NAVAL
POSTGRADUATE
SCHOOL

Assessing EARS' Ability to Locally Detect the 2009 H1N1 Pandemic

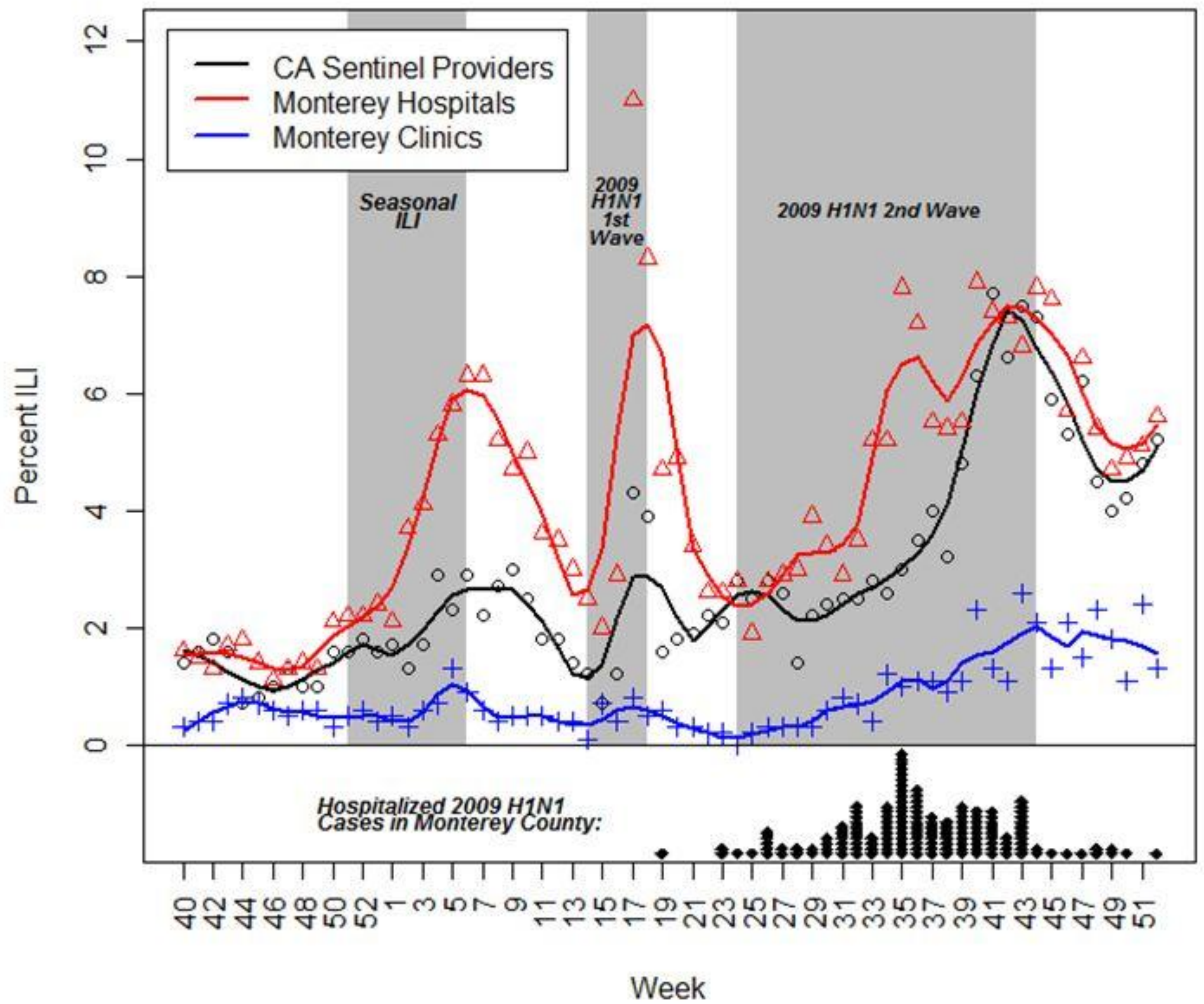
Ron Fricker, Katie Hagen,
Krista Hanni, Susan Barnes, and Kristy Michie

13th Biennial CDC Symposium on Statistical Methods
May 25, 2011

Research Questions

- How well can the Early Aberration Reporting System (v4.5) detect known outbreaks?
- Are there alternatives that improve performance?
  - ILI syndrome definitions
  - Detection algorithms

The Outbreak Periods



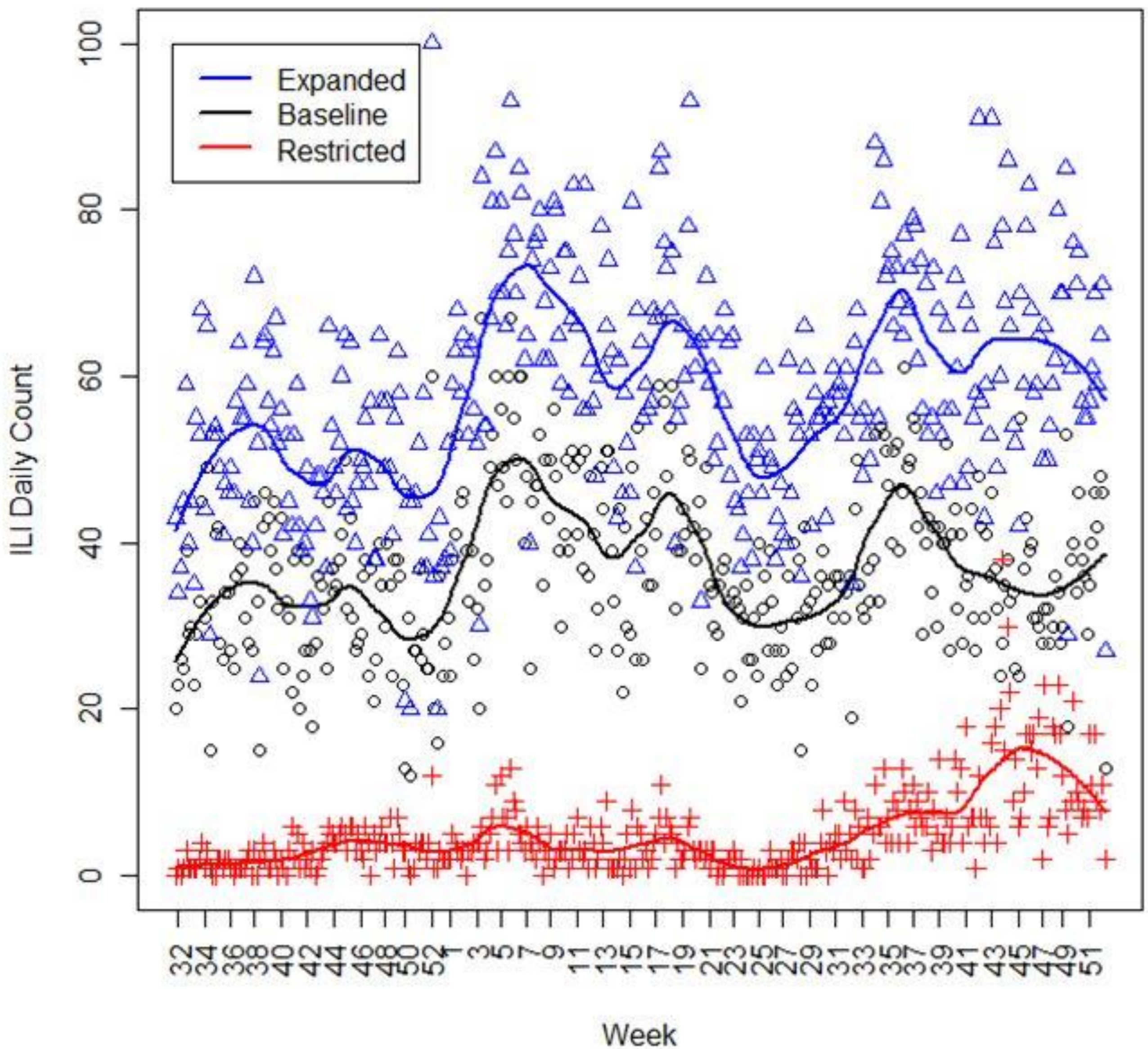
ILI Syndrome Definition Alternatives

- MCHD has used three definitions for ILI syndrome:

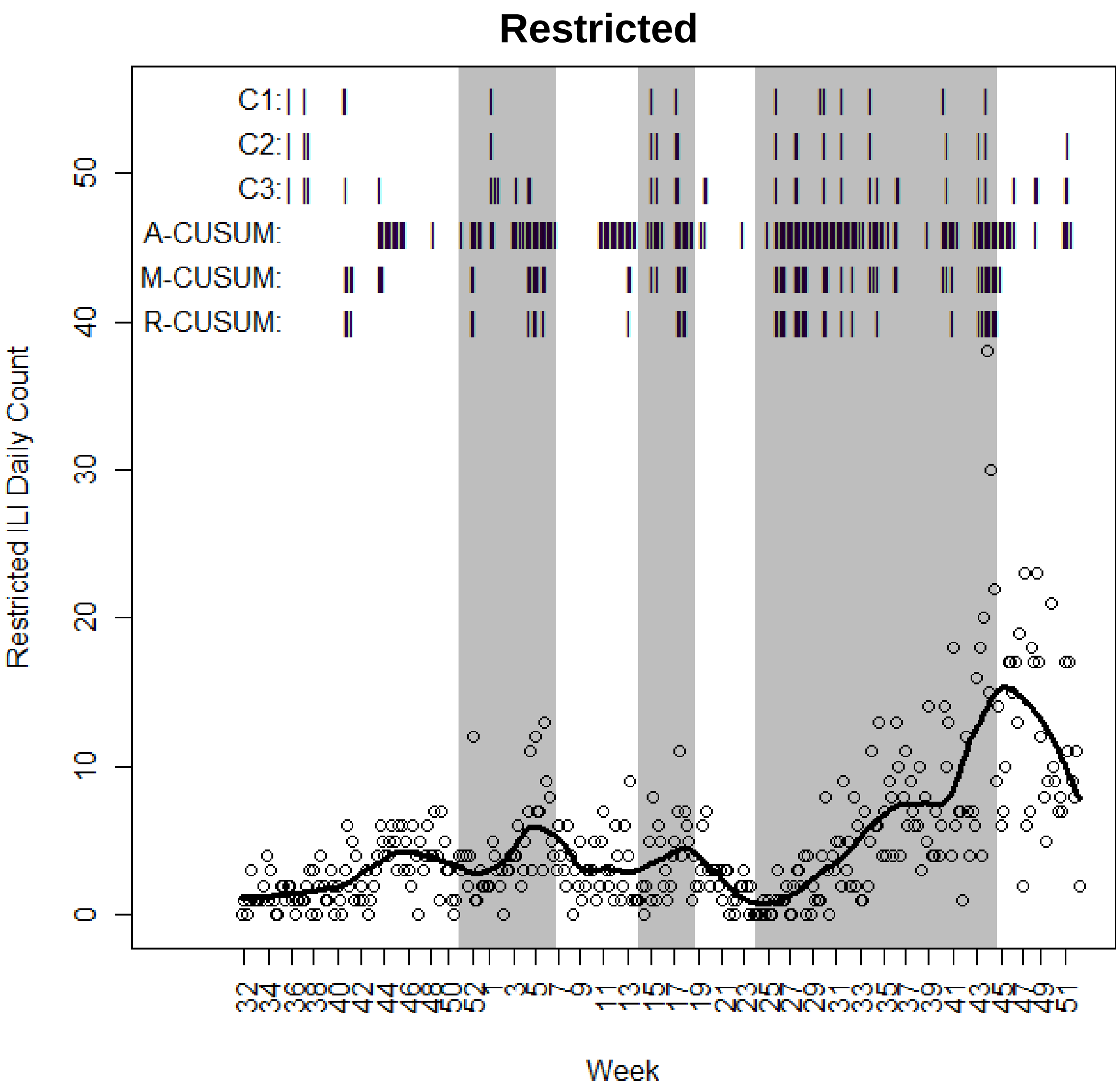
ILI Definitions	Symptom Combination Logic
EARS Baseline:	"cold" *or* "cough" *or* "sore throat"
MCHD Expanded:	"cold" *or* "cough" *or* "fever" *or* "chills" *or* "muscle pain" *or* "headache" *or* ("flu" *and not* "shot")
MCHD Restricted:	("fever" *and* "cough") *or* ("fever" *and* "sore throat") *or* ("fever" *and* "cough" *and* "sore throat") *or* ("flu" *and not* "shot")

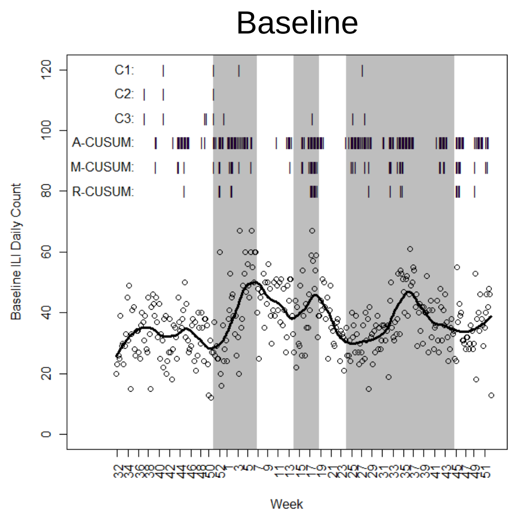
NAVAL POSTGRADUATE SCHOOL

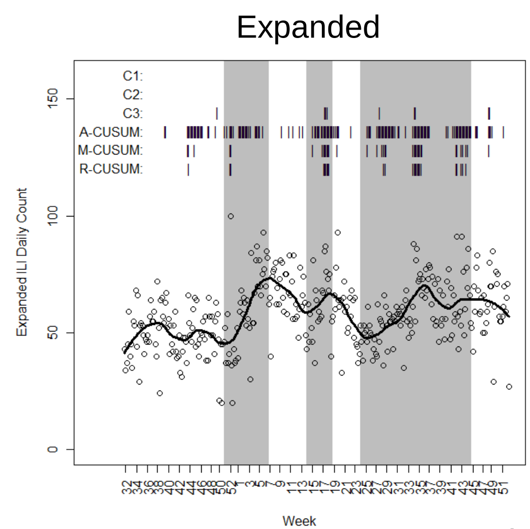
Definitions Affect Daily Counts



Restricted Definition Performed Best







Quantifying Performance

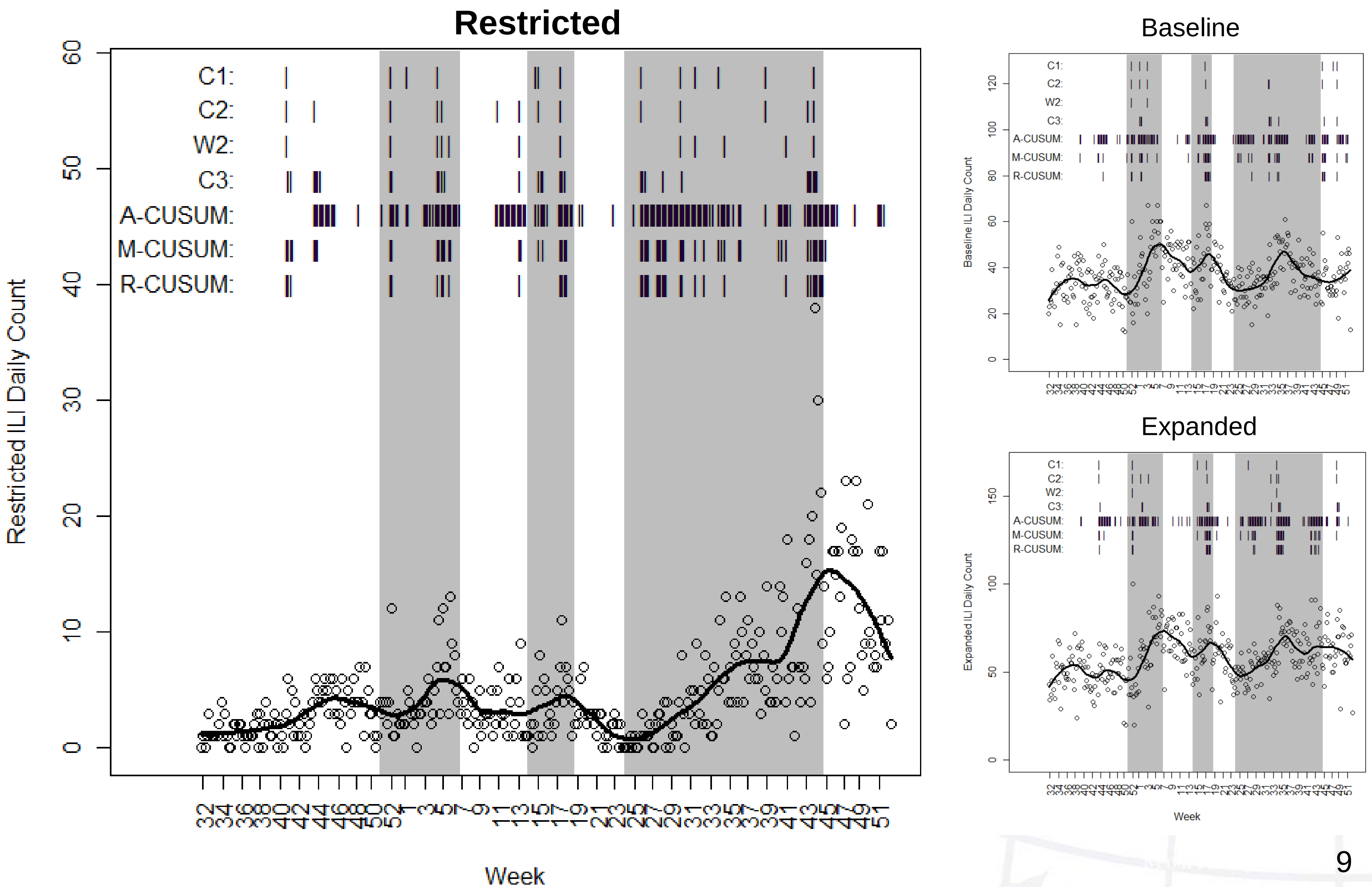
- Metrics:
  - *Sensitivity*: # outbreak days with signal / # outbreak days
  - *Specificity*: # non-outbreak days without signal / # non-outbreak days
  - *Average delay*:
    - $\bar{d}_1$ average time to signal from start of outbreak period
    - $\bar{d}_2$ average time to signal from earliest signal
- Results:

	Baseline				Expanded				Restricted			
Algorithm	**Sens.**	**Spec.**	$\bar{d}_1$	$\bar{d}_2$	**Sens.**	**Spec.**	$\bar{d}_1$	$\bar{d}_2$	**Sens.**	**Spec.**	$\bar{d}_1$	$\bar{d}_2$
C1	0.02	0.99	14+	11+	0.00	1.00	57+	52+	0.06	0.98	9.7	6.0
C2	0.01	0.99	43+	40+	0.00	1.00	57+	52+	0.08	0.98	9.7	6.0
C3	0.03	0.98	8.7	5.7	0.04	0.98	26+	21+	0.13	0.93	9.7	6.0
A-CUSUM	0.55	0.75	3.0	0.0	0.58	0.77	4.7	0.0	0.62	0.76	3.7	0.0
M-CUSUM	0.21	0.93	4.7	1.7	0.18	0.97	6.3	1.7	0.28	0.95	7.0	3.3
R-CUSUM	0.09	0.97	14.7	11.7	0.14	0.99	14.7	10.0	0.21	0.98	10.7	7.0

Results

- Restricted ILI definition gave best performance
  - For both EARS and CUSUM methods
  - For details, see Hagen, K.S., R.D. Fricker, Jr., K. Hanni, S. Barnes, and K. Michie, Assessing the Early Aberration Reporting System's Ability to Locally Detect the 2009 Influenza Pandemic, *Statistics, Politics, and Policy*
- Suggests performance gains to be had by improving syndrome definitions
  - "Low-hanging fruit"
- Results beg the question: which algorithm should be preferred?
  - Can't compare results directly – CUSUM had advantages

EARS' Methods Marginally Improved by Removing Weekend Zeros



EARS' Methods Marginally Improved by Removing Weekend Zeros

	Algorithm	Baseline Sens.	Baseline Spec.	Baseline $\bar{d}_1$	Baseline $\bar{d}_2$	Expanded Sens.	Expanded Spec.	Expanded $\bar{d}_1$	Expanded $\bar{d}_2$	Restricted Sens.	Restricted Spec.	Restricted $\bar{d}_1$	Restricted $\bar{d}_2$
Weekends Removed	C1	0.02	0.98	41+	38+	0.03	0.99	9.3	4.6	0.07	0.99	6.3	2.6
	C2	0.04	0.99	21.3	18.3	0.04	0.99	22.0	17.3	0.06	0.98	7.0	3.3
	W2	0.01	1.00	45+	42+	0.01	1.00	26+	22+	0.06	0.99	17.3	13.6
	C3	0.06	0.99	25	22	0.05	0.98	36.3	31.6	0.14	0.96	7.0	3.3
With 0s	C1	0.02	0.99	14+	11+	0.00	1.00	57+	52+	0.06	0.98	9.7	6.0
	C2	0.01	0.99	43+	40+	0.00	1.00	57+	52+	0.08	0.98	9.7	6.0
	C3	0.03	0.98	8.7	5.7	0.04	0.98	26+	21+	0.13	0.93	9.7	6.0

- Remember the metrics:
  - *Sensitivity*: # outbreak days with signal / # outbreak days
  - *Specificity*: # non-outbreak days without signal / # non-outbreak days
  - *Average delay*:
    - $\bar{d}_1$ average time to signal from start of outbreak period
    - $\bar{d}_2$ average time to signal from earliest signal

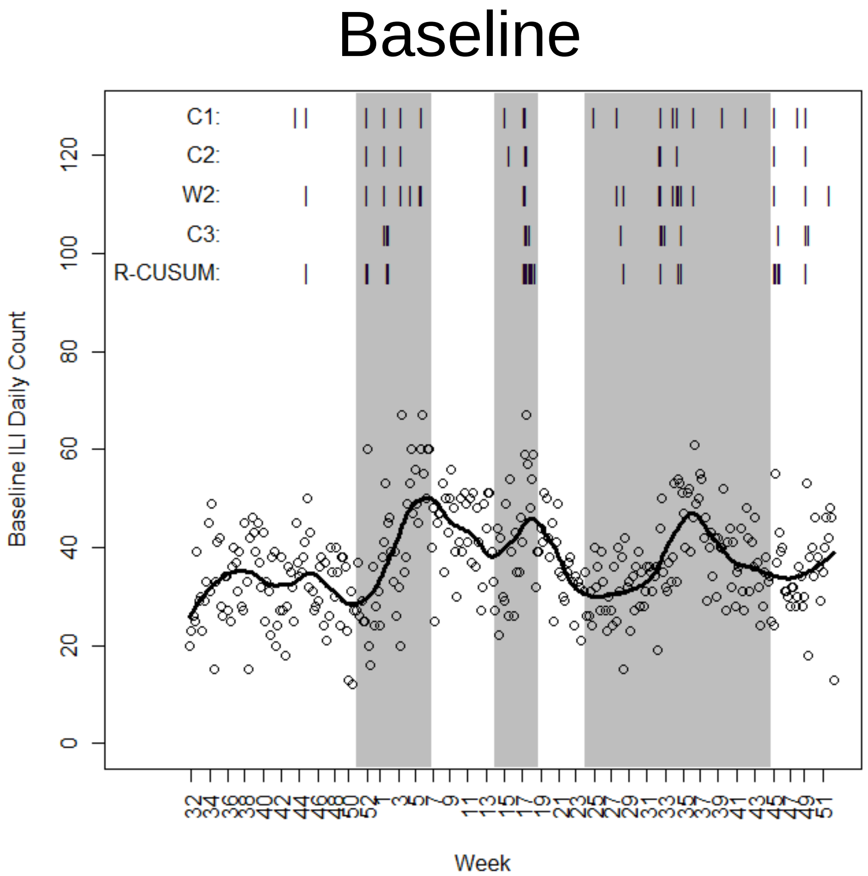
NAVAL POSTGRADUATE SCHOOL

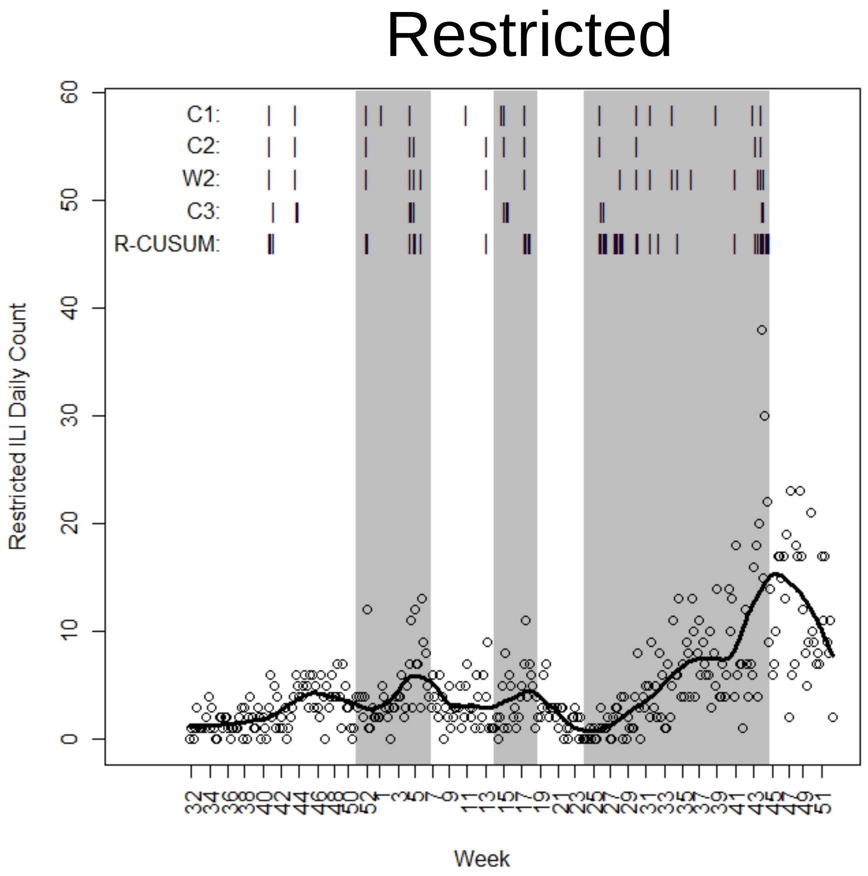
EARS Performance *Much* Improved by Adjusting Signal Thresholds

	Baseline				Expanded				Restricted			
Algorithm	**Sens.**	**Spec.**	$\bar{d}_1$	$\bar{d}_2$	**Sens.**	**Spec.**	$\bar{d}_1$	$\bar{d}_2$	**Sens.**	**Spec.**	$\bar{d}_1$	$\bar{d}_2$
C1	0.09	0.97	5.7	0.0	0.04	0.99	9.3	0.0	0.08	0.98	6.3	0.0
C2	0.09	0.97	11.3	5.6	0.05	0.99	21.3	12.0	0.05	0.98	7.0	0.7
W2	0.10	0.97	13.3	7.6	0.06	0.99	14.6	5.3	0.09	0.98	14.3	8.0
C3	0.09	0.97	10.0	4.3	0.03	0.99	37+	28+	0.06	0.98	15.3	9.0
R-CUSUM	0.09	0.97	14.7	9.0	0.14	0.99	14.7	5.4	0.21	0.98	10.7	4.4

	Baseline				Expanded				Restricted			
Algorithm	**Sens.**	**Spec.**	$\bar{d}_1$	$\bar{d}_2$	**Sens.**	**Spec.**	$\bar{d}_1$	$\bar{d}_2$	**Sens.**	**Spec.**	$\bar{d}_1$	$\bar{d}_2$
C1	0.26	0.75	2.3	0.0	0.28	0.77	3.3	0.0	0.29	0.76	4.7	1.0
C2	0.26	0.75	4.0	1.7	0.29	0.77	4.7	1.4	0.35	0.76	5.0	1.3
W2	0.39	0.75	4.0	1.7	0.41	0.77	8.3	5.0	0.41	0.76	6.3	2.6
C3	0.16	0.89	9.7	9.4	0.19	0.93	7.7	4.4	0.24	0.91	7.0	3.3
A-CUSUM	0.55	0.75	3.0	0.7	0.58	0.77	4.7	1.4	0.62	0.76	3.7	0.0

EARS Performance *Much* Improved by Adjusting Signal Thresholds

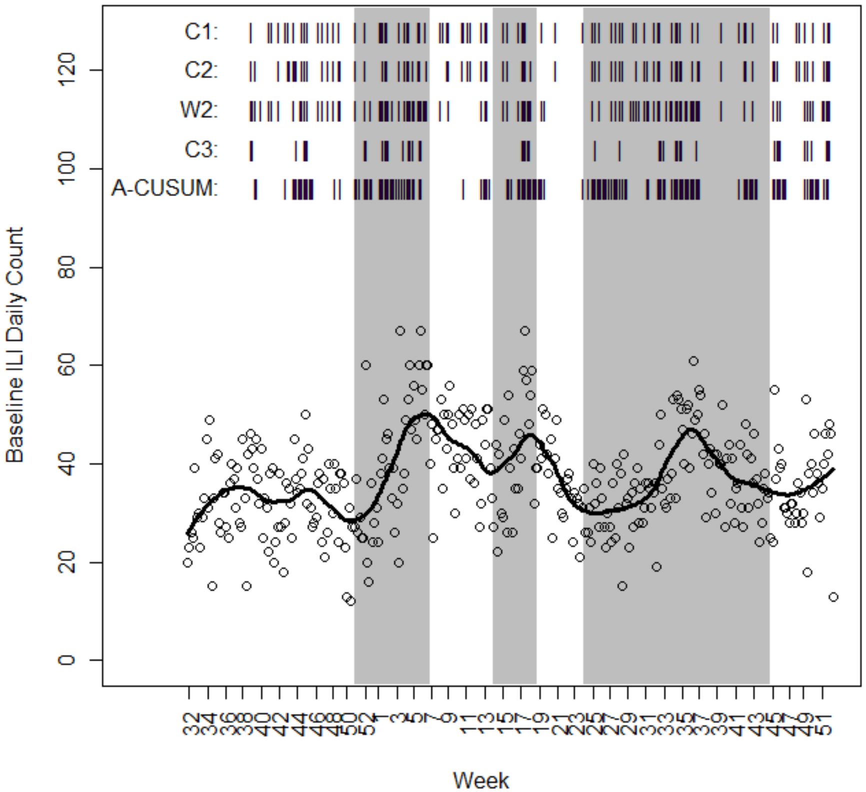




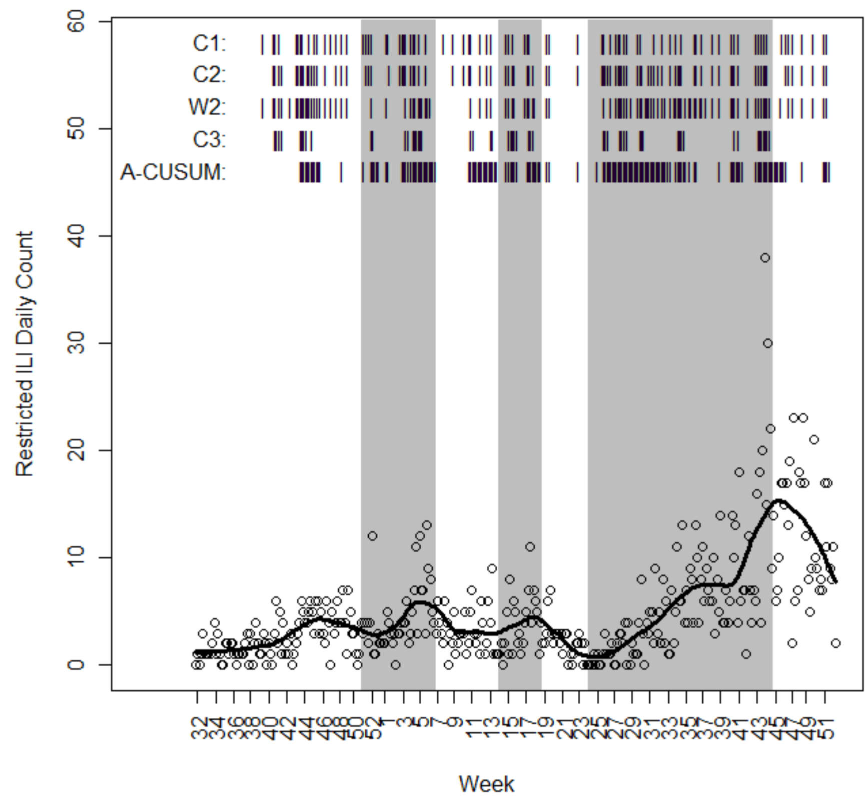
Performance when EARS thresholds set so methods match R-CUSUM specificity

EARS Performance *Much* Improved by Adjusting Signal Thresholds

Baseline



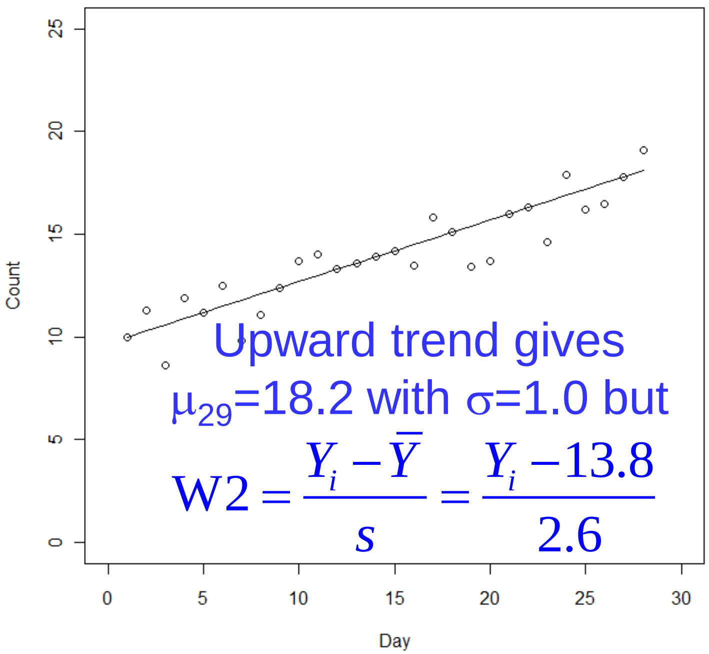
Restricted

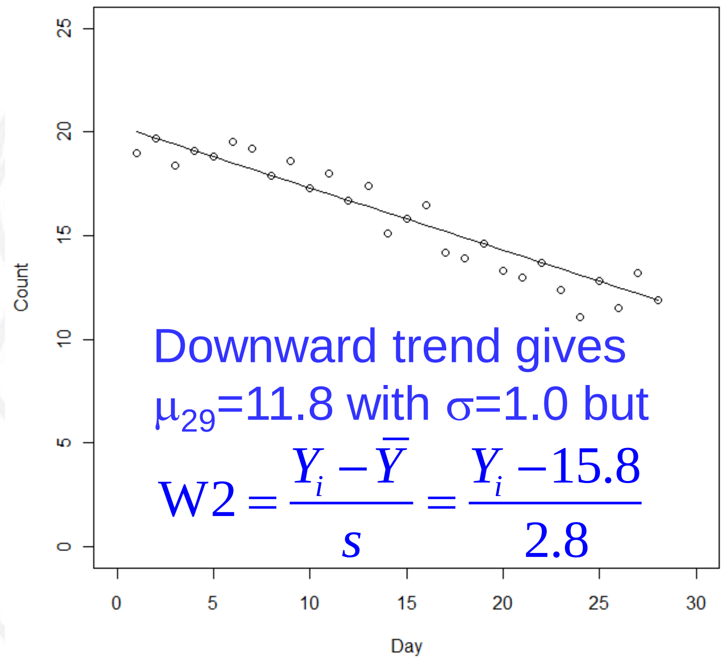


Performance when EARS thresholds set so methods match A-CUSUM specificity

Why Does W2 Average Delay Performance Lag?

- For non-stationary data, longer baselines can result in mis-estimation of mean and standard deviation
  - Thus, probability of signaling <u>for an equivalent deviation from current conditions</u> depends on past trends
- Consider:



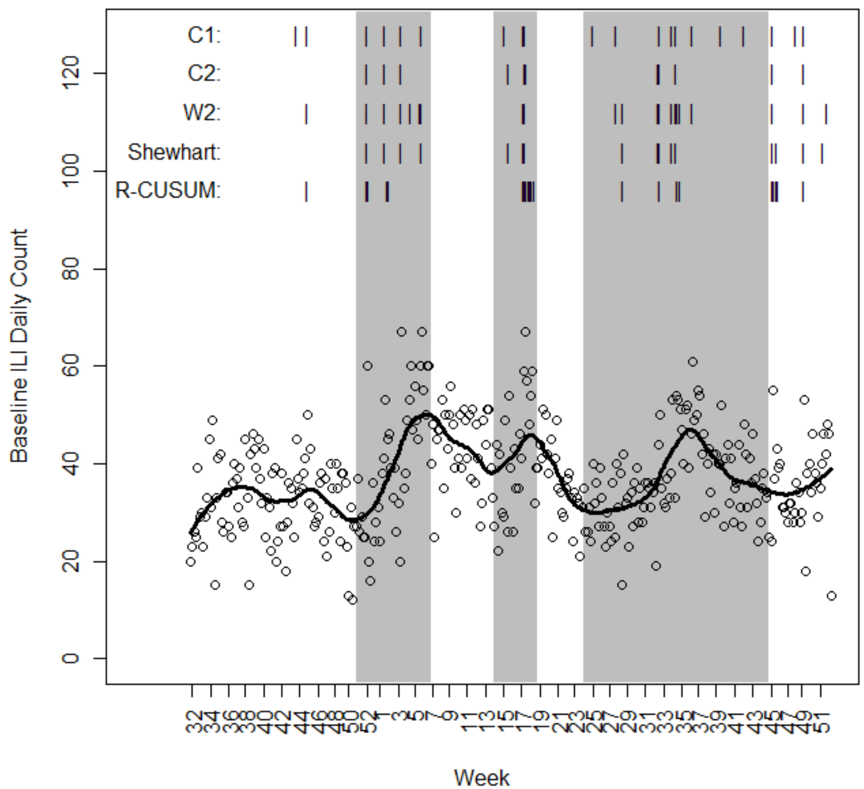


Improving on the W2 Method

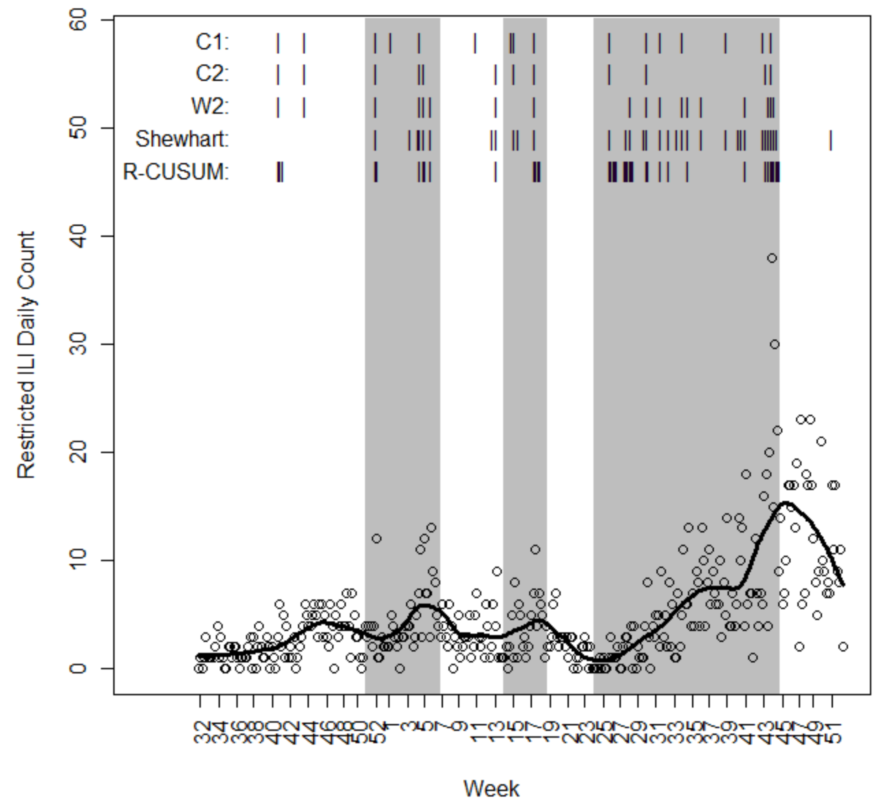
- Apply C1 and C2 methods to residuals from model (such as adaptive regression)
- Benefits:
  - Allows for longer baseline, but should give better estimation of daily means and standard deviations
  - In this work, adaptive regression residuals normally distributed, so easy to choose thresholds
- In quality control terms, it's applying Shewhart method to a model's standardized residuals
  - Model does not require years of data
  - In this work, we used 35 days (seven weeks)

Shewhart Method Applied to Adaptive Regression Residuals Performs Well

Baseline

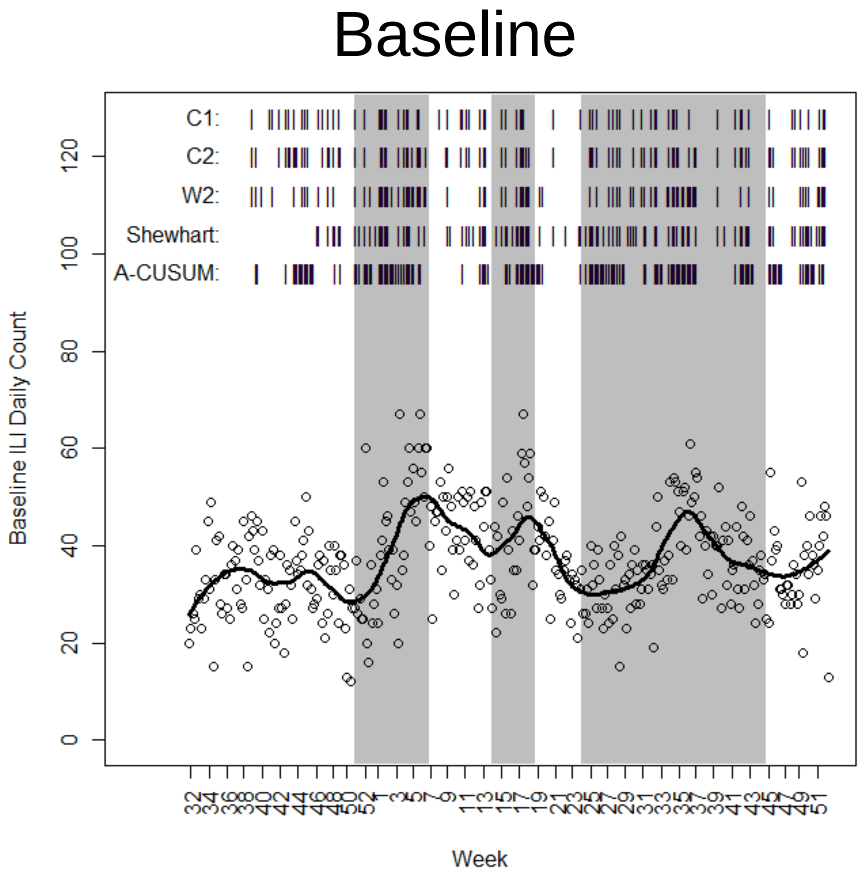


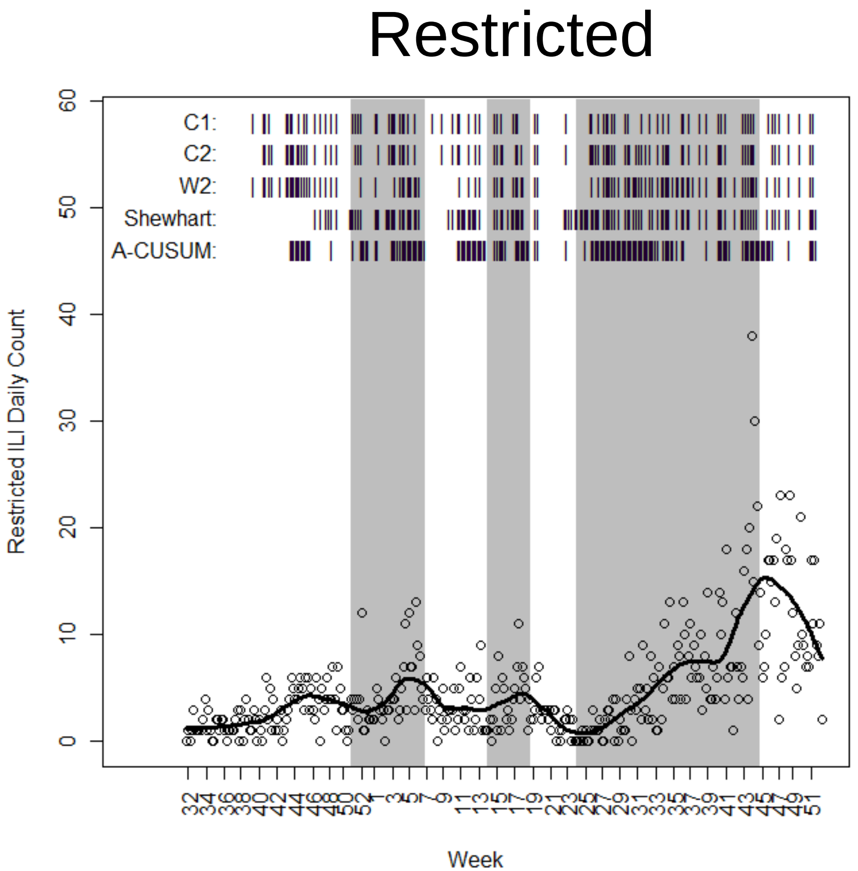
Restricted



Performance when EARS thresholds set so methods match R-CUSUM specificity

Shewhart Method Applied to Adaptive Regression Residuals Performs Well





Performance when EARS thresholds set so methods match A-CUSUM specificity

Shewhart Method Applied to Adaptive Regression Residuals Performs Well

	Baseline				Restricted			
Algorithm	**Sens.**	**Spec.**	$\bar{d}_1$	$\bar{d}_2$	**Sens.**	**Spec.**	$\bar{d}_1$	$\bar{d}_2$
C1	0.09	0.97	5.7	0.0	0.08	0.98	6.3	0.0
C2	0.09	0.97	11.3	5.6	0.05	0.98	7.0	0.7
W2	0.10	0.97	13.3	7.6	0.09	0.98	14.3	8.0
Shewhart	0.07	0.97	12.0	6.3	0.17	0.98	7.0	0.7
R-CUSUM	0.09	0.97	14.7	9.0	0.21	0.98	10.7	4.4

	Baseline				Restricted			
Algorithm	**Sens.**	**Spec.**	$\bar{d}_1$	$\bar{d}_2$	**Sens.**	**Spec.**	$\bar{d}_1$	$\bar{d}_2$
C1	0.26	0.75	2.3	1.0	0.29	0.76	4.7	3.4
C2	0.26	0.75	4.0	2.7	0.35	0.76	5.0	3.7
W2	0.39	0.75	4.0	2.7	0.41	0.76	6.3	5.0
Shewhart	0.40	0.75	1.3	0.0	0.52	0.76	1.3	0.0
A-CUSUM	0.55	0.75	3.0	1.7	0.62	0.76	3.7	2.4

Conclusions

- More research into syndrome definitions would likely provide real benefits
- EARS C1 method performed quite well with appropriately set thresholds
- W2 performance improved with better estimation of mean and std. deviation
- Shewhart methods preferred (signal fast) when outbreak is rapid
  - CUSUM will do better for gradual increases

Back-up Slides

Early Aberration Reporting System

- EARS' detection algorithms:

$$C_1(t) = \frac{Y(t) - \bar{Y}_1(t)}{s_1(t)}$$

- Sample statistics calculated from previous 7 days' data
- Signal when $C_1 > 3$

$$C_2(t) = \frac{Y(t) - \bar{Y}_3(t)}{s_3(t)}$$

- Sample statistics calculated from 7 days' of data prior to 2 day lag
- Signal when $C_2 > 3$

$$C_3(t) = \sum_{i=t}^{t-2} \max\left[0, C_2(i) - 1\right]$$

- Signal when $C_3 > 2$

- Often referred to as CUSUMs, but not true
- In SPC parlance, C_1 and C_2 are Shewhart variants

CUSUM on Adaptive Regression Forecast Errors

- Adaptive regression: regress a sliding baseline of observations on time relative to current observation
  - I.e. regress $Y(t-1),\ldots,Y(t-n)$ on $n,\ldots,1$
- Calculate standardized residuals from one day ahead forecast, $Z(t)=R(t)/\hat{\sigma}_Y$, where

$$R(t)=Y(t)-\left[\hat{\beta}_0+\hat{\beta}_1\times(n+1)+\hat{\beta}_j\right]$$

and $\hat{\sigma}^2=\frac{1}{33}\sum_{i=t-35}^{t-1} R_i^2\left(1+\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0\right)$

- CUSUM:

$$S(t)=\max\left[0,S(t-1)+Z(t)-k\right]$$

where a signal is generated if $S(t)>h$

Three CUSUMs Evaluated

- We looked at the performance of three CUSUMs based on choices of k and h:
  - Smaller k: Can detect smaller increases in mean
  - Larger h: Fewer false positive signals (i.e., larger ATFS) but slower to signal

Type	**Label**	k	h	**ATFS**
Aggressive:	CUSUM1	0.5	0.365	5
Moderate:	CUSUM2	1.0	0.695	20
Routine:	CUSUM3	1.0	1.200	60