from forecasts generated in Step 10.
12. Score forecast from Step 11 by comparison to actual \( G \) (call this score \( \sigma(q, i) \)).
13. Calculate the grand score \( \sum_{s_i} \sigma(q, i) \).
14. Choose \( q \) that maximizes the grand score.
15. Set \( D_n = q \).