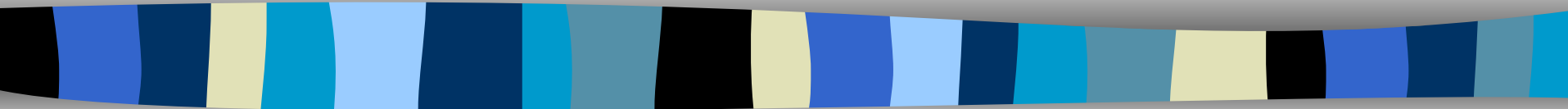


# Evaluation of school absenteeism data for early outbreak detection, New York City.



Besculides M, Heffernan R, Mostashari F, Weiss D

New York City Department of Health  
and Mental Hygiene



# Background

- Reports of unusually high school absenteeism to the Milwaukee Department of Health helped raise initial awareness of the 1993 *Cryptosporidium* outbreak.
- Monitoring school absenteeism may therefore be useful for early outbreak detection.
- The potential usefulness of a new data source should be evaluated before it is added to routine syndromic surveillance (SS) analysis.



# Objective

- To apply the CDC framework for evaluating syndromic surveillance systems to assess whether routine monitoring of school absenteeism data is useful for early outbreak detection.

1. Sosin DM. Draft framework for evaluating syndromic surveillance systems. *J Urban Health* 2003 Jun;80(2 Suppl 1):i8-13.



# Methods

- CDC framework for evaluating SS system:
  - System description
  - Data quality assessment
  - Experience with system performance
  - Outbreak detection capacity
  - Conclusions and recommendations



## Methods Cont.

- 2001-02 school absenteeism data for NYC schools were obtained from the NYC Department of Education.
  - 1.2 million students
  - 1200 schools
  - Data: number of students registered and absent by grade, school, and day
- Citywide trends in absenteeism were examined by age group (children aged 5-12, teenagers aged 13-18).



## Methods Cont:

### Temporal aberration detection

- Linear regression models predicting the expected number of students absent by age group were developed.
  - Controlled for day of week, seasonality, pre/post holiday
  - Deleted outliers (e.g., Halloween)
  - Graphed adjusted trends with confidence intervals
- The CuSum method was also applied using a 30-day baseline.



## Methods Cont:

# Spatial cluster detection

- Spatial Scan statistic with a 30-day baseline was used to assess geographic clustering of absenteeism.
  - Clusters of increased absenteeism could occur at one or more schools.
  - Excess absenteeism was calculated for the cluster as a whole and for each school in the cluster.

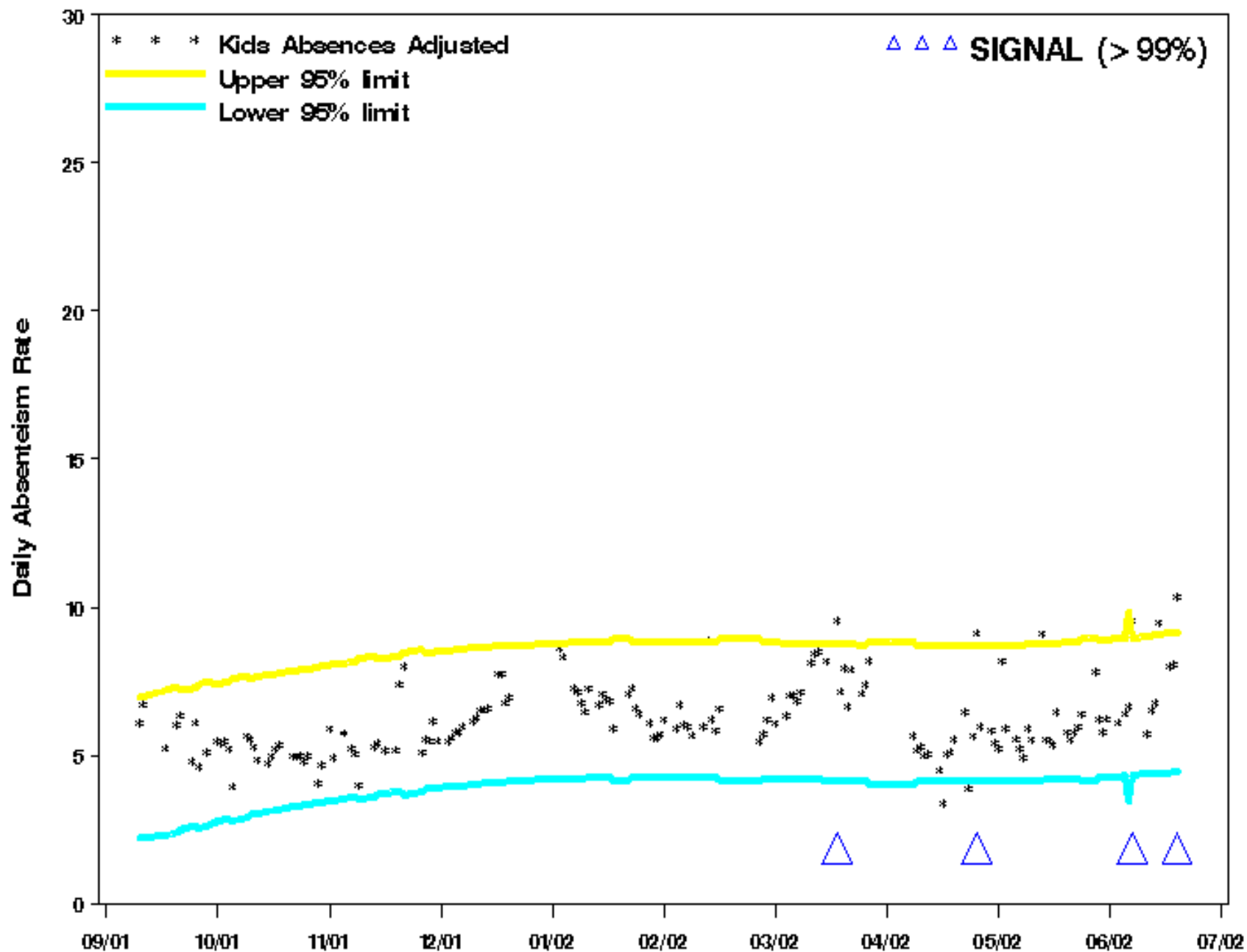


# Results

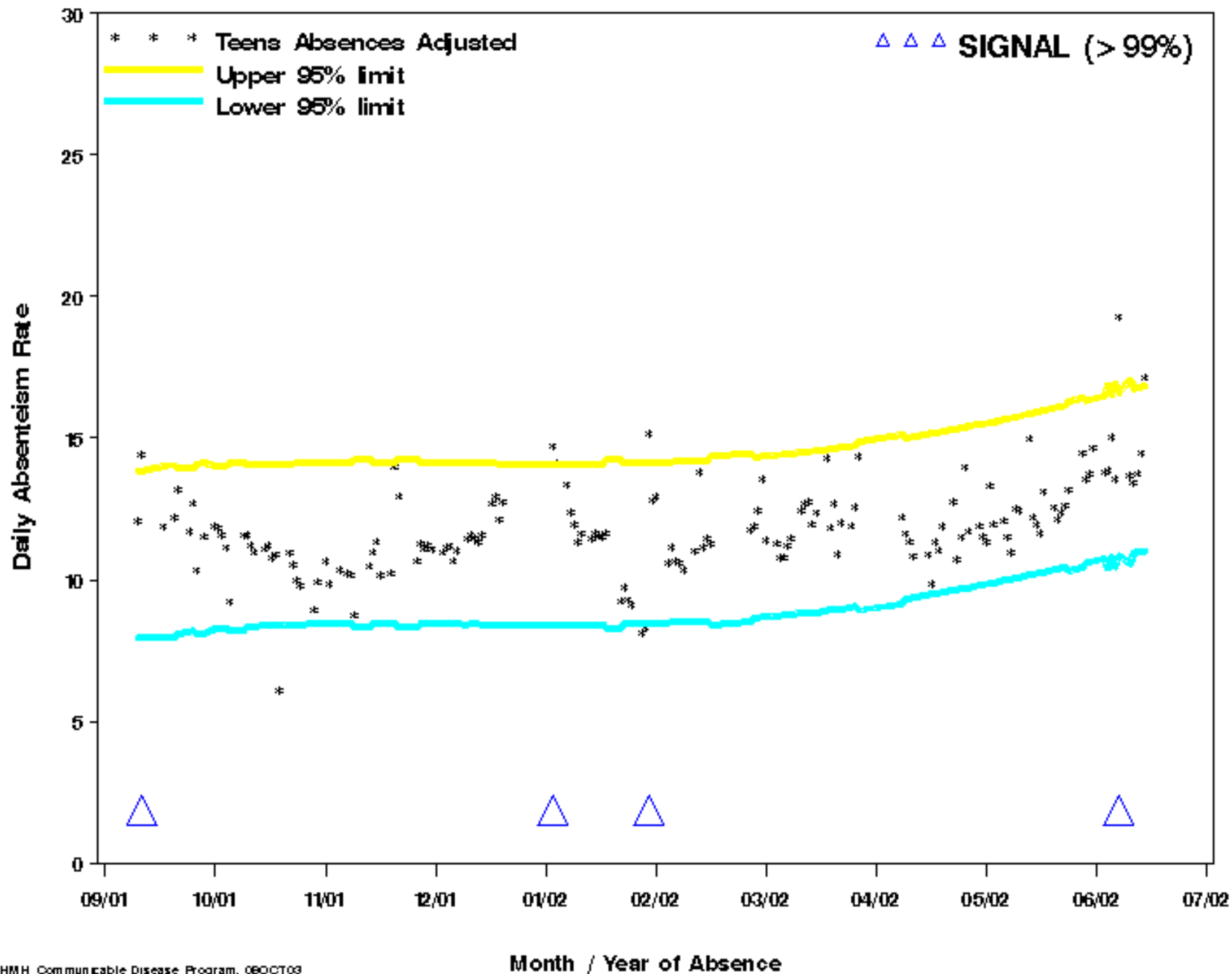
## Linear Regression Modeling

- 10.7% of students were absent each day (unadjusted)
  - Teenagers were more likely to be absent (13.8%) than children (7.5%).
- Adjusted data showed increases in absenteeism as the year progressed and around holidays. (Graphs)
- No clear citywide trends in absenteeism were evident during influenza season.

# Adjusted Absences Among Children



# Adjusted Absences Among Teenagers





# Results

## Cumulative Sums Method (CuSum)

- There were more days when children's absenteeism was above the expected compared to teenagers (25 days vs. 17 days).
- Most CuSum signals occurred around holidays and the end of the school year.



# Results

## SatScan Spatial Analysis-*Children*

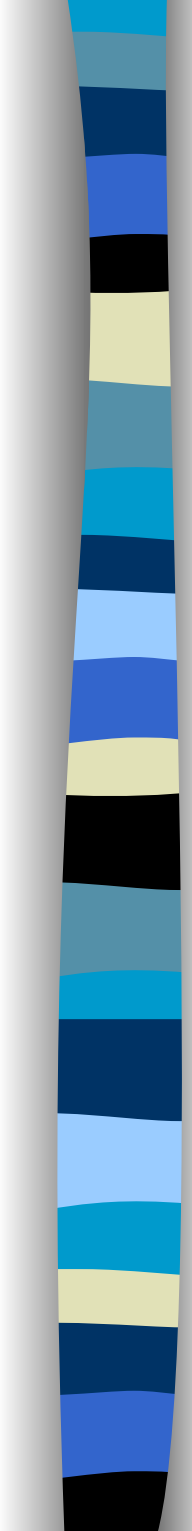
- There were 790 statistically significant ( $p \neq 0.01$ ) spatial clusters among children.
  - Average of 3 clusters per day
  - Average of 287 excess absences per cluster
  - Median of 5 schools in each cluster
  - Average of 15 excess absences per school

## Number of Times a School was in a $p \leq 0.01$ Spatial Cluster: New York City 2001-02

### cluster

- 1-9 clusters  $p \leq 0.01$
- 10-19 clusters  $p \leq 0.01$
- 20-29 clusters  $p \leq 0.01$
- 30-39 clusters  $p \leq 0.01$
- 40+ clusters  $p \leq 0.01$



- 
- The program successfully identified school X as the center of the most significant cluster during a known gastrointestinal outbreak at school X.
    - However, the size and RR of the cluster were average compared to other significant clusters during the year.
    - In looking at school level data, the excess absences in school X were well above average (but not rare) compared to the average excess in schools in other clusters during the year.



# Results

## SatScan Spatial Analysis-*Teenagers*

- There were 1018 significant ( $p \leq 0.01$ ) spatial clusters among teenagers.
  - Average of 4 clusters per day
  - 236 excess absences in each cluster
  - Median of 2 schools per cluster (mean 9)
  - Median of 15 excess absences per school (mean 44)
- The size of school did not affect the number of times a school was in a cluster for either age group.



## ■ CDC framework

### – Data quality:

- Data are representative of population
- Data are potentially available in real-time
  - Can be updated until end of school year, real-time data may be slightly inaccurate
- Data are not available during summer and holidays
- **Information on reason for absence not available**



## ■ CDC Framework

### – Experience with system:

- Not useful in detecting influenza season
- Useful in detecting a known outbreak but not earlier than it was reported to the DOHMH
- Evaluation took significant staff time but analysis developed could be used daily

### – Outbreak detection:

- System could be adapted easily to look for different size clusters (lower p-value, limit outbreak size) but doing this did not improve ability to detect true outbreaks
- Excess absences seen during outbreak at school X were not rare
  - several false signals would likely be investigated and true signals may be missed



# Conclusions

- Monitoring school absenteeism data is feasible and in some circumstances may allow detection of less severe illness or illness in earlier stages than emergency department surveillance.
- However, based on this evaluation, the DOHMH will not pursue prospective absenteeism surveillance until methods improve and/or reason for absence become available.

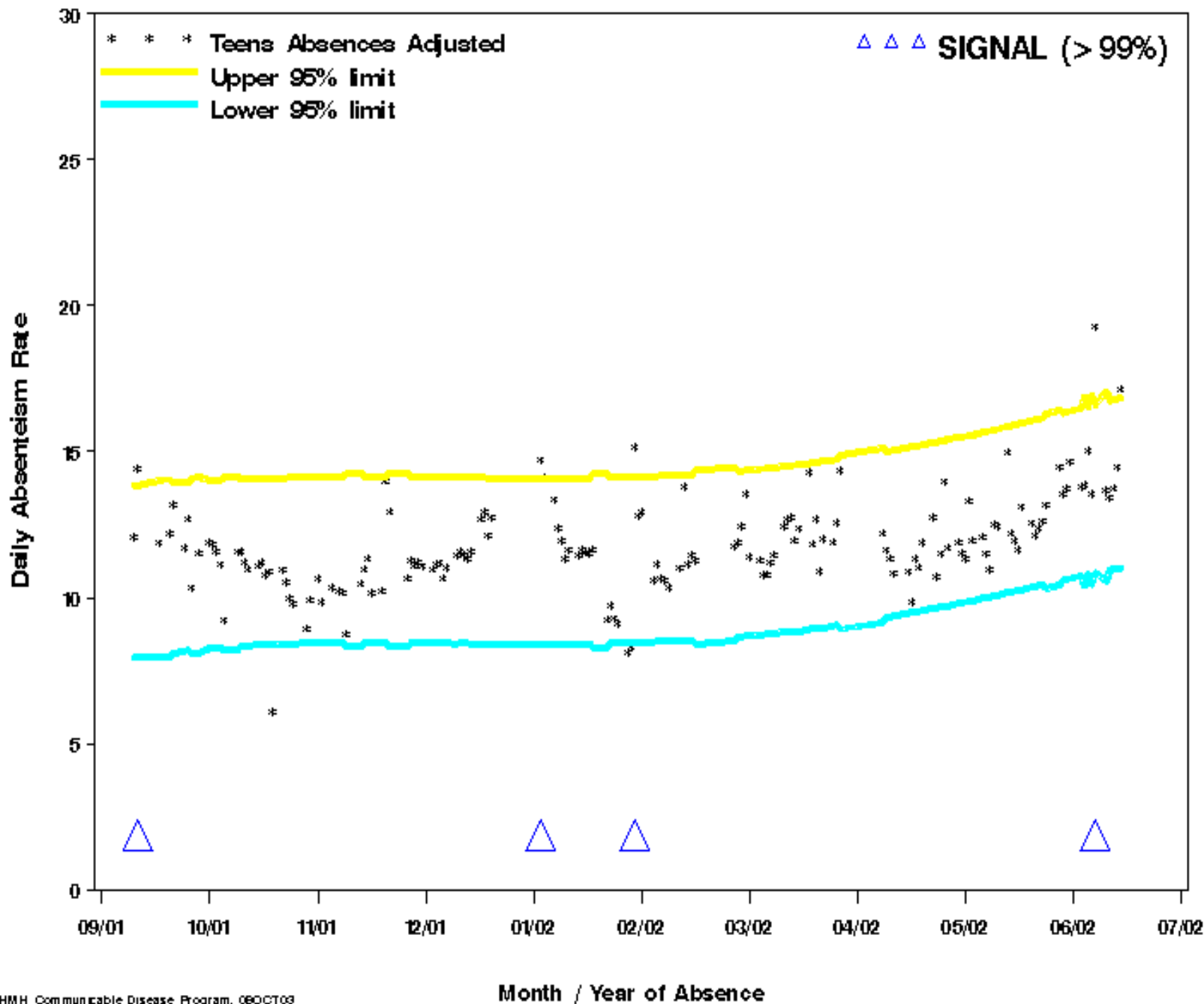
## Number of Times a School was in a $p \leq 0.01$ Spatial Cluster: New York City 2001-02

### cluster

- 1-9 clusters  $p \leq 0.01$
- 10-19 clusters  $p \leq 0.01$
- 20-29 clusters  $p \leq 0.01$
- 30-39 clusters  $p \leq 0.01$
- 40+ clusters  $p \leq 0.01$



# Adjusted Absences Among Teenagers



# Adjusted Absences Among Children

