A Bayesian hidden Markov model for notifiable disease surveillance

Rochelle Watkins

Australian Biosecurity CRC

Curtin University of Technology

Faculty of Health Sciences

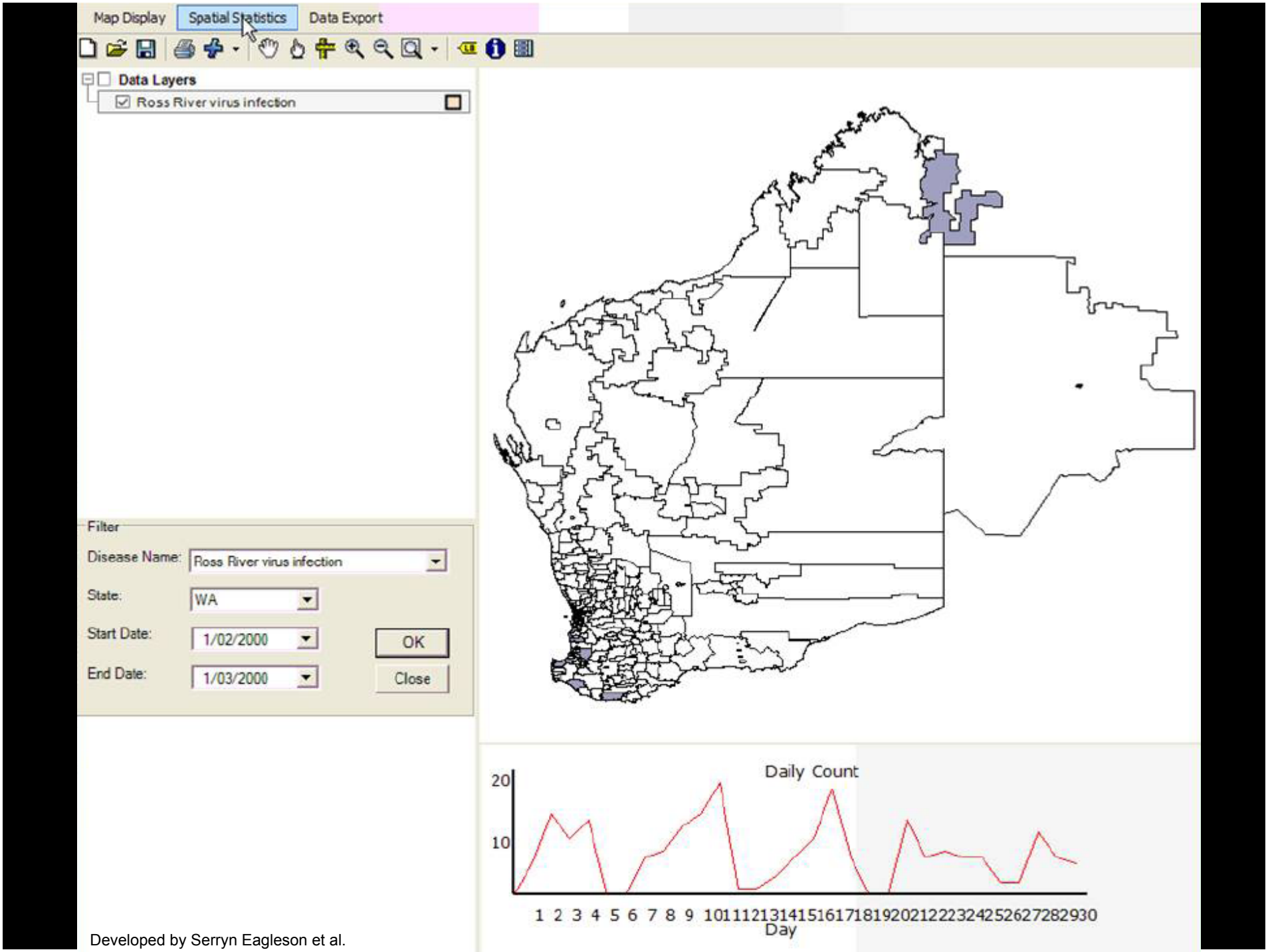
Background

Aim

- to develop an automated method to monitor routinely collected notifiable disease data
- efficient, comprehensive
- enable response, inform management

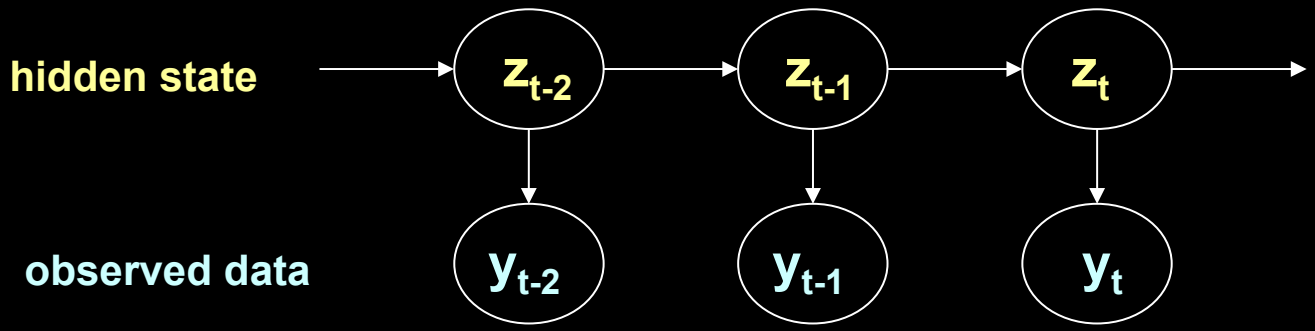
Context

- data updated daily (0,0,0,0,0,0,0,0,…)
- postcode level
- detection goal: poorly defined, highly variable



Developed by Serryn Eagleson et al.

Bayesian Hidden Markov Model



2 hidden states used to classify observed data based on pre-specified distributions

Intuitive

- model what we are interested in
- serial dependency a key component of the model

Interpretation

- what is the probability of an outbreak today, given our data?

Bayesian Approach

Provides a formal method to incorporate expert knowledge

- account for uncertainty in these unknowns

May be more robust to system changes over time

- simplify maintenance

Computationally intensive for large scale spatio-temporal data

Model

t: time (day)
i: area (postcode)
z: state (endemic, outbreak)

x[t,i] <- count[t,i] + count_neigh[t,i]

x[t,i] ~ dpois(mu[z[t,i]])

z[t,i] ~ dcat(p[z[t-1,i],1:K])

7 day model structure

- analyse 7, 14, 28 days of data

Priors

Relatively uninformative constrained priors for means:

mu[1] ~ dgamma(10,10)

mu[2] ~ dgamma(40,20)|(mu[1],)

Gamma equivalent to Dirichlet prior on transition matrix:

```
for(k in 1:K)
for (l in 1:K)
p[k,l] <- px[k,l]/sum(px[k,])
px[k,l] ~ dgamma(alpha[l],1)
```

Evaluation Scenario

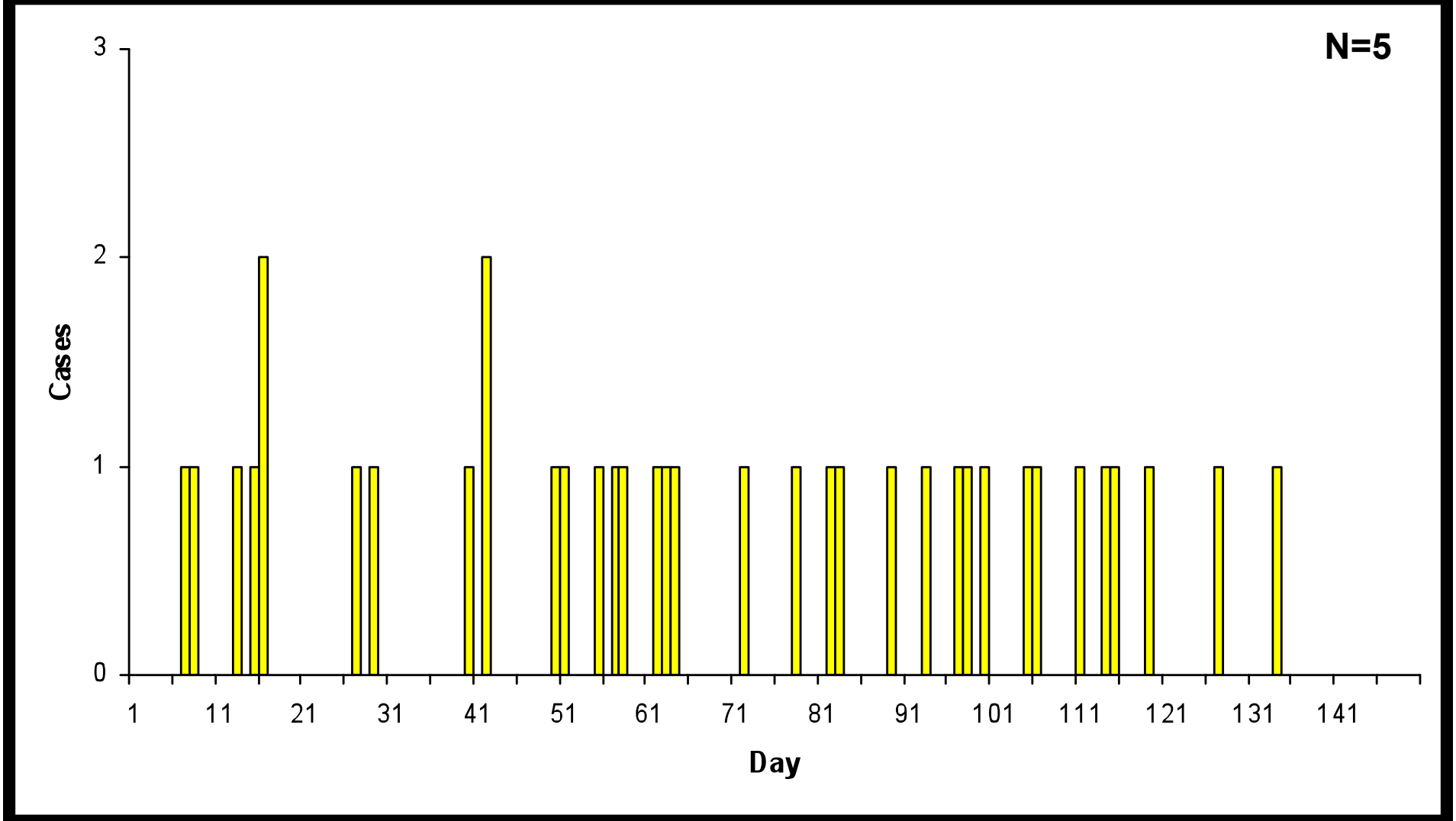
Hepatitis A in Western Australia

- simulated outbreaks
- authentic baseline

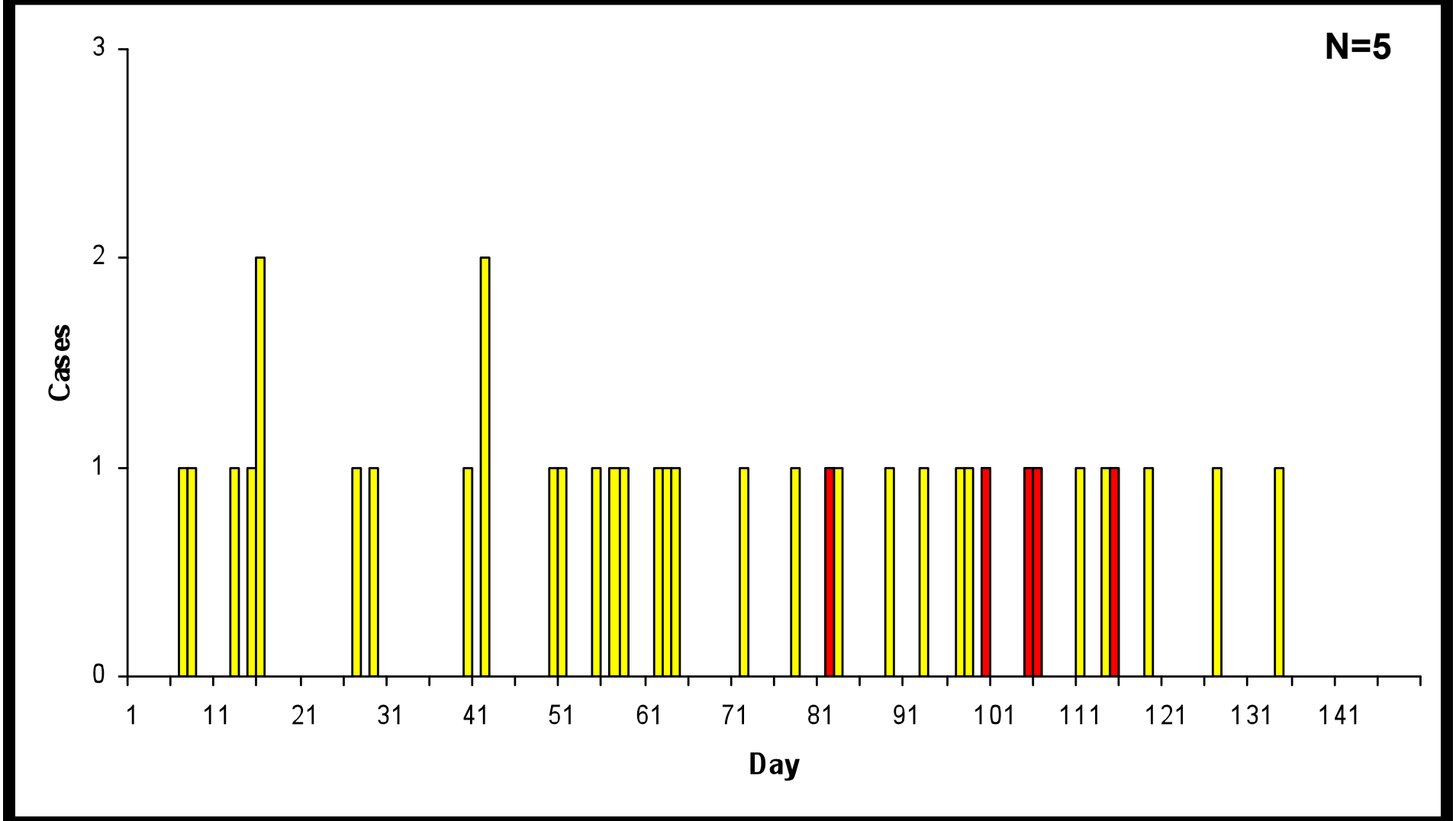
4 replications of 150 days * 60+ trials

- size of outbreak
- clustering of cases

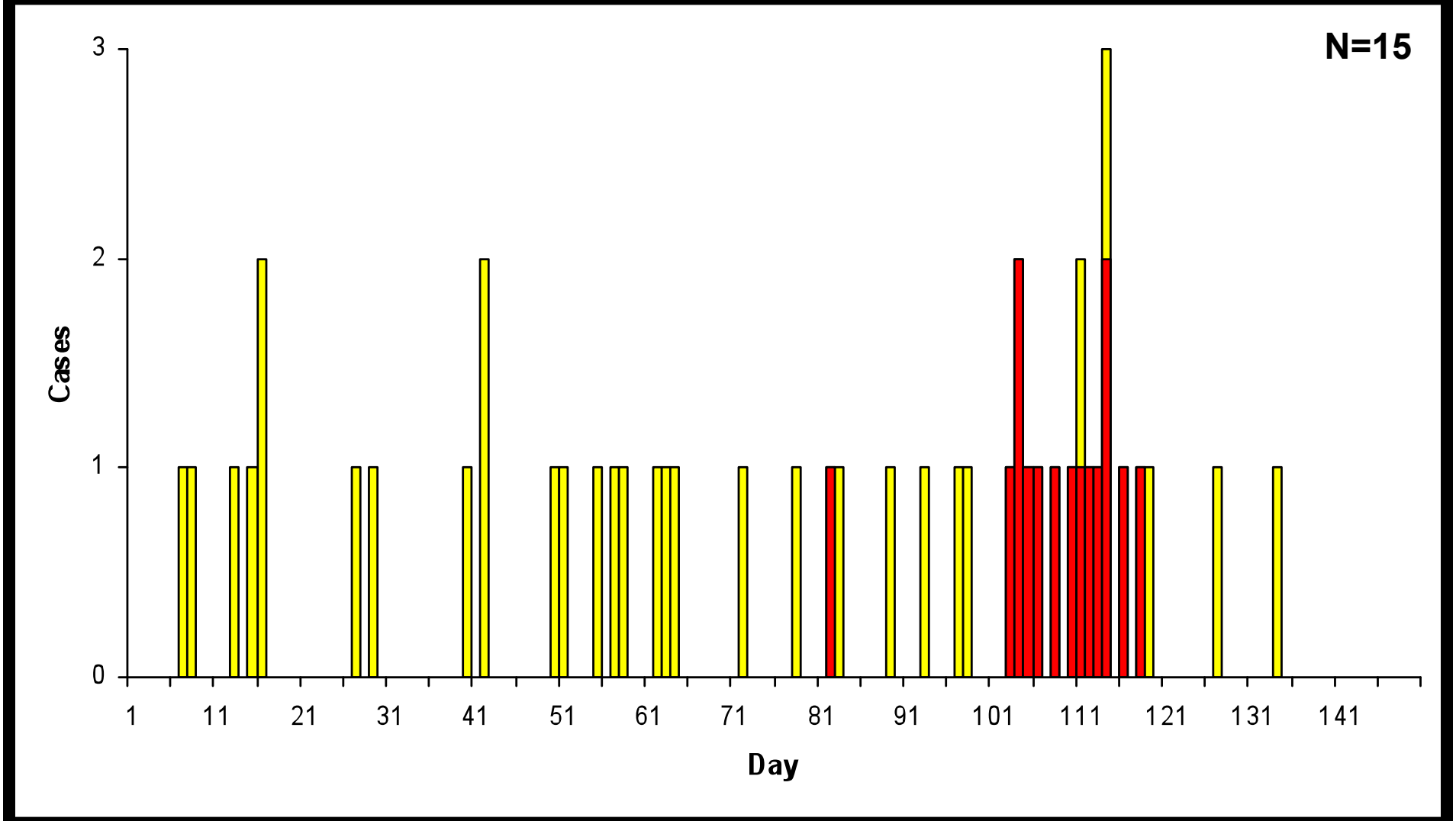
Small outbreak



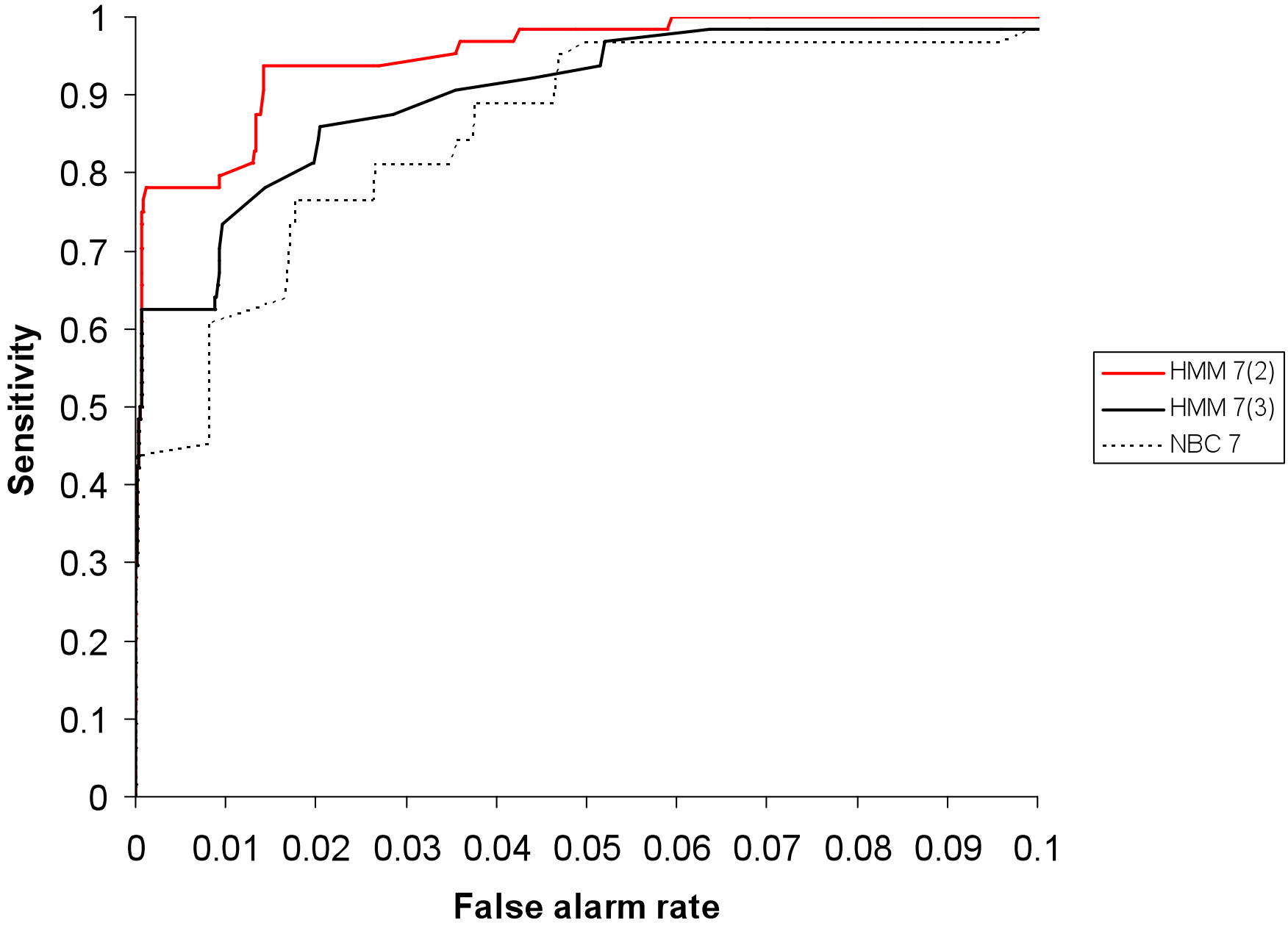
Small outbreak



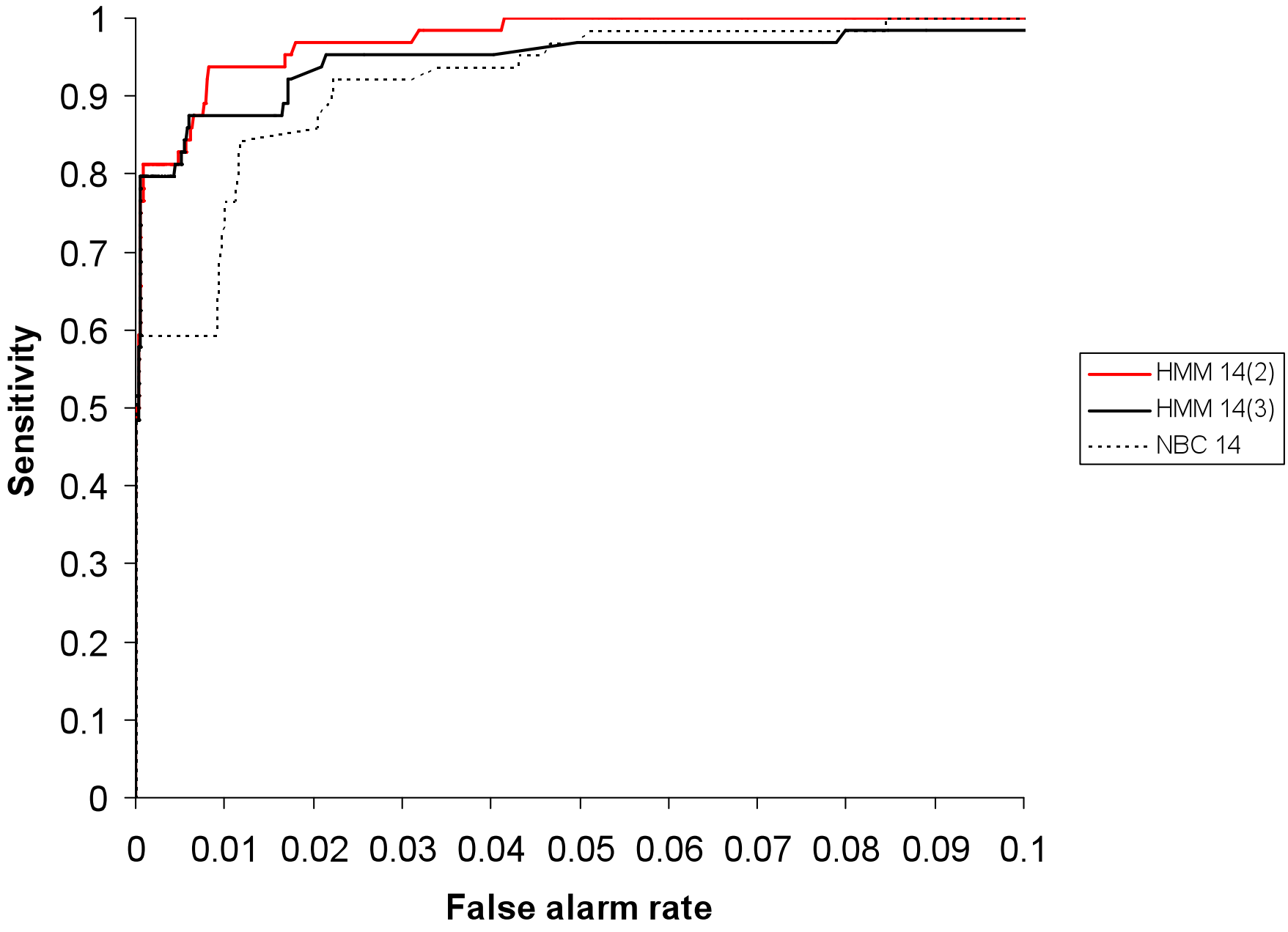
Large outbreak



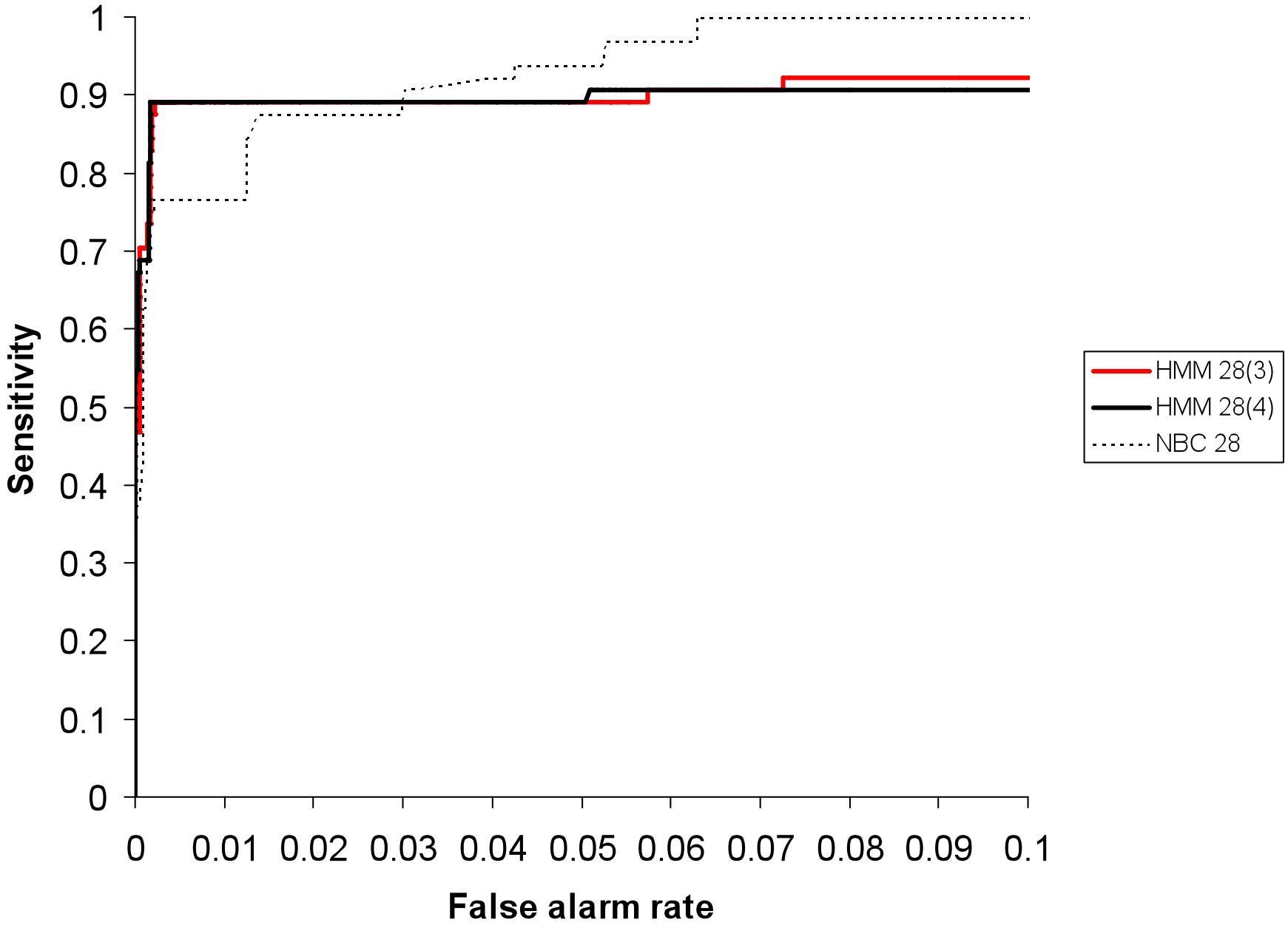
Small less clustered – 7 day



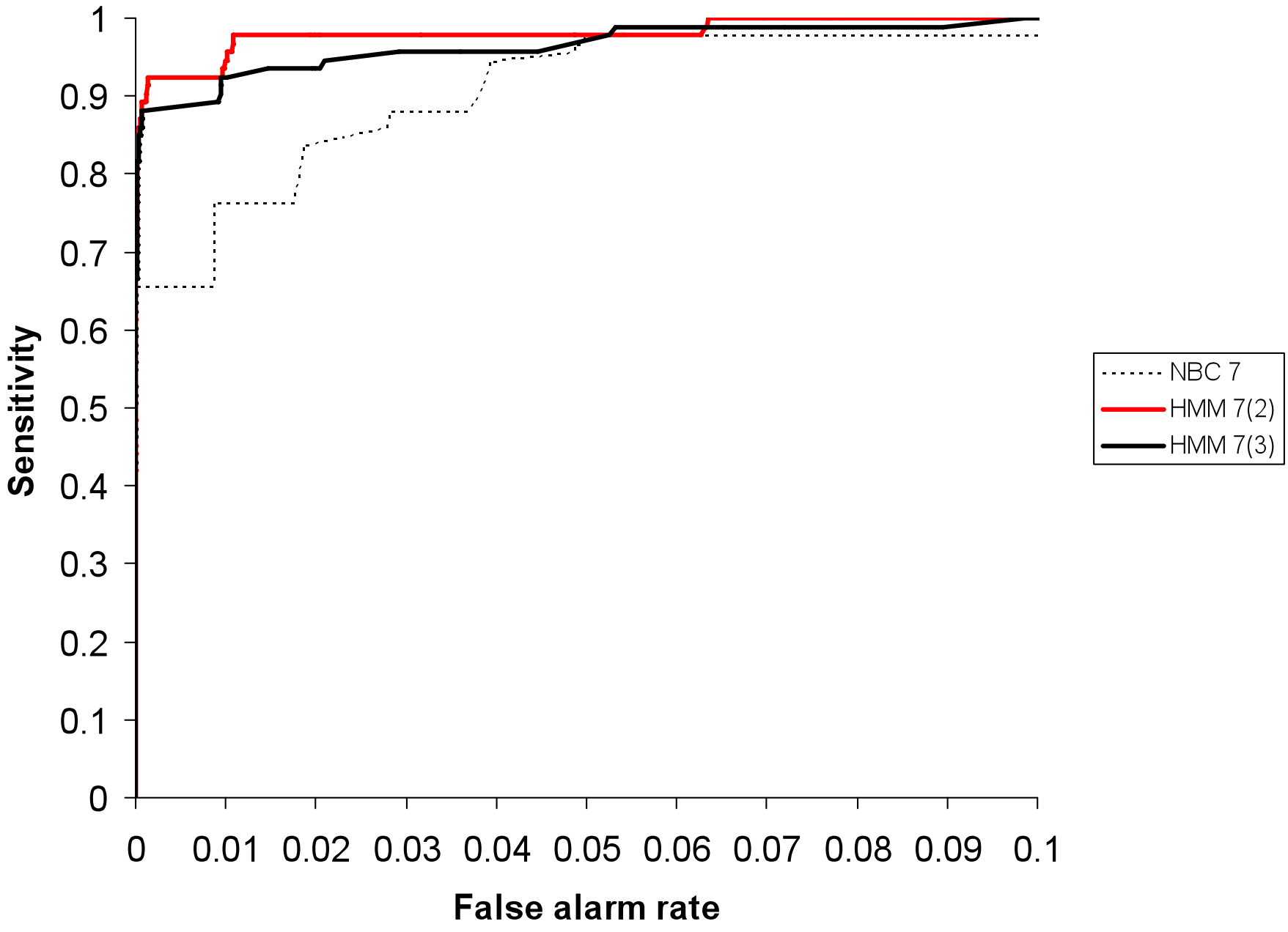
Small less clustered – 14 day



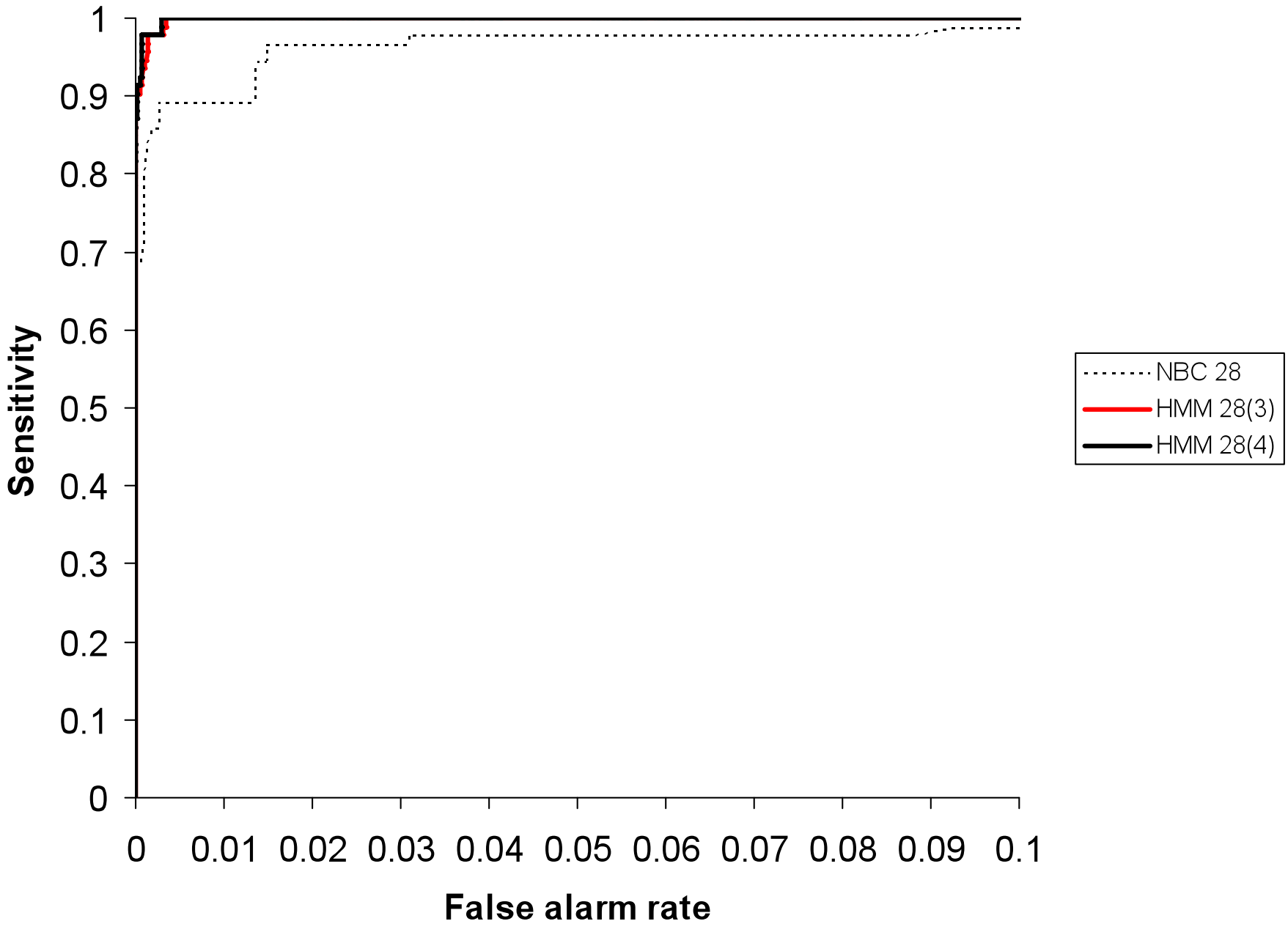
Small less clustered – 28 day



Large more clustered – 7 days



Large more clustered – 28 days



Summary

Higher prior means were generally associated with decreased model sensitivity

- depends on the amount of data analysed

Relative algorithm performance depends on the desired false alarm level

Short-baseline models are unlikely to be the best performing models

More work to be done…

- optimal analysis window length
- distance-based model
- other comparisons

further details: ‘Disease surveillance using a hidden Markov model’ www.biomedcentral.com/1472-6947/9/39

Thank you