Using Routinely Collected Surveillance Data to Predict TB Outbreaks in the United States

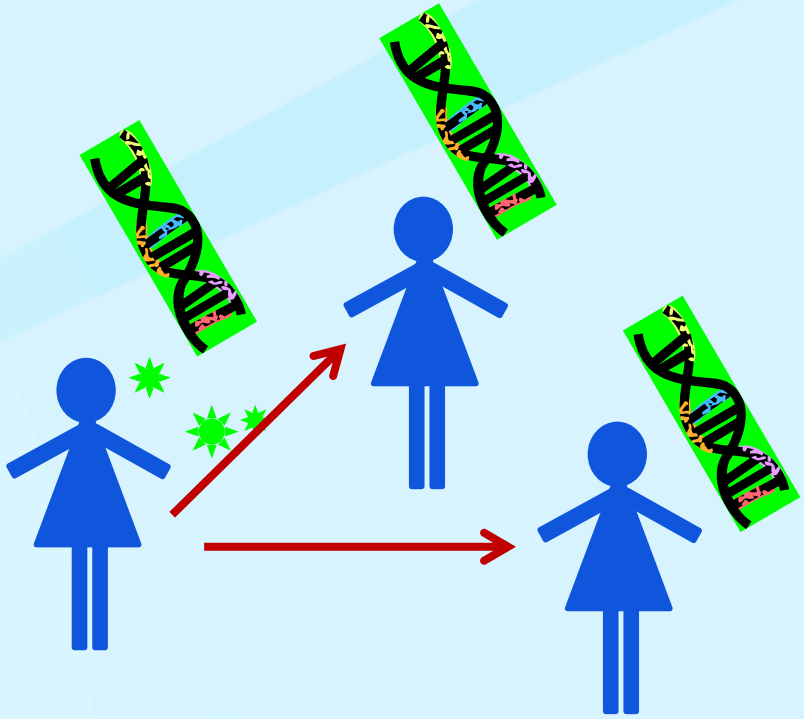
Thomas R. Navin, MD

Chief of Surveillance, Epidemiology, Outbreak Investigation Branch
Centers for Disease Control and Prevention
Atlanta, Georgia

March 2, 2013

TB genotype background

- Examines genetic variation to determine relatedness of TB strains from different cases
- Only performed with culture-positive cases
- Outbreaks occur when highly infectious TB patients transmit TB
  - "Same" place
  - "Same" time
  - Secondary cases have same genotype

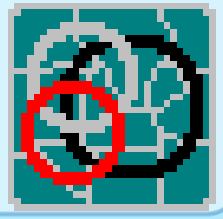


Methods overview

- **Genotyping with spoligotyping + 12-locus MIRU-VNTR**
- **SaTScan**
  - To identify TB clusters (to define our analytic cohort)
- **Decision Tree Analysis**
  - To identify clusters at high risk of becoming outbreaks

SaTScan

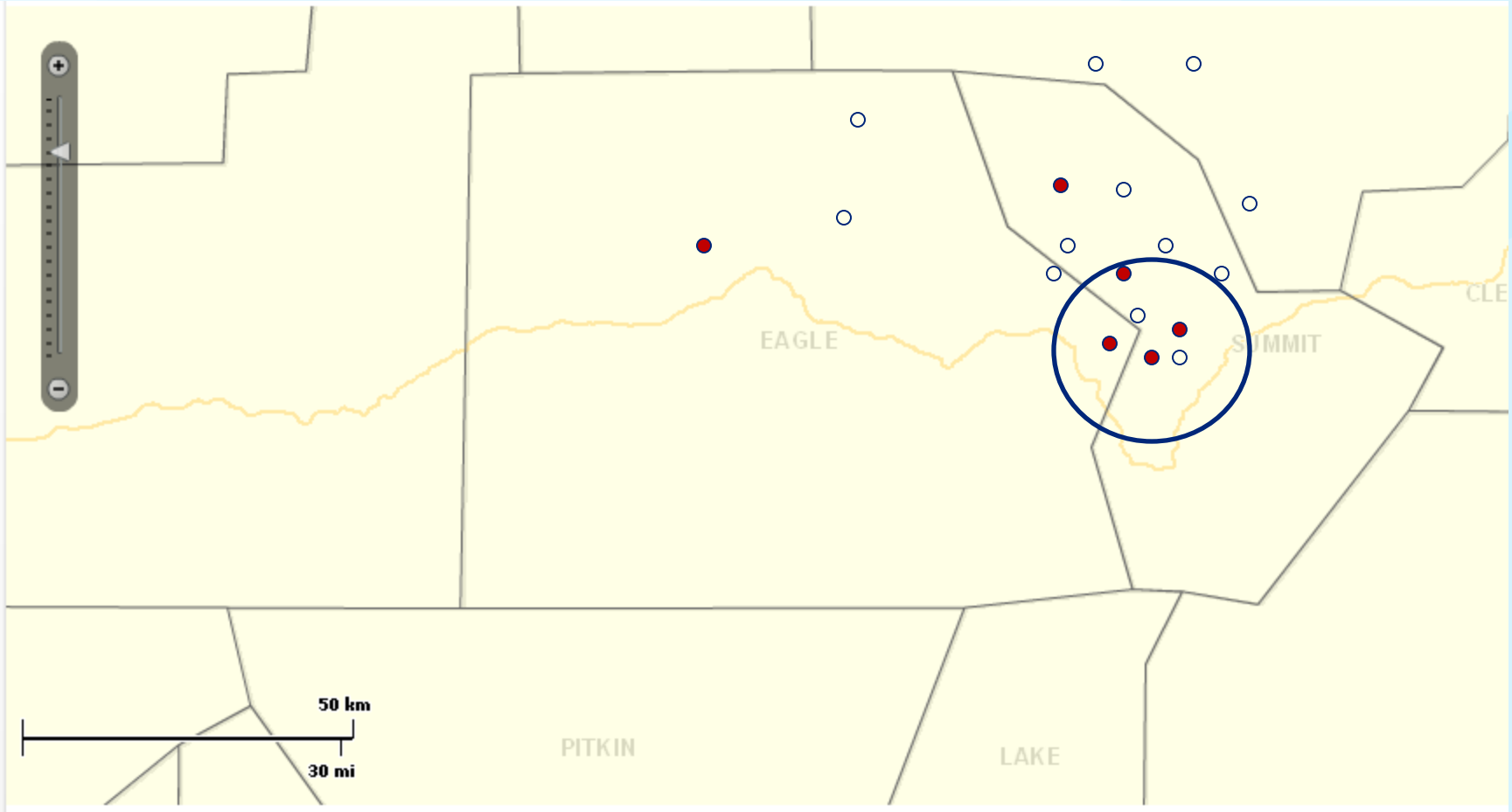
- **GIS software standard**
  - Performing geographical surveillance of disease
  - Detecting spatial disease clusters
- **Poisson-based model**
- **Monte Carlo simulation used to compute p-value for each cluster**



SaTScan parameters

- TB cases with genotyping results reported in U.S. from 2006 through 2010
- Search within 50 km radius
- Significant and non-significant

Sample SaTScan map



How we defined an incident TB cluster

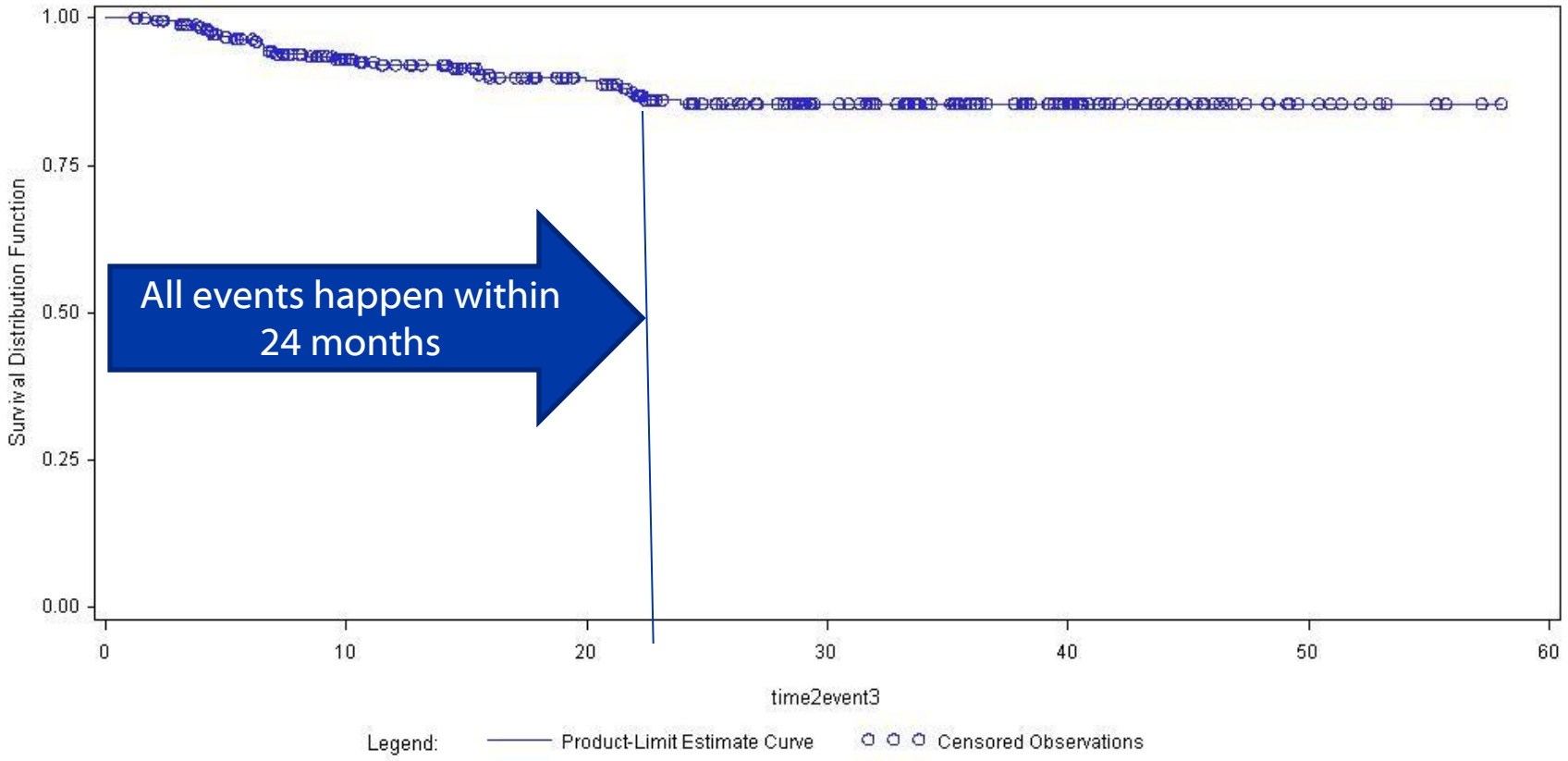
- Identified by SaTScan (2006-2010)
- Clusters ≥3 cases with same genotype
- Annual county genotyping coverage ≥75%
- No matching cases in preceding 24 months
  - Identify new (i.e., incident) clusters

How we defined our outcome: "outbreak"

From beginning (January 2006) to end (Dec 2010) of observation period:

- Grew from 3 to ≥6 cases
- Within 24 months
- Confirmed by local public health officials as an outbreak

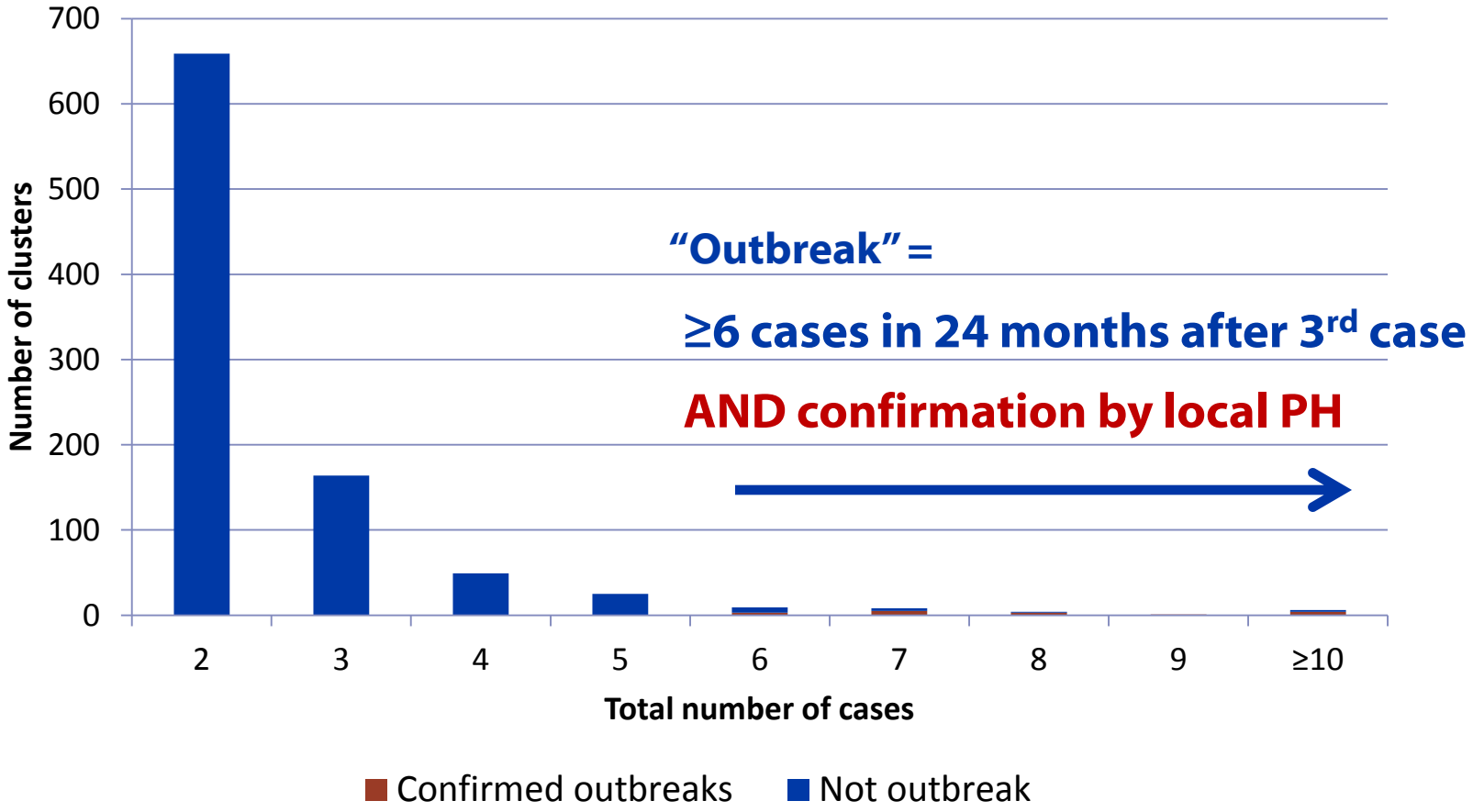
Time to reach 6 cases, N=146



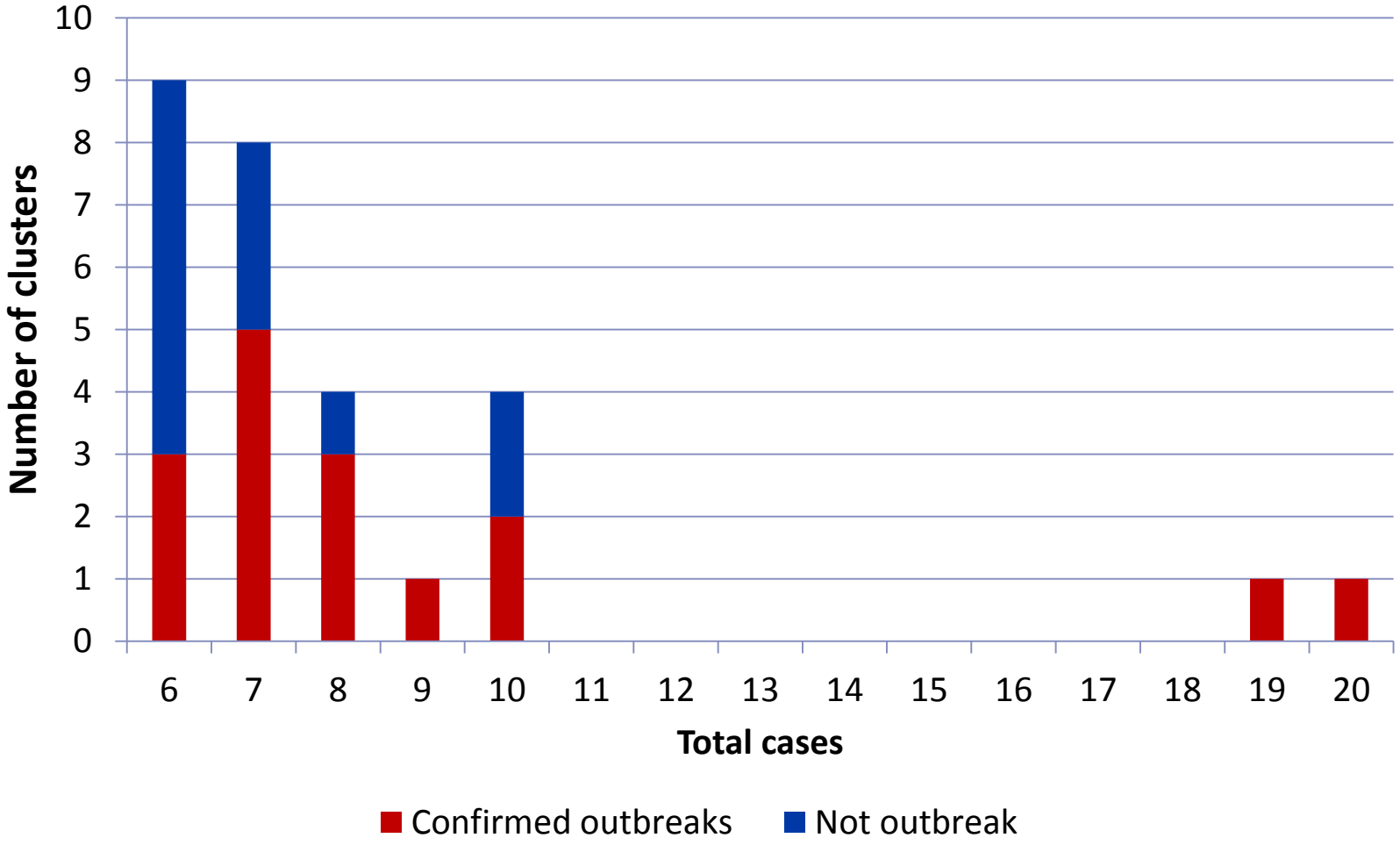
Entry into cohort: 3rd case

Event: 6th case in an outbreak cluster of n>=6, else censored at end of timeline (Dec 2010)

Distribution of clusters by final cluster size, stratified by outcome, N=925

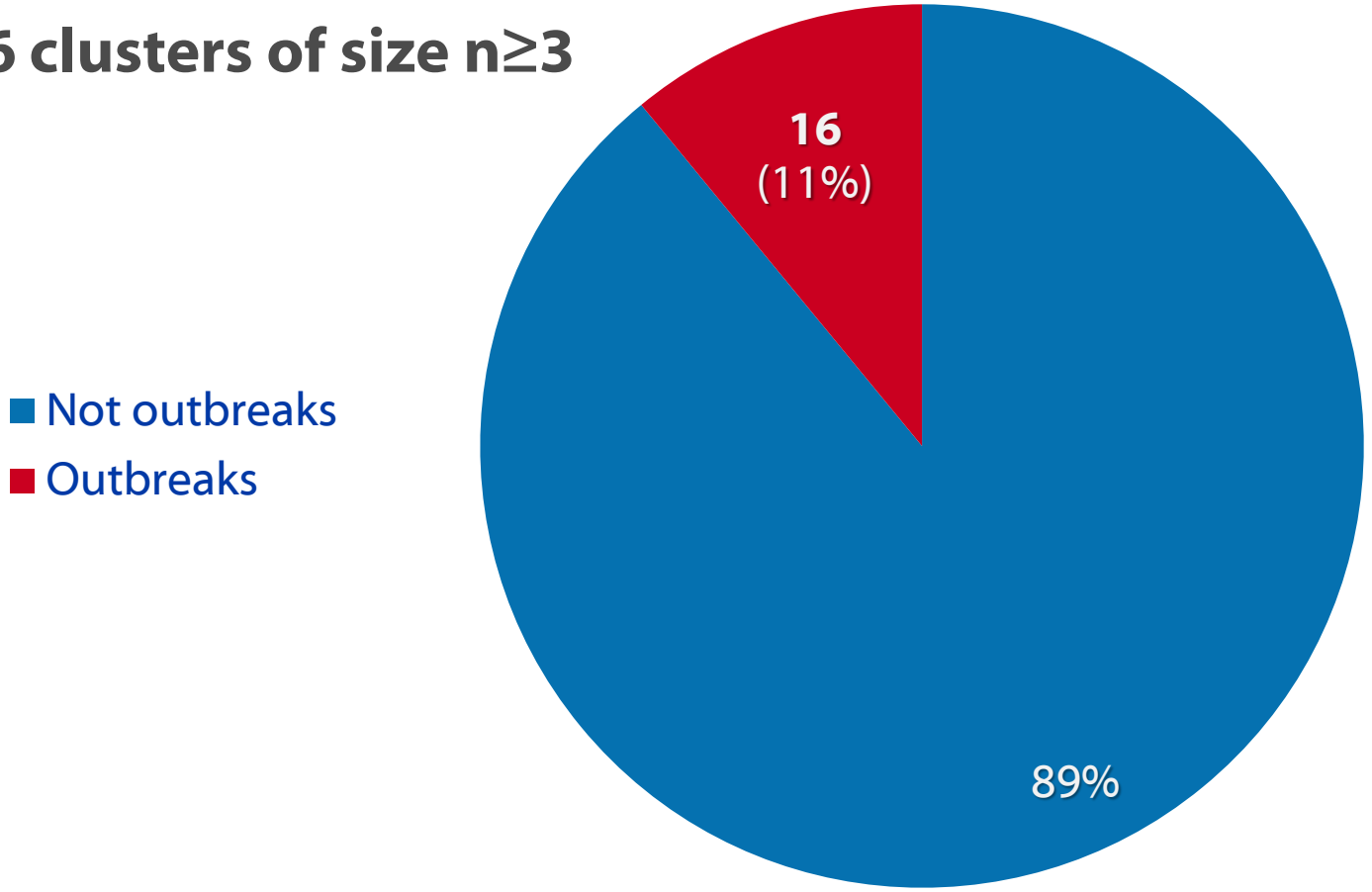


Distribution of clusters by final cluster size, stratified by outcome, n>=6



Outcomes

- 146 clusters of size n≥3



Characteristics of 146 TB clusters and associated risk factors of cluster becoming an outbreak case

Cluster characteristic	Outbreak clusters (n=16) with characteristic		Relative risk in predicting outbreaks	
Patient risk factors	n	%	RR	95% CI
Homeless or excess alcohol use or illicit drug use or incarceration at diagnosis	**15**	**22.4**	**17.7**	**2.4, 130.4**
Homeless or alcohol or drug use	**14**	**21.9**	**9.0**	**2.1, 38.0**
Drug use	**12**	**25.5**	**6.3**	**2.2, 18.6**
Alcohol use	**11**	**25.0**	**5.1**	**1.9, 13.8**
Homeless or alcohol use	**11**	**22.0**	**4.2**	**1.6, 11.5**
Homeless and alcohol use	**6**	**30.0**	**3.8**	**1.5, 9.3**
Incarceration	**5**	**29.4**	**3.4**	**1.4, 8.7**
Homeless	**6**	**23.1**	**2.8**	**1.1, 6.9**

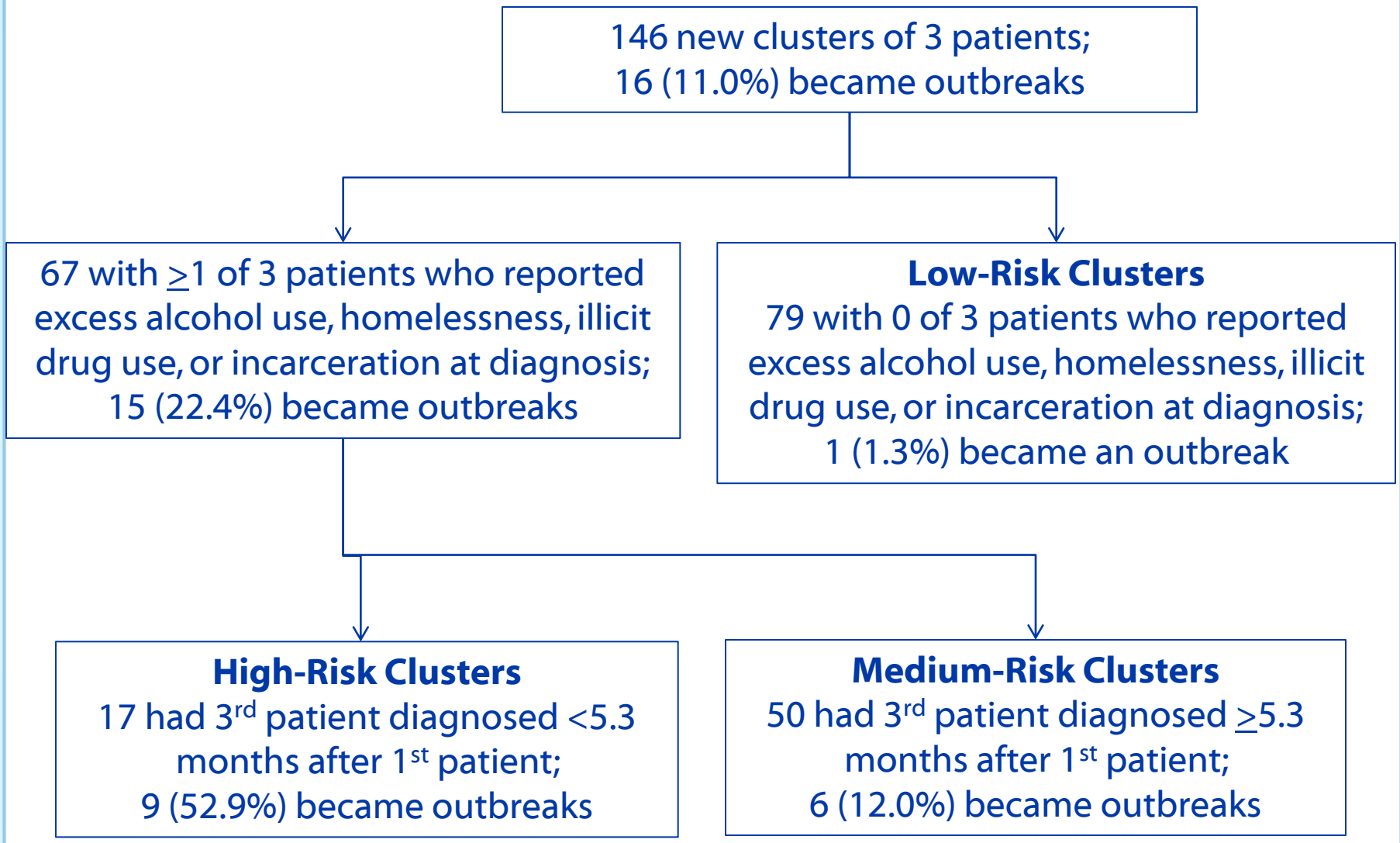
Characteristics of 146 TB clusters and associated risk factors of cluster becoming an outbreak case

Cluster characteristic	Outbreak clusters (n=16) with characteristic		Relative risk in predicting outbreaks	
Cluster risk factors	n	%	RR	95% CI
T1 to T3 <5.3 months	**9**	**21.4**	**3.2**	**1.3, 8.0**
T1 to T2 <4.4 months	**12**	**16.9**	**3.2**	**1.1, 9.4**
T2 to T3 <0.9 months	**7**	**21.2**	**2.7**	**1.1, 6.6**
Sig. LLR at 3rd case	**13**	**16.3**	**3.5**	**1.0, 11.6**

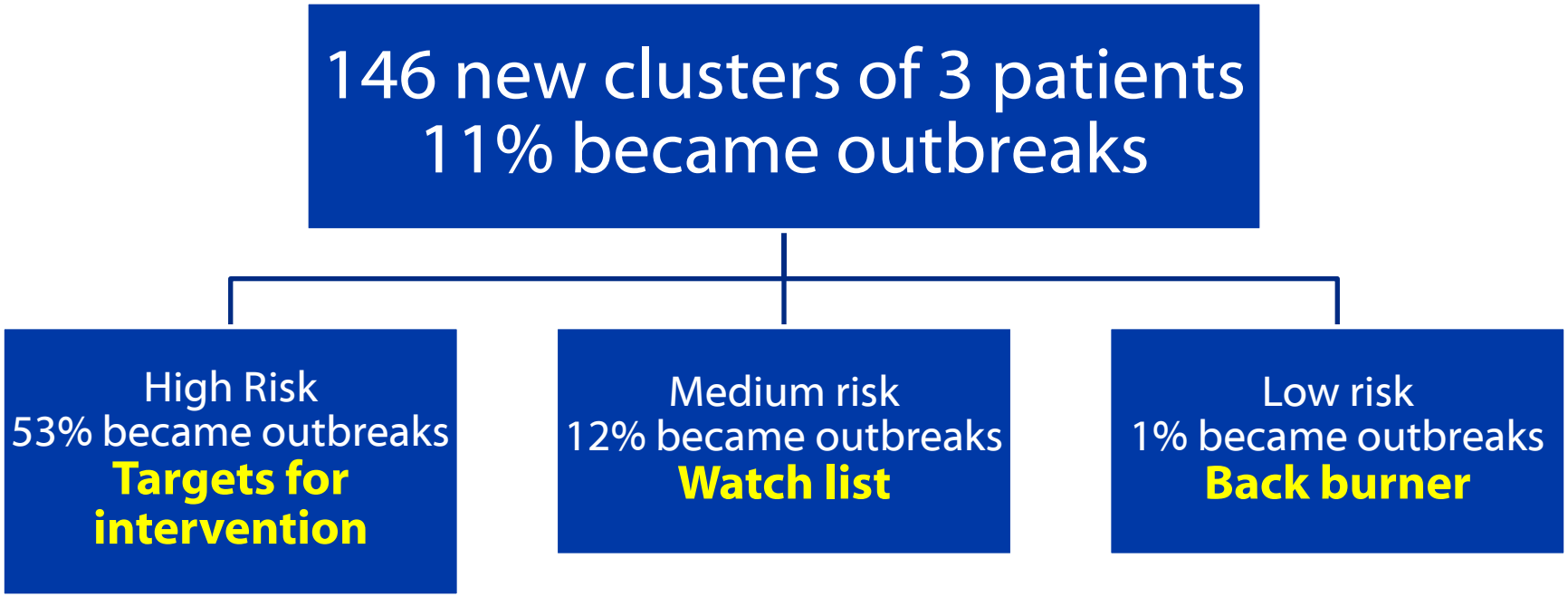
Decision Tree analysis using JMP

- **SAS Predictive modeling software**
- **Recursive partitioning function**
  - "enables users to systematically analyze large data sets to discover unsuspected or unknown relationships.
  - finds a set of cuts or groupings of X values that best predict an outcome by exhaustively searching all possible cuts or groupings, recursively forming a tree of decision rules until the desired fit is reached."

Algorithm for predicting outbreaks at 3rd case



Predicting outbreaks at 3rd case



Limitations

- Surveillance data
- Observation period of only 5 years
- 75% genotyping coverage
  (We now have >90% coverage)
- Based on 12-locus MIRU
  (We now have 24-locus MIRU data)

Conclusion

- With this methodology, we are able to predict half of clusters that would become outbreaks
- This provides an approach to prioritizing clusters for early intervention

Acknowledgements

- **CDC**
  - Sandy Althomsons
  - Steve Kammerer
  - Nong Shang
  - Juliana Grant
  - Krista Powell
  - Roque Miramontes
  - Lauren Cowan
- **State and local TB programs**