Modeling emergency department visit patterns for infectious disease complaints: results and application to disease surveillance

by Judith C. Brillman, Tom Burr, David Forslund, Edward Joyce, Rick Picard, and Edith Umland
BMC Medical Informatics and Decision Making 2005, 5:4
http://www.biomedcentral.com/1472-6947/5/4

Presented by Christopher Maier
INLS 279: Bioinformatics Research Review
2006-03-29

Public Health Surveillance

- "ongoing, systematic collection, analysis, interpretation, and dissemination of data about a health-related event for use in public health action to reduce morbidity and mortality and to improve health"

- Supporting Case Detection and Public Health Interventions
- Estimating the Impact of Disease or Injury
- Portraying the Natural History of a Health Condition
- Determining the Distribution and Spread of Illness
- Generating Hypotheses and stimulating Research
- Evaluating Prevention and Control Measures
- And...

See http://www.cdc.gov/mmwr/preview/mmwrhtml/rr5305a1.htm for more

Outbreak Detection

- Outbreaks are defined as “increase in frequency of disease above the background occurrence of the disease”
- Traditionally detected by examining collected case reports or by clinicians observing clusters of disease incidence and issuing alerts
- This can be slow, however
- Availability of electronic data provides new possibilities for quicker detection
- The threat of bioterrorism provides strong motivation

Syndromic Surveillance

- Uses pre-diagnosis health data (i.e. monitor the symptoms for characteristic patterns). Helpful for bioterrorism since the diseases caused by bioterror agents are rare, often misdiagnosed, and overlap in their presentation

- If collected electronically, can provide rapid response relative to traditional means (wait for diagnosis, or discharge evaluation)

- "Drop-In" systems denote short-term solutions, such as those put in place to detect bioterror threats for DNC, RNC, Olympics, immediate post-9/11 NYC, etc.
    - Often difficult to maintain long term because they require care providers to collect non-routine information, instead of using pre-existing data
    - Also, have no baseline knowledge

See http://www.cdc.gov/ncidod/EID/vol9no3/02-0363.htm

Using Emergency Department (ED) Data

- Main question to be answered: can emergency department data be used in syndromic surveillance to achieve more timely indicators of outbreaks?

Categorization: Defining the Syndromes

- Important task for any syndromic surveillance system
- Took chief complaints (CCs) from 9 years of records
  - recorded by triage nurse, entered into system by clerk
- Looked at all CCs that occurred over 5 times in the corpus, condensed and grouped them
- Groupings reviewed by medical experts

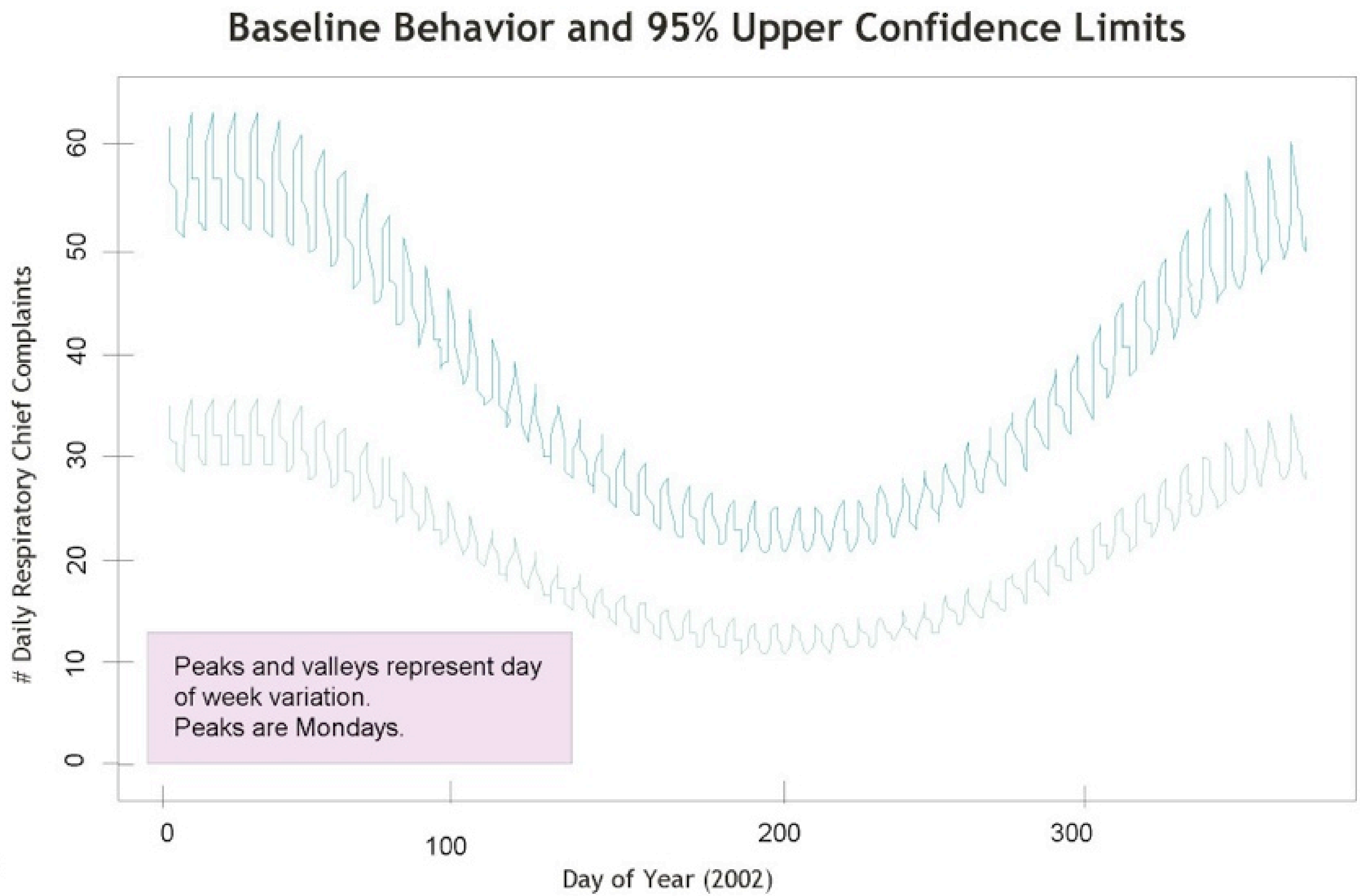
Table 1: B-SAFER dictionary for matching chief complaints to body systems

Respiratory	Gastro-intestinal	Neurologic	Skin	Lympatic	UDI (undifferentiated infection)
Breath	Abdominal Pain	Altered mental status	Abscess	Arm pit	Achy
Bronchiolitis	Abdomen/back pain	Anxious	Abnormal Skin	Glands	Body aches
Chest congestion	Abdomen pain	Confusion	Blisters	Lumps	Body sores
Chest pain	Abdominal cramps	Difficulty Talking	Bug Bites	Lumps in neck	Fatigue
Cold Congested	Abdominal pain	Difficulty thinking	Cellulitis	Neck	Fever
Congestion	Blood in stool	Difficulty Walking	Chicken pox	Nodes	Fussy
Cough	Diarrhoea	Disoriented	Dermatitis	Red streaks	Infection
Croup	Food poisoning	Drowsy	Insect bite	Skin streaks	Tired
Flu	Hepatitis	Facial droop	Itching	Weak	
Headache	Jaundice	Facial weakness	Pox Rash		
Laryngitis	Stomach pain	Hyper	Skin redness		
Pneumonia	Vomit	Loss of consciousness	Skin swelling		
Respiratory	Nausea	Mental	Tick bite		
Sinus	Non responsive	Nervous			
Stuffy nose		Numbness			
Throat		Paralysis			
		Seizure			
		Slurred speech			
		Sores			
		Stroke			
		Swallowing			
		Syncope			
		Thinking slow			
		Tingling			
		Trouble talking			
		Trouble thinking			
		Trouble walking			
		Unresponsive Weak			

Model

- Use "started log" counts
- Equation reflects weekly, seasonal periodic patterns in ED data

$$S(d) = [\sum_i c_i \times I_i(d)] + [c_8 + c_9 \times d] + [c_{10} \times \cos(2\pi d/365.25) + c_{11} \times \sin(2\pi d/365.25)]$$


Baseline Behavior and 95% Upper Confidence Limits
Daily Respiratory Chief Complaints
0
10
20
30
40
50
60
0
100
200
300
Day of Year (2002)
Peaks and valleys represent day of week variation.
Peaks are Mondays.

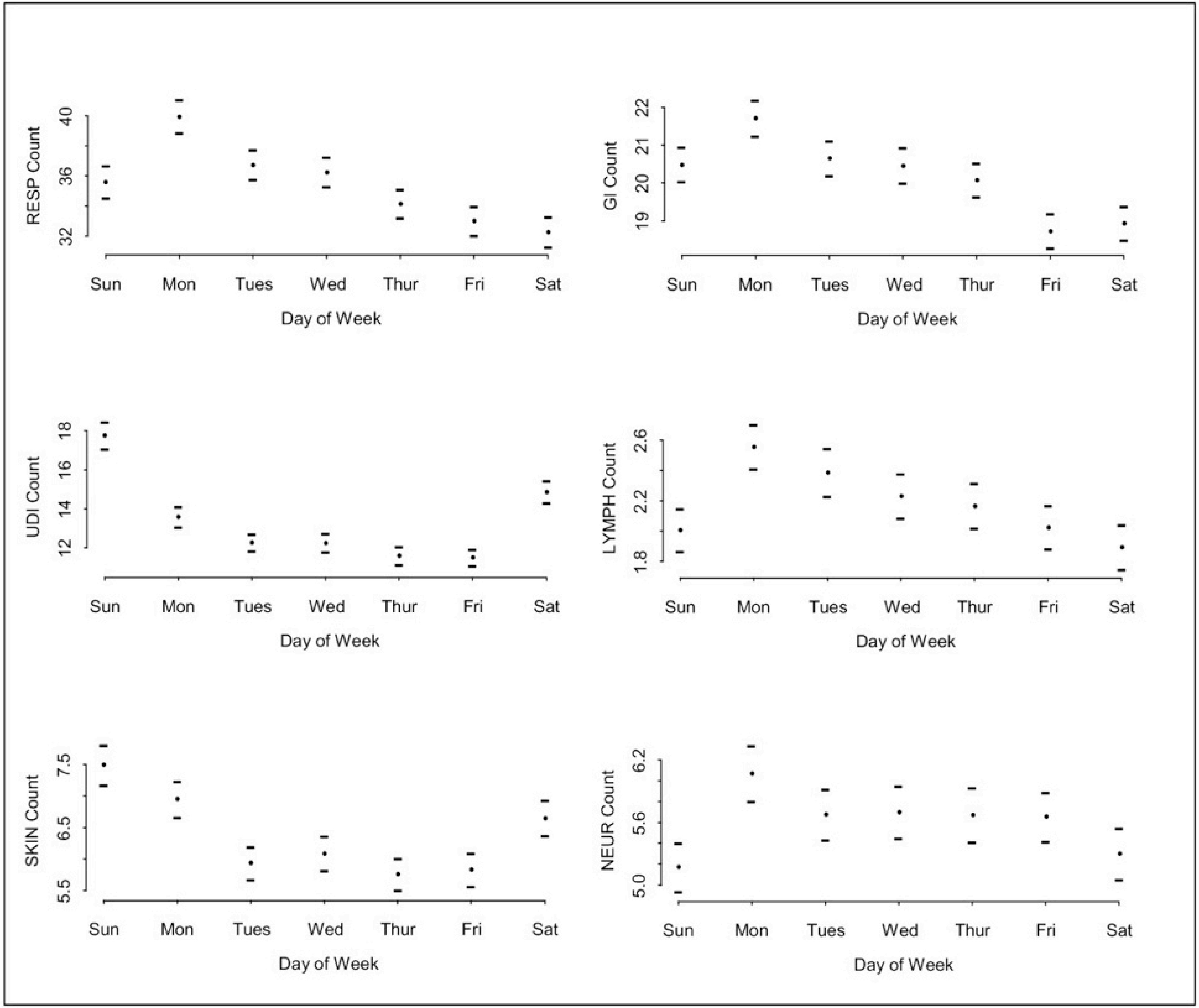
Page's Statistic

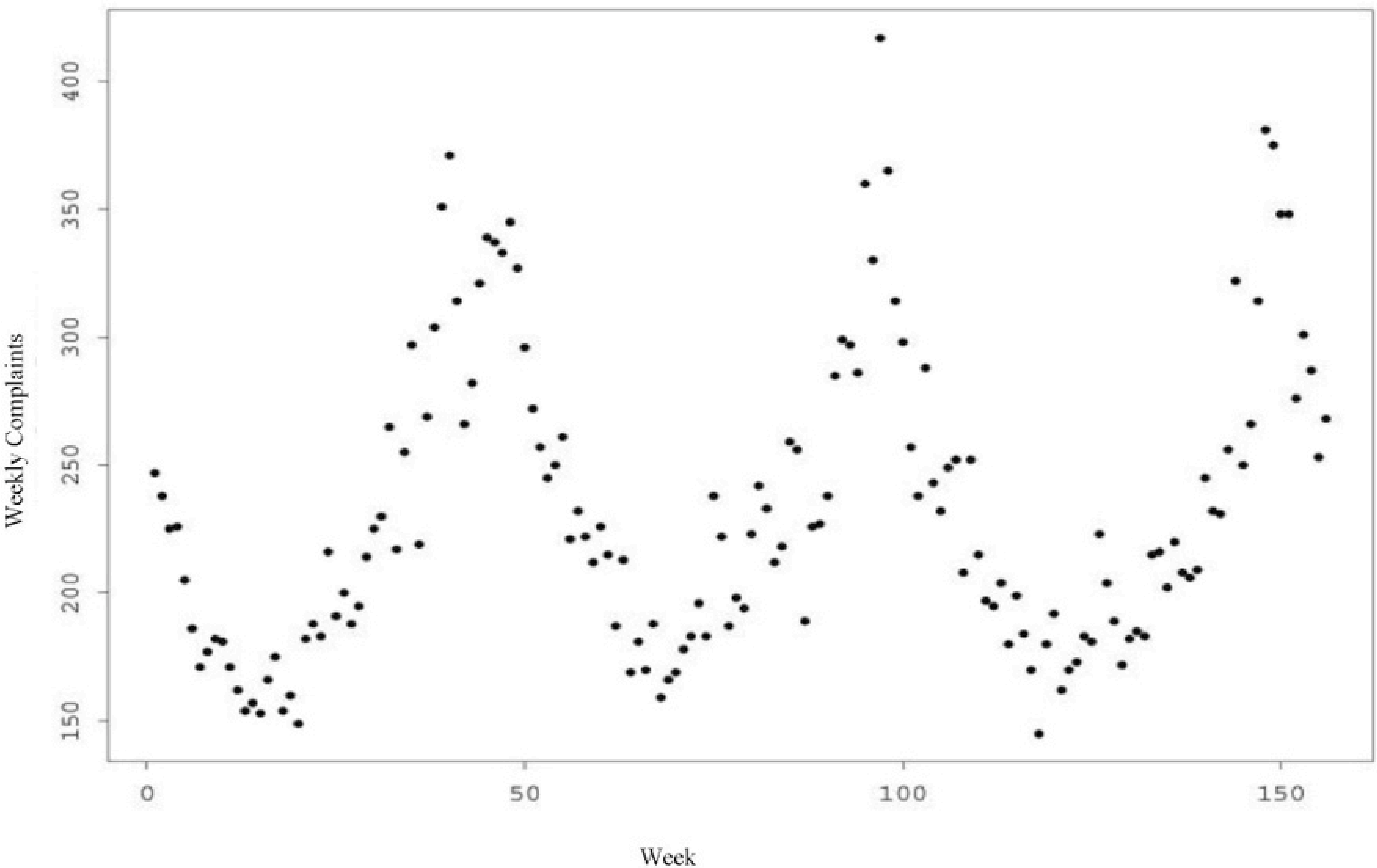
- Need to compare incoming CC counts to baseline model; if it's higher, it could signal an outbreak of some disease or condition

$$P(d) = \text{maximum of } 0 \text{ and } [P(d-1) + \epsilon_d / s_d - 1/2]$$

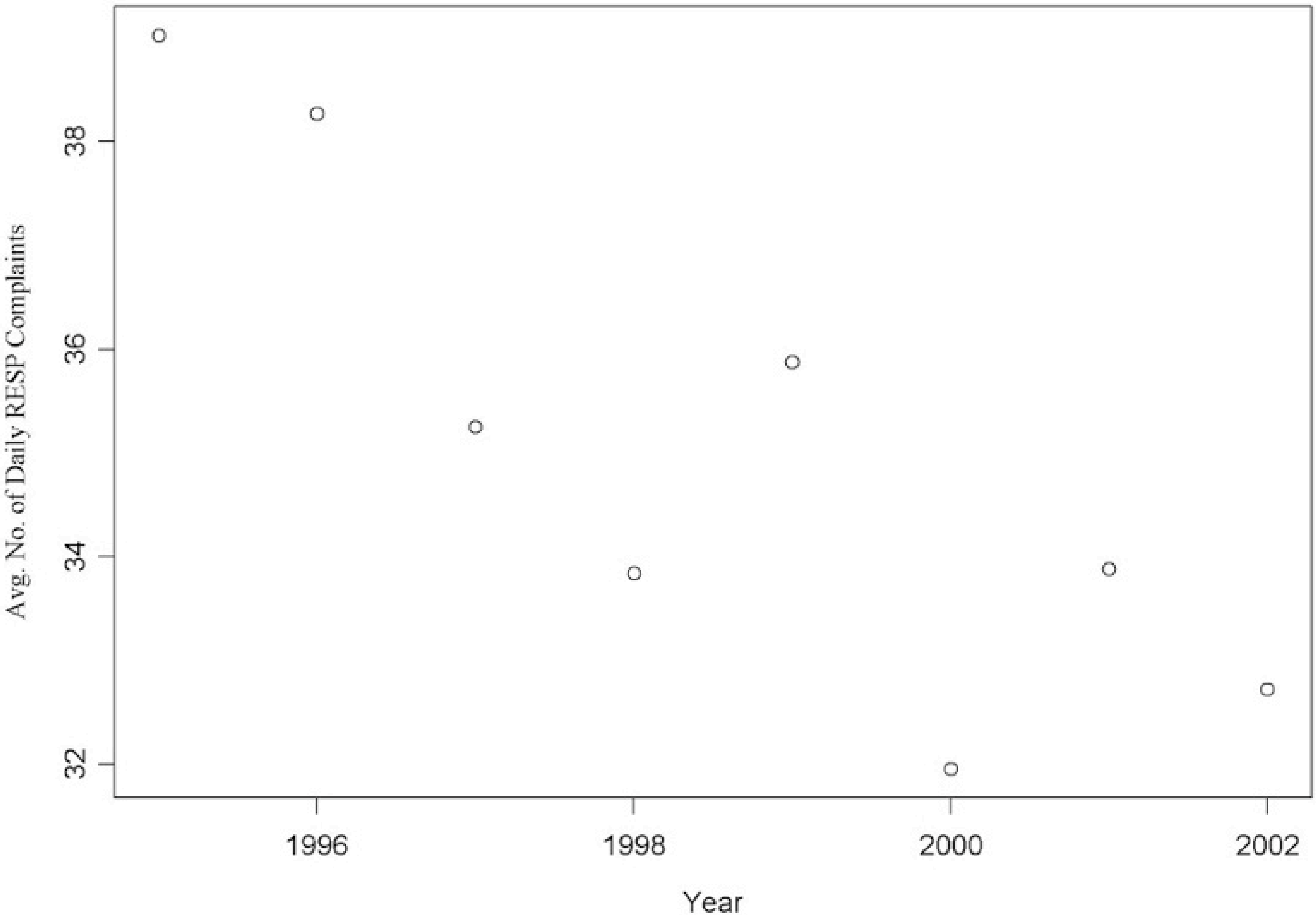
Evaluation: Comparison with traditional surveillance

- Compared respiratory data to influenza data collected by New Mexico Department of Health for 2002–2003 flu season

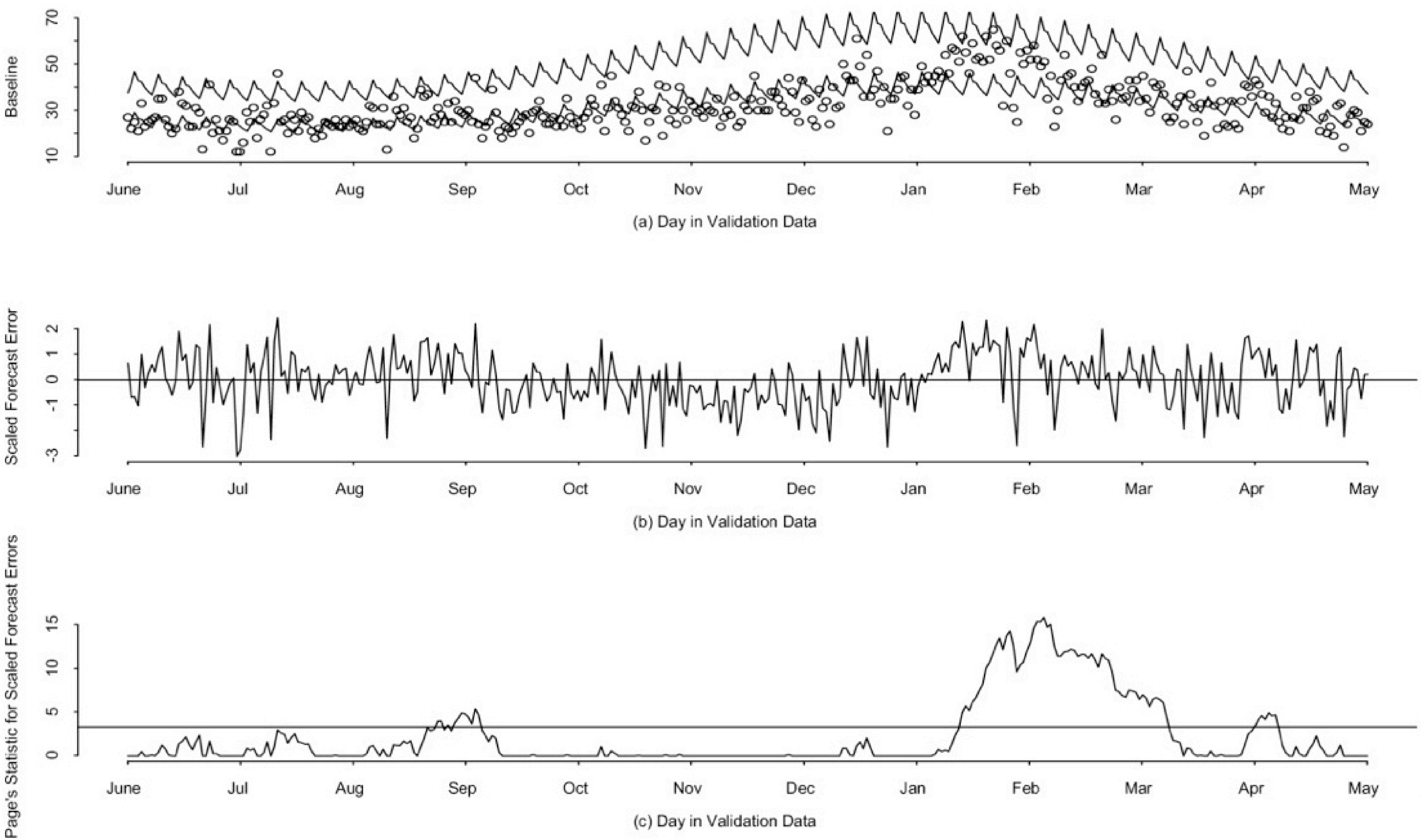

RESP Count
32
36
40
Sun
Mon
Tues
Wed
Thur
Fri
Sat
Day of Week
GI Count
19
20
21
22
Sun
Mon
Tues
Wed
Thur
Fri
Sat
Day of Week
UDI Count
12
14
16
18
Sun
Mon
Tues
Wed
Thur
Fri
Sat
Day of Week
LYMPH Count
1.8
2.2
2.6
Sun
Mon
Tues
Wed
Thur
Fri
Sat
Day of Week
SKIN Count
5.5
6.5
7.5
Sun
Mon
Tues
Wed
Thur
Fri
Sat
Day of Week
NEUR Count
5.0
5.6
6.2
Sun
Mon
Tues
Wed
Thur
Fri
Sat
Day of Week



Weekly Complaints
400
350
300
250
200
150
0
50
100
150
Week



Avg. No. of Daily RESP Complaints
38
36
34
32
1996
1998
2000
2002
Year


Baseline
70
50
30
10
June
Jul
Aug
Sep
Oct
Nov
Dec
Jan
Feb
Mar
Apr
May
(a) Day in Validation Data
Scaled Forecast Error
2
1
0
-1
-3
(b) Day in Validation Data
Page's Statistic for Scaled Forecast Errors
15
10
5
0
(c) Day in Validation Data

Outbreak Simulation

- Add in extra counts to simulate 1-day, 1–10 day, and 2–10 day outbreaks and see if the system can detect them.

Table 2: The fraction of simulations (out of 1000, so the 95% confidence limit is approximately ± 0.03) in which the Page statistic (or the one-at-a-time statistic) exceeded its 2.5% false alarm threshold during the simulated outbreak for the baseline model, the baseline model with residuals modified by EWMA, the null model, and the null model with residuals modified by EWMA.

Outbreak duration	Test	Baseline	Baseline + EWMA	Null	Null + EWMA
1–10 days	Page	0.46	0.21	0.28	0.17
1–10 days	One-at-a-time	0.37	0.31	0.17	0.31
>1 day	Page	0.42	0.20	0.30	0.14
>1 day	One-at-a-time	0.31	0.25	0.13	0.26
1 day	Page	0.44	0.36	0.30	0.45
1 day	One-at-a-time	0.71	0.70	0.47	0.68

Hierarchical Model

- Basic model presented treats all seasons equally, which isn’t necessarily true when dealing with various classes of complaints

- Hierarchical models replace the sine and cosine functions with other Gaussian functions

- Could be used in practice, but is very computationally expensive, requiring frequent updating of the model

- First order model seems to work well enough though, so stick with it

Summary

- Emergency Department information can be used for timely outbreak detection
- Page’s statistic can be used to characterize baseline data
- Simple first order model is sufficiently sensitive to changes in CC counts
- Performance of the system depends on how CCs are recorded, the categories CCs are grouped into, and the method by which CCs are assigned to categories

Stephen Colbert says:
Support your local medical information scientist! Timely outbreak detection is vitally important to our great nation! We cannot afford to wait, because....