Modelling the impact of noise on wastewater-based environmental surveillance

Jeremy Bingham

2024-06-26

Motivation

- Wastewater contains valuable information
- Mixed *practical* success
- Misalignment of reality vs expectations

"Easy"
e.g., presence/absence

Hard
e.g., estimate incidence

Modelling pipeline: steps

Epidemic

Daily incidence time series

Individual shedding profiles

Time series of total RNA shed

Flow, dilution, and degradation

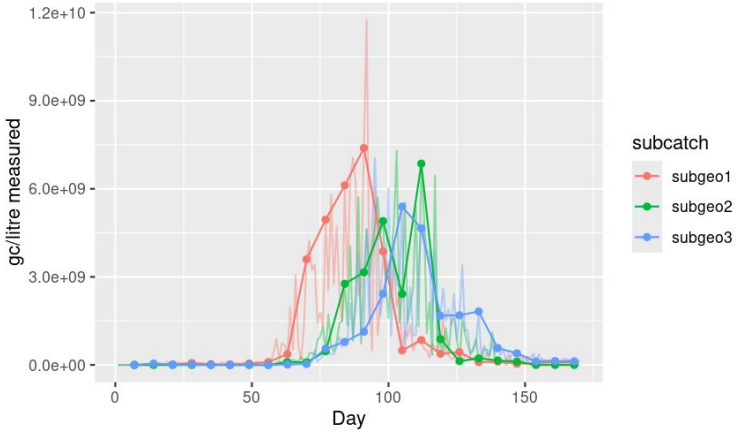
Time series of RNA concentrations

Sample collection and testing

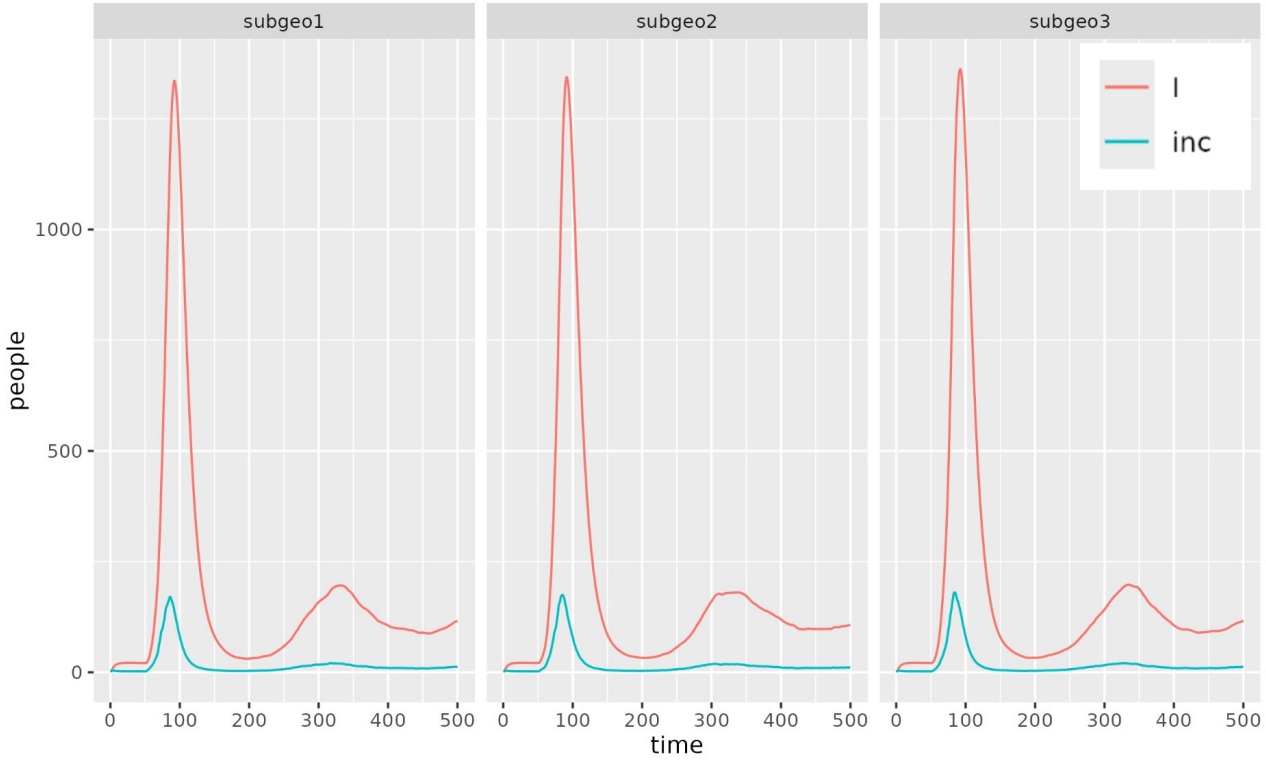
Time series of measured RNA concentrations

Analysis

Answer to public health question – e.g., is incidence increasing?



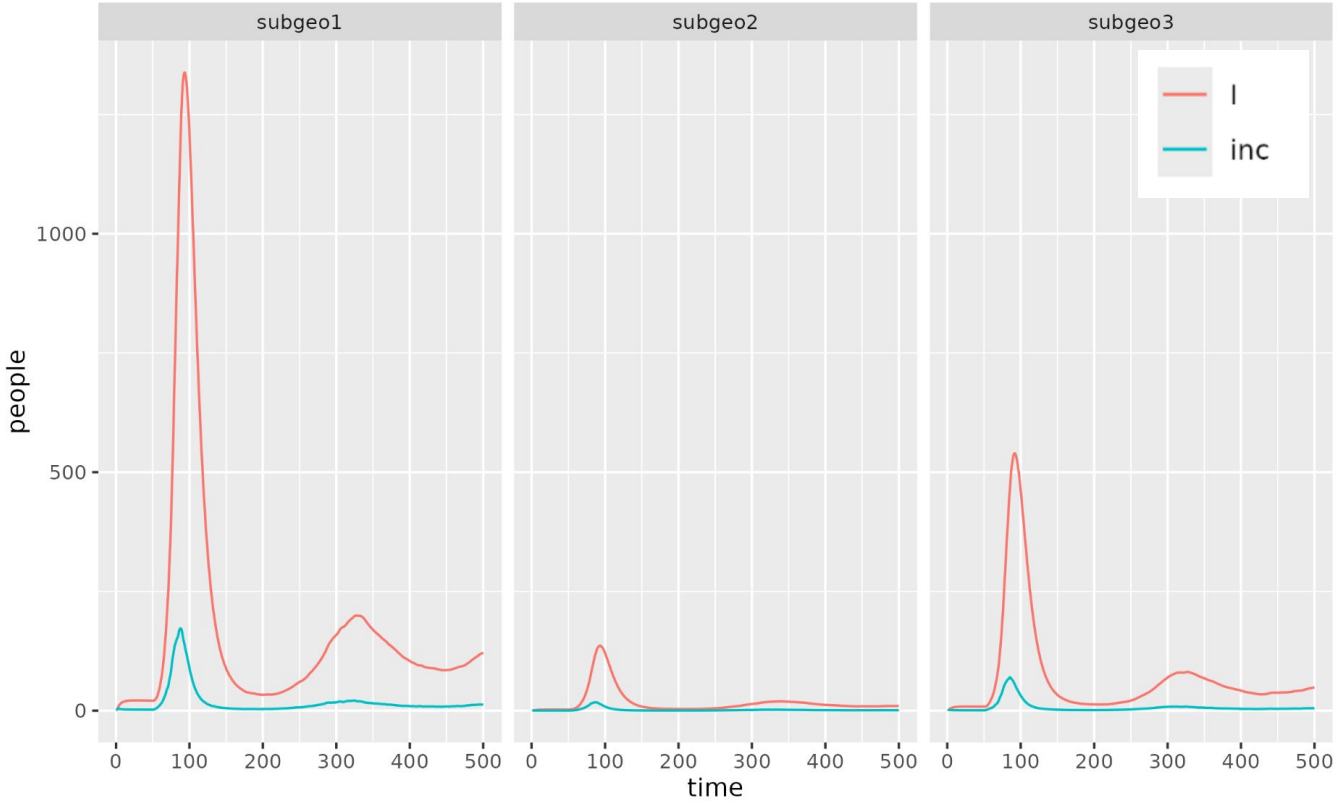
Base case: 3 nearly-identical epidemics



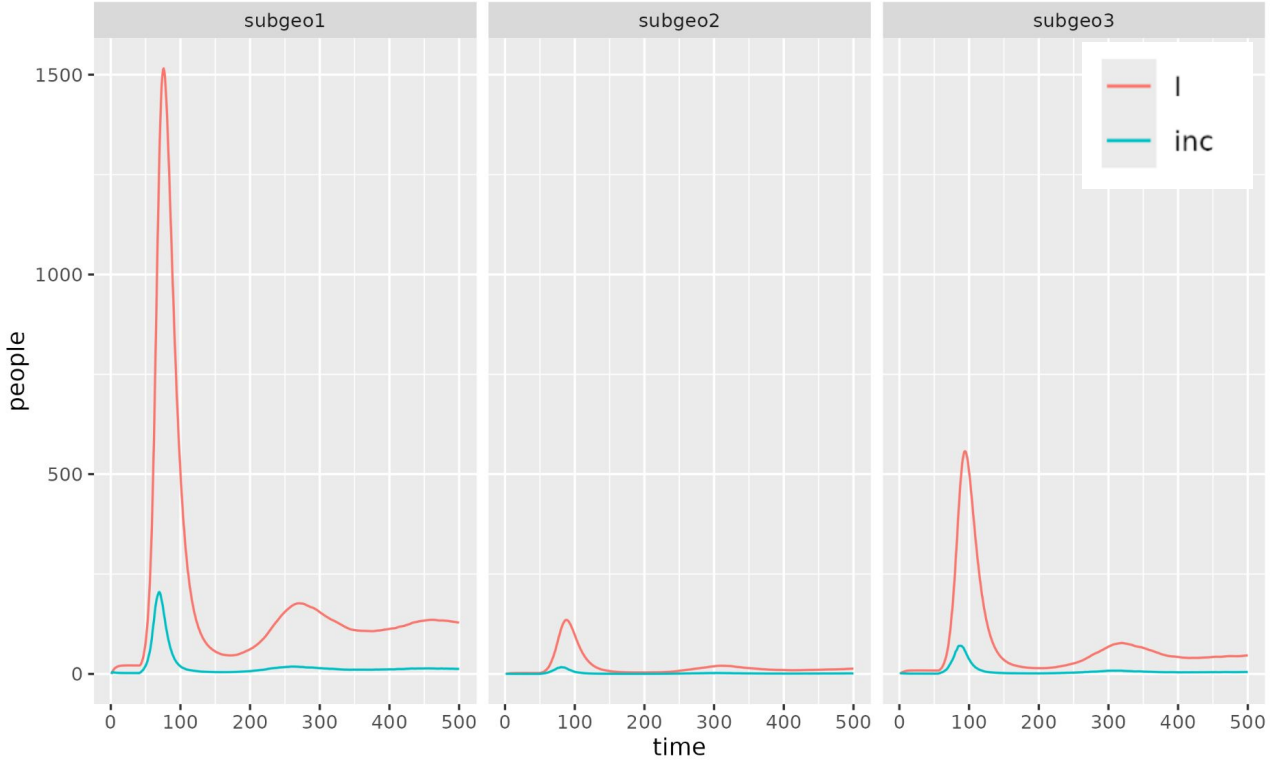
SEIRS model

Initial lull (low transmission) followed by higher transmission

Second case: same epidemic parameters, different population sizes (5000, 500, 2000)



Third case: different population sizes, start times, and transmission rates



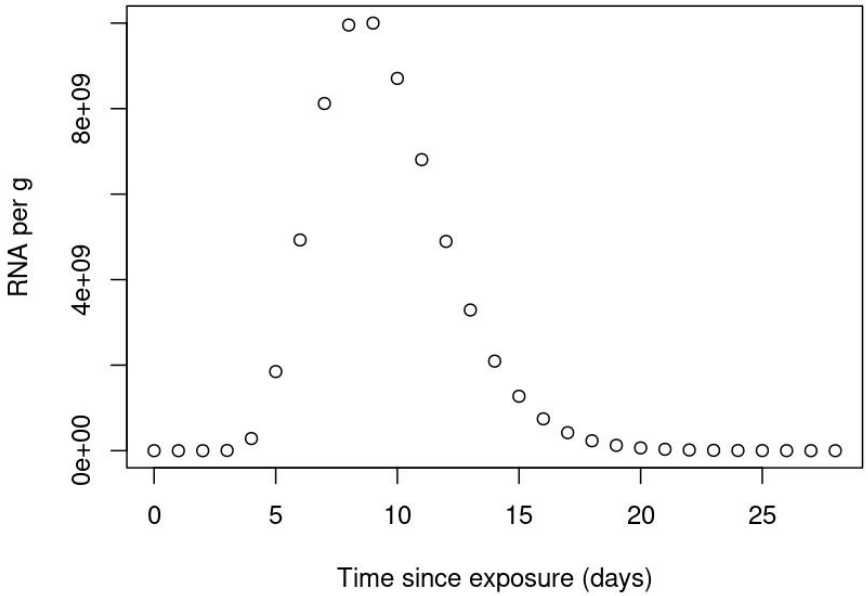
Epidemics "start" 1 & 2 weeks apart

- Large population, high transmission
- Small population, low transmission
- Medium population, moderate transmission

Shedding

Indidual-level shedding

Gamma distributed concentration over time **RNA/g**



Flow/Dilution

White noise around a mean **per-person dilution**

Input: Total RNA shed (daily # gene copies)

Output: gene copies/litre

Sample collection and measurement

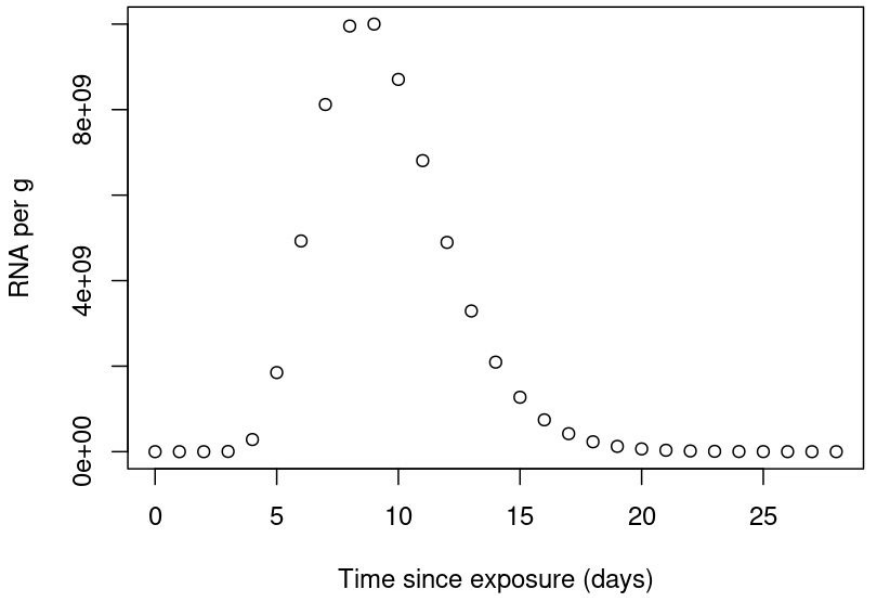
White noise around daily mean

Observation daily/4-days/weekly

Shedding

Indidual-level shedding

Gamma distributed concentration over time **RNA/g**



Flow/Dilution

White noise around a mean **per-person dilution**

Input: Total RNA shed (daily # gene copies)

Output: gene copies/litre

Sample collection and measurement

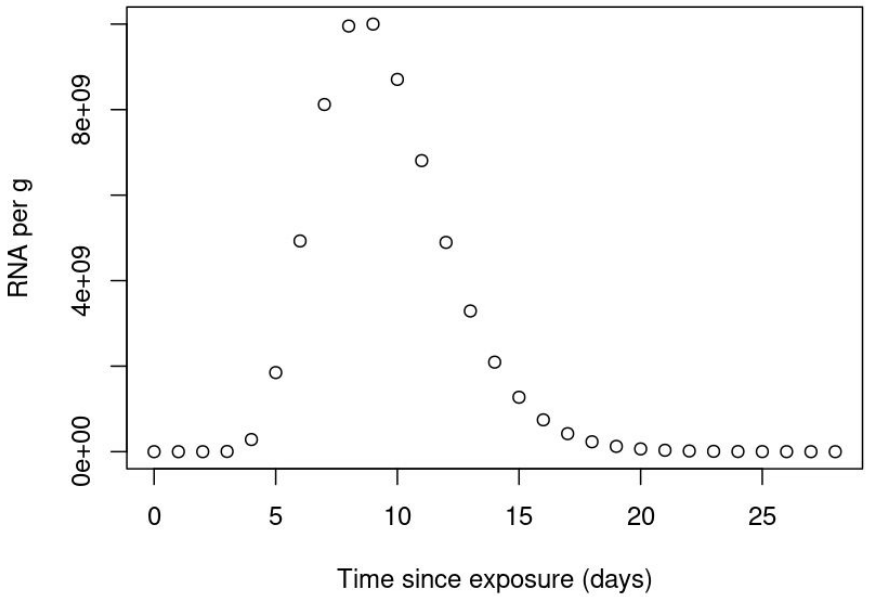
White noise around daily mean

Observation daily/4-days/weekly

Shedding

Indidual-level shedding

Gamma distributed concentration over time **RNA/g**



Flow/Dilution

White noise around a mean **per-person dilution**

Input: Total RNA shed (daily # gene copies)

Output: gene copies/litre

Sample collection and measurement

White noise around daily mean

Observation daily/4-days/weekly

Analysis

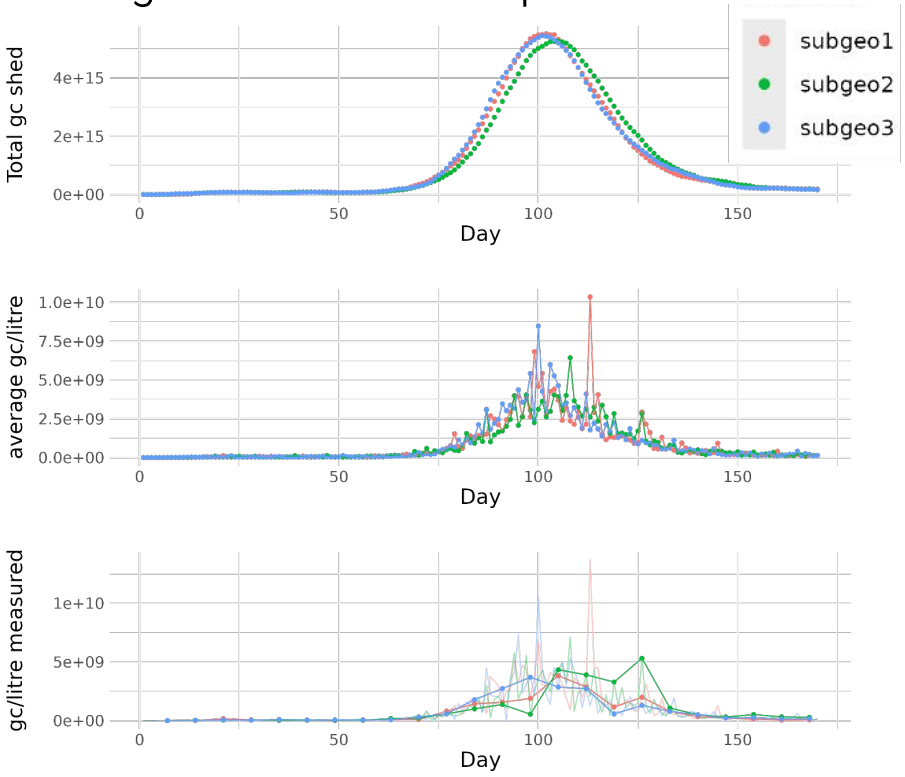
How often do we “detect” a difference between two catchment populations using a paired t-test,

- when epidemic dynamics are identical?
- when epidemic dynamics are different?

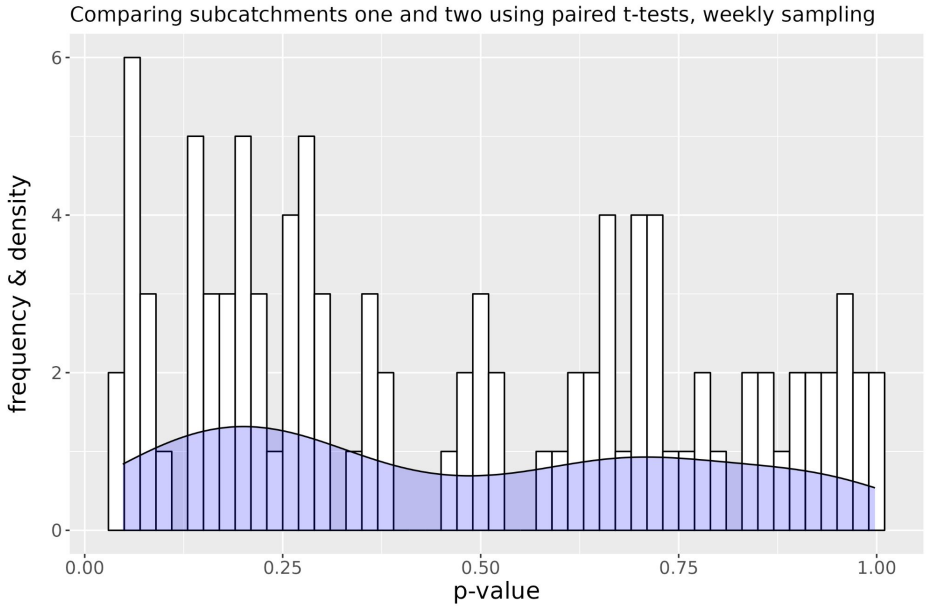
Are the different sub-catchments likely to be reliably differentiable for answering public-health-relevant questions?

Example: three identical epidemics

“High” noise in all steps

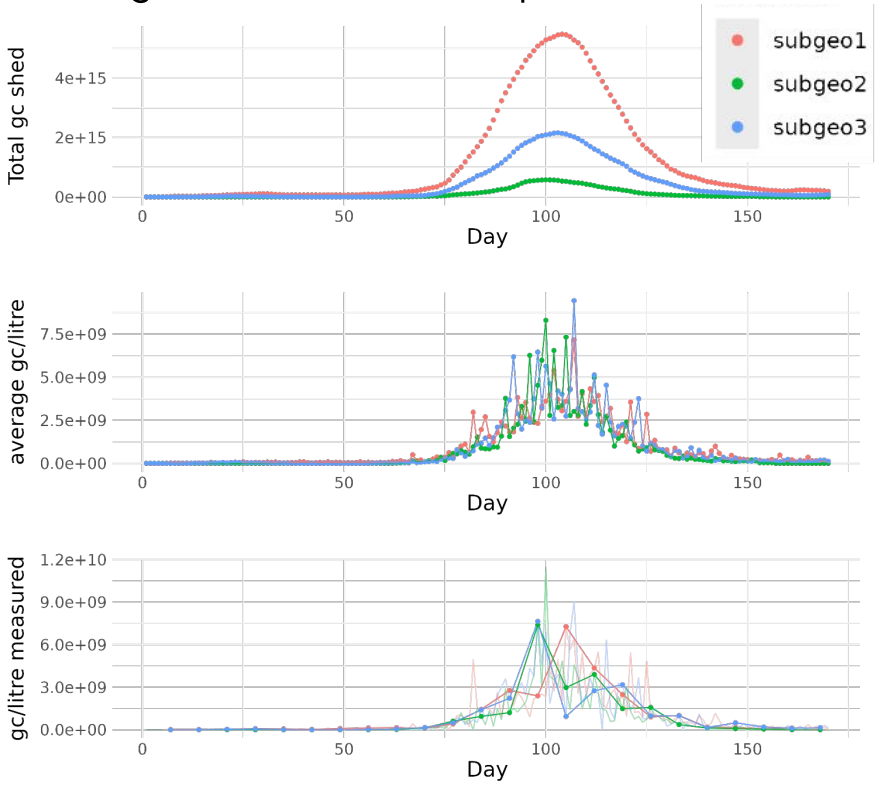


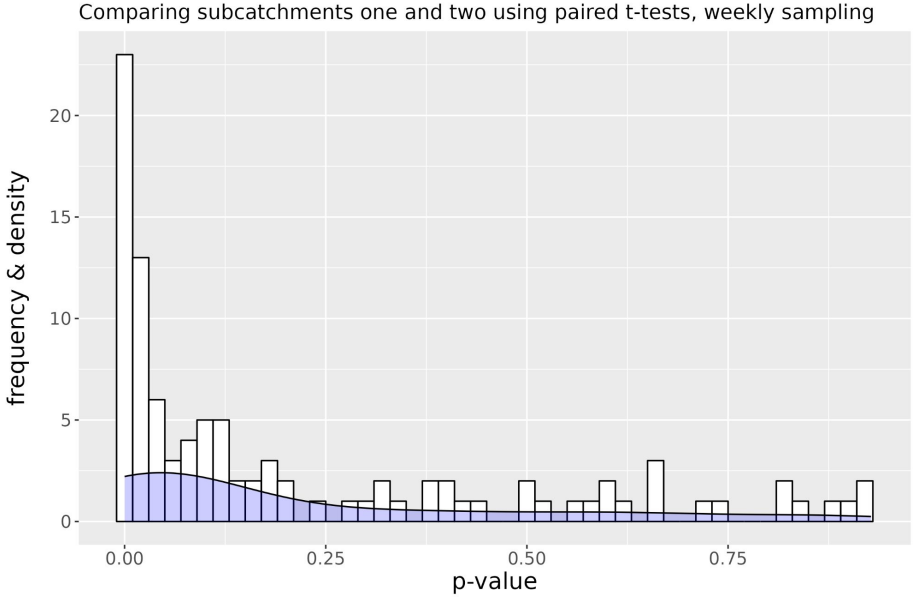
p-values roughly evenly spread
(and approx 5% under 5%)



Example: three similar epidemics (Δ pop sizes)

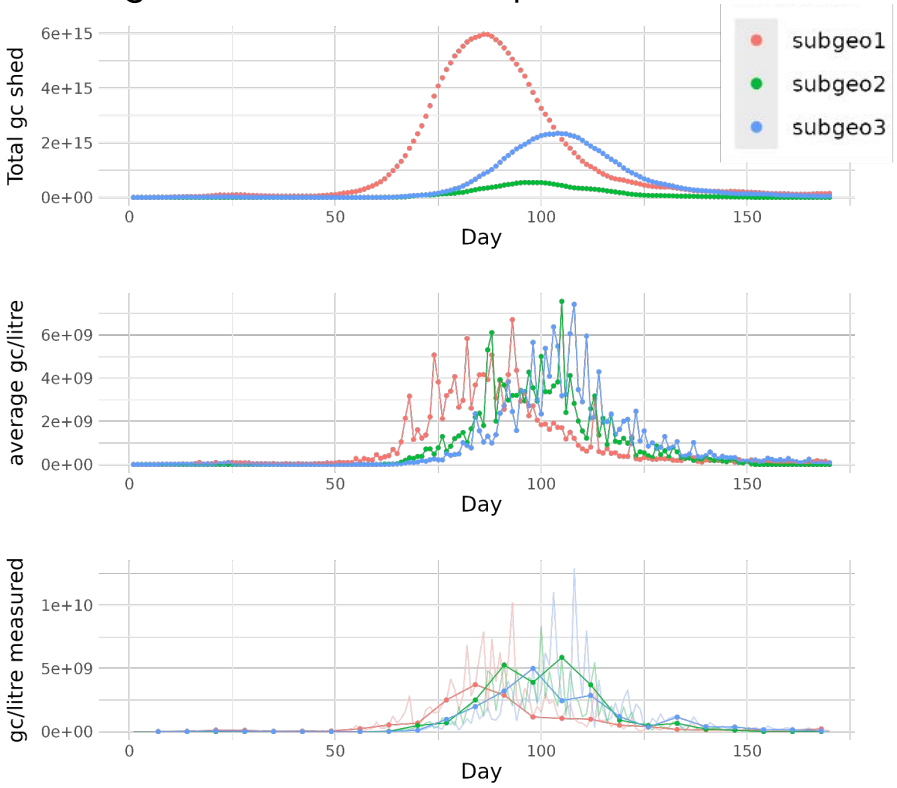
“High” noise in all steps



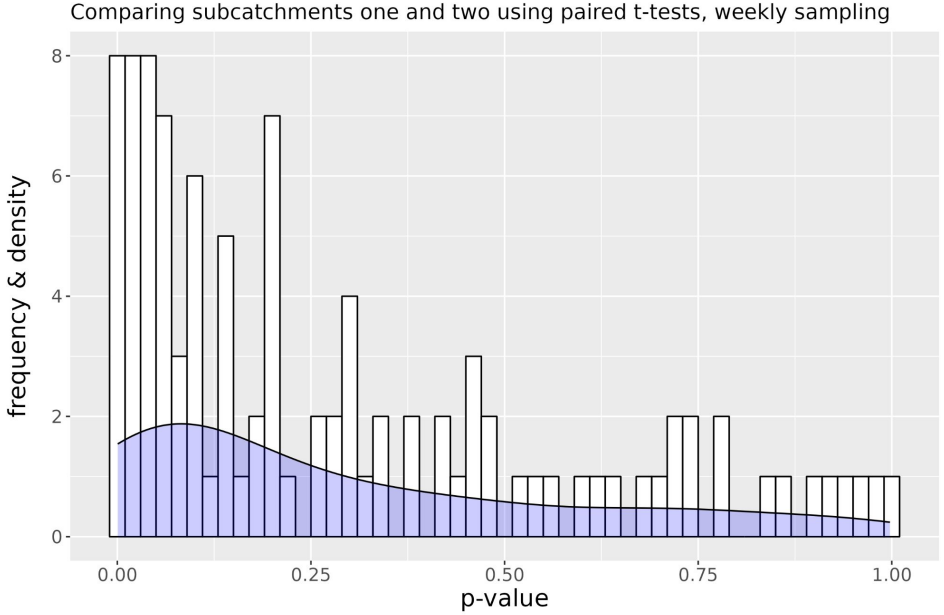


Example: three different epidemics

“High” noise in all steps

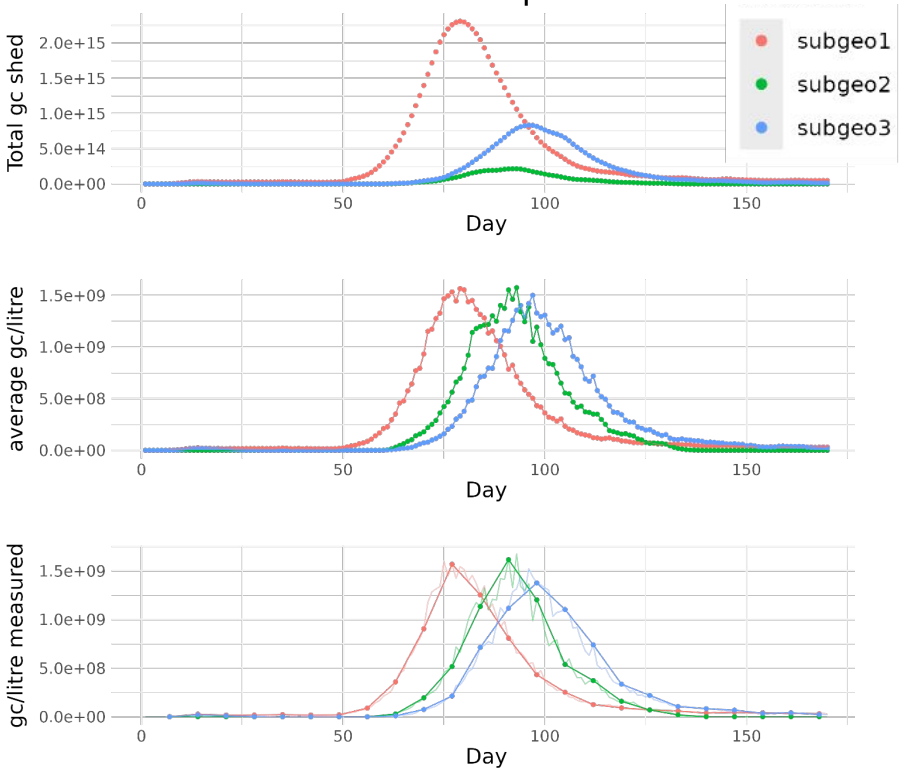


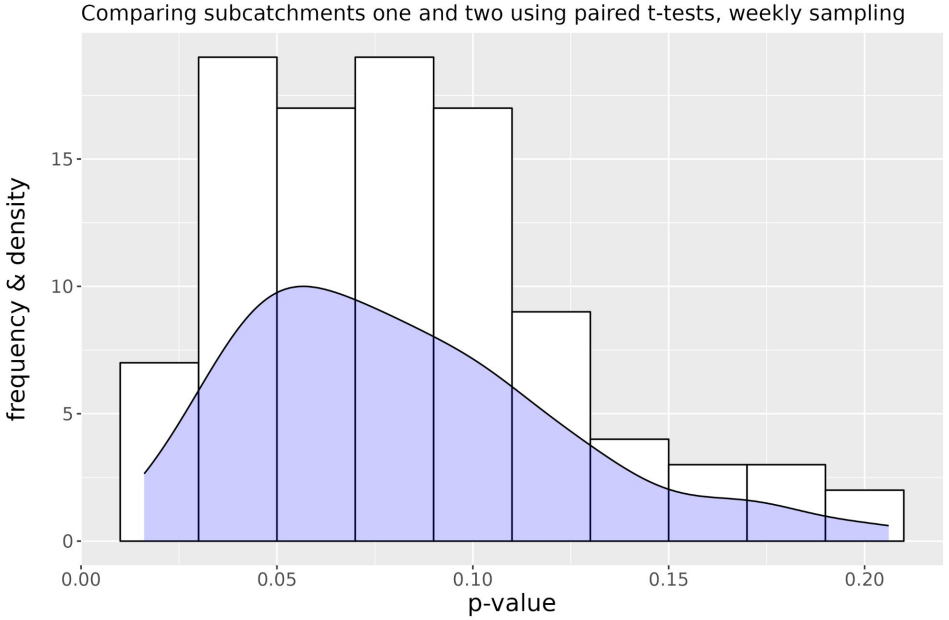
Too much noise to reliably detect differences, even over long timespan



Example: three different epidemics

“Low” noise in all steps

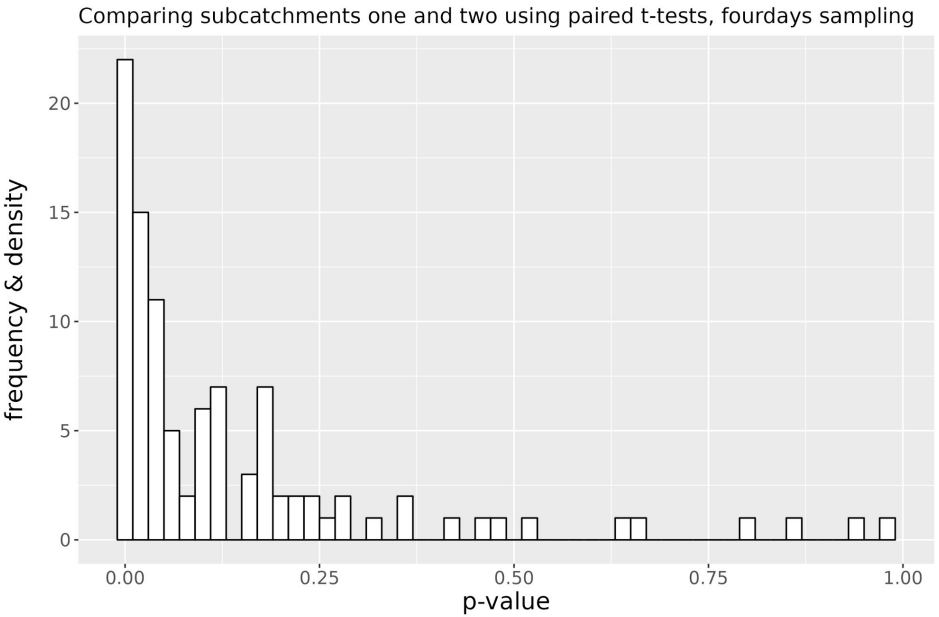


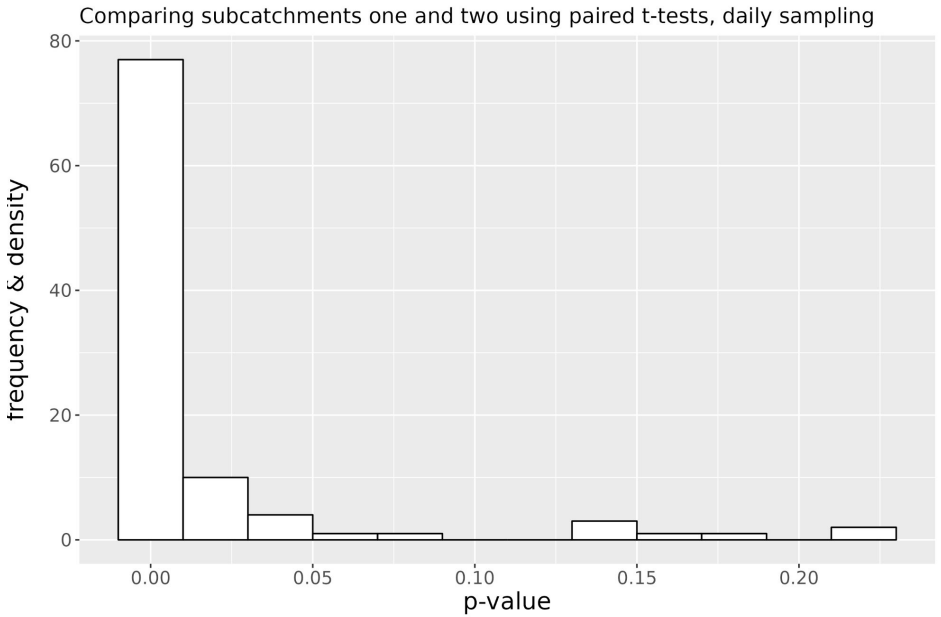


What if we can't reduce the noise?

Increase sampling frequency from daily to every four days:

With daily sampling we can reliably tell catchments apart!





Key takeaways

- Characterise noisiness of system as early as possible
- Set clear expectations
- Design surveillance for specific questions

Next steps

- Incorporate data on South African systems and catchment populations
- Test practical public health question
  - e.g. “is a new wave starting”
- Leverage nested catchments to filter outliers

Jeremy Bingham: jeremyb@sun.ac.za

Special thanks to

- Kathleen O'Reilly (LSHTM)
- Cari van Schalkwyk (SACEMA)
- Kerrigan McCarthy (NICD)
- Gauteng City Regional Observatory
- Bill and Melinda Gates Foundation

and thank *you* for listening!

Shedding

Indidual-level shedding:

Gamma distributed concentration over time **RNA/g**

Start shedding at 3 days

Shedding for 25 days

Noise via:

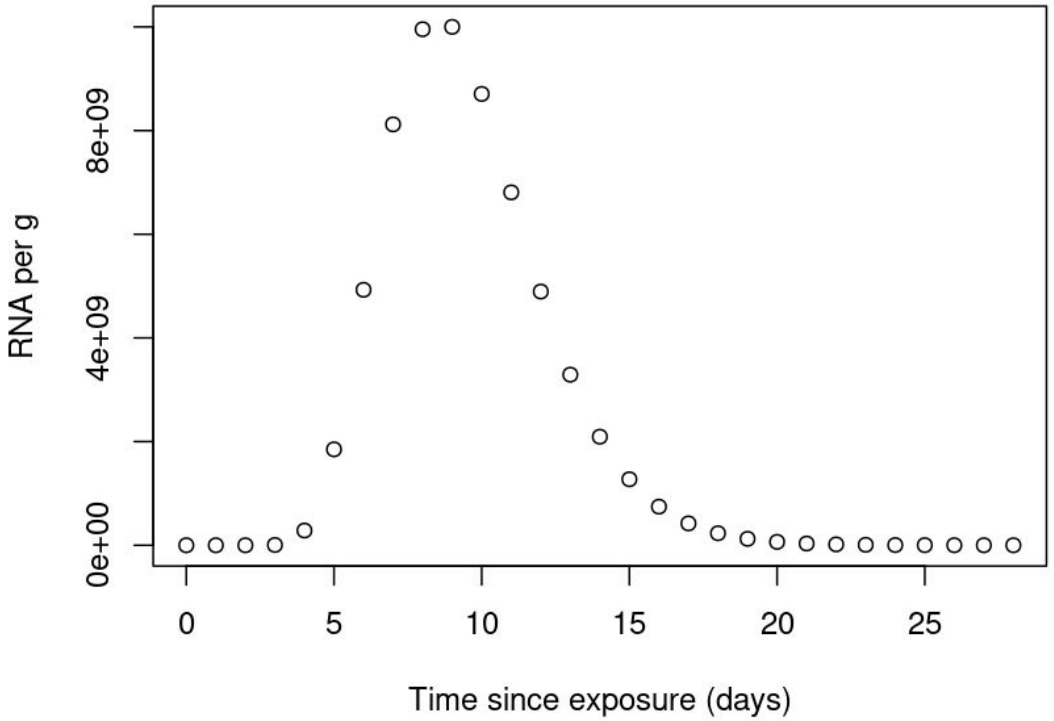
Peak RNA/g (mean 10^10)

Time to peak shedding (10 days)

“Scale” parameter of gamma distribution (mean 1)

Assume daily **200g** per person

Example of individual shedding profile



Peak RNA/g = 10^10; peak at 10 days; Start at 3 days; duration of 25 days

Flow/Dilution

White noise around a mean **per-person dilution**

Input: Total RNA shed (daily # gene copies)

Output: gene copies/litre

(**300l per person** per day, slightly above national average of 237l per person per day)

Still open question: how best to measure dilution and characterise the noisiness of dilution estimates?

“Low” noise: Standard deviation = 5% of mean

“Medium” noise: Standard deviation = 10% of mean

“High” noise: Standard deviation = 30% of mean

Sample collection and measurement

Collection and laboratory procedure currently modelled in a single step:

Input: Daily average concentration (gene copies/litre)

Output: Daily/four-days/weekly measured concentration (gene copies/litre)

White noise around **daily mean**

Parameterised via standard deviation as proportion of daily mean

Low noise: 5% of mean

Medium noise: 10% of mean

High noise: 40% of mean