



Using MSA-Specific Bayesian Predictive Distributions to Detect Outbreaks of Influenza-Like Illness

Owen J. Devine¹, Howard Burkom²

¹ Centers for Disease Control and Prevention, Atlanta, Georgia

² Johns Hopkins University, Baltimore, Maryland

SAFER • HEALTHIER • PEOPLE™

Background

ILINet – Surveillance network comprised of volunteer health care providers who provide CDC with weekly reports of the proportion of patient visits meeting specified ILI criteria.

Let e_{ij} = Number of visits meeting ILI criteria in MSA i , week j

v_{ij} = Total number of visits in MSA i , week j

Surveillance metric of interest is

$$r_{ij} = \frac{e_{ij}}{v_{ij}}$$

SAFER • HEALTHIER • PEOPLE™

Background

Comparison of observed r_{ij} to values of the proportion associated with “baseline” or non-influenza periods can lead to timely identification of MSAs exceeding baseline.

Issues:

1. Definition of baseline periods.

Weeks when percentage of laboratory confirmed influenza specimens are $\leq 10\%$ for the Census Region containing the MSA.

2. Estimation of variability of observed r_{ij} during these baseline periods

SAFER • HEALTHIER • PEOPLE™

Background

Current approach:

Let \bar{r}_i^B and s_i^B be the mean and standard deviation of the observed ILI proportions in geographic area i during weeks meeting the baseline criteria over three previous years.

If r_{ij} is the current observed proportion in week j , then flag this observation if r_{ij} is at least $k * s_i^B$ higher than \bar{r}_i^B , i.e.

$$\frac{(r_{ij} - \bar{r}_i^B)}{s_i^B} \geq k$$

SAFER • HEALTHIER • PEOPLE™

Background

Problems:

1. s_i^B can be highly unstable for areas with few visits
2. r_{ij} can vary during baseline periods due to changing composition of types of providers reporting (ER, Pediatric, General Practice, etc.)
3. Reliance on arbitrary cut-off values

Method not currently below Influenza Surveillance Regions

SAFER • HEALTHIER • PEOPLE™

Background

Solutions: Frequentist

More robust criteria for r_{ij} to be flagged to address stability issues

Adjust the observed value of r_{ij} to reflect provider mix in \bar{r}_i^B

Burkom et. al. , *Analyzing ILINet Data for Increased Spatial Resolution*

SAFER • HEALTHIER • PEOPLE™

Background

Solution: Bayesian

Model the process generating the observed e_{ij}^B including impact of time varying mix of provider types

Borrow strength across MSAs allowing estimation at finer geographic resolution

SAFER • HEALTHIER • PEOPLE™

Model

Recall e_{ik}^B is the count of ILI visits for MSA i in baseline week j and v_{ik}^B is the total number of visits that week

$$e_{ij}^B \sim f(e_{ij}^B \mid \mu_{ij}^B, v_{ij}^B, \phi)$$

$$\mu_{ij}^B \sim g(\bar{\mu}_{ij}^B, \delta)$$

$$\ln(\bar{\mu}_{ij}^B) = \ln(\mu_i^B) + \beta X_{ij} + \Gamma Z_{ij}$$

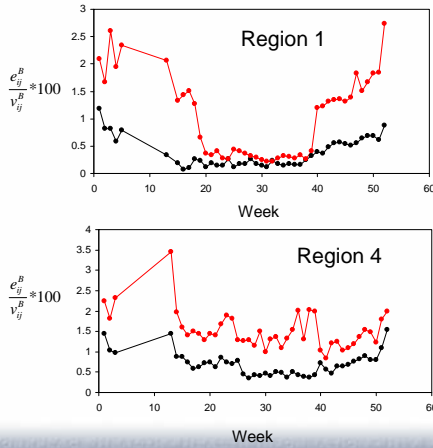
$$\ln(\mu_i^B) \sim h(\theta)$$

$$e_{ij}^B \sim f(e_{ij}^B \mid \mu_{ij}^B, v_{ij}^B, \phi) \left. \begin{array}{l} \text{Poisson} \\ \text{Zero-Inflated Poisson (ZIP)} \\ \text{Negative Binomial} \\ \text{Zero-Inflated Negative Binomial (ZINB)} \end{array} \right\}$$

SAFER • HEALTHIER • PEOPLE™

Model

Adjusting for variable provider mix



— ILI proportion among ER, Infectious Disease and Pediatric providers, 2005-2007
 — ILI proportion among other provider types, 2005-2007

p_{ij} = Proportion of visits reported by ER, Infectious Disease and Pediatric providers in MSA i week j

SAFER • HEALTHIER • PEOPLE™

Candidate Models: Poisson

$$e_{ij}^B \sim \text{Poisson}(\mu_{ij}^B v_{ij}^B)$$

$$\ln(\mu_{ij}^B) \sim N(\bar{\mu}_{ij}^B, \delta^2)$$

$$\ln(\bar{\mu}_{ij}^B) = \ln(\mu_i^B) + \beta_1 p_{ij}$$

$$\ln(\bar{\mu}_{ij}^B) = \ln(\mu_i^B) + \beta_1 p_{ij} + \gamma_{ij}$$

RE Poisson

$$\gamma_{ij} \sim N(0, \sigma^2) \quad \text{or} \quad \gamma_{ij} \sim t(0, \sigma^2, 3)$$

SAFER • HEALTHIER • PEOPLE™

Candidate Models: ZIP and ZIP RE

$$e_{ij}^B \sim \pi_{ij} * 0 + (1 - \pi_{ij}) * \text{Poisson}(\mu_{ij}^B v_{ij}^B)$$

$$\text{logit}(\pi_{ij}) = \alpha_0 + \alpha_1 I(p_{ij} > 0)$$

$$\ln(\mu_{ij}^B) \sim N(\bar{\mu}_{ij}^B, \delta^2)$$

where

$$\ln(\bar{\mu}_{ij}^B) = \ln(\mu_i^B) + \beta_1 p_{ij} \quad \text{ZIP}$$

$$\ln(\bar{\mu}_{ij}^B) = \ln(\mu_i^B) + \beta_1 p_{ij} + \gamma_{ij} \quad \text{ZIP RE}$$

SAFER • HEALTHIER • PEOPLE™

Model Evaluation – Posterior Predictive Checks

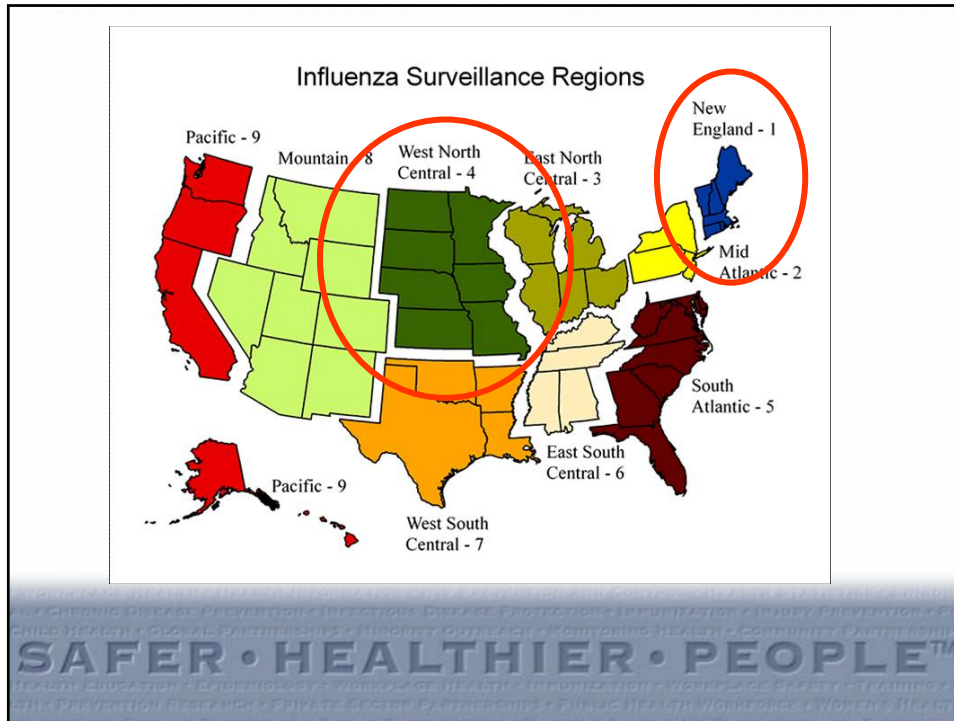
Let e_{ik}^{B-gen} be a generated datum sampled under the assumed model

$$\lambda^2 \text{ Discrepancy} \quad P \left[\frac{[e_{ik}^B - E(e_{ik}^B)]^2}{\text{var}(e_{ik}^B)} \geq \frac{[e_{ik}^{B-gen} - E(e_{ik}^B)]^2}{\text{var}(e_{ik}^B)} \right]$$

$$\text{Variance/Mean Ratio} \quad P \left[\frac{\sum (e_{ik}^B - \bar{e}_{ik}^B)^2}{\bar{e}_{ik}^B} \geq \frac{\sum (e_{ik}^{B-gen} - \bar{e}_{ik}^{B-gen})^2}{\bar{e}_{ik}^{B-gen}} \right]$$

Also compared deviance and number of generated zero ILI count weeks

SAFER • HEALTHIER • PEOPLE™



Model Evaluation

Region 1: 12 MSAs reporting

	<u>Poisson</u>	<u>Poisson RE¹</u>	<u>ZIP</u>	<u>ZIP RE¹</u>
χ^2 Discrepancy	1.0	0.75	1.0	0.74
Variance/Mean Ratio	1.0	0.49	1.0	0.54
Deviance	7962	3627	6564	3455
$e_{ik}^{B_gen} = 0$ (%) (Observed = 36%)	(17, 21)	(27, 31)	(33, 36)	(31, 35)

¹ RE - t (3)



Model Evaluation

Region 4: 23 MSAs reporting

	Poisson	Poisson RE ¹	ZIP	ZIP RE ¹
χ^2 Discrepancy	1.0	0.39	1.0	0.44
Variance/Mean Ratio	1.0	0.31	1.0	0.38
Deviance	10430	5286	9141	5151
$e_{ik}^{B-gen} = 0$ (%) (Observed = 50%)	(36, 40)	(43, 47)	(45, 48)	(45, 48)

¹ RE - Normal

SAFER • HEALTHIER • PEOPLE™

Using posterior predictive distributions to evaluate observed rates

$$e_{ij}^B \sim \pi_{ij} * 0 + (1 - \pi_{ij}) * \text{Poisson}(\mu_{ij}^B v_{ij}^B)$$

$$\text{logit}(\pi_{ij}) = \alpha_0 + \alpha_1 I(p_{ij} > 0)$$

$$\ln(\mu_{ij}^B) \sim N(\bar{\mu}_{ij}^B, \delta^2)$$

$$\ln(\mu_{ij}^B) = \ln(\mu_i^B) + \beta_1 p_{ij} + \gamma_{ij}$$

$$\gamma_{ij} \sim T(0, \sigma^2, 3)$$

Posterior distributions for $\mu_i^B, \alpha_0, \alpha_1, \beta, \delta^2, \sigma^2$ derived using observed e_{ik}^B for 2005-2007

SAFER • HEALTHIER • PEOPLE™

Using posterior predictive distributions to evaluate observed rates

Let e_{ij} be the observed ILI count out of v_{ij} total visits in MSA i week j in 2008 with an observed proportion (p_{ij}) of v_{ij} reported by ER, Infectious Disease and Pediatric providers.

Generate

$$h(e_{ij}^{B-gen} | e_{ij} = e_{ij}^B, p_{ij}, v_{ij}, \mu_i^B, \alpha_0, \alpha_1, \beta, \delta^2, \sigma^2)$$

Observed

Posterior Estimates

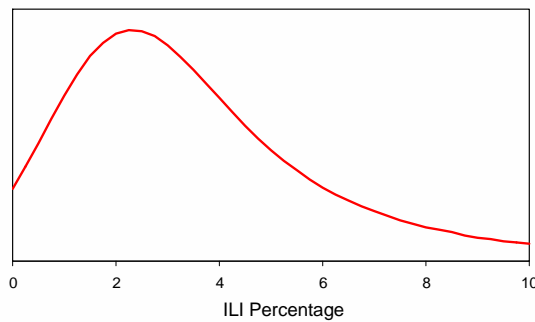
Estimate

$$P[E_{ij} \geq e_{ij} | e_{ij} = e_{ij}^B, p_{ij}, v_{ij}, \mu_i^B, \alpha_0, \alpha_1, \beta, \delta^2, \sigma^2]$$

SAFER • HEALTHIER • PEOPLE™

Using posterior predictive distributions to evaluate observed rates

$$h(e_{ij}^{B-gen} | e_{ij} = e_{ij}^B, p_{ij}, v_{ij}, \mu_i^B, \alpha_0, \alpha_1, \beta, \delta^2, \sigma^2)$$

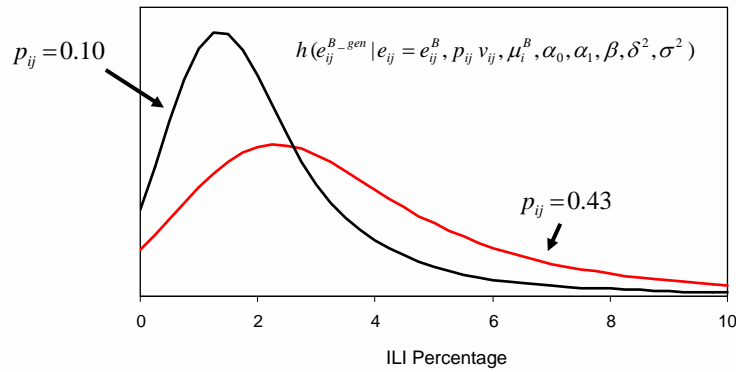


$$v_{ij} = 10,811, e_{ij} = 240 \Rightarrow \left(\frac{e_{ij}}{v_{ij}} \right) * 100 = 2.22 \text{ with } p_{ij} = 0.43$$

SAFER • HEALTHIER • PEOPLE™

Effect of Proportion of Visits From ER, Infectious Disease and Pediatric Providers

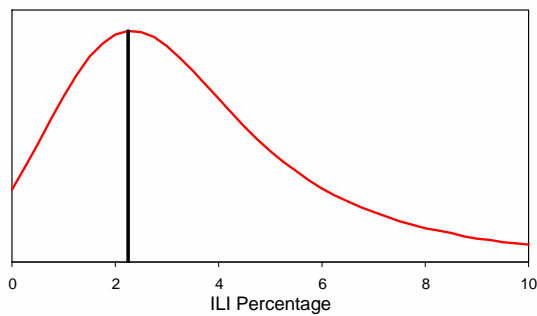
$$v_{ij} = 10,811, e_{ij} = 240 \Rightarrow \left(\frac{e_{ij}}{v_{ij}} \right) * 100 = 2.22$$



SAFER • HEALTHIER • PEOPLE™

Using posterior predictive distributions to evaluate observed rates

$$h(e_{ij}^{B-gen} | e_{ij} = e_{ij}^B, p_{ij}, v_{ij}, \mu_i^B, \alpha_0, \alpha_1, \beta, \delta^2, \sigma^2)$$

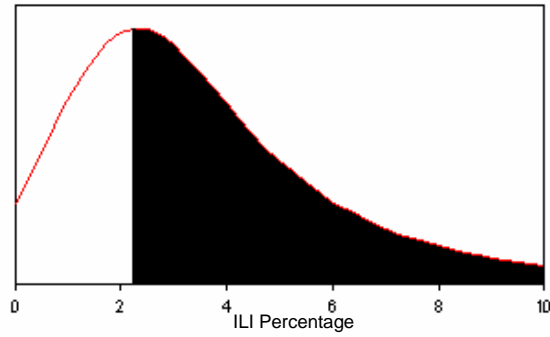


$$v_{ij} = 10,811, e_{ij} = 240 \Rightarrow \left(\frac{e_{ij}}{v_{ij}} \right) * 100 = 2.22 \text{ with } p_{ij} = 0.43$$

SAFER • HEALTHIER • PEOPLE™

Using posterior predictive distributions to evaluate observed rates

$$h(e_{ij}^{B-gen} | e_{ij} = e_{ij}^B, p_{ij}, v_{ij}, \mu_i^B, \alpha_0, \alpha_1, \beta, \delta^2, \sigma^2)$$

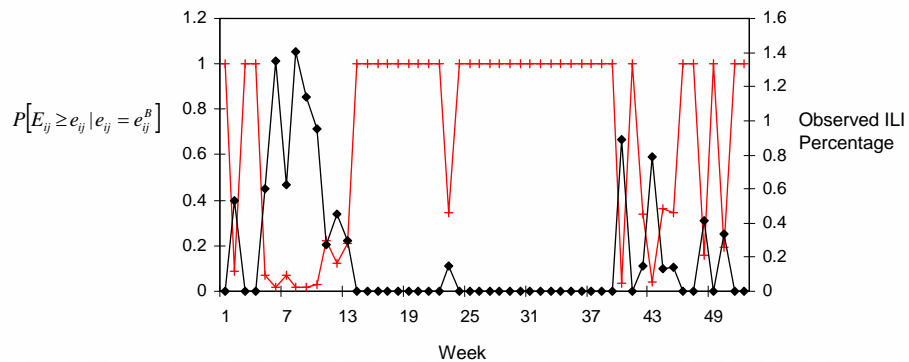


Estimated probability of observing a proportion of 0.22 or higher if this MSA was in the baseline state = 0.67.

SAFER • HEALTHIER • PEOPLE™

Using posterior predictive distributions to evaluate observed rates

MSA = 730, Year = 2008



SAFER • HEALTHIER • PEOPLE™

Advantages of Approach:

- Borrowed strength across MSAs allowing estimation of area-specific baseline ILI visit distribution.
- Continuous measure corresponding to likelihood of being in baseline state
- Not too “Black Boxy”
- Updating estimates of baseline distributions is periodic and inherent in Bayesian approach

SAFER • HEALTHIER • PEOPLE™

Future work:

- Develop models for all regions
- Develop meaningful metrics for classification as “above” baseline.
- Incorporate time trends and patient age distribution in model
- Design simulation-based assessment for comparison to current methods
- Evaluate using 2009 ILINet data
- Evaluate impact of uncertainty in current definition of baseline weeks (HMM)

SAFER • HEALTHIER • PEOPLE™