

1 Socio-environmental and 2 measurement factors drive spatial 3 variation in influenza-like illness

4 Elizabeth C. Lee^{1*}, Ali Arab², Sandra Goldlust¹, Cécile Viboud³, Shweta Bansal^{1,3*}

***For correspondence:**

ecl48@georgetown.edu (ECL);
shweta.bansal@georgetown.edu (SB)

5 ¹Department of Biology, Georgetown University, Washington, DC, USA;

6 ²Department of Mathematics & Statistics, Georgetown University, Washington, DC,
7 USA; ³Fogarty International Center, National Institutes of Health, Bethesda, MD, USA

8

9 **Abstract** The mechanisms hypothesized to drive spatial heterogeneity in reported
10 influenza activity include: environmental factors, contact patterns, population age structure,
11 and socioeconomic factors linked to healthcare access and quality of life. Harnessing the
12 large volume and high specificity of diagnosis codes in medical claims data for influenza
13 seasons from 2002-2009, we estimate the importance of socio-environmental determinants
14 and measurement-related factors on observed variation in influenza-like illness (ILI) across
15 United States counties. We found that South Atlantic states tended to have higher ILI
16 seasonal intensity, and a combination of transmission, environmental, influenza subtype,
17 socioeconomic and measurement factors explained the variation in seasonal intensity
18 across our study period. Moreover, our models suggest that sentinel surveillance systems
19 should have fixed report locations across years for the most robust inference and prediction,
20 and high volumes of data can offset measurement biases in opportunistic data samples.

22 Introduction

23 Seasonal influenza represents an important public health burden worldwide, and even within
24 a single year, there is substantial variation in disease burden across populations (*Moorthy et al.,*
25 *2012; Lee et al., 2015*). Many studies have examined the drivers and patterns influenza season-
26 ality (*Lofgren et al., 2007; Tamerius et al., 2011*), while others have focused on the large-scale
27 spatial patterns in influenza epidemic timing, suggesting for instance, spread from West to
28 East across North America due to a combination of local contact patterns and global travel
29 patterns (*Wenger and Naumova, 2010; Schanzer et al., 2011b; Grais et al., 2003; Brownstein*
30 *et al., 2006*). While there are numerous studies explaining spatial variation in seasonal in-
31 fluenza transmission and disease burden, most studies focus on very aggregated or very local
32 study areas (e.g., country-level or one school district, respectively) compare only one or two
33 hypotheses in isolation.

34 Among these, humidity and temperature have each been associated with seasonal flu
35 onset, seasonal fluctuations, and heightened morbidity and mortality in epidemiological con-
36 texts (*Shaman et al., 2010; Yu et al., 2013; Barreca and Shimshack, 2012; Deyle et al., 2016*),
37 and lower humidity and colder temperatures may increase influenza virus transmission and
38 survival (*Lowen et al., 2007; Shaman and Kohn, 2009*). Chronic illnesses such as asthma, ex-
39 acerbated by air pollution, elevated the risk for severe symptoms of pandemic H1N1 (*Van*
40 *Kerkhove et al., 2011*). Empirical evidence supports the occurrence of both aerosol and droplet

41 transmission of influenza virus (*Killingley and Nguyen-Van-Tam, 2013*), and these transmission
42 modes suggest that influenza seasons may follow both density-dependent and frequency-
43 dependent disease dynamics (per capita contact rates between susceptible and infectious
44 individuals do and do not change with population density, respectively). The high connectiv-
45 ity of school-aged children in contact surveys (*Mossong et al., 2008; Kucharski et al., 2014*) has
46 led to hypotheses that children drive local transmission and adults seed new infections across
47 longer distances (*Viboud et al., 2006; Apolloni et al., 2013*), which may manifest in shifted epi-
48 demic timings across age groups (*Lemaitre and Carrat, 2010; Peters et al., 2014; Schanzer et al.,*
49 *2010; Wallinga et al., 2006; Timpka et al., 2012*). Immune landscapes vary across locations; epi-
50 demic outcomes in one season may trickle down to subsequent years through differences
51 in cross-protective immunity, and high flu vaccination coverage may reduce morbidity and
52 incidence of severe clinical outcomes (*Kostova et al., 2013*). Finally, flu type and subtype cir-
53 culation may also drive spatial heterogeneity; A/H3-dominant flu seasons are associated with
54 greater morbidity and mortality and an older patient age distribution than A/H1 season (*Frank*
55 *et al., 1985; Simonsen et al., 1997; Khiabani et al., 2009; Peters et al., 2014*), while influenza B
56 is thought to circulate predominantly and earlier among children (*Peters et al., 2014; Hayward*
57 *et al., 2014; Beauté et al., 2015*).

58 Beyond socio-environmental mechanisms, we must consider the possibility that the mea-
59 surement of influenza disease burden plays a significant role in driving the observed spatial
60 heterogeneity. While poverty and other social determinants are thought to increase risk for
61 influenza morbidity, hospitalization, and mortality (*Lowcock et al., 2012; Kumar et al., 2015;*
62 *Hadler et al., 2016; Charland et al., 2011; Grantz et al., 2016*), these observations are often con-
63 founded by care-seeking behavior, the likelihood that sick individuals will seek treatment from
64 a health care provider. Roughly 43% of adults and 60% of elderly seek care for influenza-like
65 illness (ILI) in the United States, as many cases are too mild to warrant a visit to the doctor (*Big-*
66 *gerstaff et al., 2014b*). In addition to differences in personal choice, limited access to health
67 care and health insurance also delay or reduce care-seeking behavior, further generating bi-
68 ases in reported case severity or patient numbers among physician-based surveillance sys-
69 tems (*Biggerstaff et al., 2014b*).

70 In this study, we examine the transmission, environmental, influenza-specific, and socioe-
71 conomic mechanisms and measurement processes underlying the spatial variation in reported
72 influenza-like illness across counties in the United States. Leveraging highly resolved medical
73 claims data, we identified important drivers of spatial heterogeneity in the magnitude and
74 duration of flu seasons from 2002 to 2009 in a large-scale ecological analysis. We then used
75 our Bayesian modeling framework in new applications to probe the robustness of this ecolog-
76 ical inference with limited data availability and to assess the predictive ability of our model in
77 a more recent flu season. Our results highlight the relative contributions of surveillance data
78 collection and socio-environmental processes to disease reporting, and highlight the impor-
79 tance of considering measurement biases when using surveillance data for epidemiological
80 inference and prediction.

81 **Results**

82 We examined the socio-environmental and measurement-related drivers of spatial hetero-
83 geneity in influenza disease burden across U.S. counties for flu seasons from 2002-2003 through
84 2008-2009 using a hierarchical Bayesian modeling approach. Using medical claims data rep-
85 resenting 2.5 billion visits from upwards of 120,000 health care providers each year, our study
86 considered six disease burden response variables: two measures of influenza disease burden
87 (relative risk of seasonal intensity, which is a proxy for attack rate, and epidemic duration in
88 number of weeks) in three populations (total population, children 5-19 years old, and adults
89 20-69 years old) with multi-season and single season model structures. There were 13 county-
90 level, 2 state-level and 4 HHS region-level predictors in the final model *Table 1*; all predictors

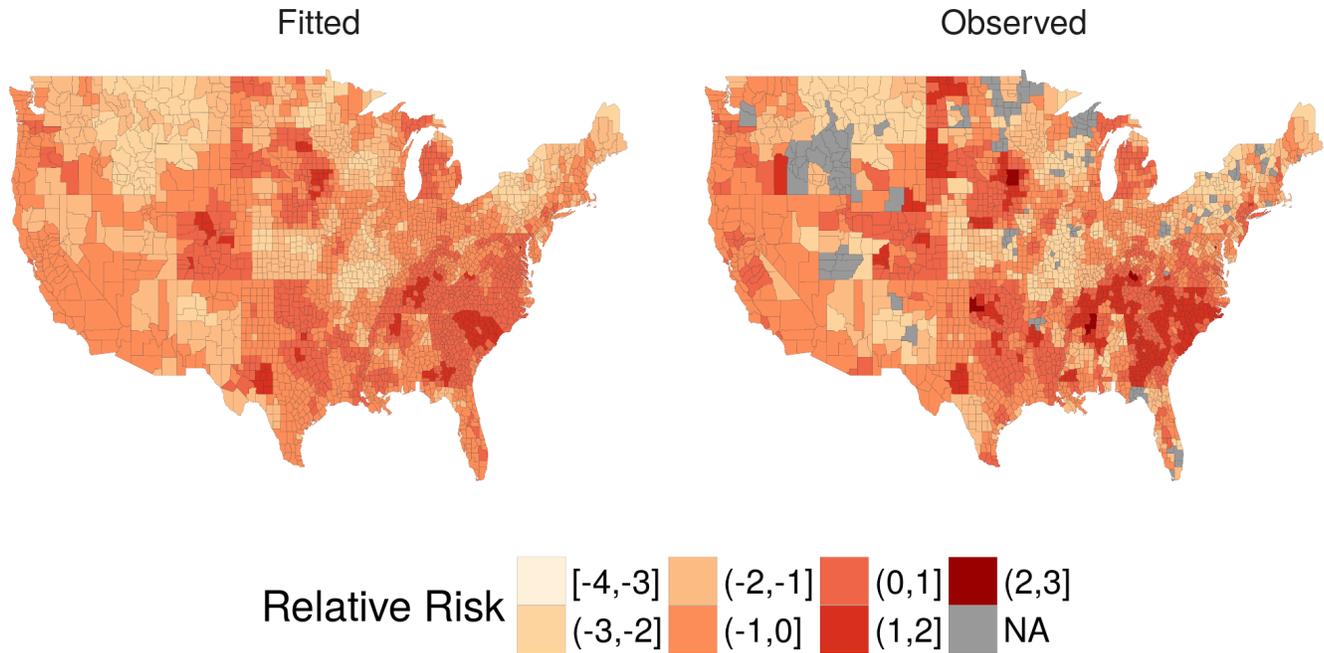


Figure 1. Continental U.S. county map for fitted and observed relative risk of seasonal intensity for an example flu season (2006-2007).

Figure 1-Figure supplement 1. Continental U.S. county maps for fitted (left) and observed (right) relative risk of seasonal intensity for remaining influenza seasons.

91 were the same across response variables except care-seeking behavior, which was specific to
92 the age group in the response. The seasonal intensity model fit the data well and the Pear-
93 son's cross-correlation coefficient between the log seasonal intensity and log prediction was
94 $R = 0.87$ (Figure 1). Results reported in the following sections are from the multi-season total
95 population seasonal intensity model unless otherwise noted.

96 **Temporal and spatial patterns of influenza-like illness**

97 Group (random) effects were used to identify consistent spatial or temporal patterns across
98 locations and study years. We found that the 2004-2005 flu season had greater seasonal inten-
99 sity, while 2008-2009 had relatively low seasonal intensity (Figure 2). For the seasonal intensity
100 model, no single region had a significant group effect, although several South Atlantic states
101 like Georgia, Maryland, North Carolina, South Carolina, and Virginia had relatively greater risk
102 than other states across the study period, while several Plains and Rocky Mountain states like
103 Kansas, Minnesota, Missouri, Montana, and Utah had relatively lower risk.

104 **Drivers of seasonal intensity**

105 Several socio-environmental drivers of seasonal intensity risk were identified in the multi-
106 season model (Figure 3). Total seasonal intensity had positive associations with the adult-flu
107 H3 and child-flu B interaction terms, estimated average household size, and a proxy for prior
108 immunity. There were negative associations with adult and child population proportions, aver-
109 age flu season specific humidity, proportion of the population in poverty, proportion of single
110 person households, and infant vaccination coverage.

111 We found that careseeking behavior and claims database coverage had strong positive as-
112 sociations with seasonal intensity (Figure 3). In considering the single-season models, the pos-
113 itive effect of claims database coverage on seasonal intensity appeared to decline in magni-

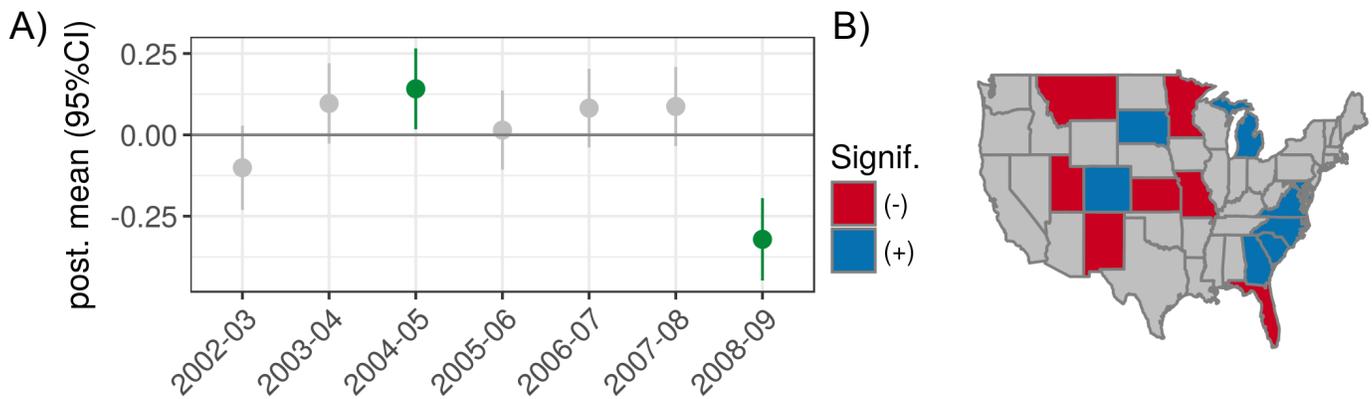


Figure 2. Temporal and spatial group effects for total population seasonal intensity. A) 95% credible intervals for group (random) effects by influenza season. B) Continental U.S. maps highlighting states with significantly greater or lower seasonal intensity across the study period.

114 tude over time (*Figure 3* supplement). This corresponded with an increase in claims database
115 coverage over time (*Appendix 5*).

116 **Drivers of age-specific seasonal intensity**

117 Children and adults comprise the largest components of the U.S. population, and many stud-
118 ies have considered shifts in epidemic timing and immunity due to differences in contact
119 patterns, shifting risk between children and adults over time, interactions between influenza
120 types/subtypes by age, and differences in vaccine effectiveness by age group (*Bansal et al.,*
121 *2010; Lee et al., 2015; Ewing et al., 2016; Schanzer et al., 2011a; Gostic et al., 2016; Khiabani-*
122 *et al., 2009*). Considering the potential to elucidate age-specific transmission mechanisms
123 and improve targeting of public health interventions, we used the multi-season model to
124 examine drivers of seasonal intensity in the child and adult populations. Full model results
125 are reported in *Appendix 2*, and for both age groups, predicted value means appeared to be
126 systematically over-estimated relative to the observed relative risk of seasonal intensity. The
127 Pearson's cross-correlation coefficient between the log observation and log predicted mean
128 was $R = 0.89$ and $R = 0.90$ for the child and adult seasonal intensity models, respectively.

129 Children had greater intensity in the 2003-2004 flu season and lower intensity in the 2002-
130 2003 and 2008-2009 flu seasons. Adults had greater intensity in the 2004-2005 flu season
131 and lower intensity in the 2008-2009 flu seasons. Similar to results for the total population,
132 several South Atlantic states had greater risk while Plains states had lower risk of seasonal
133 intensity for both children and adults.

134 Across the three age group responses (i.e., total, children, adults), child seasonal intensity
135 had a unique positive association with influenza B circulation and adult seasonal intensity
136 had a unique positive association with H3 circulation among influenza A and proportion of
137 the population in poverty. Also notable, both child and adult seasonal intensity had a negative
138 association with estimated average household size, while the total seasonal intensity model
139 had a positive effect.

140 **Drivers of epidemic duration**

141 We also considered the mechanisms associated with epidemic duration, a measure of in-
142 fluenza disease burden that captures the number of weeks with heightened ILI activity. Better
143 understanding of factors associated with longer epidemics might improve hospital prepared-
144 ness in surge capacity and staffing needs and aid local public health departments in planning

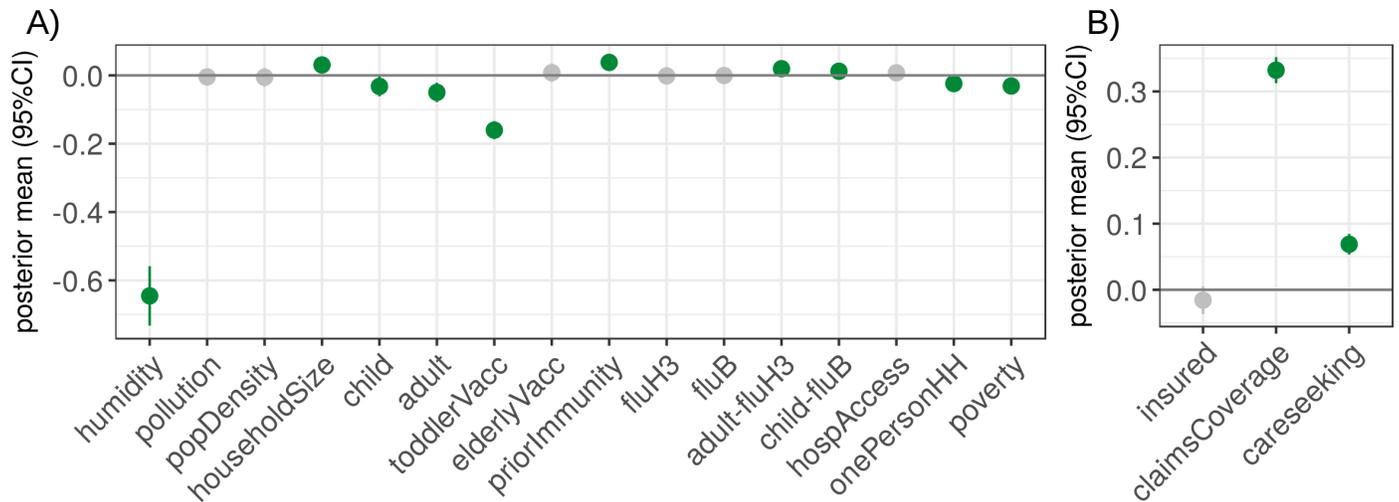


Figure 3. For the total population multi-season seasonal intensity models, these are the 95% credible intervals for the posterior distributions of the A) socio-environmental coefficients and B) measurement-related coefficients. Distributions indicated in green were statistically significant.

Figure 3-Figure supplement 1. For the total population single-season seasonal intensity models, these are the 95% credible intervals for the posterior distributions of the socio-environmental coefficients.

Figure 3-Figure supplement 2. For the total population single-season seasonal intensity models, these are the 95% credible intervals for the posterior distributions of the measurement coefficients.

145 their influenza information or vaccination campaigns. Full results for a multi-season model of
146 epidemic duration for the total population are reported in *Appendix 3*, but predicted value
147 means appeared to be systematically under-estimated relative to the observed epidemic du-
148 rations and the Pearson's cross-correlation coefficient between the observed and predicted
149 mean number of epidemic weeks was $R = 0.71$.

150 The 2004-2005 and 2007-2008 flu seasons had longer epidemics while the 2002-2003
151 and 2008-2009 seasons tended to have shorter epidemics. The Southeastern U.S. region (HHS
152 region 4) had longer epidemics than other regions, while only five states with no geographic
153 identity had significant group effects for the epidemic duration model. Epidemic duration
154 had positive associations with the interaction between adult population and influenza H3
155 circulation, influenza B circulation, estimated average household size, population density, a
156 proxy for prior immunity, and elderly vaccination coverage. There were negative associations
157 with H3 circulation among influenza A, average flu season specific humidity, and proportion
158 of the population in poverty. With regard to measurement factors, careseeking behavior and
159 claims database coverage had strong positive associations with epidemic duration.

160 **Applications to surveillance**

161 Considering the large volume and spatial resolution of our data, we sought to explore the
162 robustness of our inference and model predictions under more realistic circumstances. Two
163 sequences of models were designed to mimic different types of real-world sentinel flu surveil-
164 lance systems –*fixed-location sentinels*, where the same sentinel locations reported data ev-
165 ery year, and *moving-location sentinels*, where new sentinel locations are recruited each year.
166 A third model sequence considered the specificity of inference and model predictions to cer-
167 tain *inclusion of historical data*, thus providing insight into the generalizability of our model
168 to epidemic forecasting. We examine these applications for the total population seasonal in-
169 tensity model, and these may also serve as a sensitivity analysis to missing observations. Ten
170 replicates were performed for each model with missingness to generalize findings beyond

171 that of random chance.

172 Sentinels in fixed locations

173 In this sequence of four models, 20, 40, 60, and 80% of randomly chosen county observations
174 were removed across all years. The effect sizes of drivers were pulled towards zero as fewer
175 sentinel counties reported ILI seasonal intensity, but the primary conclusions remained robust.
176 We noted that the positive effect of care-seeking increased across most model replicates and
177 insurance coverage shifted from no effect to a slightly positive effect as sentinel reporting
178 declined (*Figure 4A*). Model predictions (county-season fitted values) remained quite robust
179 relative to the complete model, even when 80% of counties were excluded (*Figure 4B*).

180 Sentinels in moving locations

181 In this sequence of four models, 20, 40, 60, and 80% of randomly chosen seasonally-stratified
182 observations were removed. Similar to the fixed-location sequence, drivers were pulled to-
183 wards zero as fewer sentinel counties reported ILI, the drivers with the smallest means were
184 pulled towards zero and predictors with no effect in the complete model were found to be
185 significant (*Figure 4* supplement). Model predictions had good agreement with the complete
186 model up to a threshold between 60 and 80% missingness, where many county-season fits
187 suddenly became poor.

188 Inclusion of historical data

189 In this sequence of models, one, three, and five out of seven flu seasons in the study period
190 were completely removed. As hinted by the inconsistency of inference across seasons in the
191 single season model results (*Figure 3*), important drivers changed substantially when more
192 than one season was removed, particularly when they had small effect sizes in the complete
193 model (*Figure 4* supplement). Notably, medical claims coverage and care-seeking were two of
194 three predictors that remained consistent in the magnitude and direction of inference across
195 all model replicates. Model predictions were robust relative to the complete model only when
196 one season was removed. Beyond that, many seasonal fitted values were poor, particularly for
197 some seasons where data had been removed.

198 Discussion

199 Using hierarchical modeling approaches, we explored the contributions of 19 potential predic-
200 tors towards county-level variation in influenza disease burden across the United States during
201 flu seasons from 2002-2003 to 2008-2009. To our knowledge, this is the first large-scale study
202 to compare the relative importance of environmental, demographic, and socioeconomic hy-
203 potheses about influenza disease burden in addition to data reporting biases. The fine spatial
204 resolution and high coverage of our medical claims data (estimated to represent 20% of all
205 health care visits across the United States in our study period) enabled the comparison of
206 multiple hypotheses, and the inclusion of several flu seasons and sensitivity analyses enhance
207 confidence in the robustness of our findings.

208 Our model results suggest that South Atlantic states may experience flu seasons most
209 acutely because they have higher seasonal intensities relative to their baselines, and greater
210 examination of flu season surveillance and surge capacity in these areas may be warranted.
211 We also found that a mixture of factors explained the variation in our model and that these
212 factors changed across different cross-sections of time, thus highlighting the necessity of cross-
213 disciplinary approaches (e.g., from sociology to epidemiology to immunology) in future pur-
214 suits of this question. Moreover, the declining importance of claims database coverage (i.e.,
215 population representativeness of the data) as coverage increased underscores the relevance
216 of collecting and using metadata when making epidemiological inference from opportunistic
217 sources or undesigned observational samples. The ability for our model to project relatively

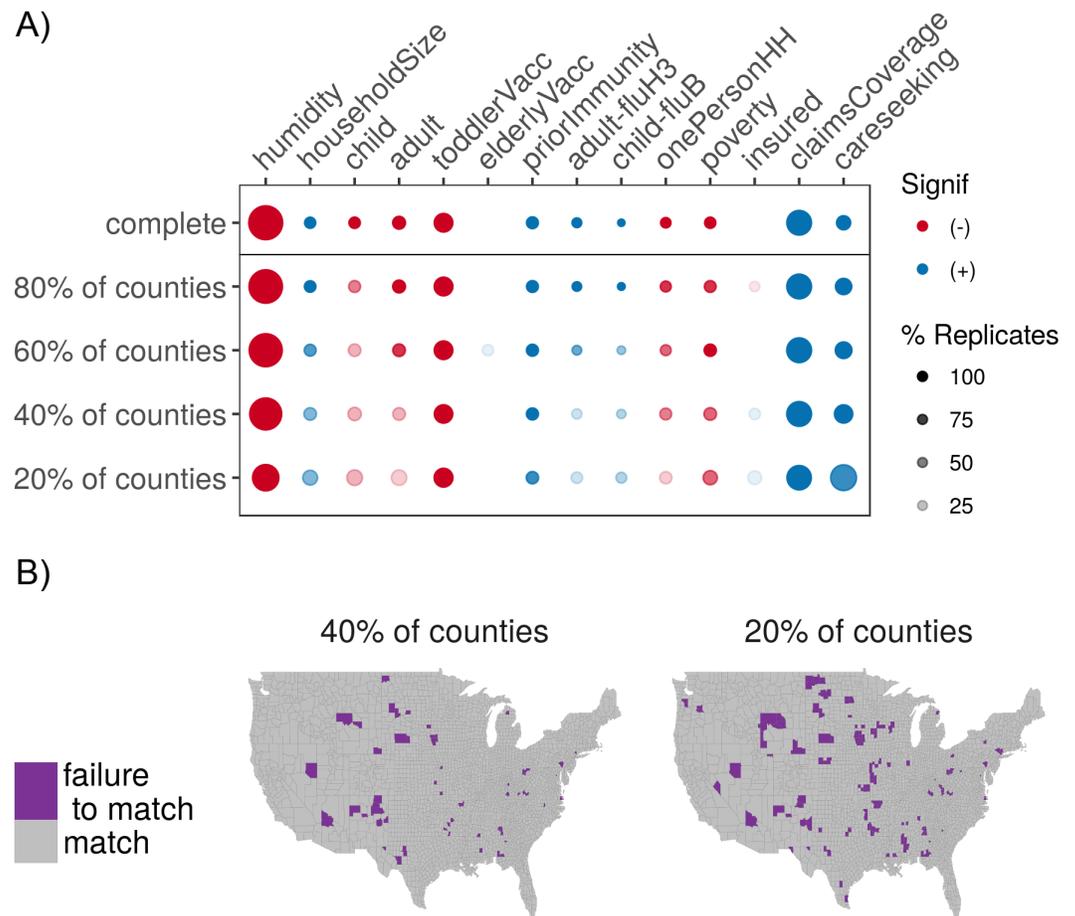


Figure 4. A) Diagram indicating changes to model inference as fewer fixed-location sentinels reported data. Color indicates directionality of the significant effect (blue is positive, red is negative) while greater transparency indicates a lower percentage of replicates with a significant effect (for models with missingness); dot size represents the magnitude of the posterior mean (or average of the posterior mean across replicates). Predictors with no significant effect across the sequence of models were removed for viewing ease, and absence of a dot means the effect was not significant across any replicates. B) Map of model prediction match between the complete model and the 40% and 20% reporting levels for fixed-location sentinels. Match between the complete and sentinel models were aggregated across 70 season-replicate combinations (7 seasons * 10 replicates). Color indicates match between posterior predictions in the missing and complete models (purple represents a failure to match in at least half of season-replicate combinations).

Figure 4-Figure supplement 1. Diagram indicating changes to model inference as fewer moving-location sentinels reported data.

Figure 4-Figure supplement 2. Map of model prediction match between the complete model and the 60% and 80% missing levels for moving-location sentinels.

Figure 4-Figure supplement 3. Diagram indicating changes to model inference as historical seasons were randomly removed from the model.

Figure 4-Figure supplement 4. Map of model prediction match between the complete model and models missing one, three, or five historical flu seasons.

218 accurate fitted values across increasingly missing data suggests that routine sentinel surveil-
219 lance in fixed locations may be more accurate for interpolating ILI disease burden among
220 uncovered areas than surveillance across changing locations, even when fewer locations may
221 be surveyed.

222 Prior studies have reported relationships between low absolute humidity and greater in-
223 fluenza transmission and survival in experimental settings, and that fluctuations in absolute
224 humidity may explain the seasonality of influenza across large geographic scales (*Tamerius*
225 *et al.*, 2011; *Lowen and Steel*, 2014). Our study adds to this literature in finding strong nega-
226 tive associations between absolute humidity and both seasonal intensity and epidemic dura-
227 tion. In addition, our results elucidate the debate about whether influenza transmits primarily
228 through frequency- or density-dependent contact. Greater seasonal intensity was associated
229 with populations with larger household sizes (a proxy for infection risk from frequent contacts),
230 while longer epidemics were associated with larger household sizes and greater population
231 density. We suspect that density-dependent transmission explained differences in epidemic
232 duration but not seasonal intensity because the calculation for seasonal intensity accounted
233 for population size; population density did not explain variation in the risk of seasonal intensity
234 after adjusting for greater transmission among larger populations.

235 Household studies of influenza transmission often examine age-specific risks of household
236 influenza introduction (*Cauchemez et al.*, 2004; *Lau et al.*, 2015), and differences in contact
237 and travel patterns between children and adults have led to the hypothesis that children
238 drive local transmission while adults drive global influenza spread (*Apolloni et al.*, 2013; *Viboud*
239 *et al.*, 2006). Contrary to these hypotheses, larger child and adult population proportions were
240 both associated with lower seasonal intensity. Rather than serving as proxies for local and
241 global transmission, the complement of these predictors together may in fact capture the
242 “high-risk” population proportion in a given location—infants, toddlers, and the elderly—which
243 typically experience greater clinical severity (*Thompson et al.*, 2006) and have higher rates of
244 care-seeking (*Biggerstaff et al.*, 2012). In examining seasonal intensity models for the child and
245 adult populations specifically, we were surprised to find negative associations with population
246 density and average household size, when there was no effect or a positive effect in the total
247 population model (*Appendix 2*). While it may be that children and adults in less connected
248 areas have greater seasonal intensity relative to their ILI baselines, these patterns may also be
249 an artifact of smaller volumes of data among age groups.

250 The positive association between influenza A/H3 and adult intensity and influenza B and
251 child intensity corroborate the results of previous epidemiological studies (*Hayward et al.*, 2014;
252 *Beauté et al.*, 2015), and agree with the positive effect of the interaction terms between chil-
253 dren and influenza B and adults and influenza A/H3 from our total seasonal intensity mod-
254 els (*Appendix 2*). Despite a positive linear correlation between the seasonal intensity and
255 epidemic duration measures (*Appendix 4*), influenza B circulation uniquely indicated longer
256 epidemics, in line with hypotheses that flu seasons are elongated when influenza B resurges
257 among children after a first wave of influenza A (*Hayward et al.*, 2014; *Beauté et al.*, 2015). We
258 acknowledge that our findings may be specific to our study period; recent research highlights
259 the importance of childhood hemagglutinin imprinting on immune responses to subsequent
260 influenza infections (*Gostic et al.*, 2016).

261 We were surprised to observe that higher estimated prior immunity was associated with
262 greater seasonal intensity and longer epidemic durations for the multi-season models and
263 most seasons in the single-season models (some years experienced no effect). One possible
264 interpretation is that some locations always tend to have high disease burden relative to their
265 epidemic baselines. Prior work suggests that larger epidemics induce more antigenic drift
266 in subsequent seasons (*Boni et al.*, 2004); building off this finding, we suggest that influenza
267 drift renews population susceptibility every flu season, even on small spatial scales. We also
268 acknowledge limitations underlying the calculation of this predictor; in using the seasonal

269 intensity measure to represent the previous flu season's attack rate, we ignore asymptomatic
270 infection, vaccination rates, and the reporting biases found to be an important component to
271 data observation. Additionally, membership in the same antigenic cluster is a simplification
272 of the immunity conferred by infection with a given strain. Beyond "pre-existing immunity",
273 we report mixed findings on the effect of flu vaccination. While higher vaccination coverage
274 among toddlers was associated with lower seasonal intensity, we note that higher vaccina-
275 tion coverage among elderly was associated with longer epidemics. We posit that vaccina-
276 tion campaigns among elderly populations may increase in anticipation of large or severe
277 flu seasons, due to their risk of severe complications from flu and clustered living in nursing
278 homes.

279 Our study found that locations with greater poverty had lower influenza disease burden, in
280 contrast with ample evidence that there are heightened rates of influenza-related hospitaliza-
281 tions, influenza-like illness, respiratory illness, neglected chronic diseases, and other measures
282 of poor health among populations with greater material deprivation (*Hadler et al., 2016; Monto*
283 *and Ullman, 1974; Tam et al., 2014; Biggerstaff et al., 2014b,a; Charland et al., 2011; Hotez, 2008;*
284 *Adler and Newman, 2002; Steptoe and Feldman, 2001*). Several possible non-exclusive ex-
285 planations for this discrepancy exist. Differences in socio-economic background may change
286 recognition and therefore reporting of disease symptoms (*Monto and Ullman, 1974*). Material
287 deprivation and lack of social cohesion have also been implicated in lower rates of health
288 care utilization for ILI, which would reduce the observation of influenza disease burden in
289 our medical claims data among the poorest populations (*Charland et al., 2011; Biggerstaff*
290 *et al., 2014a*). Indeed, higher rates of health care-seeking were associated with greater disease
291 burden, while hospitals per capita had no effect among our results, which further suggests
292 that patient-side needs and concerns captured ILI variation better than deficits in health re-
293 source availability. Future studies focused on estimation and surveillance of influenza disease
294 burden should consider collecting and incorporating data on health care utilization in their
295 populations of interest in order to account for reporting biases and limited forecasting ability
296 in poorer neighborhoods (*Scarpino et al., 2016*).

297 Building off mechanistic explanations for measurement biases, we noted that the positive
298 explanatory effect of claims database coverage declined as coverage itself increased through-
299 out our study period (*Appendix 5*). Conversely, when we artificially removed counties from
300 our model (fixed-location sentinels) or subset our data into age groups, health care-seeking
301 behavior more strongly explained the variation in seasonal intensity among the remaining
302 observations. These two results together suggest that statistical inference from opportunist-
303 ic data samples may avoid some types of reporting biases when the coverage or volume of
304 data achieves a minimum threshold, in response to concerns posed in *Lee et al. (2016)*. In our
305 specific case, increases to claims database coverage or care-seeking behavior might reduce re-
306 porting biases by increasing the representativeness of a given location's sample. Additionally,
307 we present the concept of a network of *sentinel locations*, in contrast to sentinel physicians or
308 hospitals, which may be composed of administrative units (e.g., counties) that were chosen for
309 either their representativeness of the larger population or their status as an outlier (e.g., match
310 or failure to match locations in *Figure 4*, respectively). Given the growing availability of health-
311 associated big data in infectious disease surveillance (*Bansal et al., 2016; Simonsen et al., 2016*),
312 we project the possibility that sentinel locations may report high volume digital health data
313 from disparate sources to a central public health organization and that the informed choice
314 of sentinels may improve the robustness of sentinel surveillance systems.

315 We urge caution in the interpretation of our results because they are correlative and prone
316 to invoking the ecological fallacy, where statistical inference about a group (in our case, county
317 populations) is falsely assumed to apply at the individual level (*Morgenstern, 1982; Robinson,*
318 *2009*). Future research should build off our study to design experiments that may provide
319 causal or individual-level evidence that supports or rejects these hypotheses. We also ac-

320 knowledge the limitations of the spatial and temporal resolutions of the data used in our
321 analysis. Previous work suggests that statistically-identified drivers of disease distributions dep-
322 end on the spatial scale of analysis (Cohen *et al.*, 2016), and our results may be biased by the
323 county unit observations of our disease data. In addition, we incorporated multiple scales of
324 predictors (county, state, and HHS region) according to the best available data, thus poten-
325 tially altering our statistical inference, although we did attempt to account for differences in
326 variation across these different predictors with the inclusion of group effects. In addition, we
327 note that the nature of our disease burden estimation procedure means that a given county's
328 seasonal intensity is relative to its own baseline across years. It may not be appropriate to use
329 our model predictions to inform national-level decision makers about absolute intensity of
330 the flu season in a given location, although local public health departments could use our
331 procedure to assess intensity in a given year relative to that of previous flu seasons.

332 **Methods**

333 **Medical claims data**

334 Weekly visits for influenza-like illness (ILI) and any diagnosis from October 2002 to May 2009
335 were obtained from a records-level database of CMS-1500 US medical claims managed by
336 IMS Health and aggregated to three-digit patient US zipcode prefixes (zip3s), where ILI was
337 defined with International Classification of Diseases, Ninth Revision (ICD-9) codes for: direct
338 mention of influenza, fever combined with respiratory symptoms or febrile viral illness, or pre-
339 scription of oseltamivir. Medical claims have been demonstrated to capture respiratory in-
340 fections accurately and in near real-time (Cadieux and Tamblin, 2008; Santillana *et al.*, 2016),
341 and our specific dataset was validated to independent ILI surveillance data at multiple spa-
342 tial scales and age groups and captures spatial dynamics of influenza spread in seasonal and
343 pandemic scenarios (Viboud *et al.*, 2014; Gog *et al.*, 2014; Charu *et al.*, 2017).

344 We also obtained database metadata from IMS Health on the percentage of reporting
345 physicians and the estimated effective physician coverage by visit volume; these data were
346 used to generate "measurement" predictors (Table 1). ILI reports and measurement factors
347 at the zip3-level were redistributed to the county-level according to population weights de-
348 rived from the 2010 US Census ZIP Code Tabulation Area (ZCTA) to county relationship file,
349 assuming that ZCTAs that shared the first three digits belonged to the same zip3.

350 **Defining influenza disease burden.**

351 We performed the following data processing steps for each county-level time series of ILI per
352 population: i) Fit a LOESS curve to non-flu period weeks (flu period defined as November
353 through March each year) to capture moderate-scale time trends (span = 0.4, degree = 2); ii)
354 Subtract LOESS predictions from original data to detrend the entire time series; iii) Fit a linear
355 regression model with annual harmonic terms and a time trend to non-flu period weeks (Yu
356 *et al.*, 2013); iv) Counties "had epidemics" in a given flu season if at least two consecutive weeks
357 of detrended ILI observations exceeded the ILI epidemic threshold during the flu period (i.e.,
358 epidemic period) (Denoed *et al.*, 2007). The epidemic threshold was the upper bound of
359 the 95% confidence interval for the linear model prediction. Counties with a greater number
360 of consecutive weeks above the epidemic threshold during the non-flu period than during
361 the flu period were removed from the analysis.; v) Disease burden metrics were calculated for
362 counties with epidemics.

363 Multiple measures of influenza disease burden were defined for each county. For a given
364 season: *seasonal intensity* was the one plus the sum of detrended ILI observations during
365 the epidemic period (shifted by one to accommodate the likelihood distribution); *epidemic*
366 *duration* was the number of weeks in the epidemic period and counties without epidemics
367 were assigned the value zero.

368 **Predictor data collection and variable selection**

369 Quantifiable proxies were identified for each hypothesis found in the literature, and these
370 mechanistic predictors were collected from probability-sampled or gridded, publicly avail-
371 able sources and collected or aggregated to the smallest available spatial unit among US
372 counties, states, and Department of Health and Human Services (HHS) regions for each year
373 or flu season in the study period, as appropriate (*Table 1, Appendix 5*).

374 We selected one predictor to represent each hypothesis according to the following criteria,
375 in order: i) Select for the finest spatial resolution; ii) Select for the greatest temporal coverage
376 for years in the study period; iii) Select for limited multicollinearity with predictors represent-
377 ing the other hypotheses, as indicated by the magnitude of Spearman rank cross-correlation
378 coefficients between predictor pairs. We also compared the results of single predictor models
379 and our final multivariate models as another check of multicollinearity (*Appendix 5*). For the
380 modeling analysis, if a predictor had missing data at all locations for an entire year, data from
381 the subsequent or closest other survey year were replicated to fill in that year. If a predictor
382 data source was available only at the state or region-level, all inclusive counties were assigned
383 the corresponding state or region-level predictor value (e.g., assign estimated percentage of
384 flu vaccination coverage for state of California to all counties in California). Predictors were
385 centered and standardized prior to all exploratory analyses and modeling, as appropriate. In-
386 teraction terms comprised the product of their component centered and standardized pre-
387 dictors. Data cleaning and exploratory data analysis were conducted primarily in R (*R Core*
388 *Team, 2015*). Final model predictors are described below, and our hypotheses for each predic-
389 tor are described in *Table 1*.

390 Environmental data

391 Daily specific humidity data on a 2m grid were collected from the National Oceanic and Atmo-
392 spheric Administration (NOAA) North American Regional Reanalysis (NARR), provided by the
393 NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their website at <http://www.esrl.noaa.gov/psd/>.
394 Values were assigned to the grid point nearest to the county centroid.

395 Readings of fine particulate matter, defined as pollutants with aerodynamic diameter less
396 than 2.5 micrometers, were collected from the CDC WONDER database at the county and
397 daily scales from their website at <https://wonder.cdc.gov/>.

398 Social contact and population data

399 Annual total and age-specific population data were taken from the intercensal population
400 estimates and land area and number of housing units were reported during the 2000 and
401 2010 Census; both datasets were available at the county scale from the U.S. Census Bureau.
402 These data were used to calculate proportion of total population that are children (5-19 years
403 old) and adults (20-69 years old), population density by land area, and estimated average
404 household size.

405 Flu-specific data

406 Annual flu vaccination rates for toddlers (19-35 months old) and the elderly (≥ 65 years old)
407 were estimated at the state-level from the Centers for Disease Control and Prevention (CDC)
408 National Immunization Survey and Behavioral Risk Factor Surveillance System, respectively.
409 Annual proportion of A-typed flu samples subtyped as H3 and annual proportion of confirmed
410 flu samples typed as B across U.S. Department of Health and Human Services (HHS) regions
411 were collected by WHO/NREVSS Collaborating Labs and available at the CDC FluView website
412 at <http://www.cdc.gov/flu/weekly/fluviewinteractive.htm>.

413 Prior immunity

414 For a given county, a proxy for prior immunity was derived from the following data: 1) the
415 previous flu season's total population seasonal intensity; the proportion of positive flu strains

416 identified as A/H3, A/H1, and B in the broader HHS region during 2) the previous flu season
417 and 3) the current flu season; 4) the most prominently circulating flu strain for each cate-
418 gory (A/H3, A/H1, or B) for each flu season; 5) antigenic clusters for A/H3 and A/H1 strains as
419 identified in *Du et al. (2012)*; *Liu et al. (2015)*; and 6) Victoria- or Yamagata-like lineages for B
420 strains as noted in *Bedford et al. (2014)*. Data for items 1-3 are described above in “Defining in-
421 fluenza disease burden” and “Flu-specific data.” We obtained the antigenic characterizations
422 for circulating strains (item 4) from CDC influenza season summaries, which are available at
423 <https://www.cdc.gov/flu/weekly/pastreports.htm>.

424 Using these data, we calculated a proxy of prior immunity that captures “the proportion
425 of individuals infected in the previous flu season that would have protection during the cur-
426 rent flu season, accounting for the distribution of circulating flu strains.” For each flu category
427 among A/H3, A/H1, and B, we calculated the product of the previous and current year’s propor-
428 tion of total circulation and a binary value to indicate if previous and current strains were from
429 the same antigenic cluster or lineage (1 = same cluster/lineage, 0 = different cluster/lineage).
430 For a given county, these products were summed across A/H3, A/H1, and B, and multiplied by
431 the previous year’s seasonal intensity.

432 Socioeconomic and access to care data

433 Annual data on number of hospitals were obtained at the county-level from the Health Re-
434 sources and Services Administration (HRSA) Area Health Resources Files (AHRF). County-level
435 data on proportion of households with a single person were obtained from five-year averages
436 of American Community Survey (ACS) estimates, which were available starting in 2005. An-
437 nual estimates on proportion of the population in poverty was obtained at the county-level
438 from the model-based Small Area Income and Poverty Estimates (SAIPE). Annual estimates
439 on proportion of the population with health insurance was obtained at the county-level from
440 the model-based Small Area Health Insurance Estimates (SAHIE). SAIPE and SAHIE are both
441 products of the U.S. Census Bureau that were derived from the Current Population Survey or
442 ACS.

443 Medical claims measurement factors

444 IMS Health provided us with weekly aggregated data on visits for any diagnosis by age group
445 and location. Care-seeking behavior was defined as the total visits per population size from
446 November through April of a given flu season. Claims database coverage was the estimated
447 physician coverage among all physicians registered by the American Medical Association in
448 the IMS Health medical claims database.

449 Model structure

450 We present the most common version of our model structure here. The generic model for
451 county-year observations (for i counties and t years) of influenza disease burden y_{it} is:

$$y_{it} | \mu_{it}, \tau \sim f(\mathbf{y} | \boldsymbol{\mu}, \boldsymbol{\tau}) \quad (1)$$

452 where $\mathbf{y} = (y_1, \dots, y_n)'$ denotes the vector of all observations (*Equation 1*). We modeled the
453 mean of the observed disease burden magnitude (μ_i), where $f(\mathbf{y} | \boldsymbol{\mu}, \boldsymbol{\tau})$ is the distribution of the
454 likelihood of the disease burden data, parameterized with mean $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)'$ and precision
455 $\boldsymbol{\tau}$, as appropriate to the likelihood distribution (N.B., for the Poisson likelihood, $\boldsymbol{\mu} = 1/\boldsymbol{\tau}$).

456 The mechanisms driving disease burden were modeled:

$$g(\mu_i) = g(E_i) + \alpha + \sum_1^m X_i \beta + \gamma_i + \zeta_{j[i]} + \eta_{k[i]} + \nu_i + \phi_i + \epsilon_{it} \quad (2)$$

457 where $g(\cdot)$ is the link function, α is the intercept, there are m socio-environmental and mea-
458 surement predictors (X_i 's), and E_i is an offset of the expected disease burden, such that *Equa-*

459 tion 2 models the relative risk of disease (μ_i/E_i) in county i , common in disease mapping (Law-
460 son, 2013; Banerjee et al., 2015; Waller and Carlin, 2010). Group terms at the county, state,
461 region, and season levels ($\gamma_i, \zeta_{j[l]}, \eta_{k[l]}, \nu_t$, respectively) and the error term (ϵ_{it}) are independent
462 and identically distributed (*iid*).

463 Geographical proximity appears to increase the synchrony of flu epidemic timing (Schanzer
464 et al., 2011b; Stark et al., 2012), while connectivity between cities has been linked with spa-
465 tial spread in the context of commuting and longer distance travel (Charaudeau et al., 2014;
466 Brownstein et al., 2006; Crépey and Barthélemy, 2007; Lemey et al., 2014). We modeled county
467 spatial dependence ϕ_i with an intrinsic conditional autoregressive (ICAR) model, which smooths
468 model predictions by borrowing information from neighbors (Besag et al., 1991):

$$\phi_i | \phi_j, \tau_\phi \sim \text{Normal}\left(\frac{1}{\xi_i} \sum_{i \sim j} \phi_j, \frac{1}{\xi_i \tau_\phi}\right), \quad (3)$$

469 where ξ_i represents the number of neighbors for node i , ϕ_j is a vector indicating the neigh-
470 borhood relationship between node i and all nodes j ($i \sim j$), and τ_ϕ is the precision parameter
471 (Equation 3).

472 Model fit, sensitivity, and validation

473 To assess model fit, we examined scatterplots and Pearson’s cross-correlation coefficients be-
474 tween observed and fitted values for the relative risk of total population seasonal intensity and
475 for epidemic duration. We also examined scatterplots of standardized residuals and fitted val-
476 ues; standardized residuals were defined as $(y - \mu_{\hat{y}})/\sigma_{\hat{y}}$, where $\mu_{\hat{y}}$ is the fitted value posterior
477 mean and $\sigma_{\hat{y}}$ is the fitted value standard deviation. Model sensitivity was assessed by compar-
478 ing model fits and inference robustness when observations were randomly removed from the
479 model, as described below under “Applications to missing data & inference robustness.”

480 For each disease burden measure, we compared models with no spatial dependence,
481 county-level dependence only, state-level dependence only, and both county and state-level
482 dependence. The goal of the county-level dependence was to capture local population flows,
483 while state-level dependence attempted to capture state-level flight passenger flows (details
484 in Appendix 1). We determined that models with only county-level spatial neighborhood
485 structure best fit the data after examining the Deviance Information Criteria (DIC) values and
486 spatial dependence coefficients of the four model structures. County-level spatial structure
487 was subsequently used in all final model combinations. We report results from models with
488 county-level dependence only.

489 For model validation, we compared model fitted values for seasonal intensity with CDC ILI
490 and laboratory surveillance data (details in Appendix 1).

491 Statistical analysis

492 The goals of our modeling approach were to i) estimate the contribution of each predictor
493 to influenza disease burden, ii) predict disease burden in locations with missing data, and
494 iii) improve mapping of influenza disease burden. We performed approximate Bayesian infer-
495 ence using Integrated Nested Laplace Approximations (INLA) with the R-INLA package
496 (www.r-inla.org) (Rue et al., 2009; Martins et al., 2013). INLA has demonstrated computational
497 efficiency for latent Gaussian models and produced similar estimates for fixed parameters
498 as established implementations of Markov Chain Monte Carlo (MCMC) methods for Bayesian
499 inference (Carroll et al., 2015). Extensions to INLA have enabled its application to spatial, spatio-
500 temporal, and zero-inflated models (Lindgren et al., 2011; Arab, 2015), which is implicated in
501 INLA’s growing use in the disease mapping and spatial ecology communities (Schrödle and
502 Held, 2011; Blangiardo et al., 2013).

503 Seasonal intensity was modeled with a lognormal distribution, and epidemic duration
504 was modeled with a Poisson distribution and log link and excluded the offset term in Equa-

505 *tion 2*. Consequently, we note that all seasonal intensity models examine the relative risk of
506 seasonal intensity, while epidemic duration models directly examine the duration in weeks.
507 Multi-season models included all terms in *Equation 2*, while single-season models included
508 all terms in *Equation 2* except the season grouping (v_t). Model coefficients were interpreted
509 as statistically significant if the 95% credible interval for a parameter's posterior distribution
510 failed to include zero.

511 **Applications to missing data & inference robustness**

512 We considered the robustness of our total population model results by refitting models where
513 20%, 40%, 60% and 80% of all county observations were replaced with NAs (*sentinels in fixed*
514 *locations*), and where 20%, 40%, 60% and 80% of model observations were stratified by sea-
515 son and randomly replaced with NAs (*sentinels in moving locations*). We also refit three mod-
516 els where one, three, and five of seven flu seasons were randomly chosen and completely re-
517 placed with NAs (*inclusion of historical data*). To account for variability due to random chance,
518 models were replicated ten times each with different random seeds. For each sequence of
519 missingness, we compared the magnitude and significance of socio-environmental and mea-
520 surement drivers, and the posterior distributions of county-season fitted values. Fitted value
521 distributions were noted as significantly different if the interquartile ranges for two fitted val-
522 ues failed to overlap with each other.

523 **Acknowledgments**

524 The authors thank IMS Health for kindly sharing the medical claims data.

525 **Funding**

526 This work was supported by the Jayne Koskinas Ted Giovanis Foundation for Health and Policy
527 (JKTG) [dissertation support grant to ECL]; and the RAPIDD Program of the Science & Tech-
528 nology Directorate, Department of Homeland Security and the Fogarty International Center,
529 National Institutes of Health. The content is solely the responsibility of the authors and does
530 not necessarily reflect the official views of JKTG, the National Institutes of Health, or IMS Health.

531 **References**

- 532 **Adler NE**, Newman K. Socioeconomic disparities in health: Pathways and policies. *Health Aff.* 2002;
533 21(2):60–76. doi: 10.1377/hlthaff.21.2.60.
- 534 **Apolloni A**, Poletto C, Colizza V. Age-specific contacts and travel patterns in the spatial spread of 2009
535 H1N1 influenza pandemic. *BMC Infect Dis.* 2013 jan; 13:176. doi: 10.1186/1471-2334-13-176.
- 536 **Arab A**. Spatial and Spatio-Temporal Models for Modeling Epidemiological Data with Excess Zeros. *Int J*
537 *Environ Res Public Health.* 2015; 12(9):10536–10548. doi: 10.3390/ijerph120910536.
- 538 **Banerjee S**, Carlin BP, Gelfand AE. *Hierarchical Modeling and Analysis for Spatial Data*. Second ed. Boca
539 Raton (FL): CRC Press; 2015.
- 540 **Bansal S**, Chowell G, Simonsen L, Vespignani A, Viboud C. Big Data for Infectious Disease Surveillance and
541 Modeling. *J Infect Dis.* 2016; 214(suppl 4):S375–S379. doi: 10.1093/infdis/jiw400.
- 542 **Bansal S**, Pourbohloul B, Hupert N, Grenfell B, Meyers LA. The shifting demographic landscape of pan-
543 demic influenza. *PLoS One.* 2010 jan; 5(2):e9360. doi: 10.1371/journal.pone.0009360.
- 544 **Barreca AI**, Shimshack JP. Absolute humidity, temperature, and influenza mortality: 30 years of
545 county-level evidence from the United States. *Am J Epidemiol.* 2012 oct; 176 Suppl(7):S114–22. doi:
546 10.1093/aje/kws259.
- 547 **Beauté J**, ZUCS P, KORSUN N, BRAGSTAD K, ENOUF V, KOSSYVAKIS A, et al. Age-specific differences in in-
548 fluenza virus type and subtype distribution in the 2012/2013 season in 12 European countries. *Epidemiol*
549 *Infect.* 2015 oct; 143(14):2950–2958. doi: 10.1017/S0950268814003422.

- 550 **Bedford T**, Suchard Ma, Lemey P, Dudas G, Gregory V, Hay AJ, et al. Data from: Integrat-
551 ing influenza antigenic dynamics with molecular evolution. Dryad Digit Repos. 2014; doi:
552 <http://dx.doi.org/10.5061/dryad.rc515>.
- 553 **Besag J**, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. *Ann Inst*
554 *Stat Math*. 1991; 43(1):1-20. doi: 10.1007/BF00116466.
- 555 **Biggerstaff M**, Jhung Ma, Reed C, Garg S, Balluz L, Fry aM, et al. Impact of medical and behavioural factors
556 on influenza-like illness, healthcare-seeking, and antiviral treatment during the 2009 H1N1 pandemic:
557 USA, 2009-2010. *Epidemiol Infect*. 2014 jan; 142(1):14-25. doi: 10.1017/S0950268813000654.
- 558 **Biggerstaff M**, Jhung M, Kamimoto L, Balluz L, Finelli L. Self-reported influenza-like illness and receipt
559 of influenza antiviral drugs during the 2009 pandemic, United States, 2009-2010. *Am J Public Health*.
560 2012 oct; 102(10):e21-26. doi: 10.2105/AJPH.2012.300651.
- 561 **Biggerstaff M**, Jhung MA, Reed C, Fry AM, Balluz L, Finelli L. Influenza-like illness, the time to seek health-
562 care, and influenza antiviral receipt during the 2010-11 influenza season - United States. *J Infect Dis*.
563 2014; 210(4):535-44.
- 564 **Blangiardo M**, Cameletti M, Baio G, Rue H. Spatial and spatio-temporal models with R-INLA. *Spat Spa-*
565 *tiotemporal Epidemiol*. 2013; 4:33-49.
- 566 **Boni MF**, Gog JR, Andreasen V, Christiansen FB. Influenza drift and epidemic size: the race between gen-
567 erating and escaping immunity. *Theor Popul Biol*. 2004 mar; 65(2):179-91. doi: 10.1016/j.tpb.2003.10.002.
- 568 **Brownstein JS**, Wolfe CJ, Mandl KD. Empirical evidence for the effect of airline travel on inter-regional in-
569 fluenza spread in the United States. *PLoS Med*. 2006 sep; 3(10):e401. doi: 10.1371/journal.pmed.0030401.
- 570 **Cadieux G**, Tamblyn R. Accuracy of physician billing claims for identifying acute respiratory infections in
571 primary care. *Health Serv Res*. 2008; 43(6):2223-2238. doi: 10.1111/j.1475-6773.2008.00873.x.
- 572 **Carroll R**, Lawson AB, Faes C, Kirby RS, Aregay M, Watjou K. Comparing INLA and OpenBUGS for hierar-
573 chical Poisson modeling in disease mapping. *Spat Spatiotemporal Epidemiol*. 2015; 14-15:45-54. doi:
574 10.1016/j.sste.2015.08.001.
- 575 **Cauchemez S**, Carrat F, Viboud C, Valleron AJ, Boëlle PY. A Bayesian MCMC approach to study transmis-
576 sion of influenza: application to household longitudinal data. *Stat Med*. 2004 nov; 23(22):3469-87. doi:
577 10.1002/sim.1912.
- 578 **Charaudeau S**, Pakdaman K, Boëlle PY. Commuter mobility and the spread of infectious diseases: appli-
579 cation to influenza in France. *PLoS One*. 2014 jan; 9(1):e83002. doi: 10.1371/journal.pone.0083002.
- 580 **Charland KM**, Brownstein JS, Verma A, Brien S, Buckeridge DL. Socio-economic disparities in the burden
581 of seasonal influenza: The effect of social and material deprivation on rates of influenza infection. *PLoS*
582 *One*. 2011; 6(2):1-5. doi: 10.1371/journal.pone.0017207.
- 583 **Charu V**, Zeger S, Gog J, Bjørnstad ON, Kissler S, Simonsen L, et al. Human mobility and the spatial trans-
584 mission of influenza in the United States. *PLOS Comput Biol*. 2017; 13(2):e1005382. doi: 10.1371/jour-
585 nal.pcbi.1005382.
- 586 **Cohen JM**, Civitello DJ, Brace AJ, Feichtinger EM, Ortega CN, Richardson JC, et al. Spatial scale modulates
587 the strength of ecological processes driving disease distributions. *Proc Natl Acad Sci*. 2016; p. 201521657.
588 doi: 10.1073/pnas.1521657113.
- 589 **Crépey P**, Barthélemy M. Detecting robust patterns in the spread of epidemics: a case study of influenza
590 in the United States and France. *Am J Epidemiol*. 2007 dec; 166(11):1244-51. doi: 10.1093/aje/kwm266.
- 591 **Denoëud L**, Turbelin C, Ansart S, Valleron AJ, Flahault A, Carrat F. Predicting pneumonia and influenza
592 mortality from morbidity data. *PLoS One*. 2007 jan; 2(5):e464. doi: 10.1371/journal.pone.0000464.
- 593 **Deyle ER**, Maher MC, Hernandez RD, Basu S, Sugihara G. Global environmental drivers of influenza. *Proc*
594 *Natl Acad Sci*. 2016; doi: 10.1073/pnas.1607747113.
- 595 **Du X**, Dong L, Lan Y, Peng Y, Wu A, Zhang Y, et al. Mapping of H3N2 influenza antigenic evolution
596 in China reveals a strategy for vaccine strain recommendation. *Nat Commun*. 2012; 3:709. doi:
597 10.1038/ncomms1710.

- 598 **Ewing A**, Lee EC, Viboud C, Bansal S. Contact, travel, and transmission: The impact of winter holidays on
599 influenza dynamics in the United States. *J Infect Dis.* 2016; doi: <https://doi.org/10.1093/infdis/jiw642>.
- 600 **Frank AL**, Taber LH, Wells JM. Comparison of Infection Rates and Severity of Illness for Influenza A Sub-
601 types H1N1 and H3N2. *J Infect Dis.* 1985; 151(1):73–80.
- 602 **Gog JR**, Ballesteros S, Viboud C, Simonsen L, Bjornstad ON, Shaman J, et al. Spatial Transmission of
603 2009 Pandemic Influenza in the US. *PLoS Comput Biol.* 2014 jun; 10(6):e1003635. doi: 10.1371/jour-
604 nal.pcbi.1003635.
- 605 **Gostic KM**, Ambrose M, Worobey M, Lloyd-Smith JO. Potent protection against H5N1 and H7N9 in-
606 fluenza via childhood hemagglutinin imprinting. *Science (80-).* 2016; 354(6313):722–726. doi: 10.1126/sci-
607 ence.aag1322.
- 608 **Grais RF**, Ellis JH, Glass GE. Assessing the impact of airline travel on the geographic spread of pandemic
609 influenza. *Eur J Epidemiol.* 2003; 18(11):1065–1072. doi: 10.1023/A:1026140019146.
- 610 **Grantz KH**, Rane MS, Salje H, Glass GE, Schachterle SE, Cummings DAT. Disparities in influenza mortality
611 and transmission related to sociodemographic factors within Chicago in the pandemic of 1918. *Proc
612 Natl Acad Sci.* 2016; 113(48):13839–13844. doi: 10.1073/pnas.1612838113.
- 613 **Hadler JL**, Yousey-Hindes K, Pérez A, Anderson EJ, Bargsten M, Bohm SR, et al. Influenza-Related Hospi-
614 talizations and Poverty Levels – United States, 2010–2012. *Morb Mortal Wkly Rep.* 2016; 65(05):101–105.
615 doi: 10.15585/mmwr.mm6505a1.
- 616 **Hayward AC**, Fragaszy EB, Bermingham A, Wang L, Copas A, Edmunds WJ, et al. Comparative community
617 burden and severity of seasonal and pandemic influenza: Results of the Flu Watch cohort study. *Lancet
618 Respir Med.* 2014; 2(6):445–454. doi: 10.1016/S2213-2600(14)70034-7.
- 619 **Hotez PJ**. Neglected Infections of Poverty in the United States of America. *PLoS Negl Trop Dis.* 2008;
620 2(6):e256. doi: 10.1371/journal.pntd.0000256.
- 621 **Khiabanian H**, Farrell GM, St George K, Rabadan R. Differences in patient age distribution between in-
622 fluenza A subtypes. *PLoS One.* 2009 jan; 4(8):e6832. doi: 10.1371/journal.pone.0006832.
- 623 **Killingley B**, Nguyen-Van-Tam J. Routes of influenza transmission. *Influenza Other Respi Viruses.* 2013;
624 7(SUPPL.2):42–51. doi: 10.1111/irv.12080.
- 625 **Kostova D**, Reed C, Finelli L, Cheng PY, Gargiullo PM, Shay DK, et al. Influenza Illness and Hospitalizations
626 Averted by Influenza Vaccination in the United States, 2005–2011. *PLoS One.* 2013 jan; 8(6):e66312. doi:
627 10.1371/journal.pone.0066312.
- 628 **Kucharski AJ**, Kwok KO, Wei VWI, Cowling BJ, Read JM, Lessler J, et al. The Contribution of So-
629 cial Behaviour to the Transmission of Influenza A in a Human Population. *PLoS Pathog.* 2014 jun;
630 10(6):e1004206. doi: 10.1371/journal.ppat.1004206.
- 631 **Kumar S**, Piper K, Galloway DD, Hadler JL, Grefenstette JJ. Is population structure sufficient to gener-
632 ate area-level inequalities in influenza rates? An examination using agent-based models. *BMC Public
633 Health.* 2015; 15(1):947. doi: 10.1186/s12889-015-2284-2.
- 634 **Lau MSY**, Cowling BJ, Cook AR, Riley S. Inferring influenza dynamics and control in households. *Proc Natl
635 Acad Sci.* 2015; p. 201423339. doi: 10.1073/pnas.1423339112.
- 636 **Lawson AB**. Bayesian Disease Mapping: hierarchical modeling in spatial epidemiology. 2 ed. New York:
637 CRC Press; 2013.
- 638 **Lee EC**, Asher JM, Goldlust S, Kraemer JD, Lawson AB, Bansal S. Mind the Scales : Harnessing Spatial Big
639 Data for Infectious Disease Surveillance and Inference. *J Infect Dis.* 2016; 214(Suppl 4):S409–S413. doi:
640 10.1093/infdis/jiw344.
- 641 **Lee EC**, Viboud C, Simonsen L, Khan F, Bansal S. Detecting Signals of Seasonal Influenza Severity through
642 Age Dynamics. *BMC Infect Dis.* 2015; 15(587). doi: 10.1186/s12879-015-1318-9.
- 643 **Lemaitre M**, Carrat F. Comparative age distribution of influenza morbidity and mortality during seasonal
644 influenza epidemics and the 2009 H1N1 pandemic. *BMC Infect Dis.* 2010 jan; 10(April 2009):162. doi:
645 10.1186/1471-2334-10-162.

- 646 **Lemey P**, Rambaut A, Bedford T, Faria N, Bielejec F, Baele G, et al. Unifying Viral Genetics and Human
647 Transportation Data to Predict the Global Transmission Dynamics of Human Influenza H3N2. *PLoS*
648 *Pathog*. 2014; 10(2). doi: 10.1371/journal.ppat.1003932.
- 649 **Lindgren F**, Rue H, Lindström J. An explicit link between Gaussian fields and Gaussian Markov random
650 field: The stochastic partial differential equations approach. *J R Stat Soc Ser B Stat Methodol*. 2011;
651 73:423–498. doi: 10.1111/j.1467-9868.2011.00777.x.
- 652 **Liu M**, Zhao X, Hua S, Du X, Peng Y, Li X, et al. Antigenic Patterns and Evolution of the Human Influenza A
653 (H1N1) Virus. *Sci Rep*. 2015; 5:14171. doi: 10.1038/srep14171.
- 654 **Lofgren E**, Fefferman NH, Naumov YN, Gorski J, Naumova EN. Influenza seasonality: underlying causes
655 and modeling theories. *J Virol*. 2007 jun; 81(11):5429–36. doi: 10.1128/JVI.01680-06.
- 656 **Longini IM**, Koopman JS, Monto aS, Fox JP. Estimating household and community transmission param-
657 eters for influenza. *Am J Epidemiol*. 1982 may; 115(5):736–51.
- 658 **Lowcock EC**, Rosella LC, Foisly J, McGeer A, Crowcroft N. The social determinants of health and pandemic
659 H1N1 2009 influenza severity. *Am J Public Health*. 2012; 102(8):51–58. doi: 10.2105/AJPH.2012.300814.
- 660 **Lowen AC**, Mubareka S, Steel J, Palese P. Influenza virus transmission is dependent on relative humidity
661 and temperature. *PLoS Pathog*. 2007 oct; 3(10):1470–6. doi: 10.1371/journal.ppat.0030151.
- 662 **Lowen AC**, Steel J. Roles of humidity and temperature in shaping influenza seasonality. *J Virol*. 2014;
663 88(14):7692–5. doi: 10.1128/JVI.03544-13.
- 664 **Martins TG**, Simpson D, Lindgren F, Rue H. Bayesian computing with INLA: New features. *Comput Stat*
665 *Data Anal*. 2013; 67:68–83. doi: 10.1016/j.csda.2013.04.014.
- 666 **Monto AS**, Ullman BM. Acute Respiratory Illness in an American Community: The Tecumseh Respiratory.
667 *JAMA*. 1974; 227(2):164–169.
- 668 **Moorthy M**, Castronovo D, Abraham A, Bhattacharyya S, Gradus S, Gorski J, et al. Deviations in influenza
669 seasonality : odd coincidence or obscure consequence? *Clin Microbiol Infect*. 2012; 18(10):955–962.
- 670 **Morgenstern H**. Uses of ecologic analysis in epidemiologic research. *Am J Public Health*. 1982; 72(12):1336–
671 1344. doi: 10.2105/AJPH.72.12.1336.
- 672 **Mossong J**, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social contacts and mixing pat-
673 terns relevant to the spread of infectious diseases. *PLoS Med*. 2008 mar; 5(3):e74. doi: 10.1371/jour-
674 nal.pmed.0050074.
- 675 **Peters TR**, Snively BM, Suerken CK, Blakeney E, Vannoy L, Poehling KA. Relative timing of influenza disease
676 by age group. *Vaccine*. 2014; 32(48):6451–6456. doi: 10.1016/j.vaccine.2014.09.047.
- 677 **R Core Team**. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation
678 for Statistical Computing; 2015.
- 679 **Robinson WS**. Ecological correlations and the behavior of individuals. *Int J Epidemiol*. 2009; 40(4):1134.
680 doi: 10.1093/ije/dyr082.
- 681 **Rue H**, Martino S, Chopin N. Approximate Bayesian Inference for Latent Gaussian Models Using Integrated
682 Nested Laplace Approximations. *J R Stat Soc Ser B*. 2009; 71(2):319–392.
- 683 **Santillana M**, Nguyen AT, Louie T, Zink A, Gray J, Sung I, et al. Cloud-based Electronic Health Records for
684 Real-time, Region-specific Influenza Surveillance. *Sci Rep*. 2016; 6(April):25732. doi: 10.1038/srep25732.
- 685 **Scarpino SV**, Dimitrov NB, Meyers LA. Optimizing provider recruitment for influenza surveillance networks.
686 *PLoS Comput Biol*. 2012; 8(4). doi: 10.1371/journal.pcbi.1002472.
- 687 **Scarpino SV**, Scott JC, Eggo R, Dimitrov NB, Meyers LA. Data Blindspots: High-Tech Disease Surveillance
688 Misses the Poor. *Online J Public Health Inform*. 2016; 8(1):2579. doi: 10.5210/OJPHI.V8I1.6451.
- 689 **Schanzer D**, Vachon J, Pelletier L. Age-specific differences in influenza A epidemic curves: do children
690 drive the spread of influenza epidemics? *Am J Epidemiol*. 2011 jul; 174(1):109–17. doi: 10.1093/aje/kwr037.
- 691 **Schanzer DL**, Langley JM, Dummer T, Aziz S. The geographic synchrony of seasonal influenza: a waves
692 across Canada and the United States. *PLoS One*. 2011 jan; 6(6):e21471. doi: 10.1371/journal.pone.0021471.

- 693 **Schanzer DL**, Langley JM, Dummer T, Viboud C, Tam TWS. A composite epidemic curve for seasonal
694 influenza in Canada with an international comparison. *Influenza Other Respi Viruses*. 2010 sep; 4(5):295-
695 306. doi: 10.1111/j.1750-2659.2010.00154.x.
- 696 **Schrödle B**, Held L. Spatio-temporal disease mapping using INLA. *Environmetrics*. 2011; 22(6):725-734. doi:
697 10.1002/env.1065.
- 698 **Shaman J**, Kohn M. Absolute humidity modulates influenza survival, transmission, and seasonality. *Proc*
699 *Natl Acad Sci U S A*. 2009 mar; 106(9):3243-8. doi: 10.1073/pnas.0806852106.
- 700 **Shaman J**, Pitzer VE, Viboud C, Grenfell BT, Lipsitch M. Absolute humidity and the seasonal onset
701 of influenza in the continental United States. *PLoS Biol*. 2010 feb; 8(2):e1000316. doi: 10.1371/jour-
702 nal.pbio.1000316.
- 703 **Simonsen L**, Clarke MJ, Williamson GD, Stroup DF, Arden NH, Schonberger LB. The impact of influenza
704 epidemics on mortality: introducing a severity index. *Am J Public Health*. 1997 dec; 87(12):1944-50.
- 705 **Simonsen L**, Gog JR, Olson D, Viboud C. Infectious Disease Surveillance in the Big Data Era : Towards Faster
706 and Locally Relevant Systems. *J Infect Dis*. 2016; 214(Suppl 4):S380-S3385. doi: 10.1093/infdis/jiw376.
- 707 **Stark JH**, Cummings DaT, Ermentrout B, Ostroff S, Sharma R, Stebbins S, et al. Local variations in spatial
708 synchrony of influenza epidemics. *PLoS One*. 2012 jan; 7(8):e43528. doi: 10.1371/journal.pone.0043528.
- 709 **Steptoe A**, Feldman PJ. Neighborhood Problems as Sources of Chronic Stress : Development of a Measure
710 of Neighborhood Problems , and Associations With Socioeconomic Status and Health. *Ann Behav Med*.
711 2001; 23(3):177-185.
- 712 **Tam K**, Yousey-Hindes K, Hadler JL. Influenza-related hospitalization of adults associated with low census
713 tract socioeconomic status and female sex in New Haven County, Connecticut, 2007-2011. *Influenza*
714 *Other Respi Viruses*. 2014 may; 8(3):274-81. doi: 10.1111/irv.12231.
- 715 **Tamerius J**, Nelson MI, Zhou SZ, Viboud C, Miller Ma, Alonso WJ. Global influenza seasonality: Reconcil-
716 ing patterns across temperate and tropical regions. *Environ Health Perspect*. 2011; 119(4):439-445. doi:
717 10.1289/ehp.1002383.
- 718 **Thompson WW**, Comanor L, Shay DK. Epidemiology of Seasonal Influenza : Use of Surveillance Data and
719 Statistical Models to Estimate the Burden of Disease. *J Infect Dis*. 2006; 194(Suppl 2):S82-S91.
- 720 **Timpka T**, Eriksson O, Spreco A, Gursky Ea, Strömngren M, Holm E, et al. Age as a determinant for dissemina-
721 tion of seasonal and pandemic influenza: An open cohort study of influenza outbreaks in Östergötland
722 county, Sweden. *PLoS One*. 2012; 7(2). doi: 10.1371/journal.pone.0031746.
- 723 **Van Kerkhove MD**, Vandemaële KAH, Shinde V, Jaramillo-gutierrez G, Koukounari A, Donnelly CA, et al.
724 Risk Factors for Severe Outcomes following 2009 Influenza A (H1N1) Infection : A Global Pooled Analysis.
725 *PLOS Med*. 2011; 8(7):e1001053. doi: 10.1371/journal.pmed.1001053.
- 726 **Viboud C**, Bjørnstad ON, Smith DL, Simonsen L, Miller MA, Grenfell BT. Synchrony, Waves, and Spatial
727 Hierarchies in the Spread of Influenza. *Science (80-)*. 2006 apr; 312(April):447-451. doi: 10.1126/sci-
728 ence.1125237.
- 729 **Viboud C**, Charu V, Olson D, Ballesteros S, Gog J, Khan F, et al. Demonstrating the use of high-volume
730 electronic medical claims data to monitor local and regional influenza activity in the US. *PLoS One*.
731 2014 jan; 9(7):e102429. doi: 10.1371/journal.pone.0102429.
- 732 **Waller LA**, Carlin BP. Disease Mapping. In: Gelfand AE, Diggle P, Guttorp P, Fuentes M,
733 editors. *Handbook of Spatial Statistics* Boca Raton (FL): CRC Press; 2010.p. 217-243. <https://www.crcpress.com/Handbook-of-Spatial-Statistics/Gelfand-Diggle-Guttorp-Fuentes/9781420072877>, doi:
734 10.1201/9781420072884-c14.Disease.
735
- 736 **Wallinga J**, Teunis P, Kretzschmar M. Using data on social contacts to estimate age-specific transmission
737 parameters for respiratory-spread infectious agents. *Am J Epidemiol*. 2006 nov; 164(10):936-44. doi:
738 10.1093/aje/kwj317.
- 739 **Wenger JB**, Naumova EN. Seasonal synchronization of influenza in the United States older adult popula-
740 tion. *PLoS One*. 2010 jan; 5(4):e10187. doi: 10.1371/journal.pone.0010187.
- 741 **Yu H**, Alonso WJ, Feng L, Tan Y, Shu Y, Yang W, et al. Characterization of regional influenza seasonality
742 patterns in china and implications for vaccination strategies: spatio-temporal modeling of surveillance
743 data. *PLoS Med*. 2013 dec; 10(11):e1001552. doi: 10.1371/journal.pmed.1001552.

Table 1. Final model predictors and hypotheses.

Factor	Index	Plot Label	Spatial Scale	Hypothesized Effect
Environmental factors				
Flu transmission	Specific humidity	humidity	county	-
Respiratory disease risk	Fine particular matter	pollution	county	+
Transmission mechanisms				
Density-dependent	Population density	popDensity	county	+
Frequency-dependent	Average household size	householdSize	county	+
Diffusion mechanisms				
Local spread	% child population	child	county	+
Importation risk	% adult population	adult	county	+
Immunity				
Vaccine-acquired	Toddler vacc. coverage	toddlerVacc	state	-
	Elderly vacc. coverage	elderlyVacc	state	-
Prior exposure	Population protected due to prior season exposure	priorImmunity	county	-
Influenza circulation				
Dominant A subtype	% H3 subtype among flu type A samples	fluH3	HHS region	+
B circulation	% B type among positive flu samples	fluB	HHS region	+
H3 has older age distribution	adult population x dominant A subtype	adult-fluH3	HHS region	+
B circulates primarily in children	child population x B circulation	child-fluB	HHS region	+
Socioeconomic factors and access to care				
Health care availability	Hospitals per capita	hospAccess	county	+
Social deprivation	% single-person households	onePersonHH	county	+
Material deprivation	% in poverty	poverty	county	+
Claims-reporting population	% with health insurance	insured	county	+
Measurement factors				
Claims database coverage	% physicians reporting to claims database	claimsCoverage	county	+
Care-seeking behavior in claims database	All visits per capita reported in database	careseeking	county	+

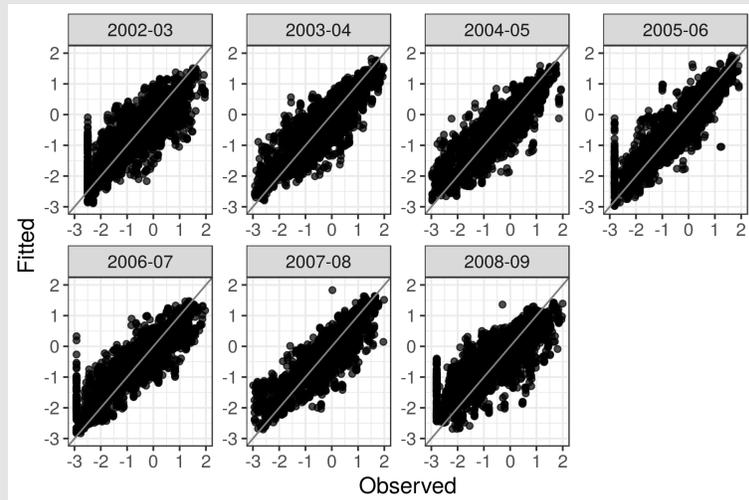
744 **Appendix 1**

745

Seasonal intensity model fit and validation

746

Model fit

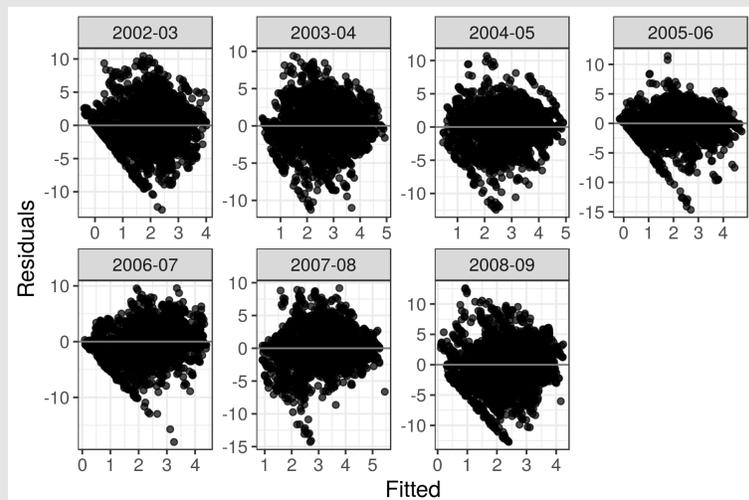


747

748

749

Appendix 1 Figure 1. Observed vs. fitted values for the relative risk of total population seasonal intensity.



751

752

Appendix 1 Figure 2. Residuals vs. fitted values for the total population log seasonal intensity.

754

Selection for spatial dependence terms

755

756

757

758

759

760

761

762

763

To determine county-level spatial neighbors, we started with the 2010 U.S. Census Bureau 500k resolution county shapefile, and connected abutting counties that were separated by bodies of water. We then used the clean shapefile to identify neighbors as counties that shared borders.

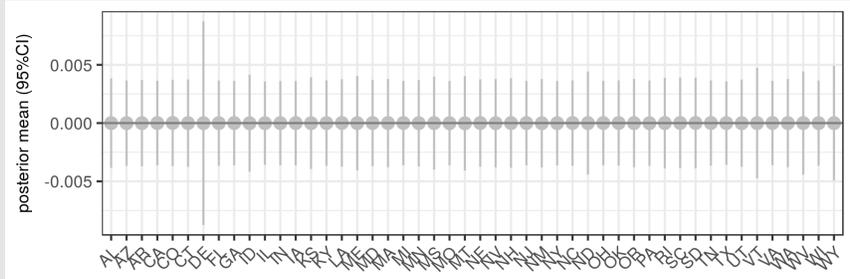
To define state-level spatial neighbors, monthly air travel passenger flows were collected from the Bureau of Transportation Statistics T-100 Domestic Market (U.S. Carriers) table from their website at <http://www.transtats.bts.gov/>. Airport flows were aggregated to the state-level and states were neighbors if passengers traveled between them from November 2007 through April 2008.

764
766

Appendix 1 Table 1. Comparison of total seasonal intensity models with different spatial dependence structures according to Deviance Information Criterion (DIC).

Spatial dependence structure	DIC
None (no ϕ terms)	40,932
County only (bordering neighbors, ϕ_i)	40,070
State only (flight passenger flows, ϕ_j)	40,933
County and state together (ϕ_i and ϕ_j)	40,070

767



768

Appendix 1 Figure 3. 95% credible intervals for the state-level spatially structured coefficients when modeling seasonal intensity with state-level spatial dependence (ϕ_j). None of the spatially structured state coefficient distribution were significant.

770
772

Validation to CDC surveillance data

773

774

775

776

777

778

779

780

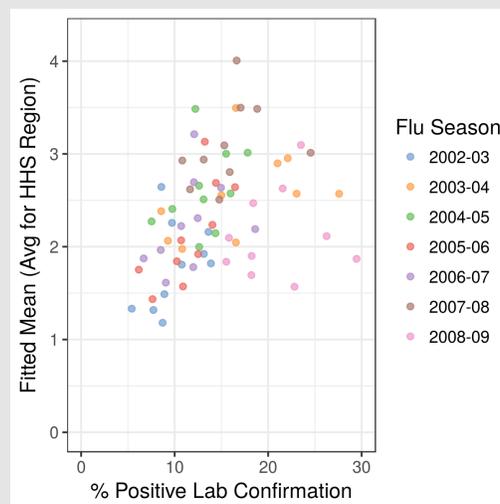
781

782

783

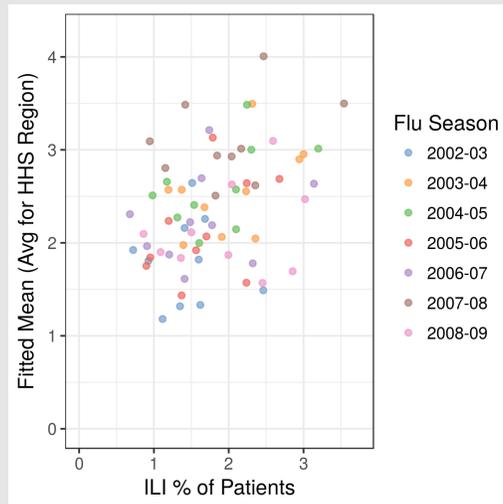
We collected a) the percentage of ILI out of all patient visits among the total population, and child and adult populations as reported by CDC's ILINet, and b) the percentage of positive influenza laboratory confirmations as reported by CDC laboratory surveillance. We note that child and adult ILI percentage was calculated with a denominator of patient visits across all age groups due to limited data availability. Both CDC surveillance systems were reported at the HHS region level and aggregated cumulatively for each flu season in our study period. We then examined scatterplots and Pearson cross-correlation coefficients (double-sided test where $H_0 =$ no difference) between the mean model fits (where we took the mean across all counties in a given HHS region) and each CDC surveillance dataset.

784



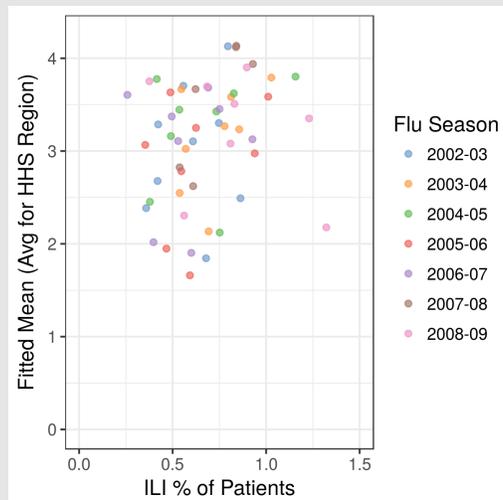
785
786
787
788

Appendix 1 Figure 4. Mean model fit averaged across counties in a given HHS region vs. percentage of positive influenza laboratory confirmations in a given HHS region and flu season. The Pearson cross-correlation coefficient was 0.35 with a p-value of 0.003 for a double-sided hypothesis test.



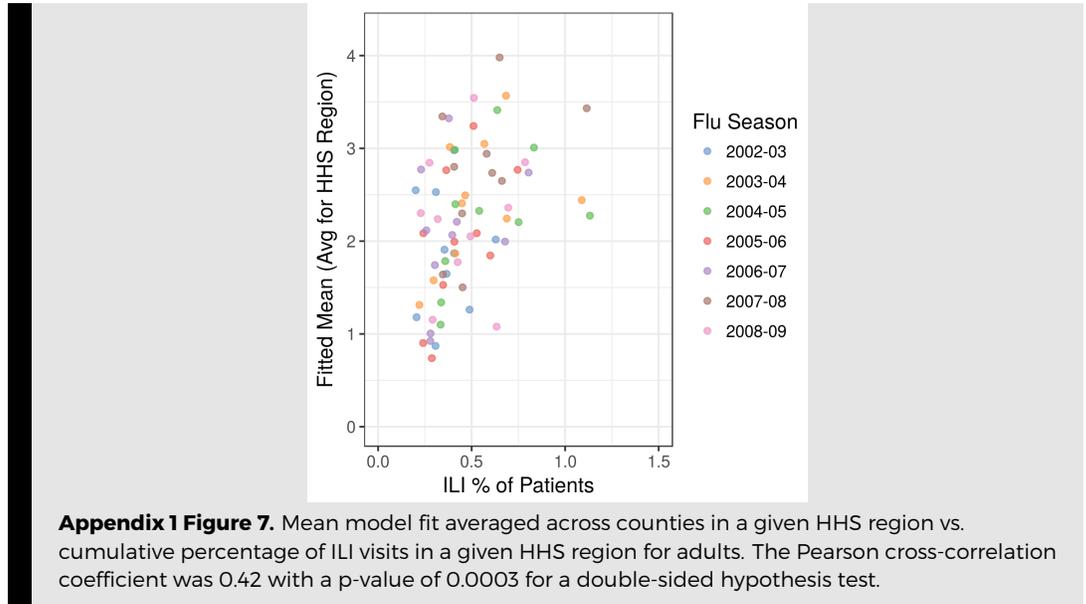
790
791
792
793

Appendix 1 Figure 5. Mean model fit averaged across counties in a given HHS region vs. cumulative percentage of ILI visits in a given HHS region for all age groups. The Pearson cross-correlation coefficient was 0.38 with a p-value of 0.001 for a double-sided hypothesis test.



795
796
797
798

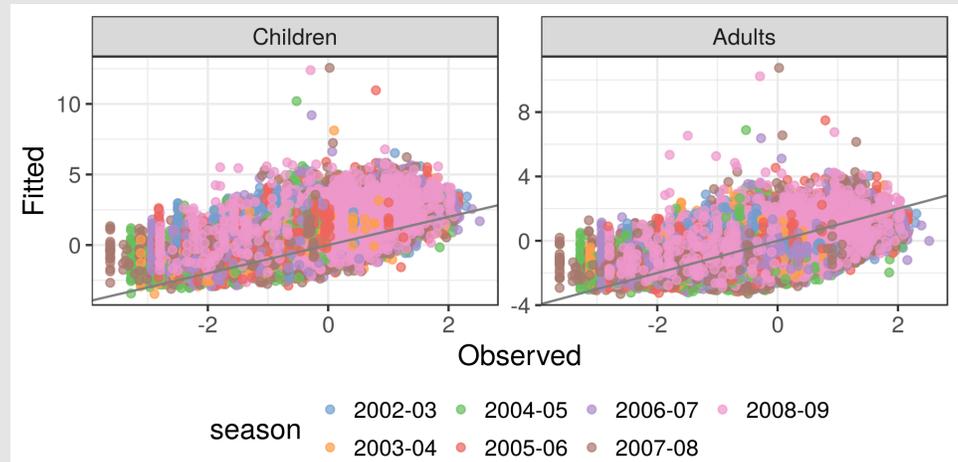
Appendix 1 Figure 6. Mean model fit averaged across counties in a given HHS region vs. cumulative percentage of ILI visits in a given HHS region for children. The Pearson cross-correlation coefficient was 0.42 with a p-value of 0.0002 for a double-sided hypothesis test.



805 **Appendix 2**

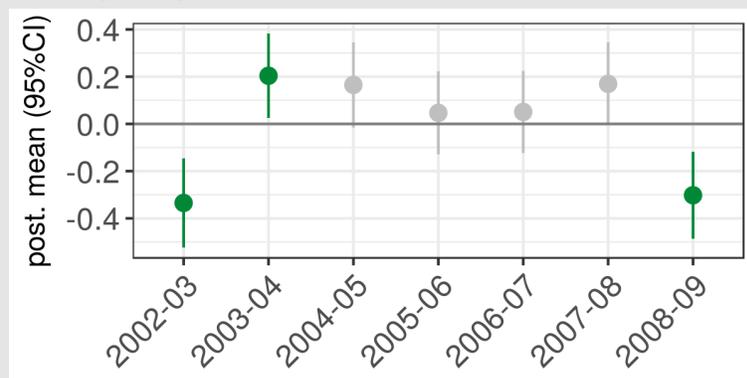
806 **Age-specific drivers of seasonal intensity**

807 **Model Fit**

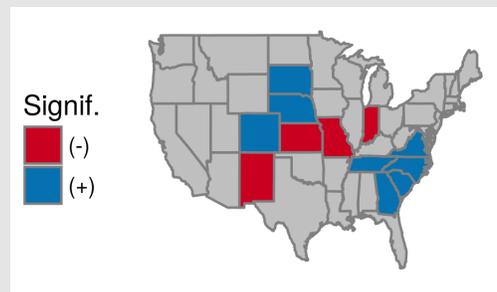


808 **Appendix 2 Figure 1.** Comparison of observed and predicted relative risk of seasonal intensity
809 across flu seasons from 2002-2003 through 2008-2009 for children and adults.
810

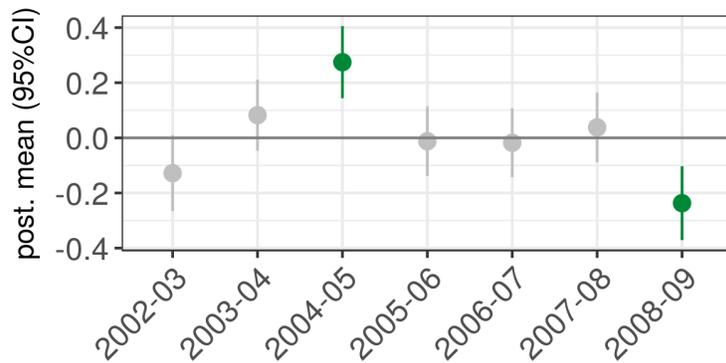
812 **Spatial and temporal patterns**



813 **Appendix 2 Figure 2.** Temporal group effects for seasonal intensity among children. 95%
814 credible interval for flu season coefficients in child population seasonal intensity.
815

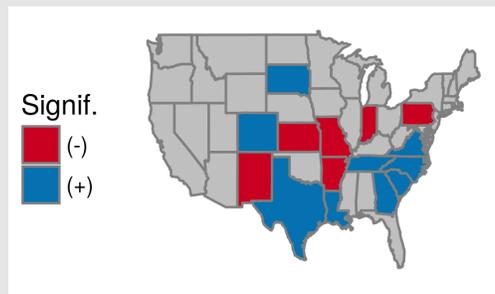


817 **Appendix 2 Figure 3.** Spatial group effects for seasonal intensity among children. Continental
818 U.S. maps highlighting states with significantly greater or lower child seasonal intensity.
819
820



821
822
823

Appendix 2 Figure 4. Temporal group effects for seasonal intensity among adults. 95% credible interval for flu season coefficients in adult population seasonal intensity.



825
826
827

Appendix 2 Figure 5. Spatial group effects for seasonal intensity among adults. Continental U.S. maps highlighting states with significantly greater or lower adult seasonal intensity.

829

Socio-environmental and measurement drivers

830

831

832

833

834

835

836

837

838

839

840

841

842

843

844

845

846

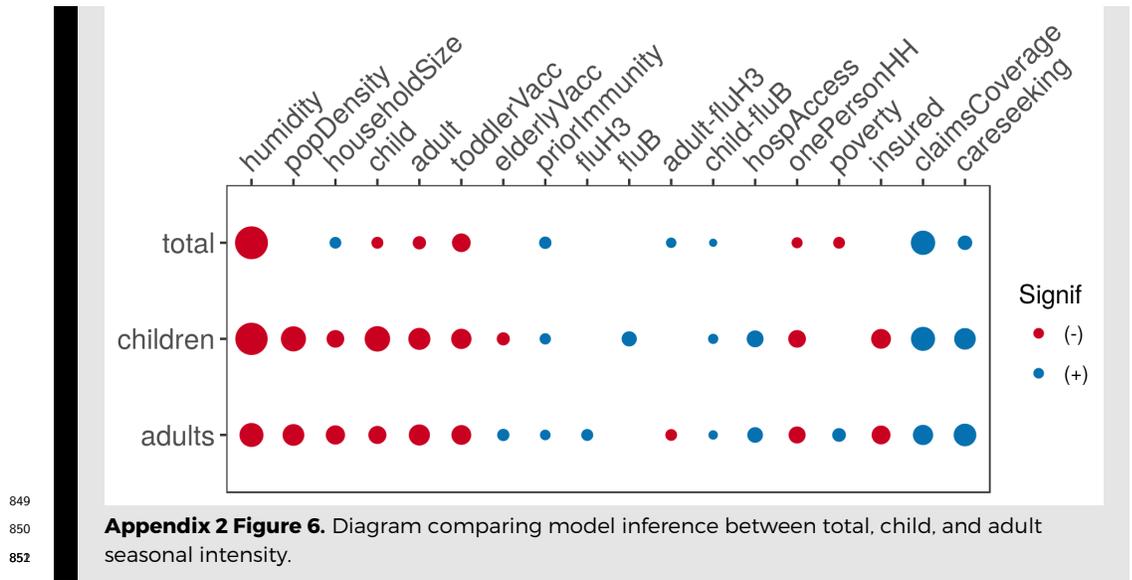
847

848

In reference to the total seasonal intensity results, the child and adult models shared the same significant positive associations for the interaction term between child population and influenza B circulation and a proxy for prior immunity, and the same significant negative associations for adult and child population sizes, average flu season specific humidity, proportion of single person households, and infant vaccination coverage. The child and adult models shared a positive association with hospitals per capita where the total population model had no effect, and a negative association with estimated average household size where the total population model had a positive effect.

Child population seasonal intensity had a unique positive association with influenza B circulation and a unique negative association with elderly vaccination coverage. Adult population seasonal intensity had unique positive associations with H3 circulation among influenza A, proportion of the population in poverty, and elderly vaccination coverage, and a unique negative association with the interaction between adult and influenza H3.

Similar to the total population models, the child and adult seasonal intensity models had significant positive associations with careseeking behavior and claims database coverage. However, both the child and adult seasonal intensity models had significant negative associations with proportion of the population with health insurance, where the total population model demonstrated no effect.



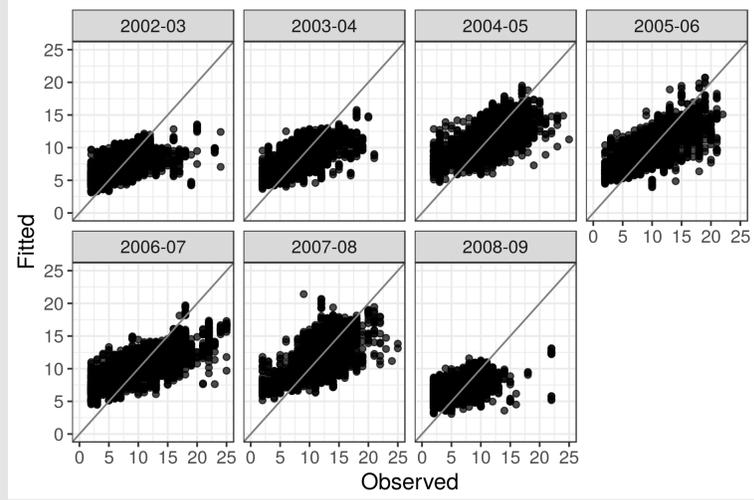
853 **Appendix 3**

854

Drivers of epidemic duration

855

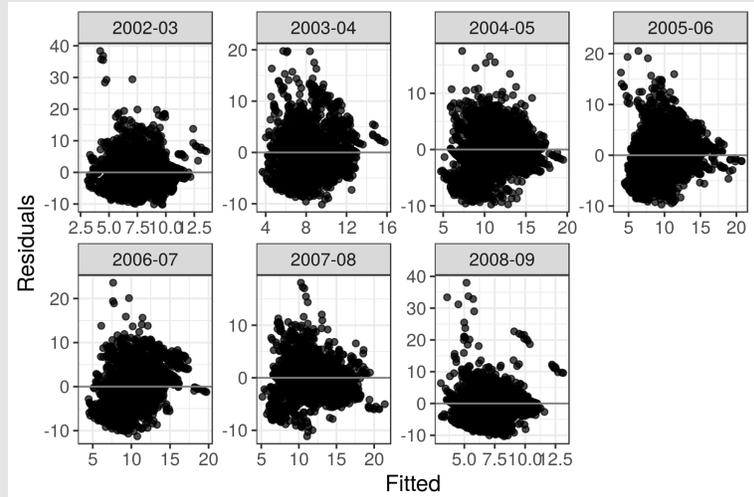
Model fit



856

857

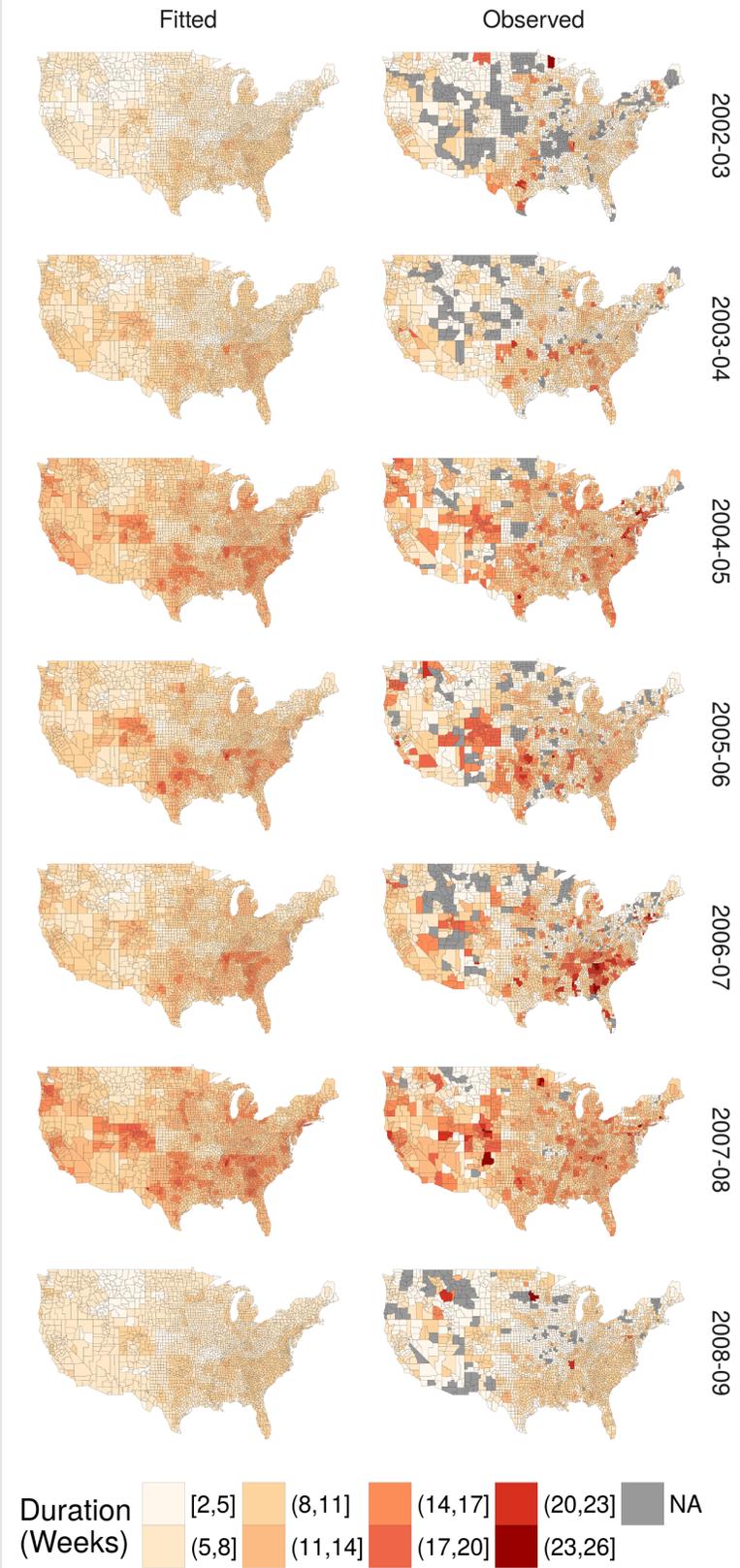
Appendix 3 Figure 1. Observed versus fitted values for epidemic duration.



859

860

Appendix 3 Figure 2. Residuals versus fitted values for epidemic duration.

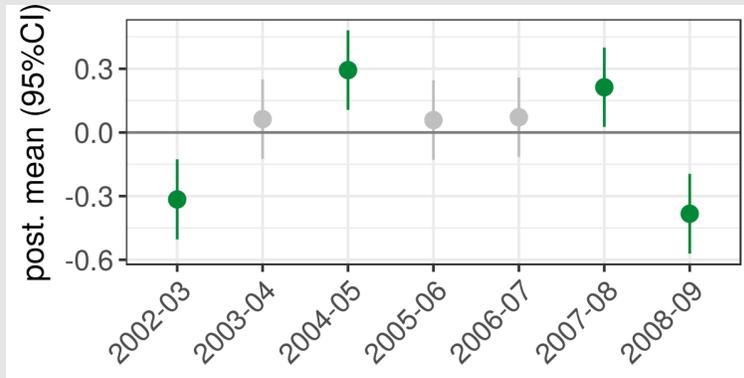


862
863
865

Appendix 3 Figure 3. Continental U.S. county maps for fitted (left) and observed (right) epidemic duration in weeks from 2002-03 through 2008-09.

866

Spatial and temporal patterns

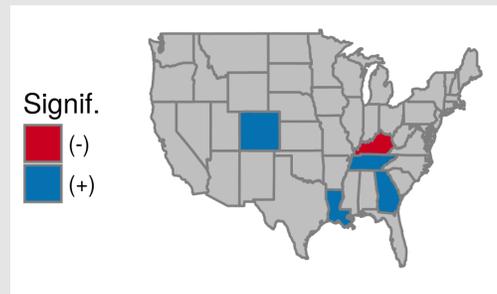


867

868

869

Appendix 3 Figure 4. Temporal group effects for influenza-like illness. 95% credible interval for flu season coefficients in epidemic duration.



871

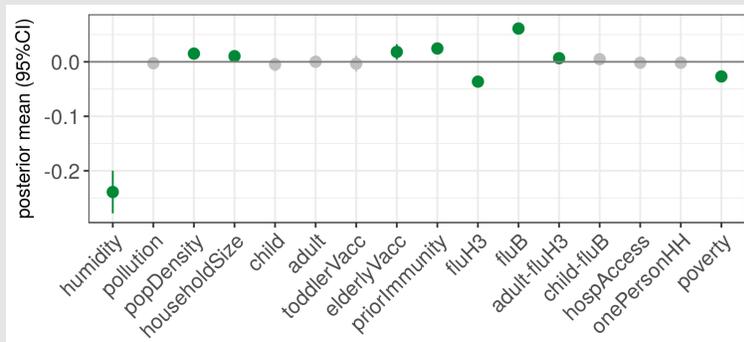
872

873

Appendix 3 Figure 5. Spatial group effects for influenza-like illness. Continental U.S. map highlighting states with significantly longer or shorter epidemic durations.

875

Socio-environmental and measurement drivers



876

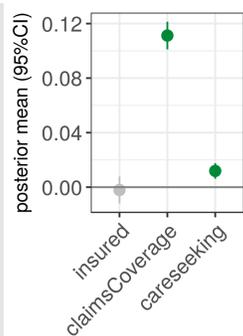
877

878

880

Appendix 3 Figure 6. For the total population multi-season epidemic duration models, these are the 95% credible intervals for the posterior distributions of the socio-environmental coefficients and B) measurement-related coefficients.

881
882
883

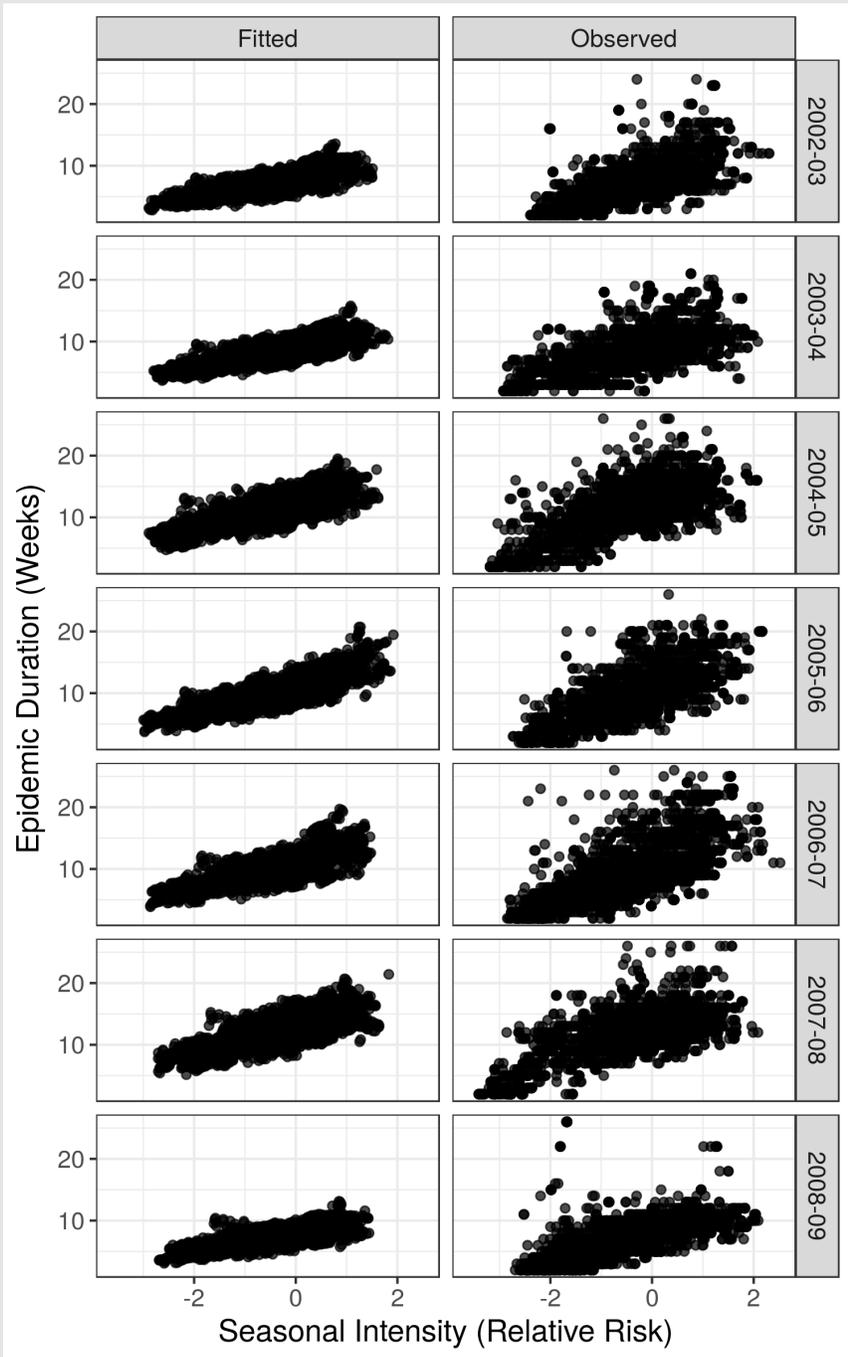


Appendix 3 Figure 7. For the total population multi-season epidemic duration models, these are the 95% credible intervals for the posterior distributions of measurement-related coefficients.

885 **Appendix 4**

886

Comparison of disease burden metrics



887

888

889

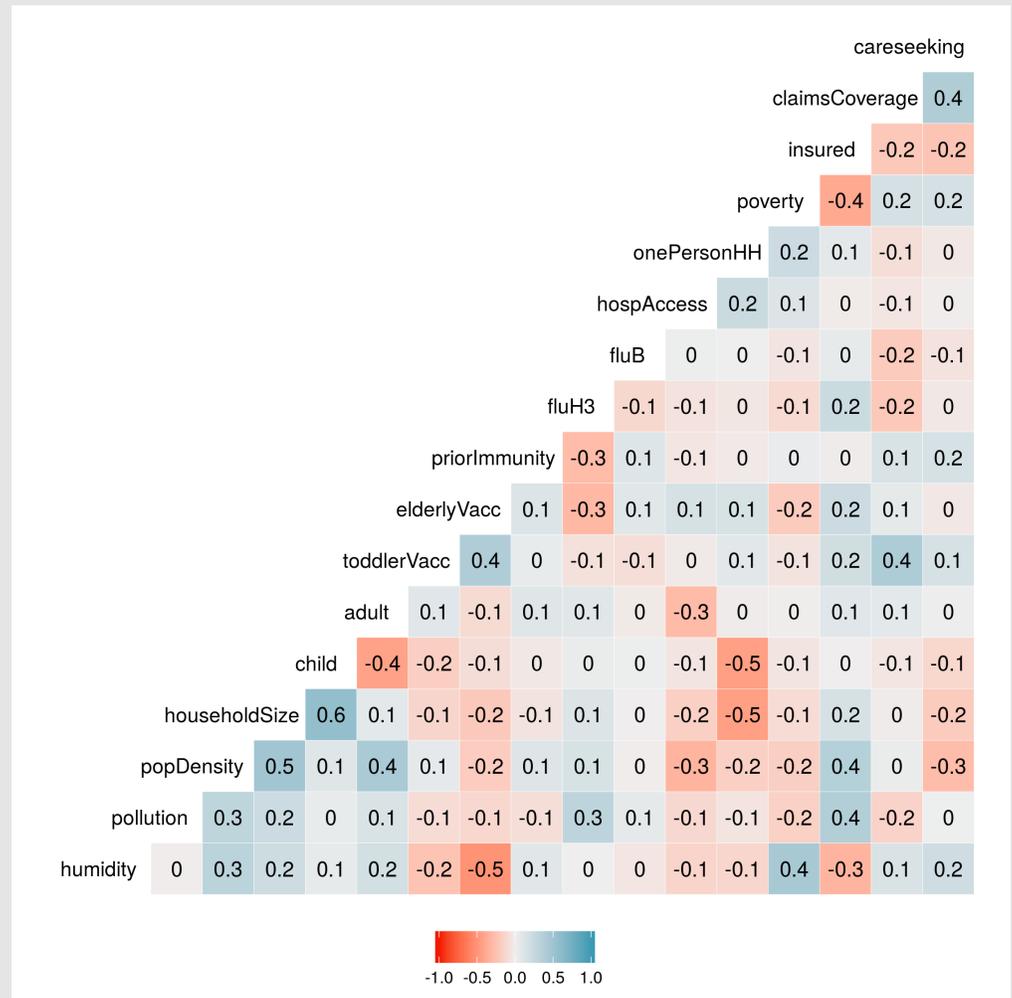
Appendix 4 Figure 1. Comparison of epidemic duration and relative risk for seasonal intensity among fitted (left) and observed (right) values.

891 **Appendix 5**

892 **Model predictors**

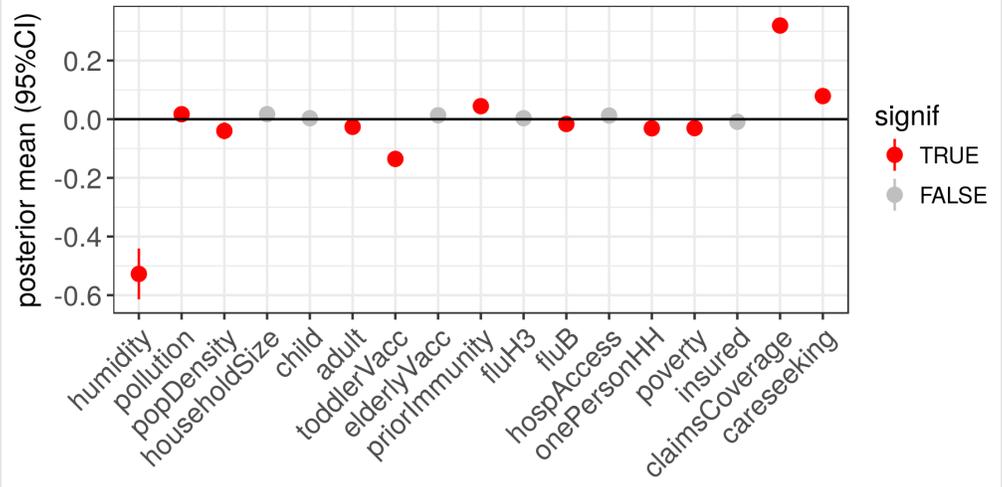
893 **Checks for multicollinearity**

894 We checked for multicollinearity among predictors by examining Spearman rank cross-
 895 correlation coefficients between all pairs of final model predictors (excluding interac-
 896 tion terms). No single pair had a linear correlation coefficient that exceeded a magni-
 897 tude of 0.6.



898 **Appendix 5 Figure 1.** Spearman rank cross-correlation matrix for all pairs of final model
 899 predictors.
 900

902 Additionally, we ran our multi-season seasonal intensity model with each coefficient
 903 individually. Multicollinearity between predictors may sometimes be detected when a
 904 predictor significantly deviates from zero in the single predictor model, but does not
 905 appear to have an effect in a multivariate context. Some predictors (pollution, popDen-
 906 sity, fluB) that were significant in the single predictor context no longer had an effect
 907 in our complete model (and vice versa for householdSize and child). Nevertheless, all
 908 of these predictors had small effect sizes in both single and multivariate models, and
 909 the other predictors that were significant in both models retained effect sizes with the
 910 same order of magnitude and directionality.



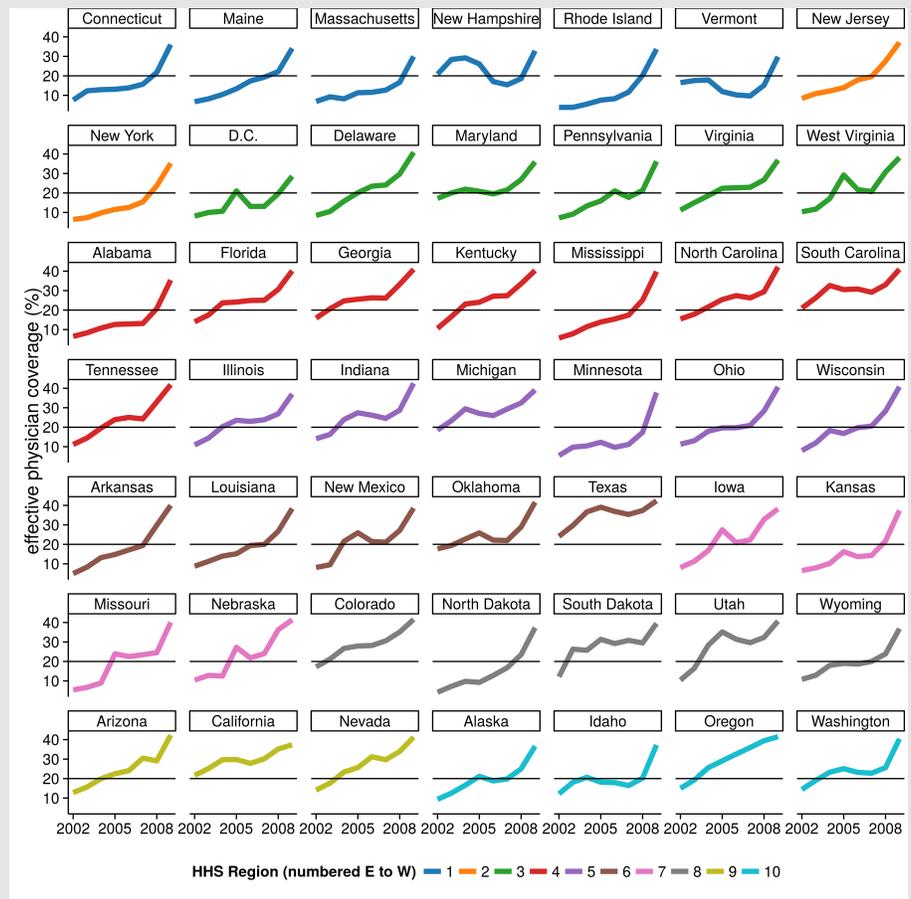
911
912
913

Appendix 5 Figure 2. These are the 95% credible intervals among multi-season models with a single predictor for seasonal intensity.

915
916

Medical claims coverage

Medical claims database coverage increased over time across each state.



917
918
919
920

Appendix 5 Figure 3. Medical claims database coverage by year and state. Colors represent states that belong to the same HHS region. The black horizontal line at 20% effective physician coverage is a visual guide to ease the comparison of data across panels.

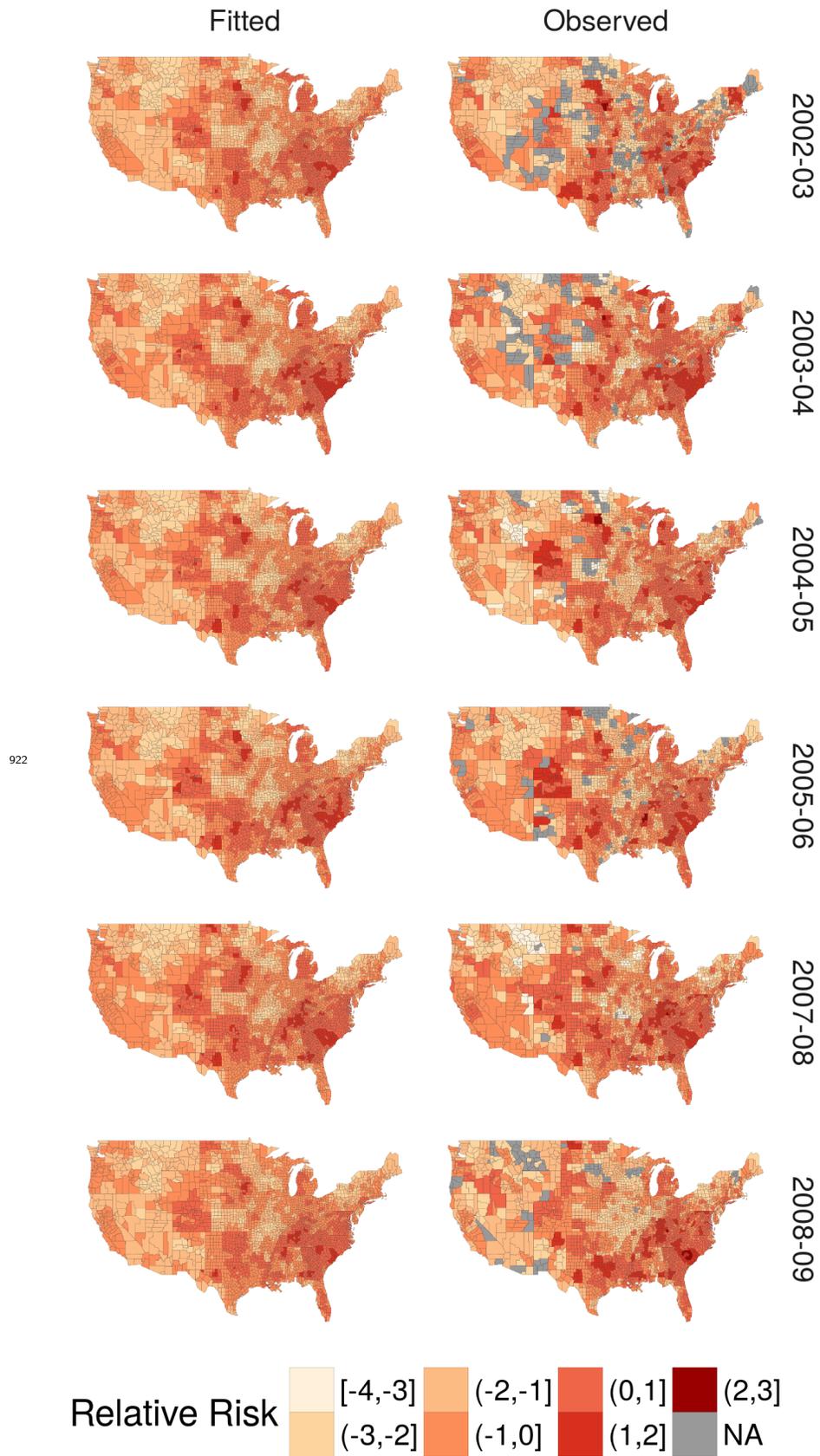


Figure 1-Figure supplement 1. Continental U.S. county maps for fitted (left) and observed (right) relative risk of seasonal intensity for remaining influenza seasons.

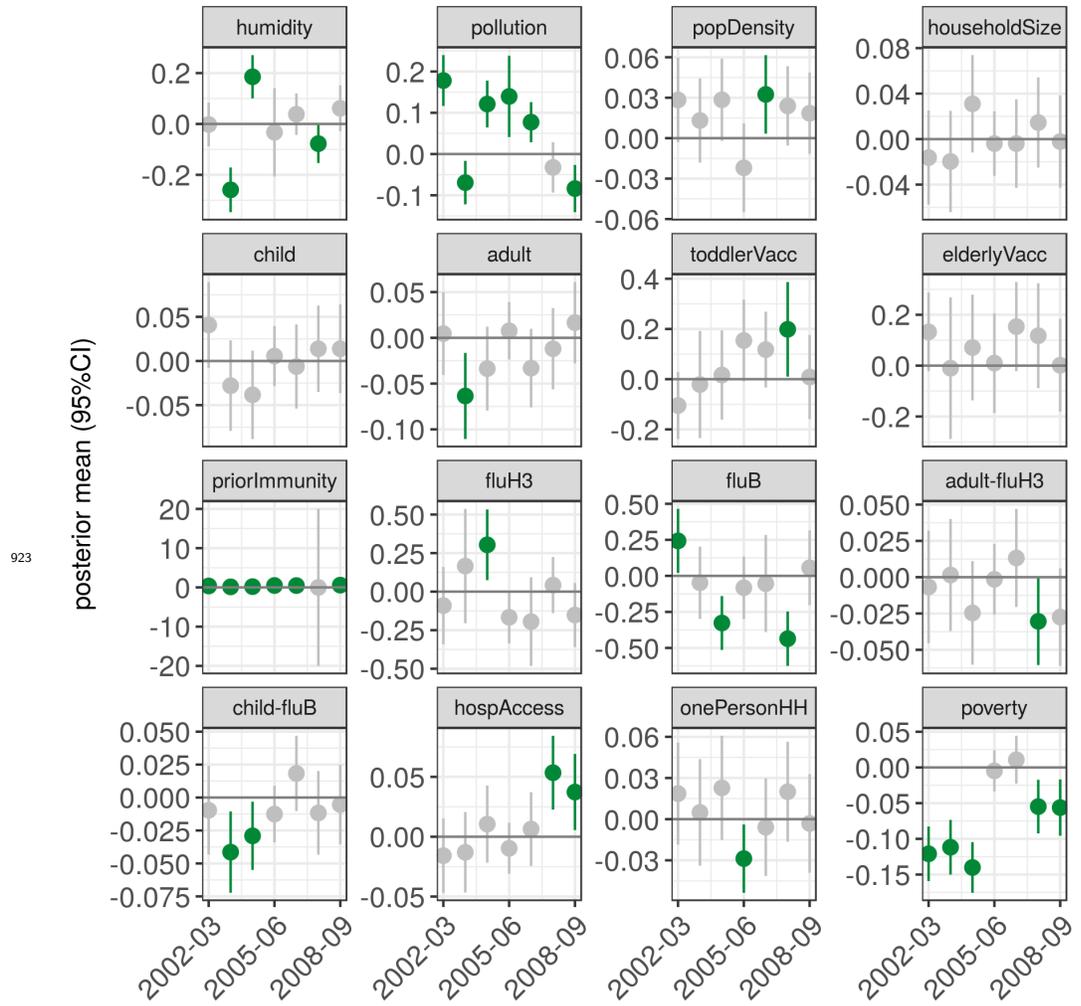


Figure 3-Figure supplement 1. For the total population single-season seasonal intensity models, these are the 95% credible intervals for the posterior distributions of the socio-environmental coefficients.

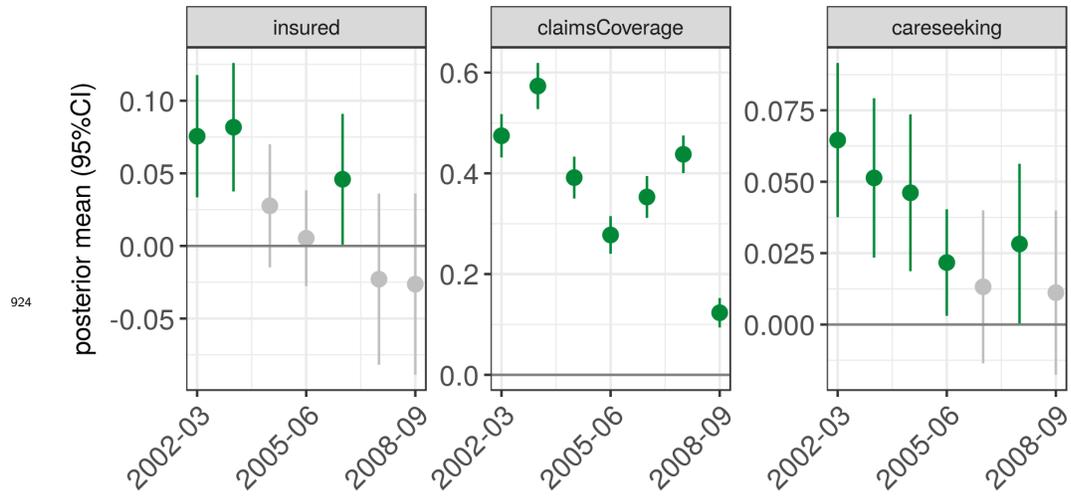


Figure 3-Figure supplement 2. For the total population single-season seasonal intensity models, these are the 95% credible intervals for the posterior distributions of the measurement coefficients.

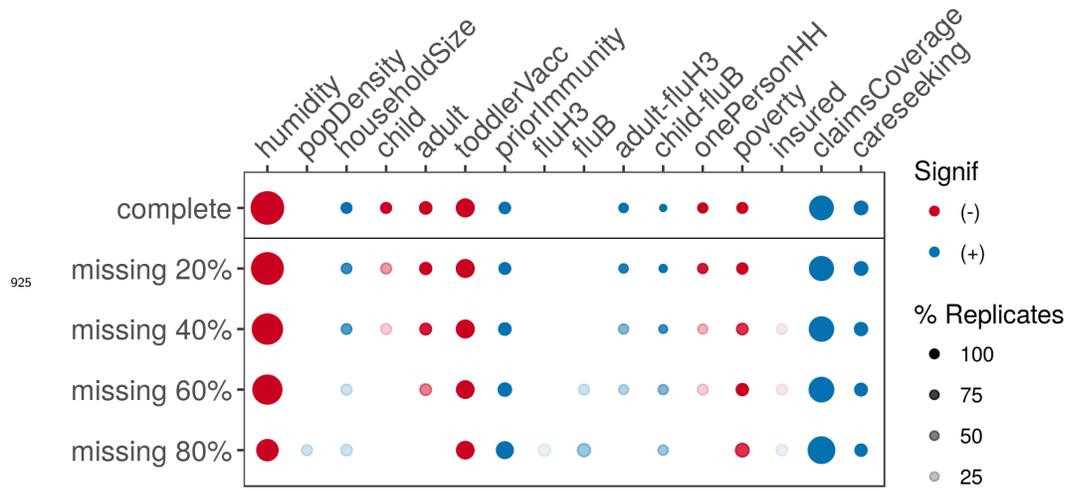


Figure 4-Figure supplement 1. Diagram indicating changes to model inference as fewer moving-location sentinels reported data.

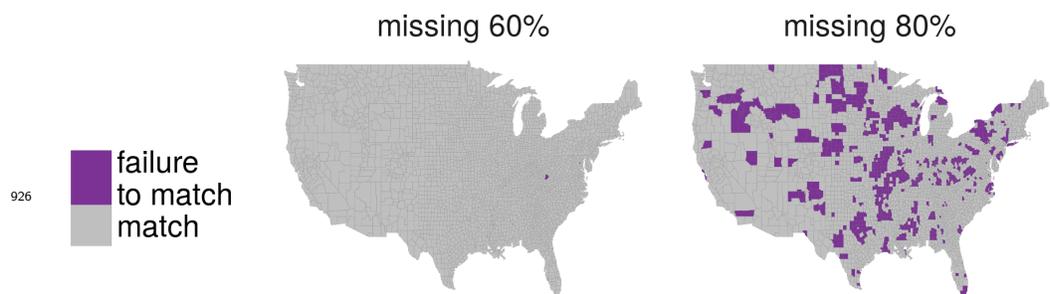


Figure 4-Figure supplement 2. Map of model prediction match between the complete model and the 60% and 80% missing levels for moving-location sentinels.

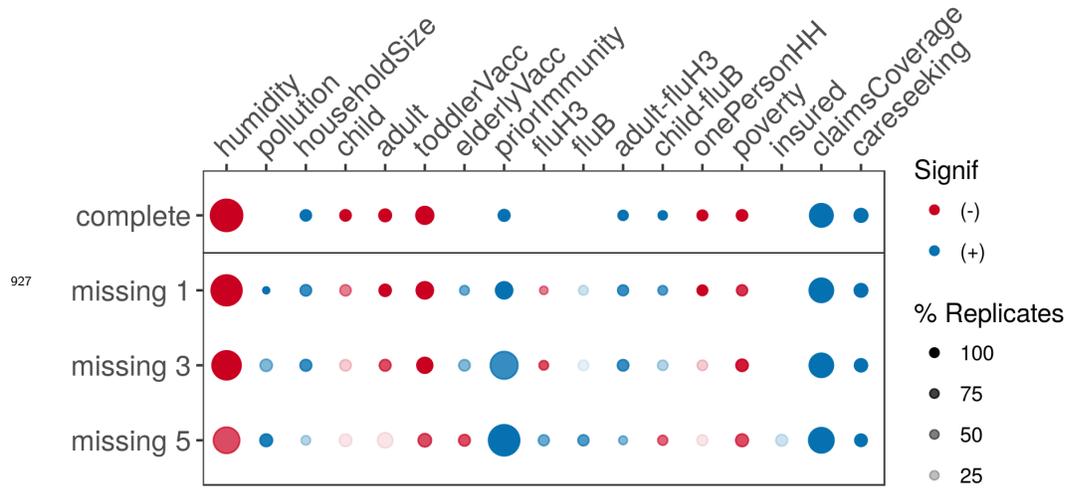


Figure 4-Figure supplement 3. Diagram indicating changes to model inference as historical seasons were randomly removed from the model.

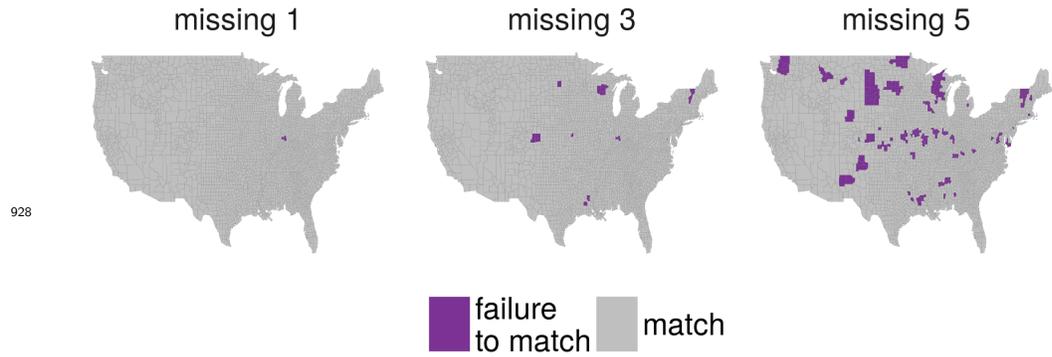


Figure 4-Figure supplement 4. Map of model prediction match between the complete model and models missing one, three, or five historical flu seasons.