The Clue is in the Poo…

utilising wastewater data for infectious diseases

IDD conf – 6th Sept 2023

Kath M O'Reilly
Associate Professor
LSHTM

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Overview

- What happens?
- Examples of using environmental surveillance (ES) data* for infectious diseases
- Questions to ask of ES data…
- Example of COVID-19
- From research to routine use

** Also called, sewage data, wastewater data (depends on context and research specialism)*

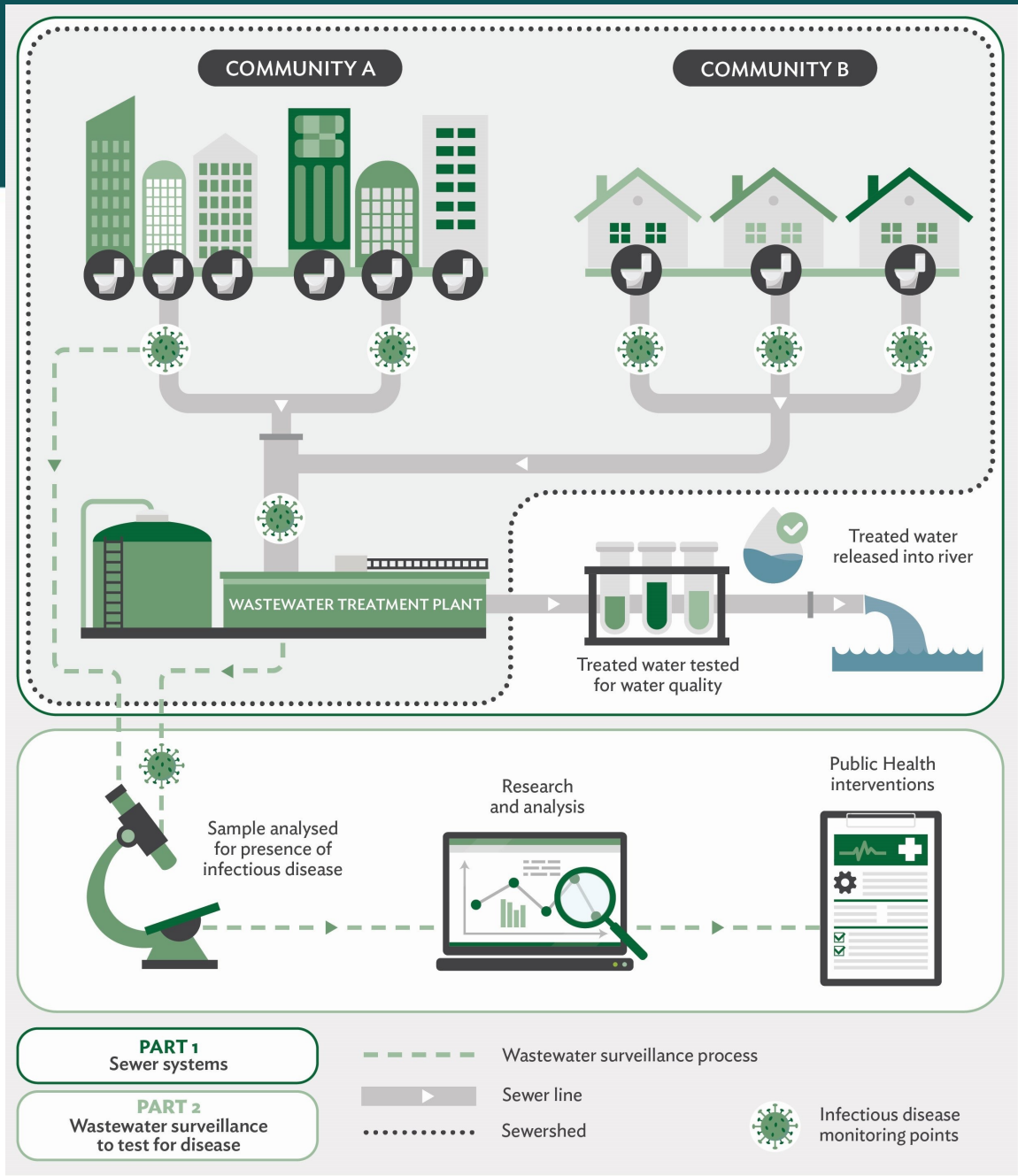
How?

- Many pathogens are shed in feces and urine
  - Part of the transmission cycle (fecal-oral)
  - And RNA/DNA, body removing waste
- Polio eradication
  - Shed live virus (fecal-oral transmission)
  - Virus culture & RNA detection
  - Detection in London sewage in 2022*
- Typhoid
  - Water & food contamination
- Others
  - COVID-19 (SARS-CoV-2), 'flu, norovirus, measles, cholera, mpox, …

* https://tinyurl.com/wbepolio

What happens?

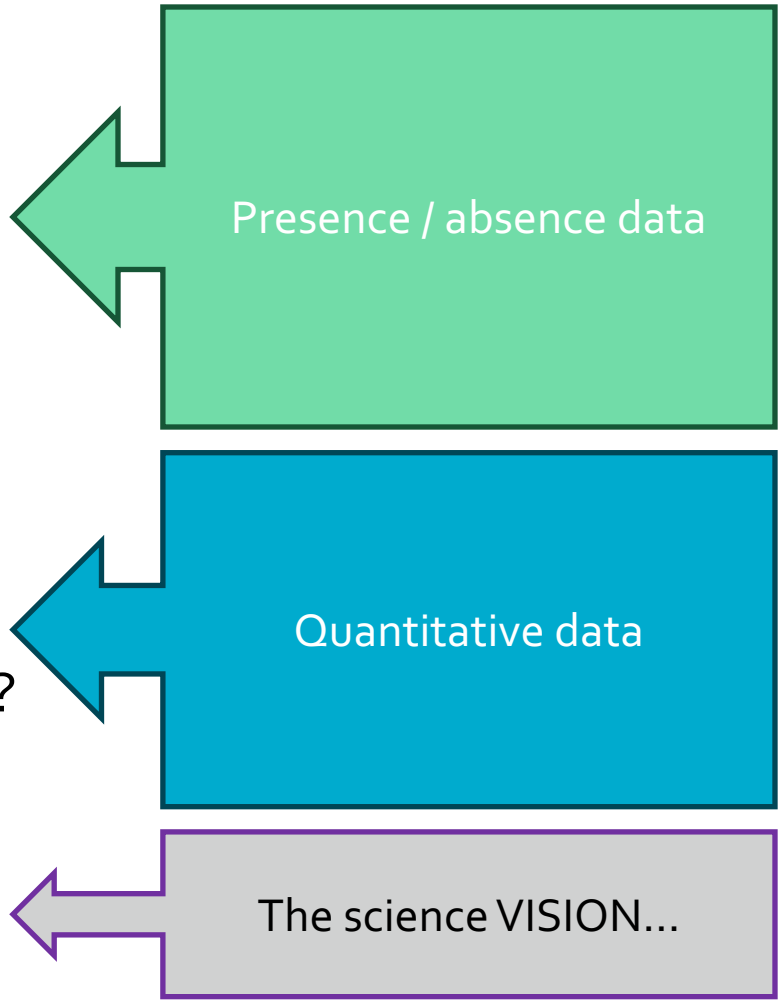
- Select a site for regular sampling
- 'Grab' a sample, or use composite sampler
    - Maybe record site characteristics (eg. flow)
- Take sample to the lab
- Carry out lab analysis:
    1) Record 'meta-data' (eg. ammonia)
    2) Concentrate and/or extract the sample
    3) Identify what you are looking for
        - PCR based analysis (primers…)
        - Virus isolation (cell culture)
        - WGS (eg. nanopore / illumina sequencing)
        - Meta-genomics
- Make sense of the data collected



More details: http://tinyurl.com/wbegcro

Questions to ask of ES data

1. Is pathogen A present?
    - Single or multiple sites
    - For a defined duration
2. Is pathogen A absent?
    - Single/multiple sites, and over a defined duration
3. How has the prevalence of pathogen A changed in time? (1 site)
4. How has incidence of pathogen A changed in time? (1 site)
5. How does prevalence/incidence compare between locations? (multiple sites)
6. What is the incidence of infection (or disease) with pathogen A, estimated by combining clinical and ES data together



Change in pathogen prevalence

Using a simple model,

Morvan et al (2022)

- Prevalence estimation
- Sensitivity analysis

Concentration (GC/L)

Quantity wastewater (L/day/person)

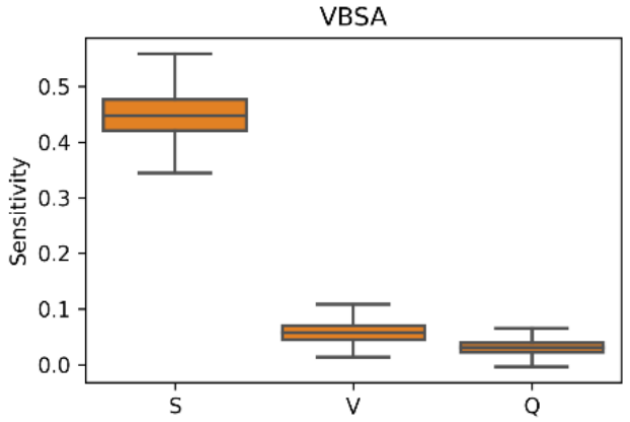
$$P = \frac{C \times Q_p}{S \times V}$$

Shedding rate (GC/ml)

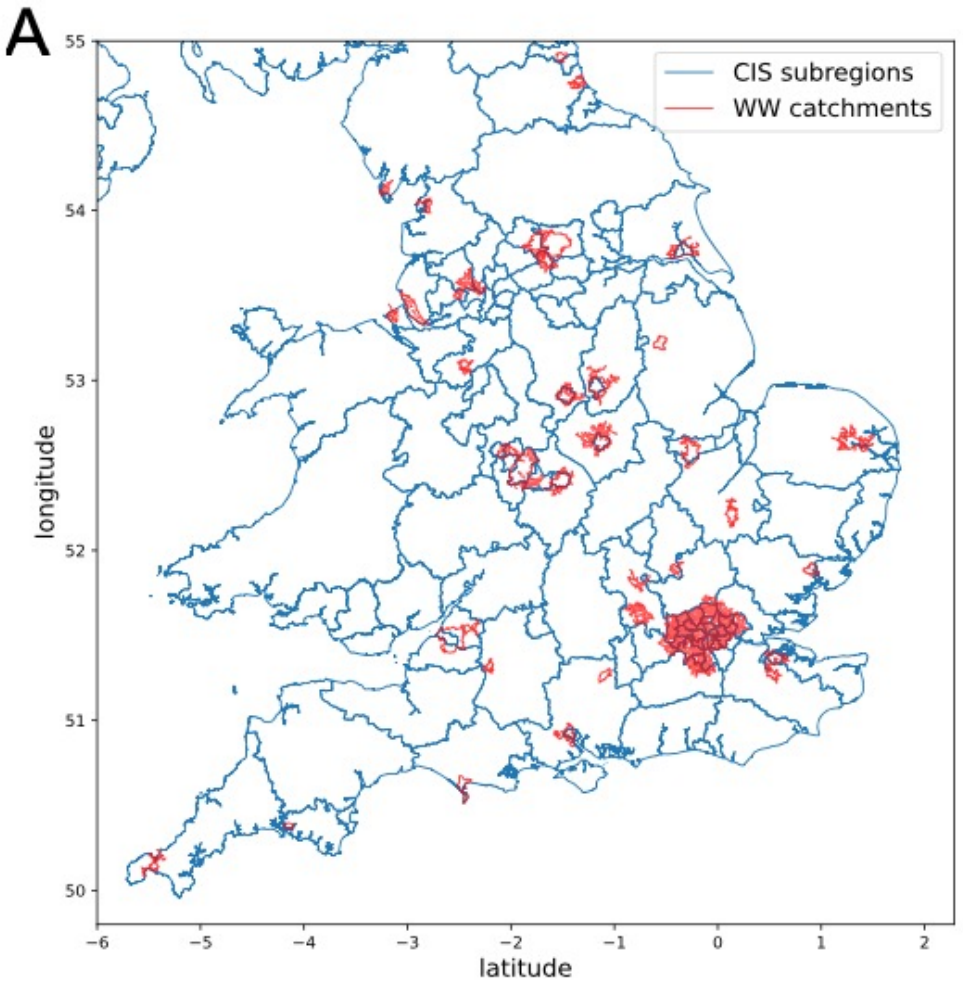
Volume of stool (ml/day)

For accurate estimates of prevalence, need to know;

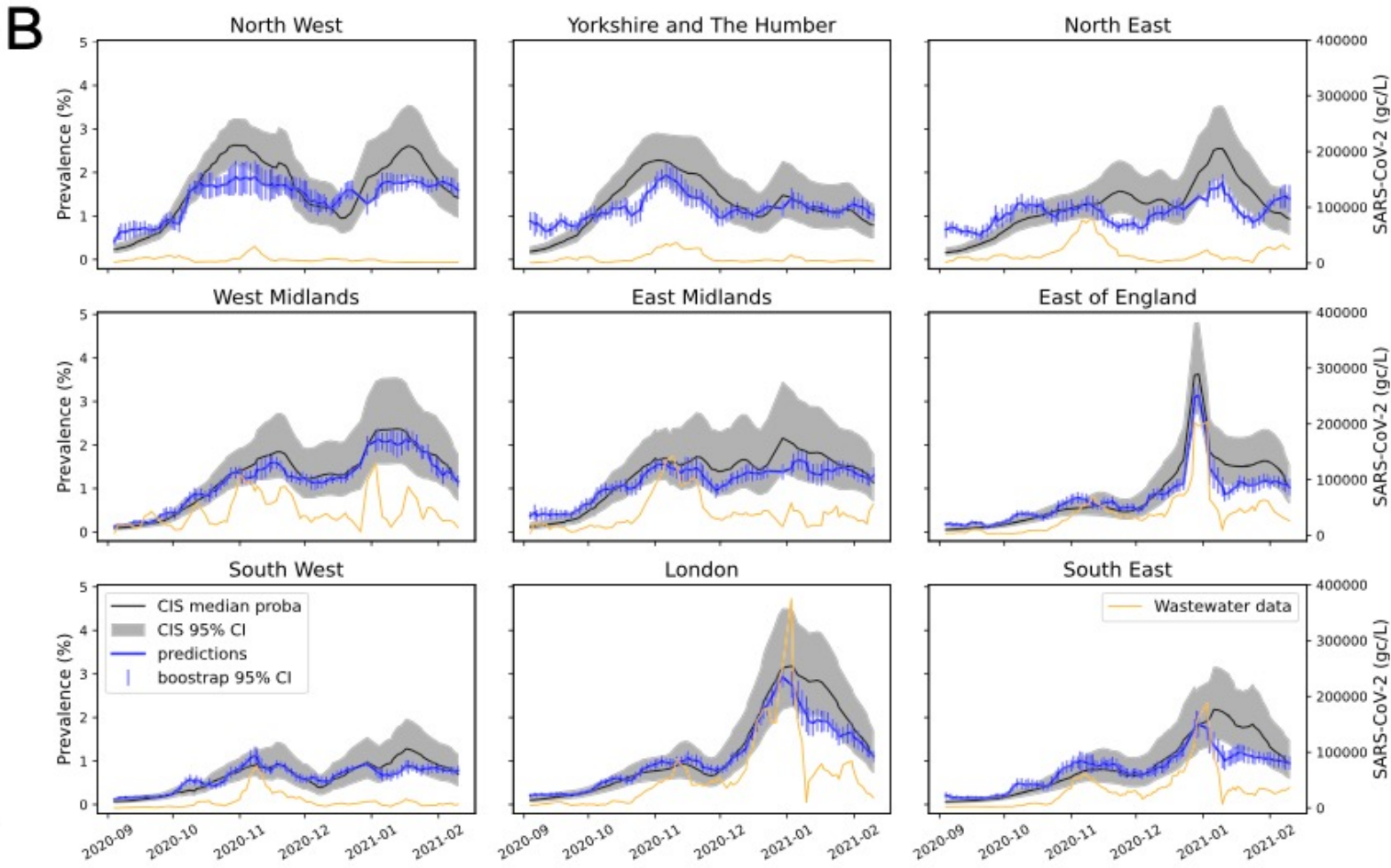
- Concentration in sample
- Wastewater flow (if it changes a lot)
- Shedding in stool (mean and var) *
- Also be sure that virus shedding = being 'infected'



Validation of Prevalence Estimates (England)



Locations of WW sites and comparisons to CIS regions

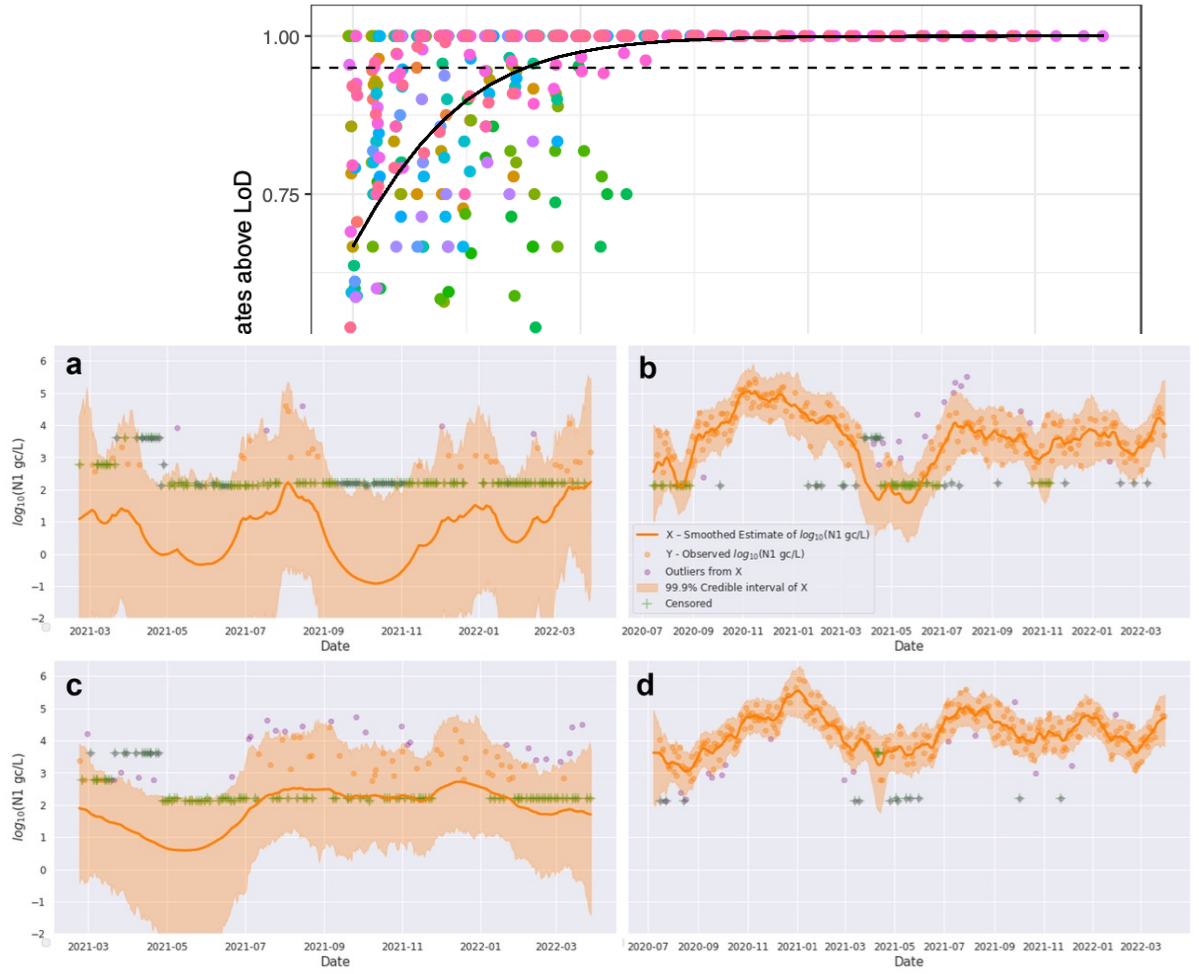


Blue: Estimates of prevalence from WW data & model
Black & grey: CIS prevalence estimates
Yellow: 'raw' estimate of WW data

Morvan et al (2022)
https://doi.org/10.1038/s41467-022-31753-y

Messy Data…

- SARS-CoV-2 prevalence (via CIS) was often high
  - Proportion of ES samples >LOD also high
  - Some sites, and some periods had many below LOD
- Explored effect of 'left censoring'
  - **Lewis-Borrell et al (2023)** 10.3934/math.2023859
  - Large effect in some sites
  - Data augmentation improved correspondence to CIS



Green: raw data (with LOD)
Orange: Inferred gc/l measure

Change in Pathogen Incidence

A measure of prevalence can be used to infer incidence,

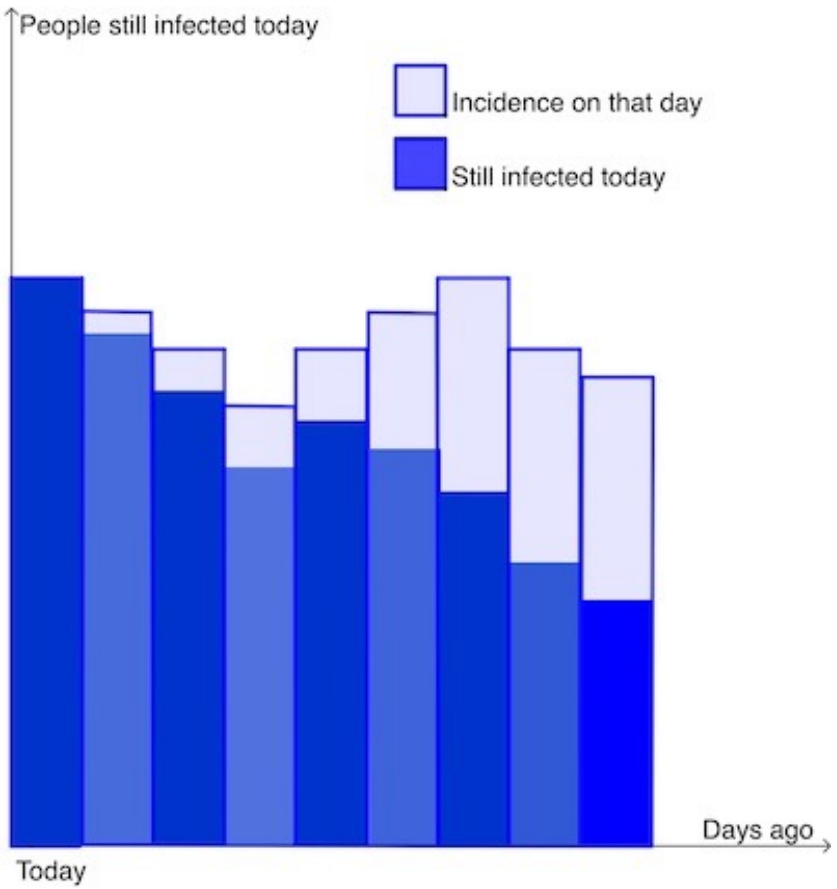
$$Prev(t) = \int_{T=0}^{\infty} Inc(t-T)Dur(T)dT$$

But we also want to account for,

- Changing shedding profile during course of infection

$$\text{Shedding}(t) = \int_{T=0}^{\infty} Inc(t-T)Shed(T)dT$$

- Deconvolution and then *usual methods to estimate Re*



https://plus.maths.org/content/keeping-covid-19
(handy!)

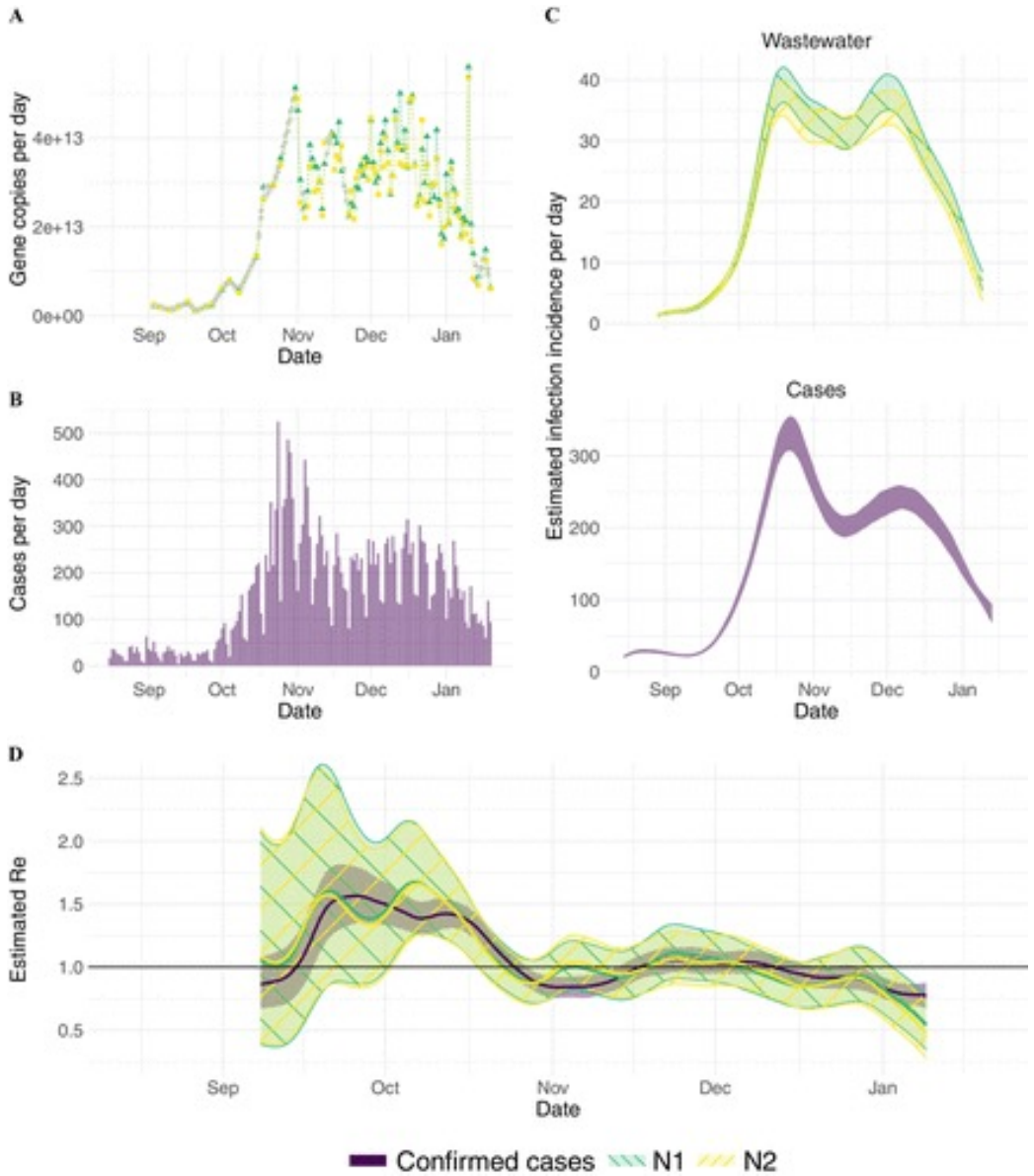
Validation of Incidence and Re Estimates (Switzerland)

Described in full in **Huisman et al (2022)** https://doi.org/10.1289/EHP10050

- Infection incidence inferred from WW and case data broadly agreed
- Ribbons of Re estimates had good overlap
    - More uncertainty from WW data

In addition to WW data (measure, flow), require shedding profile of the pathogen

- Some flexibility, depending on accuracy of output



Back to the Questions

	Pathogen measure	LOD effects	Shedding - average	Shedding - profile	Water flow
3. How has the prevalence of pathogen A changed in time?	X	X	X	-	X
4. How has the prevalence of pathogen A changed in time?	X	X	X	X	X
5. How does prevalence compare between locations?	X	X	X	X	XX

But also need to consider further questions:

- How representative is the ES sample of the population I'm interested in?
- What is the sensitivity / recovery of pathogen?
- Are inhibitors present in the sample?
- Has the sample been affected by degradation?

So ES Data has Potential…

- Data collection that can be *designed* to answer specific questions
  - This hasn't happened much yet, but will happen soon
- Sample collection is cheap!
  - *~£500/sample for lab costs*
  - *Well, cheaper than CIS…*
- Much ES data are publicly available (eg. dashboards)
  - Few issues of identifiability
- For the data and analysis to be useful, clear uses / interventions need to be actionable from ES data
  - Are the data reliable enough?
  - Sampling the 'general population' means limited knowledge about vulnerable groups
  - What actions/interventions are appropriate?

From Research to Routine Use

- Increased interest in using ES since COVID-19
  - A very cross-disciplinary field
  - Evolving technologies
  - Evolving techniques (inc. analysis)
- Research needs for routine use **Shaw et al (2023)** doi/10.1038/s41591-023-02457-7:
  - Translation from HIC to LMICs – non-sewered settings
  - Designing a sampling scheme *
  - Minimal criteria for reporting
  - Validation for new pathogens *
  - Integration with clinical data *
  - Compare costs of data collection to benefits of interventions *
  - Best practice for communication

* More modelling & analysis needed!

Acknowledgements

People:

- Leon Danon
- Matt Wade
- Mario Morvan
- Anna Lo Jacamo
- Koen Pouwels
- Chris Lilley
- Andrew Singer
- David Allen
- David Walker
- Alex Shaw
- Jillian Gauld
- Supriya Kumar
- Kerrigan Macarthy
- Juliet Pulliam
- Jeremy Bingham
- Zi Mithombothi
- Gillian Maree

And many more…

Funding:

- UKHSA
- Alan Turing Institute
- BMGF

Some of the workshop attendees in March 2023 - **Wastewater Analytics** - Funded by Alan Turing, continued support from Juniper Consortium

Slides will appear on Github.com/kath-o-oreilly/presentations