

Continuously monitoring a small incidence rate

Landon Sego
Bill Woodall
Marion Reynolds, Jr.

Department of Statistics
Virginia Polytechnic Institute and State University

Contact: Landon Sego <sego@vt.edu>

Continuously monitoring a small incidence rate (proportion)

Examples:

- Monitoring the proportion of defective parts in a manufacturing assembly line, where each item is inspected.
- Monitoring the rate of a congenital malformation, where each infant in the region of monitoring is examined and diagnosed.

Historical Development

- Prompted by the Thalidomide tragedy, the discussion of how best to monitor congenital birth defects began in the late 1960's.
- Many surveillance methods have been proposed, most of which rely on the assumption that the number of incidences observed per unit time follows a Poisson distribution.

Poisson Methods

- Several authors have discussed the idea of counting the number of incidences per unit time (X) and then plotting the standardized Poisson variate

$$Z = \frac{X - E[X]}{\sqrt{E[X]}}$$

in a Shewhart chart with the traditional control limits at ± 3 .

Poisson Methods

- Others have suggested using the standard CUSUM chart based on the Poisson distribution with reference value k , threshold h , and X is the number of incidences observed per unit time.

$$C_i = \max(0, C_{i-1} + X_i - k)$$

- The chart signals when $C_i \geq h$

Assumptions

1. Monitoring a specific congenital malformation in a specified region.
2. All infants are examined and data are recorded sequentially.
3. We are interested in detecting increases in the baseline incidence rate.
4. The base-line (or in-control) probability of incidence is small:
i.e. $p_0 = 1/100, 1/1000, 1/10000$

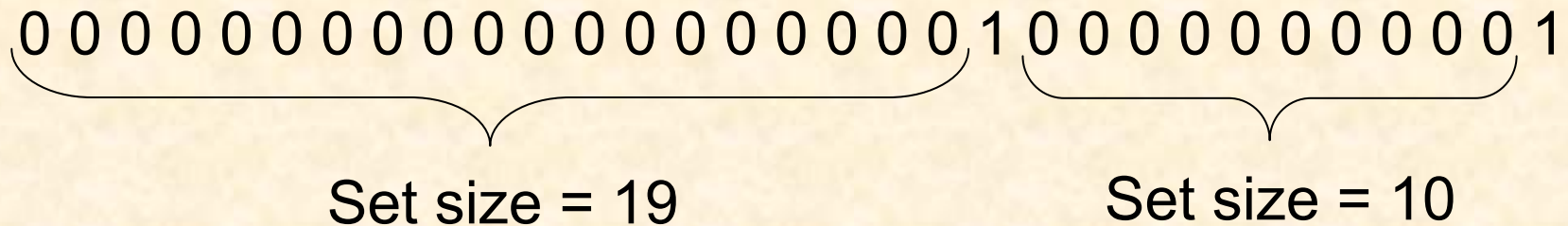
Bernoulli-Geometric methods

Since we assume the data are observed sequentially, we can consider and will compare the following methods:

- **Sets Method** Chen (1978) JASA
- **CUSCORE Method** Wolter (1987) Meth. Inform. Med.
- **Sitter et. al. Method** Sitter, et. al. (1990) Am J Epidemiol
- **Bernoulli CUSUM** Reynolds & Stoumbos (1999) JQT

Sets Method

- The Sets method was devised by Rina Chen in 1978 specifically for the surveillance of birth defects.
- A set is defined as the number of healthy babies that are born consecutively between those born with malformations.



Sets Method

- Set Size = number of normal births between malformation $X \sim Geo(p)$, $X \in \{0,1,2,\dots\}$ and $p = P(\text{observing a malformation})$

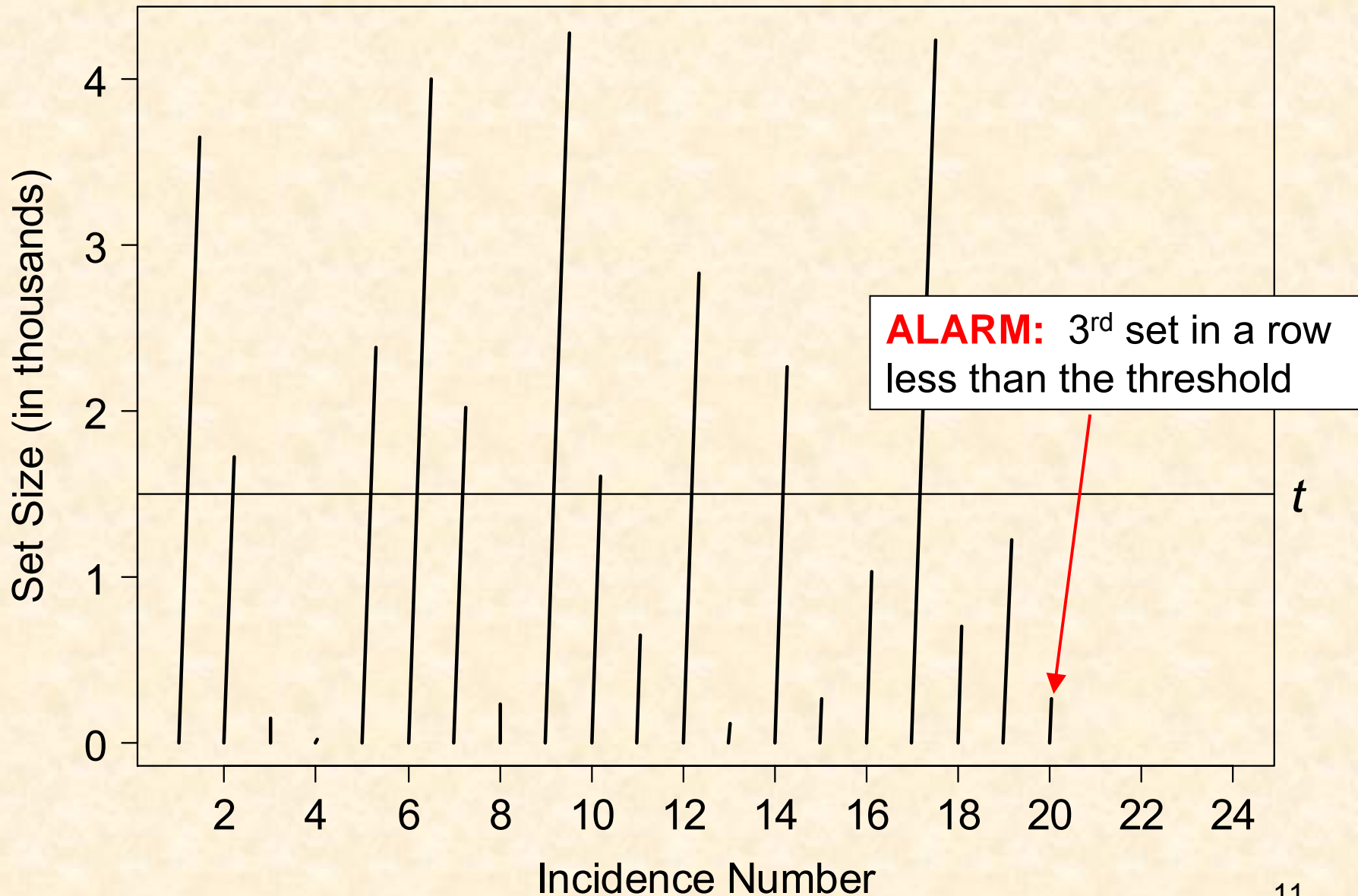
- Alternatively, the size of the set can be determined by the amount of time elapsed between malformed cases if the birth rate is constant, in which case

$$X \sim Exp\left(\frac{1}{\lambda}\right) \quad \text{with } \lambda = -\log(p)$$

Sets Method

The Sets method signals an alarm if n_s sets in a row have a size that is less than a threshold, t .

Sets Method: $n_s = 3$, $t = 1500$

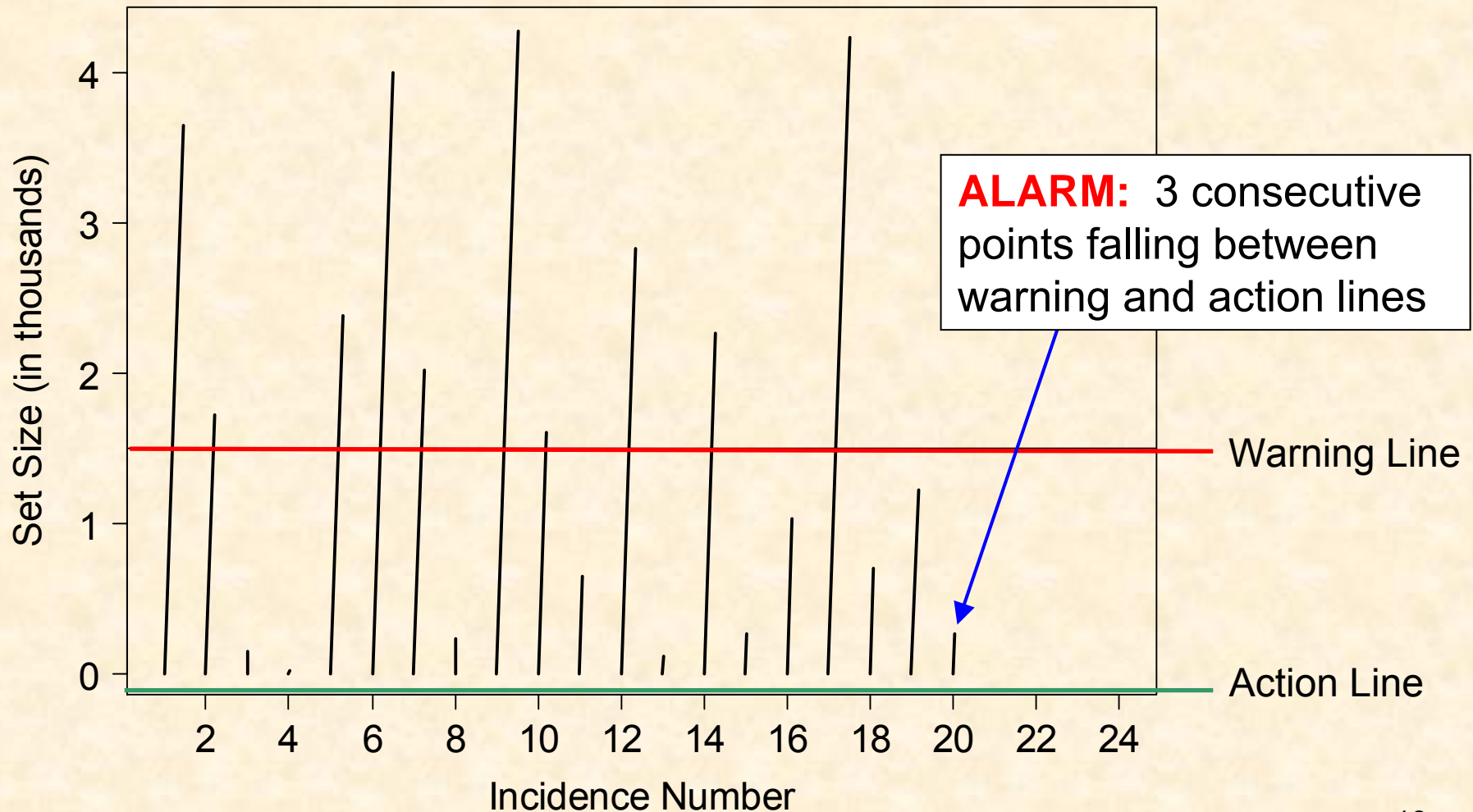


Sets method a special case

- The Sets method is a special case of a runs rule monitoring scheme given by Page in 1955:
- Signal an alarm if n consecutive points fall between the warning and action lines or if any point falls outside the action lines.

Sets method a special case

Sets Method: $n_s = 3$, $t = 1500$



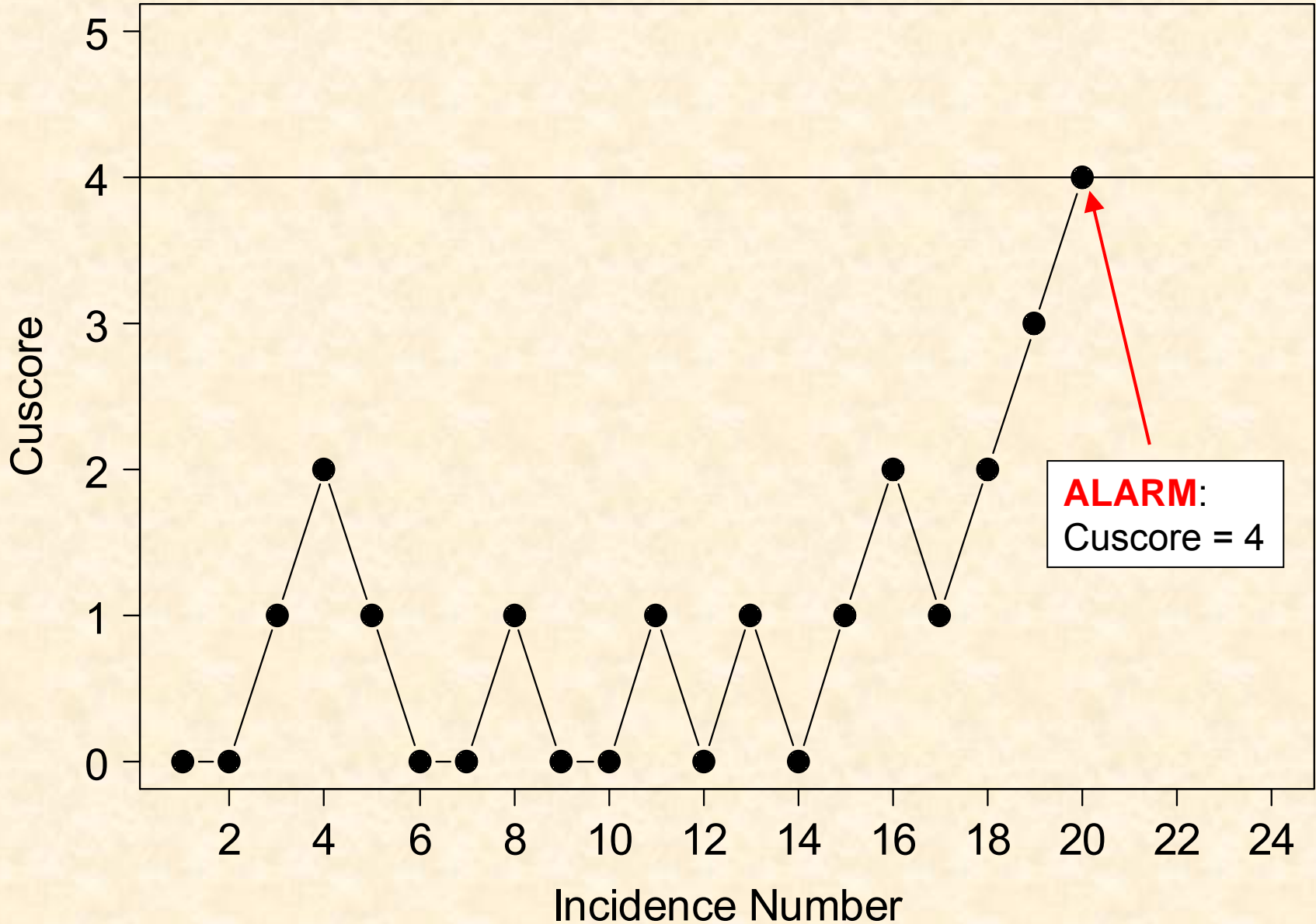
CUSCORE Method

- Same basic setup as the Sets method.
- Assign a score to each set:

$$score(X) = \begin{cases} 1 & \text{if } X < t \\ -1 & \text{otherwise} \end{cases}$$

- Accumulate the scores: $C_0 = 0$ and
 $C_i = \max(0, C_{i-1} + score(X_i))$ $i = 1, 2, \dots$
- Signal when $C_i = n_c$

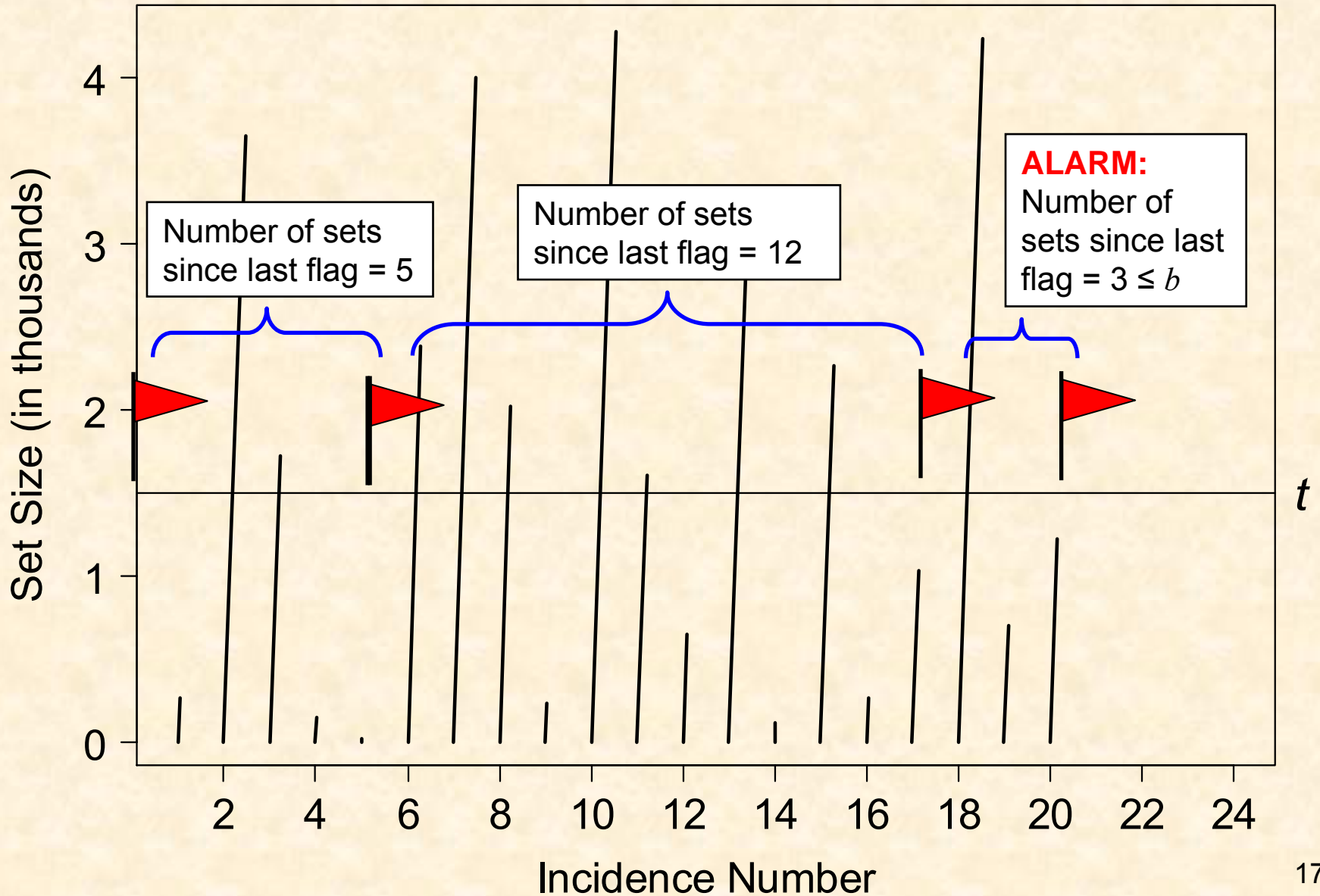
CUSCORE Method: $n_c = 4$, $t = 1500$



Sitter et. al. method

- Same basic setup as Sets method.
- Raise a flag if n_t sets in a row have size less than t .
- If the number of sets since the last flag is less than or equal to b , sound an alarm.
- After each flag, the number of sets in a row with size less than t resets to 0.

Sitter Method: $n_t = 2$, $b = 4$, $t = 1500$



Bernoulli CUSUM

$$Y = \begin{cases} 1, & \text{if the infant has the malformation} \\ 0, & \text{if the infant does not} \end{cases}$$

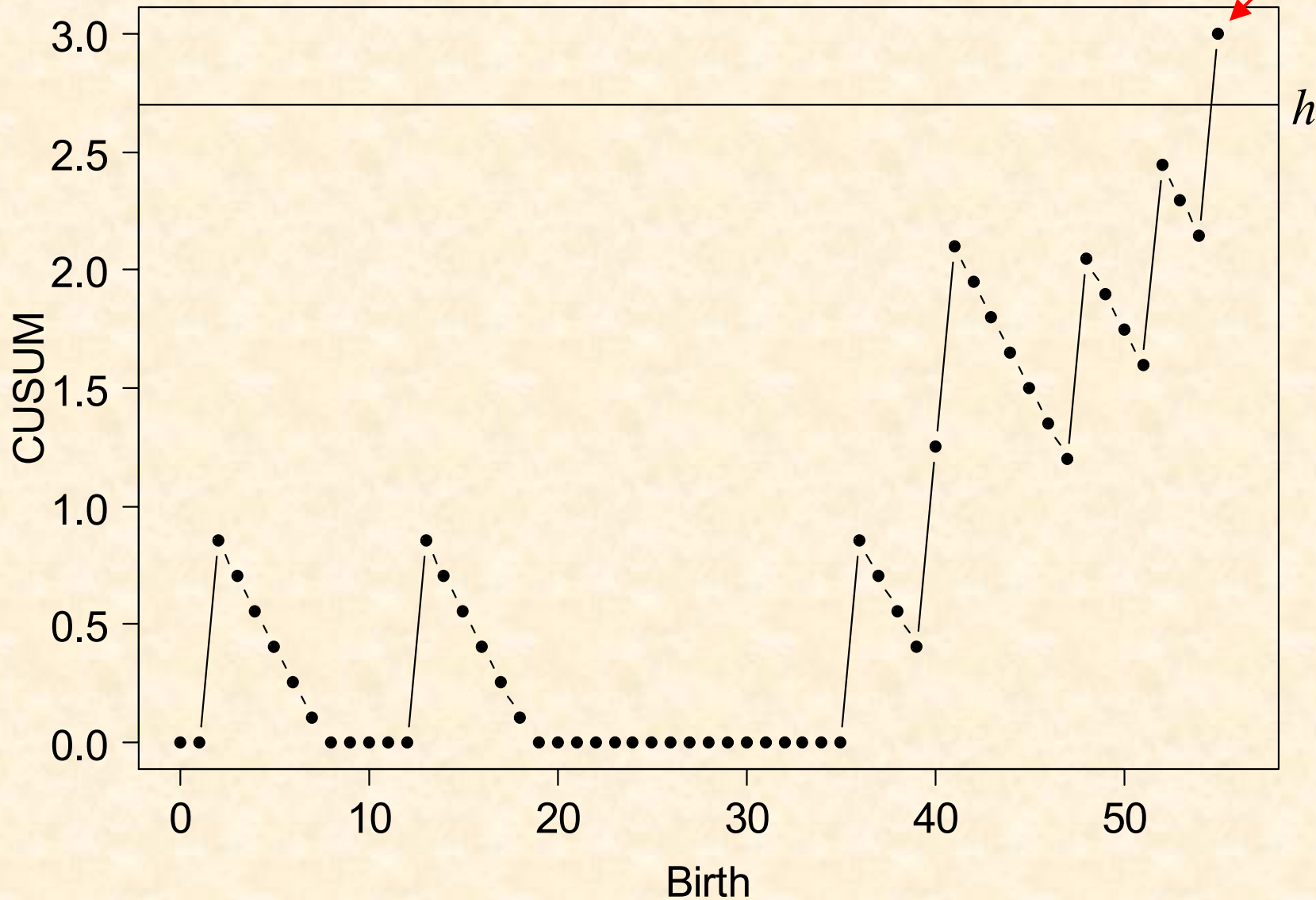
$$B_0 = 0$$

$$B_i = \max(0, B_{i-1} + Y_i - k) \quad i = 1, 2, \dots$$

Alarm signaled if $B_i \geq h$ (We'll define k later).

CUSUM Method: $k = 0.15$, $h = 2.7$

ALARM:
CUSUM ≥ 2.7



Objectives of chart design

- We want the chart to run for a long time without signaling a false alarm if baseline incidence rate remains unchanged at p_0 . (Large in-control average run length, ARL_0).
- Want the chart to signal quickly if the incidence rate increases to $p_1 = \gamma p_0$ ($\gamma > 1$). (Small out-of-control average run length ARL_1)
- If an alarm is signaled, further investigation is necessary.

Performance Criteria

- We measure the “speed” with which a chart signals with the Average Run Length (ARL)
- For the Sets, CUSCORE, and Sitter et. al. methods, the ARL could be the average number of malformations until signal.
- However, in order to make valid comparisons with the Bernoulli CUSUM, we will define the average run length as the average number of births until signal.

Performance Criteria

- One possible advantage for defining the ARL as the average number of births until signal is that it can be intuitively interpreted in terms of units of time if the birth rate is relatively constant.
- However, if the ARL is based on the number of malformations until signal, interpreting the ARL in terms of units of time is not as straightforward since the frequency of the malformations depends on the incidence rate which we are monitoring.

Choosing the parameters for the CUSUM

- The typical choice for the reference value, k , is based on Wald's Sequential Probability Ratio Test (SPRT):

$$k = \frac{-\log\left(\frac{1-p_1}{1-p_0}\right)}{\log\left(\frac{p_1(1-p_0)}{p_0(1-p_1)}\right)}$$

To find k , we need an estimate of the baseline incidence rate p_0 and we need to pick a shift γ that we would like detect. (Recall $p_1 = \gamma p_0$)

Choosing the parameters for the CUSUM

- To calculate the ARL of the CUSUM using Markov Chain theory, it is necessary to have a finite number of states. We can accomplish this by adjusting our choice of γ by a small amount to γ_a such that

$$k_a = \frac{1}{m}$$

where m is a positive integer.

Choosing the parameters for the CUSUM

- For example, suppose we are interested in detecting a doubling in the incidence rate ($\gamma = 2$) when $p_0 = 0.001$. Then

$$k = -\log\left(\frac{1-p_1}{1-p_0}\right) / \log\left(\frac{p_1(1-p_0)}{p_0(1-p_1)}\right) = 0.001442778$$

But if we let $\gamma_a = 2.000556$, then

$$k_a = 0.001443001 = \frac{1}{693}$$

Choosing the parameters for the CUSUM

- Pick, M , the target in-control average run length.
- After choosing γ and finding the corresponding values of γ_a and m , we search for the value of h that gives the smallest $ARL_0 \geq M$.
- Incidentally, the number of states in the Markov Chain is $m \times h$.

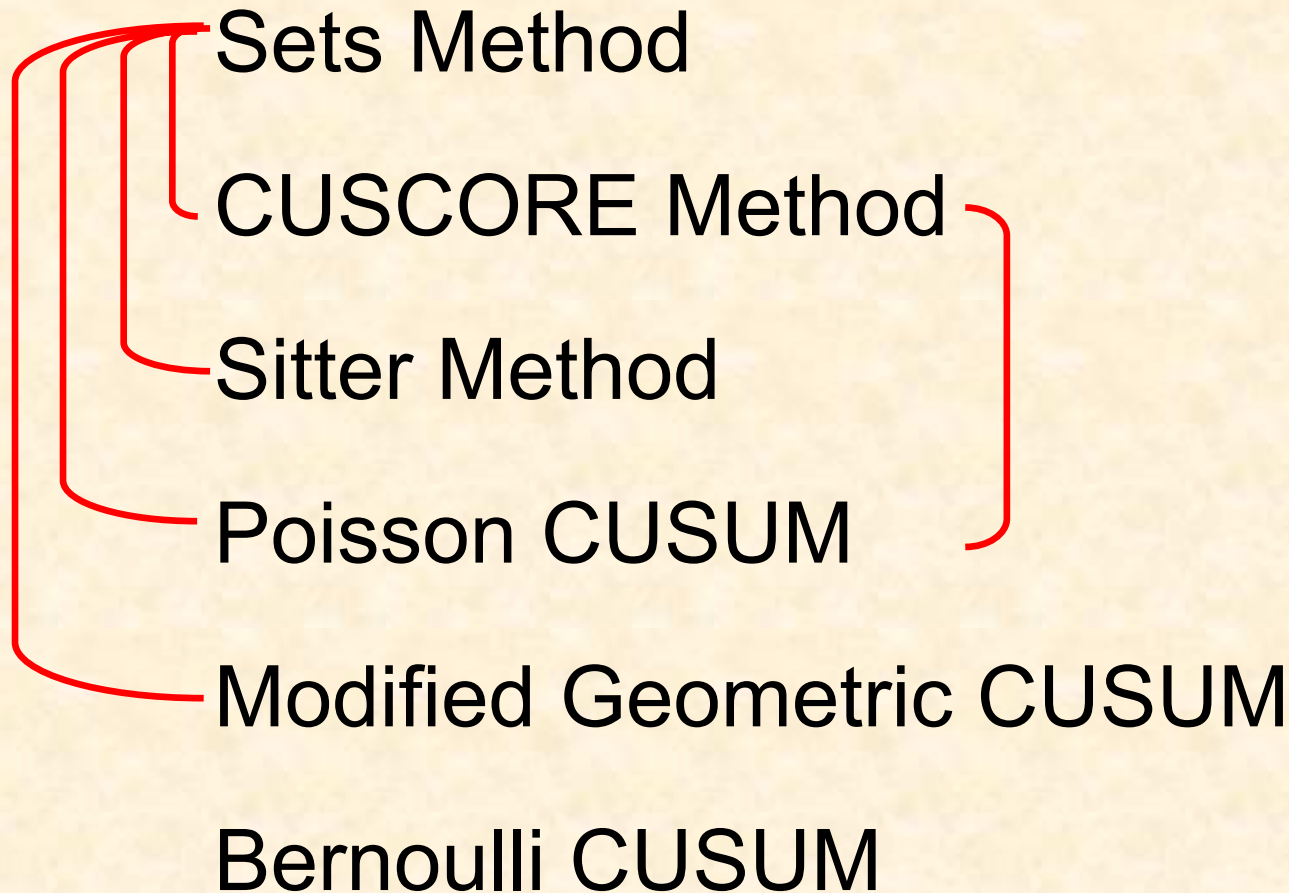
Choosing the parameters for Sets, CUSORE, and Sitter et. al. Methods

1. Pick a shift size γ that we would like to detect.
2. Find γ_a (in order to make equivalent comparisons to the CUSUM).
3. Choose the parameters n , t , (and b) that minimize ARL_1 subject to $ARL_0 \geq M$.

Choosing the parameters for Sets, CUSORE, and Sitter et. al. Methods

- Since n , t , and b are integers, we find the optimal parameter choices using grid searches:
- For each n (and b), we adjust t until we find the smallest $ARL_0 \geq M$.
- Then we identify the n (and b) that give the smallest ARL_1 .

Comparisons in the literature



Results of comparisons reported in the literature

- Generally, all other methods perform better than the Sets method.
- Under certain conditions, the Sets method performs better than the Poisson CUSUM.
- With a few exceptions, most comparisons were based on the initial state ARL_1 performance (where ARL is the number of malformations until signal).
- No one has compared the methods using the steady state ARL_1 (where ARL is the number of births until signal).

Steady State ARL

- Suppose you have been monitoring for a long period of time under baseline (in-control) conditions, AND that you have not yet had an alarm.
- Then suppose at time T , the incidence rate shifts.
- The average run length since the shift at time T is the steady state ARL_1

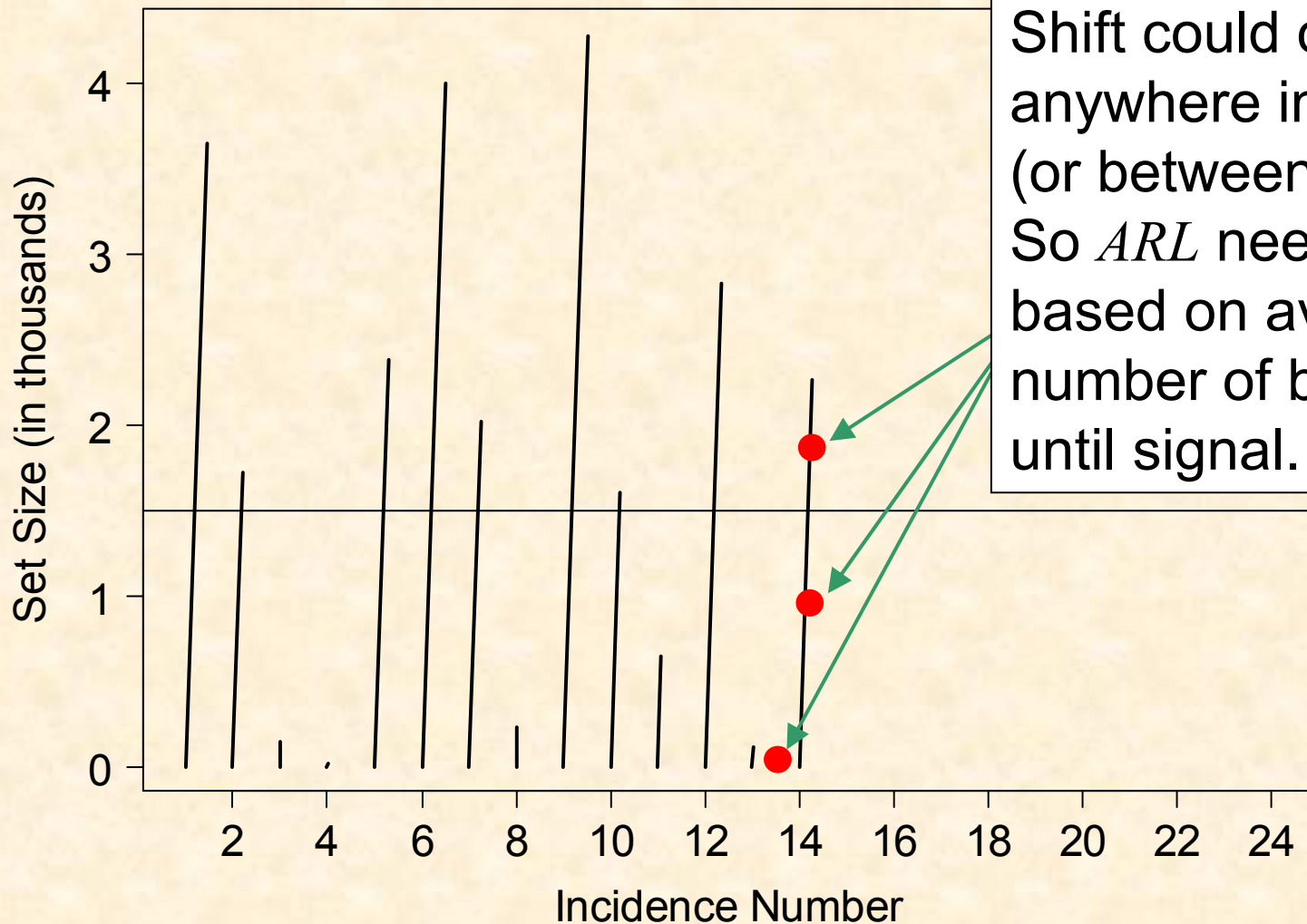
Steady State ARL

The steady state *ARL* (based on the average number of births until signal) is better than the performance criteria used by authors in previous comparisons for these two reasons:

1. The shift in the incidence rate can occur anytime within a set. \implies Need *ARL* based on Avg. Number of Births till signal.
2. The Sets, Cuscore, and Sitter et. al. methods have a built in “head-start” feature if monitoring begins after the shift has taken place. \implies Need steady state *ARL*.

Steady State ARL: Point 1

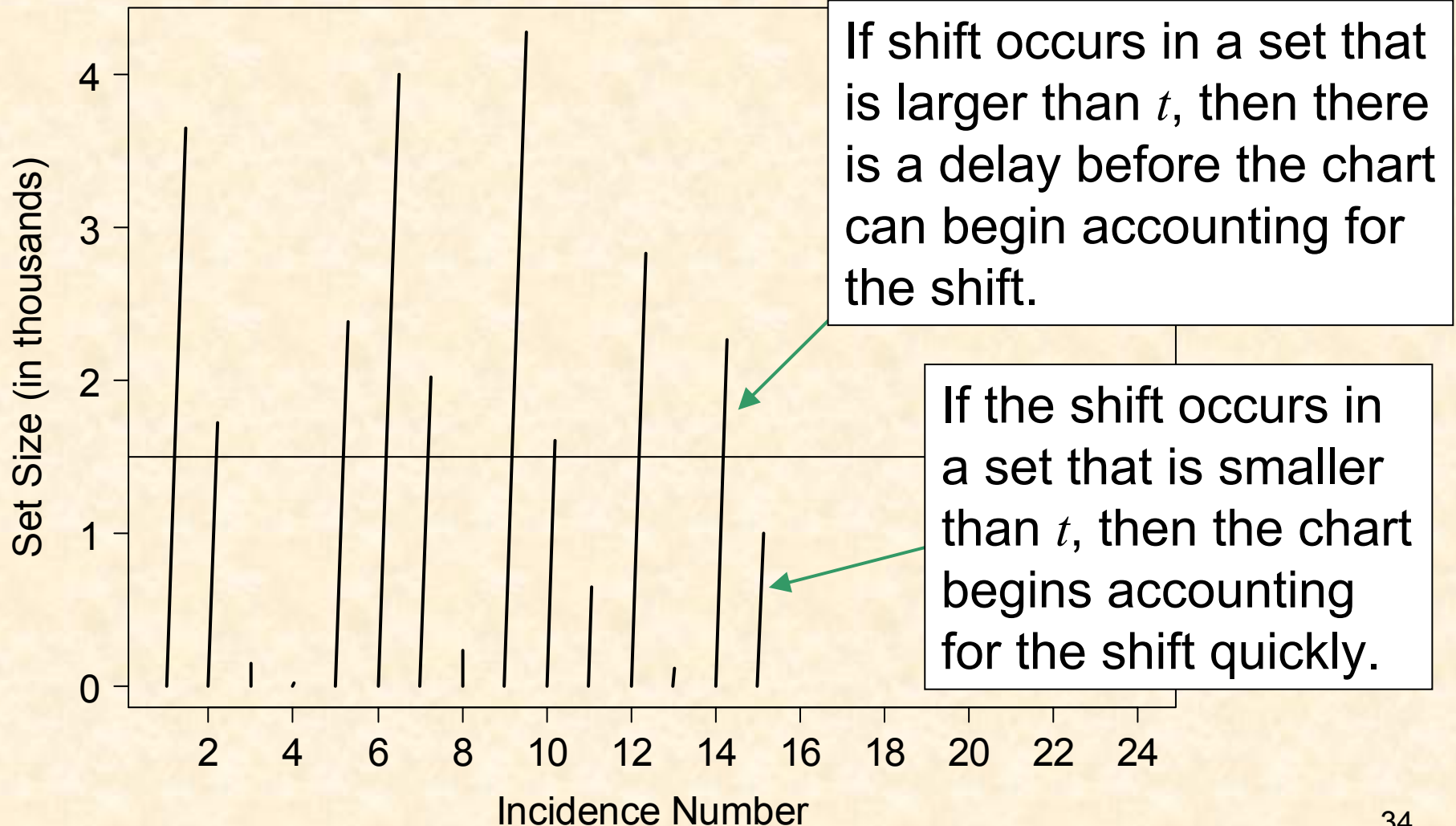
Sets Method: $n = 3, t = 1500$



Shift could occur anywhere in a set (or between 2 sets). So *ARL* needs to be based on average number of births until signal.

Steady State ARL: Point 2

Sets Method: $n = 3, t = 1500$



Steady State ARL: Point 3

- Steady state *ARL* is a more realistic measure of performance than initial state *ARL*:
- Construction of the chart assumes we have a reliable estimate of p_0 . If an epidemic is already underway when we begin monitoring, chances are our estimate of p_0 is biased.
- The typical monitoring paradigm is that one begins monitoring during a non-epidemic period (in-control) with the hope of quickly detecting an emerging epidemic if one occurs.

Calculating the Steady-State ARL

- For the Sets, CUSCORE, and Sitter et. al. methods, exact formulas were derived for the steady-state average number of births until signal by expanding the Markov Chain results.
- For the CUSUM, exact steady-state *ARL* calculations were made when the number of states was not prohibitively large (5,000 or less). Otherwise, simulation with 1 million replications was used.

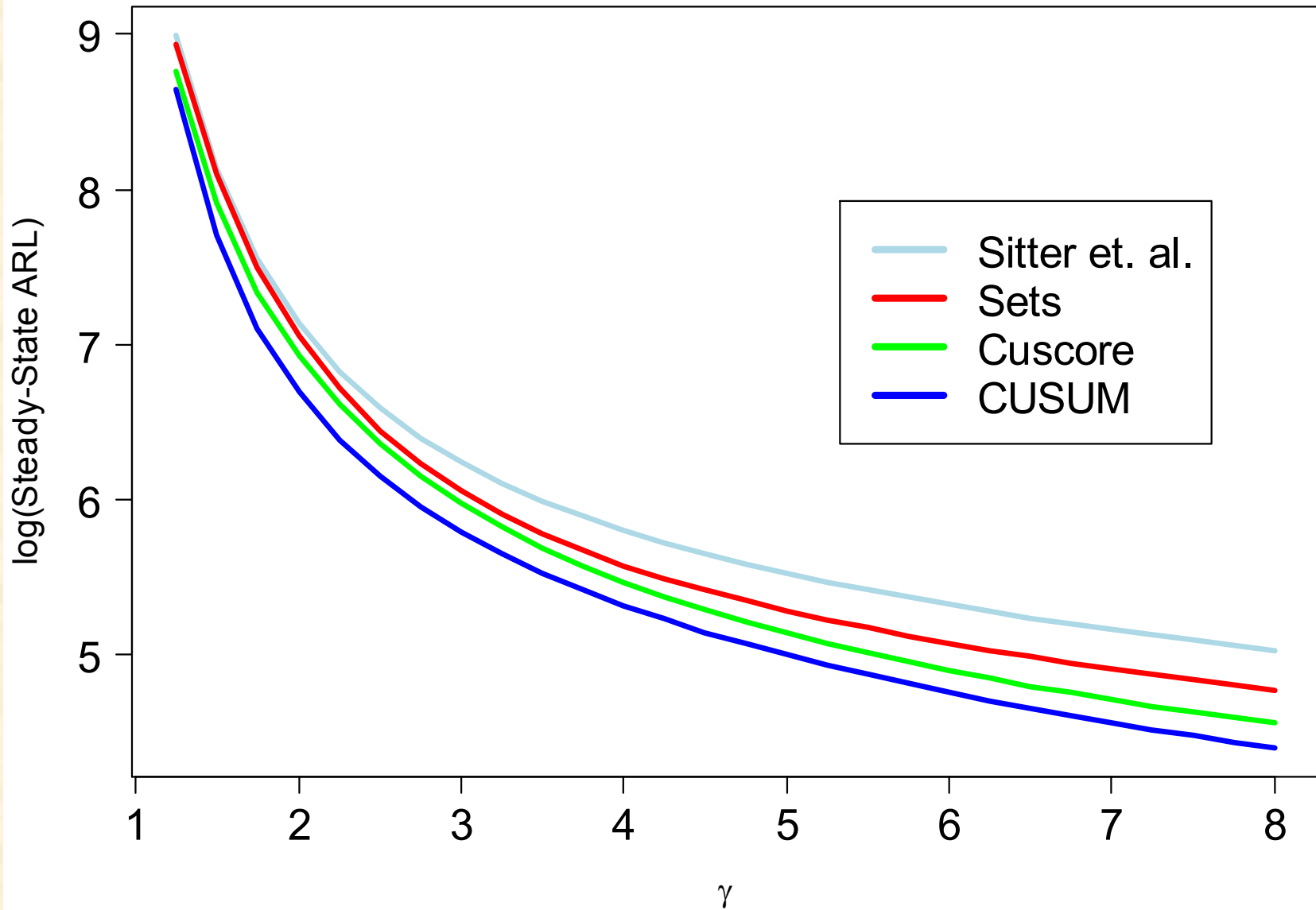
Comparisons

- Compared 23 different combinations of the baseline incidence rate, target in-control ARL , and shift size:
 - Values of p_0 : 0.01, 0.005, 0.001, 0.0005, 0.0001
 - Values of M : range from 10,000 to 500,000
 - Values of γ_{design} : 2, 4, 6

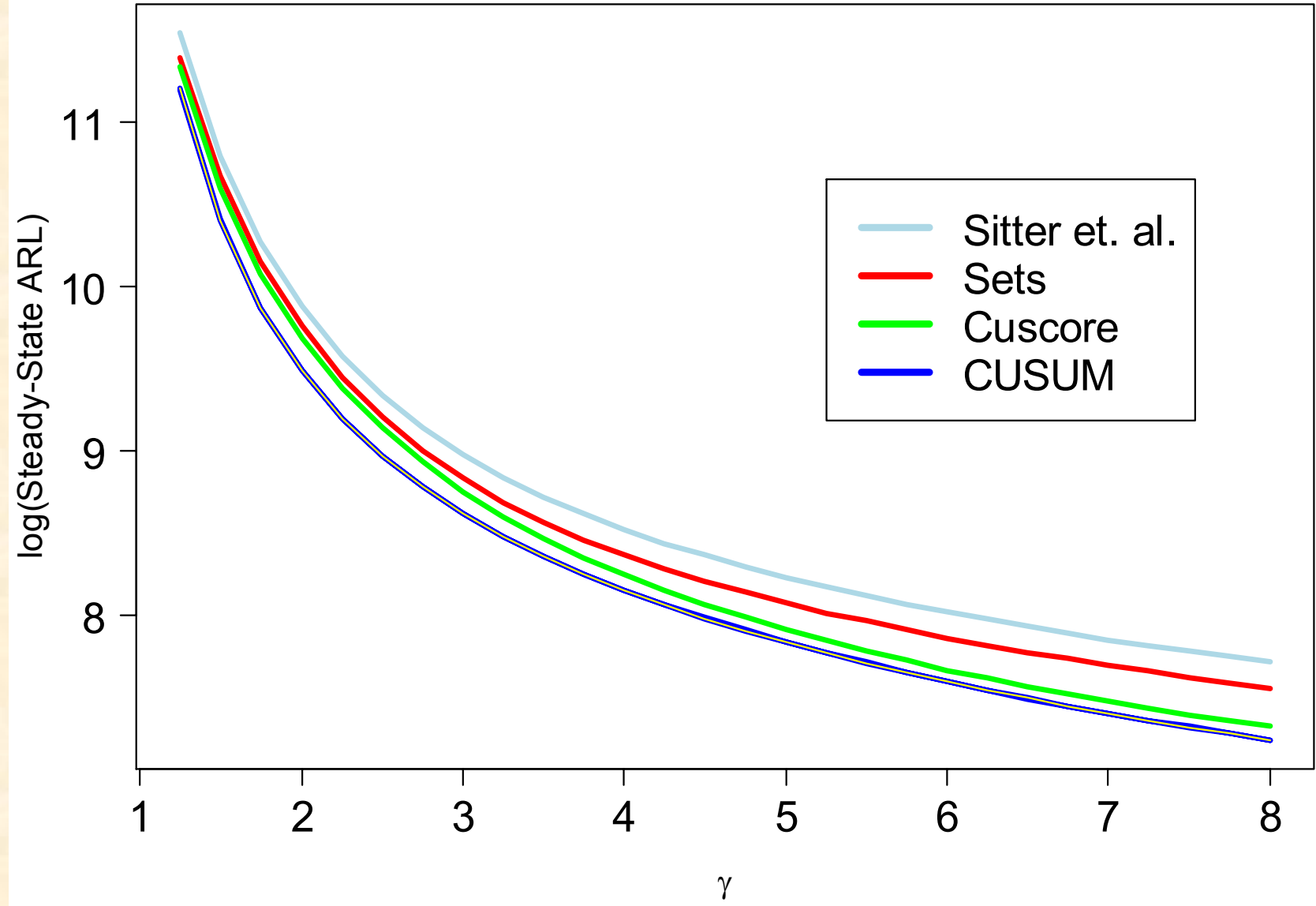
Comparisons

- Charts were designed as previously described, choosing the parameters that minimized the initial state ARL_1 subject to the initial state $ARL_0 \geq M$.
- In practice, one doesn't know the actual value of the shift, hence we calculated (or simulated) steady state ARL for actual shifts of 1.25, 1.5, 1.75, ..., 7.75, 8.

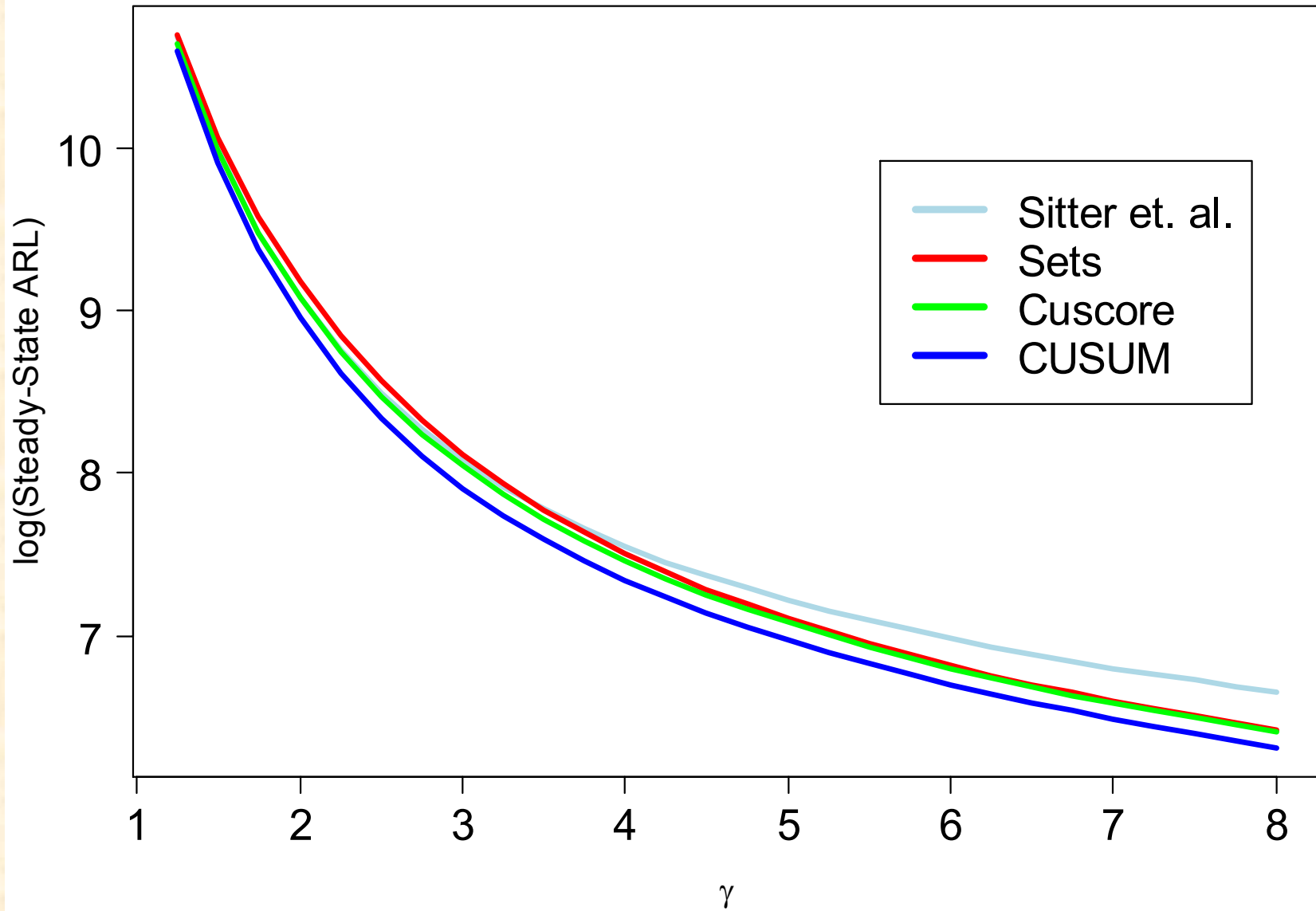
$p_0 = 0.01$ $M = 25,000$ $\gamma_{design} = 2$



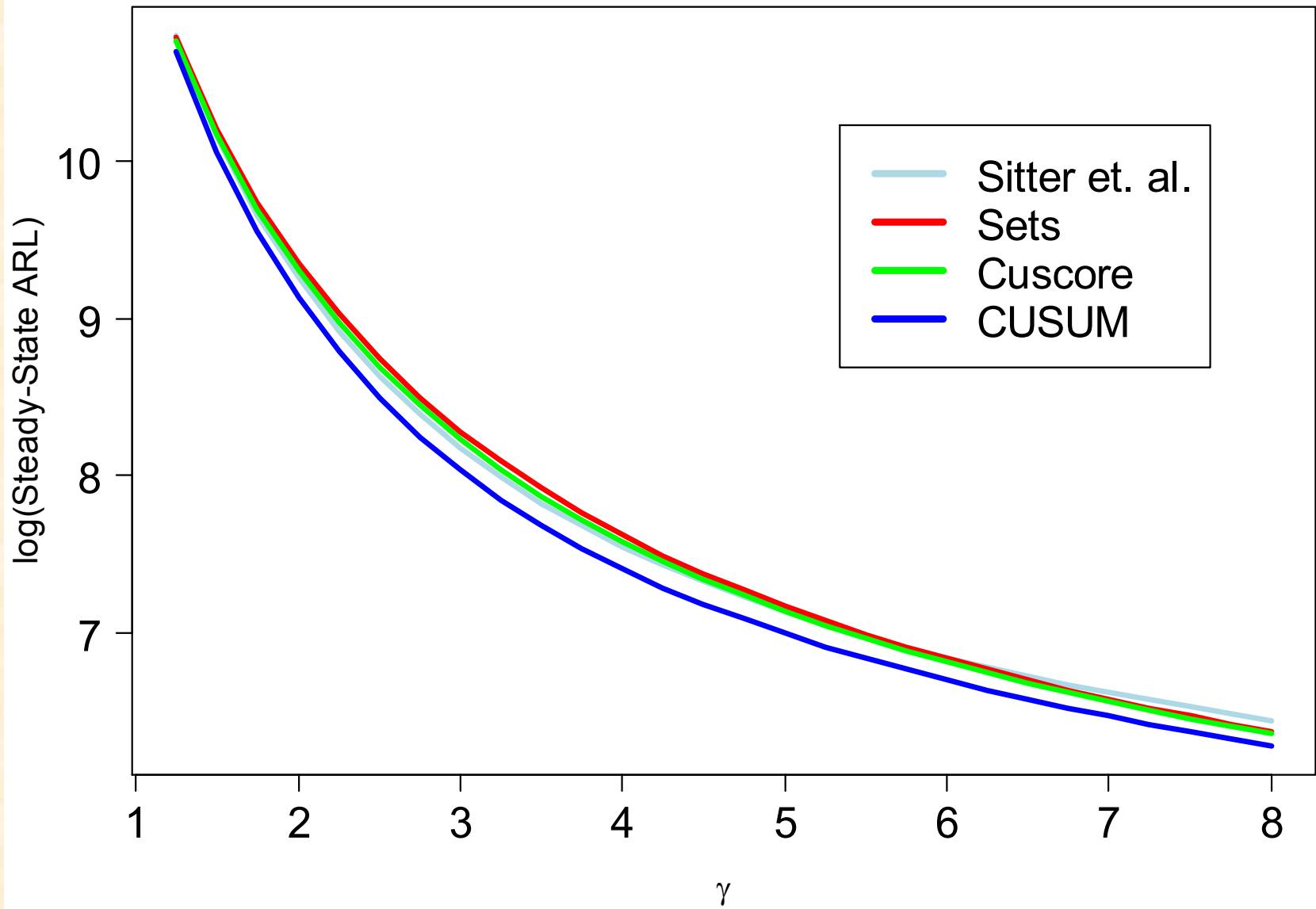
$p_0 = 0.0005$ $M = 250,000$ $\gamma_{design} = 2$



$p_0 = 0.001$ $M = 100,000$ $\gamma_{design} = 4$



$p_0 = 0.001$ $M = 100,000$ $\gamma_{design} = 6$



Results

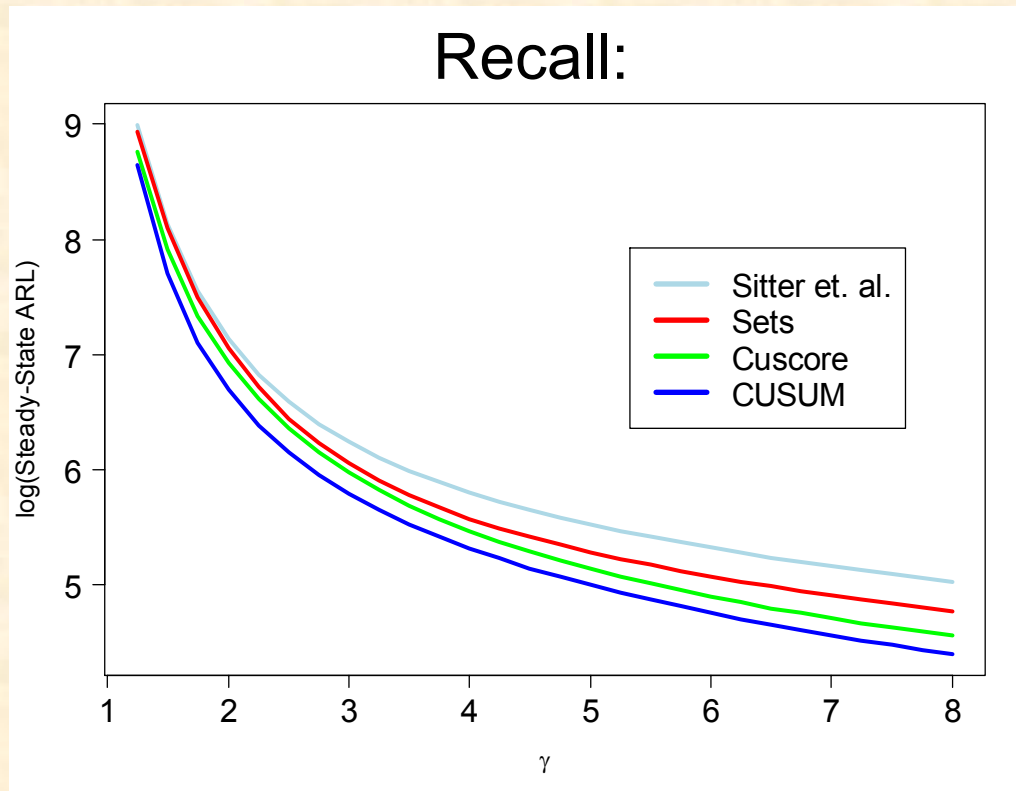
Of the $23 \times 28 = 644$ combinations of p_0 , M , γ_{design} , and γ for which we calculated the *SS-ARL*, the CUSUM outperformed all the other methods except at the following 13 combinations:

Method	p_0	M	γ_{design}	γ
Cuscore	0.001	50,000	6	1.25, 1.5, 1.75
Cuscore	0.0001	500,000	6	1.25, 1.5, 1.75
Sitter et.al.	0.01	25,000	6	1.25, 1.5, 1.75, 2.0
Sitter et. al.	0.005	50,000	6	1.25, 1.5, 1.75

Results

In 21 of the 23 combinations of p_0 , M , and γ_{design} , when $\gamma = \gamma_{design}$, the following inequality holds:

$$SS.ARL_{Sitter} \geq SS.ARL_{Sets} \geq SS.ARL_{Cuscore} \geq SS.ARL_{Cusum}$$



Results

In the 2 cases where the inequalities above do not hold, we have

$$SS.ARL_{Sets} \geq SS.ARL_{Sitter} \quad \text{and} \quad SS.ARL_{Cuscore} \geq SS.ARL_{Sets}$$

These differences can be explained by the fact that the ARL_0 of the apparently “better” method was closer to the target in-control ARL , M .

Conclusions

- Sitter et. al, Sets, and CUSCORE methods are all prone to delay when a shift occurs because the set in which shift occurs may exceed the threshold t and therefore not move the chart toward signaling an alarm.
- Sitter et. al. method typically shows the worst steady state performance because it is prone to the longest delays in reacting to a shift.

Conclusions

The CUSCORE method generally outperforms the Sets and Sitter et. al. methods because at any given point in time, the CUSCORE method is more likely to have accumulated more information about the process than the other two methods.

Conclusions

- Of the four methods considered, the CUSUM is best at accumulating information about the process.
- The CUSUM performs uniformly better than the other three methods except when the chart is designed for a large shift ($\gamma_{design} = 6$) and the actual shift is small ($\gamma \leq 2$).
- The *SS-ARL* performance of the CUSUM is always best when the actual shift matches the shift for which the chart was designed ($\gamma_{design} = \gamma$).

Recommendations

- For monitoring a process where Bernoulli observations can be observed sequentially (such as surveillance of birth defects), use the Bernoulli CUSUM.
- The Geometric CUSUM could also be used, since it is equivalent to the Bernoulli CUSUM (see Reynolds & Stoumbos (1999) JQT).

Acknowledgements

In addition to my advisors/co-authors listed at the beginning:

- Heidi Sego—for proofreading, listening, and patience.
- Mike Box and Jeff Norris—for excellent computer support.