A Space-Time Permutation Scan Statistic for Spatial Disease Surveillance

Developed jointly with:
Farzad Mostashari, Rick Heffernan, Jessica Hartman
New York City Department of Health

Problematic Issues

- What if there are no denominator data?
- Can we adjust for purely spatial clusters?
- Can we adjust for purely temporal clusters?
- What about missing data?

NYC Emergency Department Syndromic Surveillance System

- Daily reports of visits to hospital emergency departments
- Information about nature of the visit and residential zip-code
- Daily analyses for outbreaks

Electronic Log

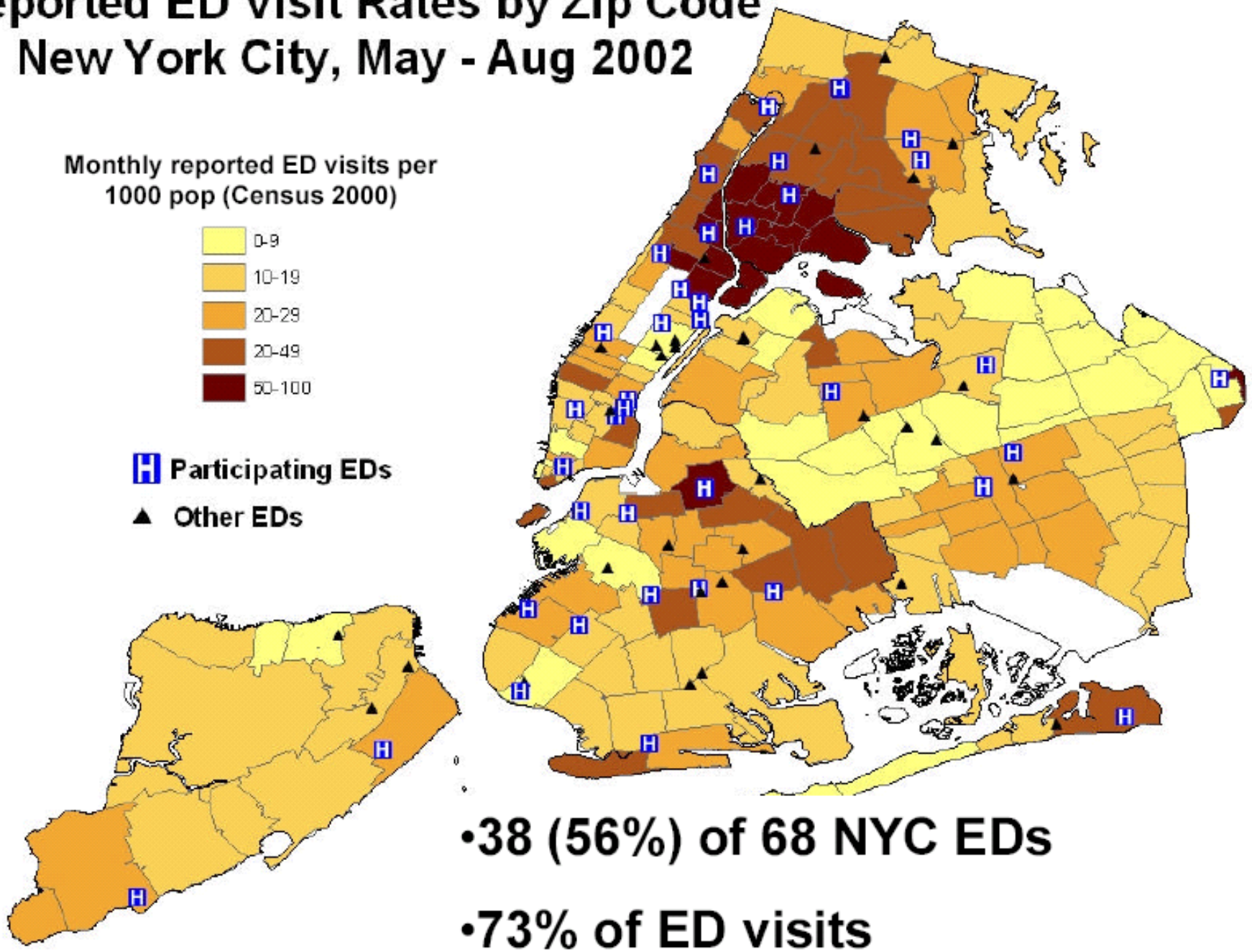
Admission List For 01/28/2002

AGE	SEX	TIME	CHIEF COMPLAINT	ZIP
15	M	01:04	ASSAULTED YESTERDAY, RT EYE REDDENED.	11691
1	M	01:17	FEVER 104 AS PER MOTHER.	11455
42	F	03:20		11220
4	F	01:45	FEVER, COUGH, LABORED BREATHING.	11507
62	F	22:51	ASTHMA ATTACK.	10013
48	M	13:04	SOB AT HOME.	10027
26	M	06:02	C/O DIFFICULTY BREATHING.	
66	M	17:01	PT. MOTTLED AND CYANOTIC.	10031

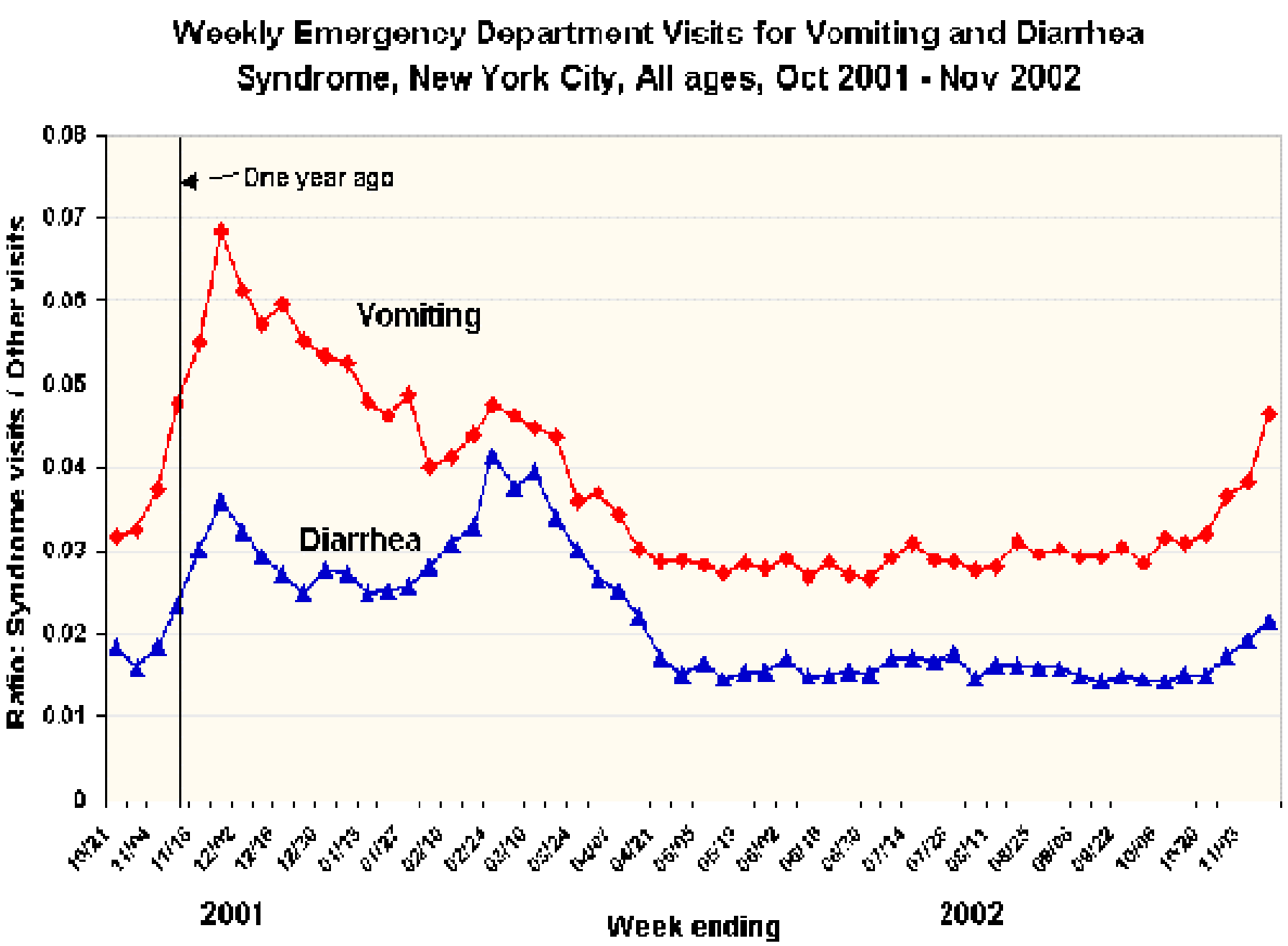
~4% have missing/uninformative information

~5% have missing zip-code

Reported ED Visit Rates by Zip Code New York City, May - Aug 2002



- 38 (56%) of 68 NYC EDs
- 73% of ED visits


Weekly Emergency Department Visits for Vomiting and Diarrhea Syndrome, New York City, All ages, Oct 2001 - Nov 2002
Ratio: Syndrome visits / Other visits
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
One year ago
Vomiting
Diarrhea
2001
2002
Week ending

Data Availability

- From October 2001, until present.
- Reliable spatial data starts October 26.

Why Use a Space-Time Scan Statistic?

With disease outbreaks:

- We do not know when they will occur.
- We do not know how rapidly they will emerge.
- We do not know where they will occur.
- We do not know their geographical size.

Space-Time Scan Statistic

Use a cylindrical window, with the circular base representing space and the height representing time.

We will only consider cylinders that reach the present time.

No Denominator Data

- Often, only case data are available.
- Impossible to detect purely spatial clusters
- Impossible to detect purely temporal clusters
- We can still look for space-time interaction (Knox, 1964)
- Space-time permutation scan statistic

Space-Time Permutation Scan Statistic

For each cylinder, calculate the expected number of cases conditioning on the marginals

$$\mu_{st} = \Sigma_s c_{st} \times \Sigma_t c_{st} \; / \; C$$

where c_{st} = # cases at time t in location s and C = total number of cases

Space-Time Permutation Scan Statistic

For each cylinder, calculate

$$T_{st} = [c_{st} / \mu_{st}]^{c_{st}} \times [(C-c_{st})/(C- \mu_{st})]^{C-c_{st}}$$

Test statistic $T = \max_{st} T_{st}$

Statistical Inference

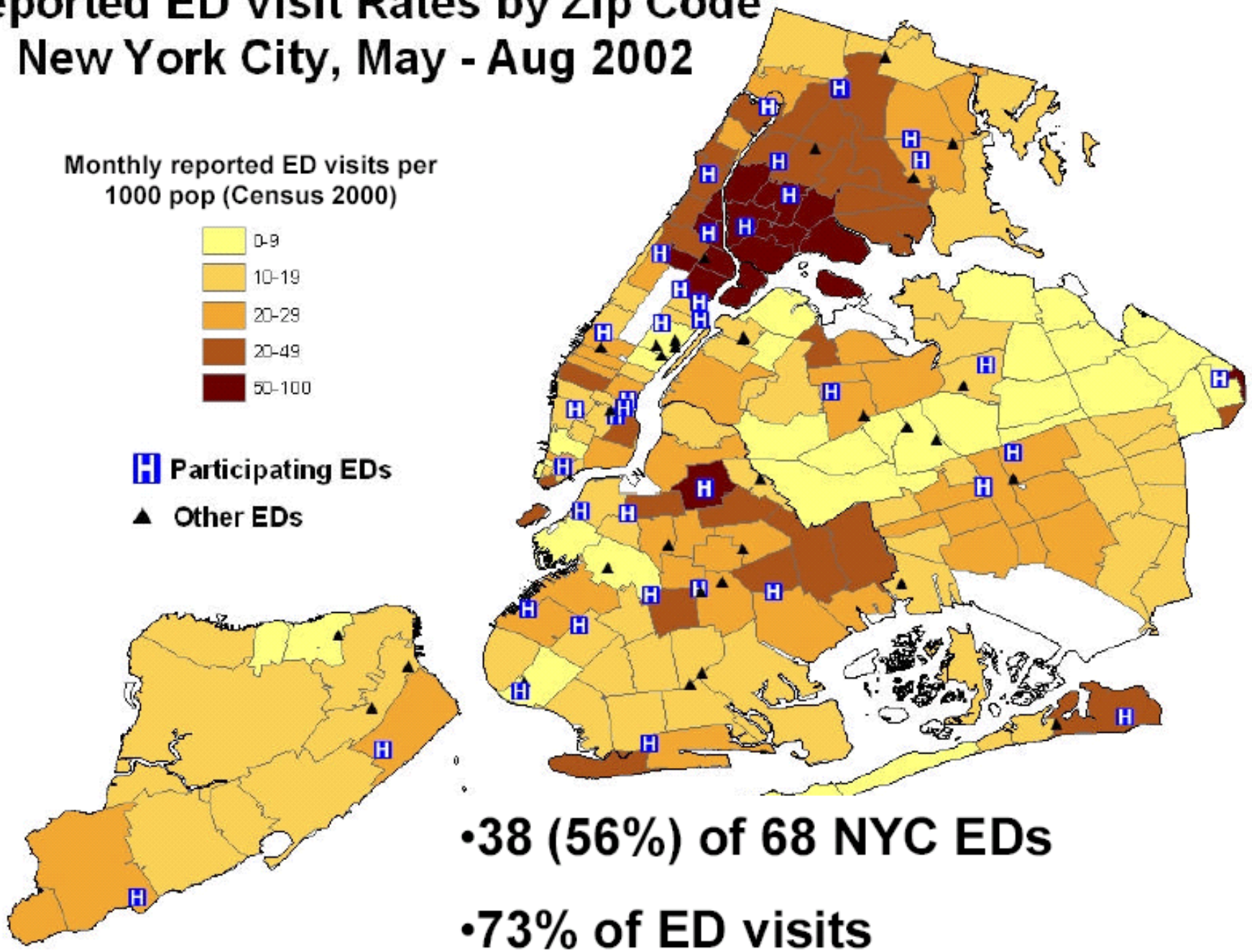
- Generate random replicas of the data set conditioned on the marginals, by permuting the pairs of spatial locations and times.
- Compare test statistic in real and random data sets using Monte Carlo hypothesis testing (Dwass, 1957):

$$p = \text{rank}(T_{real}) / (1+\#\text{replicas})$$

Space-Time Permutation Scan Statistic: Properties

- Adjusts for purely geographical clusters.
- Adjusts for purely temporal clusters.
- Simultaneously tests for clusters of any size and any location, by using circular windows with continuously variable radius.
- Accounts for multiple testing.
- Aggregated or non-aggregated data (counties, zip-code areas, census tracts, individuals, etc).

Reported ED Visit Rates by Zip Code New York City, May - Aug 2002



- 38 (56%) of 68 NYC EDs
- 73% of ED visits

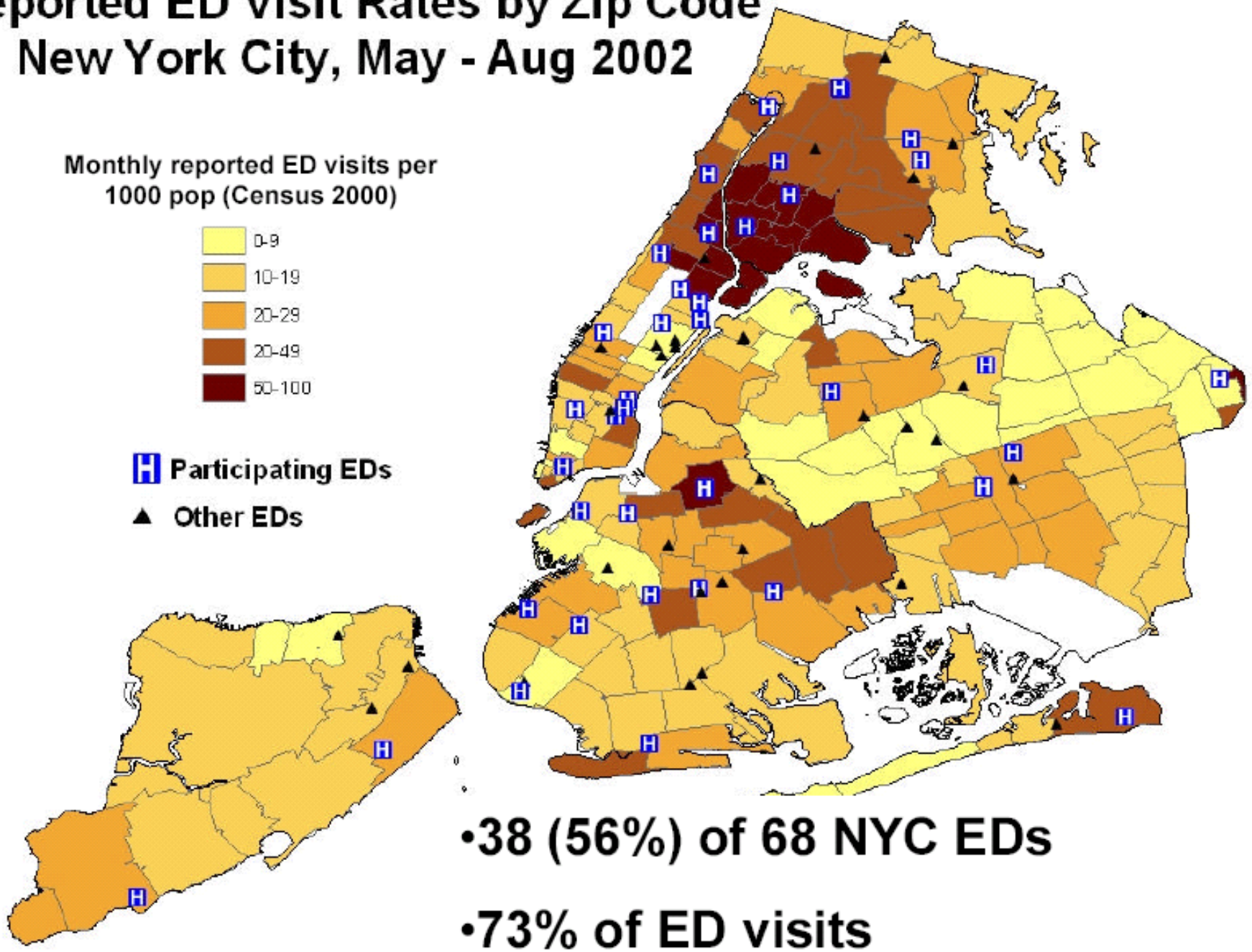
Application

- Diarrhea, all age groups
- Data available from October 26
- Surveillance starts 2 weeks later: November 8
- Spatial window size: 0-10 kilometers
- Temporal window size: 1-7 days

Results

Day	Primary Cluster	#Days	Radius	Cases	Expected	RR	p=
Nov 8	11201+5more	2	2.2	14	5.1	2.8	0.07
Nov 9	11377 + 10 more	1	3.1	15	6.6	2.3	0.44
Nov 10	11215 + 8 more	2	3.2	21	10.0	2.2	0.14
Nov 11	11207 + 28 more	3	9.0	141	113	1.2	0.71
Nov 12	11210, 11230, 11226, 11229	1	2.7	14	5.4	2.6	0.23
Nov 13	10451	1	0	6	1.2	4.8	0.24
Nov 14	11421	1	0	6	0.8	8.0	0.009
Nov 15	11385, 11379, 11421, 11208	3	2.7	36	18.4	2.0	0.04
Nov 16	10306 + 7 more	2	7.9	9	2.0	4.4	0.04
Nov 17	11414, 11421, 11208 + 6 more	5	4.5	67	42.7	1.6	0.08
Nov 18	11218 + 7 more	1	3.0	19	7.1	2.7	0.02
Nov 19	11218 + 6 more	2	2.7	25	12.0	2.1	0.17
Nov 20	11225 + 4 more	3	2.0	46	27.9	1.6	0.25
Nov 21	11225 + 6 more	4	2.3	66	44.0	1.5	0.35

Reported ED Visit Rates by Zip Code New York City, May - Aug 2002



- 38 (56%) of 68 NYC EDs
- 73% of ED visits

Outbreak Investigation

- School Children
- ~ 150 total cases
- 19 vomited at school after lunch
- 18 visited hospital emergency departments
- Calicivirus

What If?

- Hospital A tends to have many ED visits on weekdays
- Hosptial B tends to have many ED visits on weekends.

Adjustment for Day-of-Week/Space Interaction

Problem: Weekly variation in rates may differ in different areas

Solution: Randomize stratified by day-of-week

Warning: Less amount of baseline data may create problems

Missing Data

Three Options:

- Remove geographical area (or data source) for all days.
- Remove day for all geographical areas
- If day-of-week is a covariate, remove geographical area only for selected days of the week.

Year Long Run

- Daily analyses of ER diarrhea visits
- From November 2001 to November 2002
- Thirty days of data used in each run.
- Max # days = 7, Max radius = 5km
- Adjusted for spatial / day-of-week interaction.
- Adjusted for missing data, using all three methods at different times.
- Cut-off: One false signal per year ($p<0.0028$)

Results: Two Signals

February 9, 2002:

Bronx, 15 zip-codes, 2 days

63 observed, 34.7 expected

p=0.0001, occurrence rate = 1 / 27.5. years

March 7, 2002:

Northern Manhattan, 8 zip-codes, 2 days

63 observed, 37.3 expected

p=0.0027, occurrence rate = 1 / 12.2 months

Limitations

- Space-time clusters may occur for other reasons than disease outbreaks
- Automated detection systems does not replace the observant eyes of physicians and other health workers.
- Epidemiological investigations by public health department are needed to confirm or dismiss the signals.

Conclusions

- The space-time permutation scan statistic can serve as an important tool in prospective systematic time-periodic geographical surveillance for the early detection of disease outbreaks.
- Only case data are needed.

Software

Kulldorff M and IMS Inc. SaTScan v.3.1: Software for the spatial and space-time scan statistics, 2003. http://www.satscan.org/

Acknowledgement

Research funded by the Alfred P Sloan Foundation.