National Institute for Public Health
and the Environment
Ministry of Health, Welfare and Sport

Modelling SARS-CoV-2 in wastewater: Inference of virus loads and relation to hospital admissions

Michiel van Boven

June 2023

Introduction

- Surveillance for SARS-CoV-2 circulation is largely based on testing data, hospital and intensive care admissions, tailored studies (e.g., serological and household studies) and analyses of wastewater data

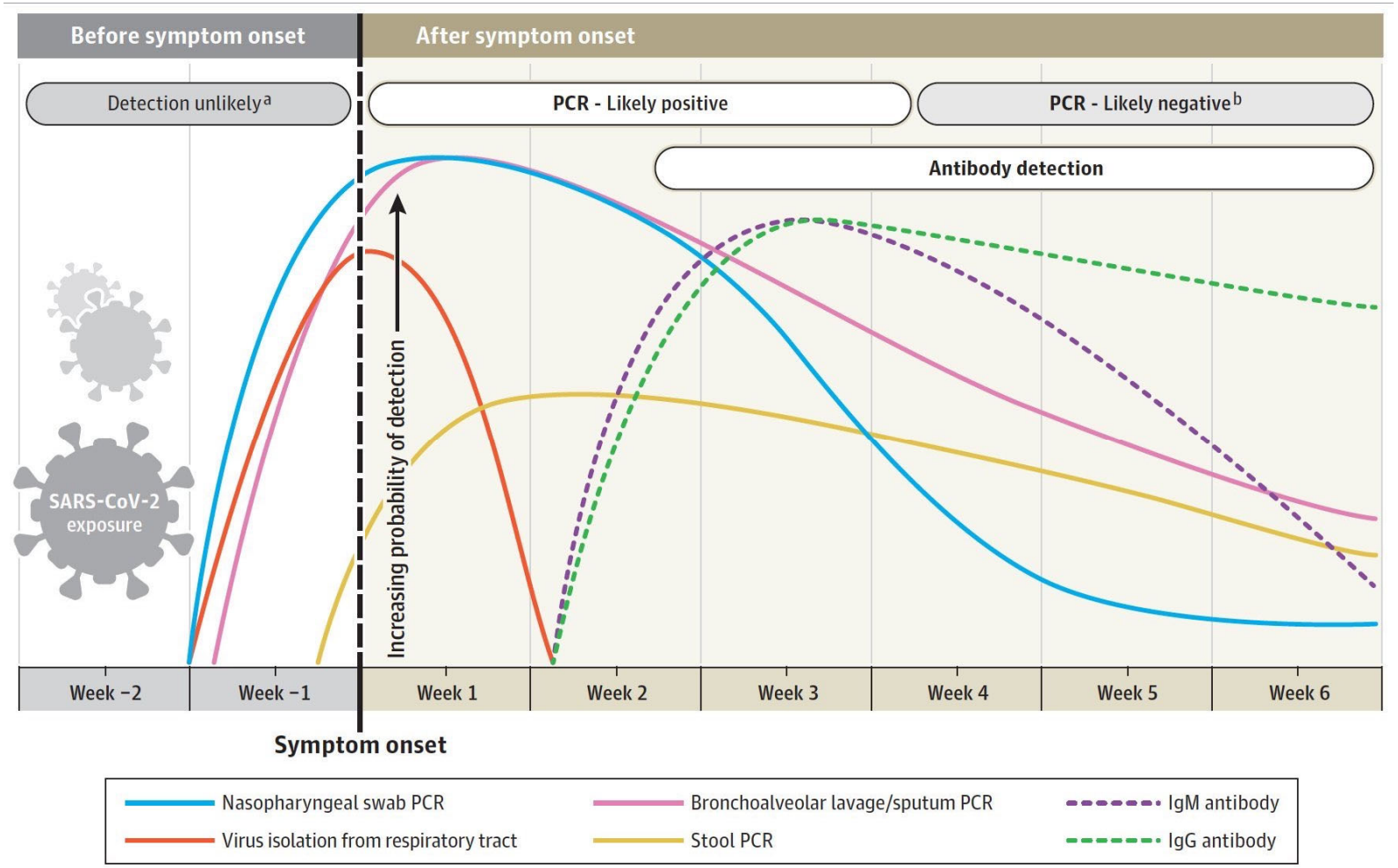
- Wastewater data is noisy and cannot readily be translated to numbers of infections. It does provide unbiased near real-time data with high temporal and spatial resolution

- Provide integrative description of the wastewater data that can be used for assessment of trends at various resolutions, and steps to relate wastewater data to hospital admission data

A quiz

- How many sewage treatment plants (STPs) are there in the Netherlands?
  313 and decreasing
- Which fraction of the Dutch population is connected to the sewerage?
  >99%
- How many samples have been taken since early 2020?
  >85,000
- Does feces of households connected to the sewerage always end up at a STP?
  No. Overflows by heavy rainfall (>10 mm/h) and diapers (babies and elderly)

SARS-CoV-2 in feces

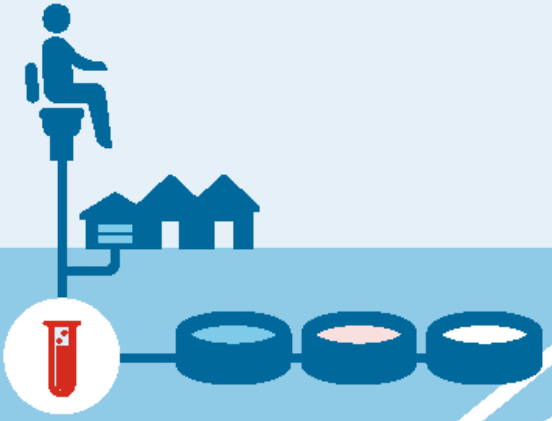


Saturaman 2020

Sewage as an indicator of public health

National Institute for Public Health and the Environment
Ministry of Health, Welfare and Sport

Sewage research is now being used to detect the coronavirus (SARS-CoV-2). This type of monitoring has been used for some time to detect poliovirus and antibiotic-resistant bacteria. RIVM is investigating options for expanding sewage research to include other data that could provide information about public health.



A healthy lifestyle
Oxidative stress*

Personal hygiene and care products
Microplastics, parabens*, UV filters*

Obesity
Biomarkers correlated to BMI

Nutrition
Mycotoxins, vitamins*, isoflavones*

Medication use
Regimen compliance*, insight into over-the-counter medicines*

Non-communicable diseases
Diabetes*, cancer*, allergies*

Use of muscle-building supplements
Anabolic steroids, sequestering agents*, weight-loss supplements*

Drug use
Heroin, cocaine, new party drugs

Forensic research
Drug production

Smoking and drinking
Nicotine, alcohol consumption

Dangerous substances
PFAS, substances of very high concern (SVHC / ZZS), pesticides, flame retardants*, new substances*

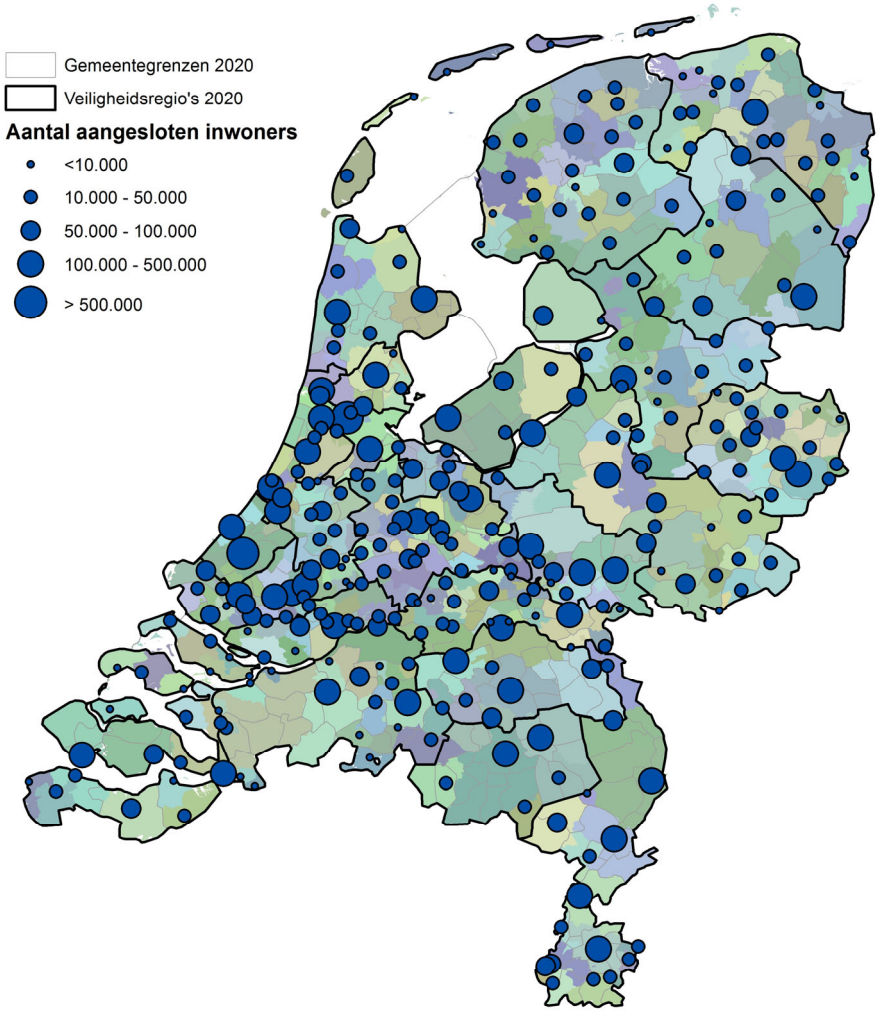
Pathogens
SARS-CoV-2, poliovirus, norovirus, antibiotic-resistant bacteria

*These parameters are not yet used in sewage research. Further research is needed to determine whether this is possible.

Data

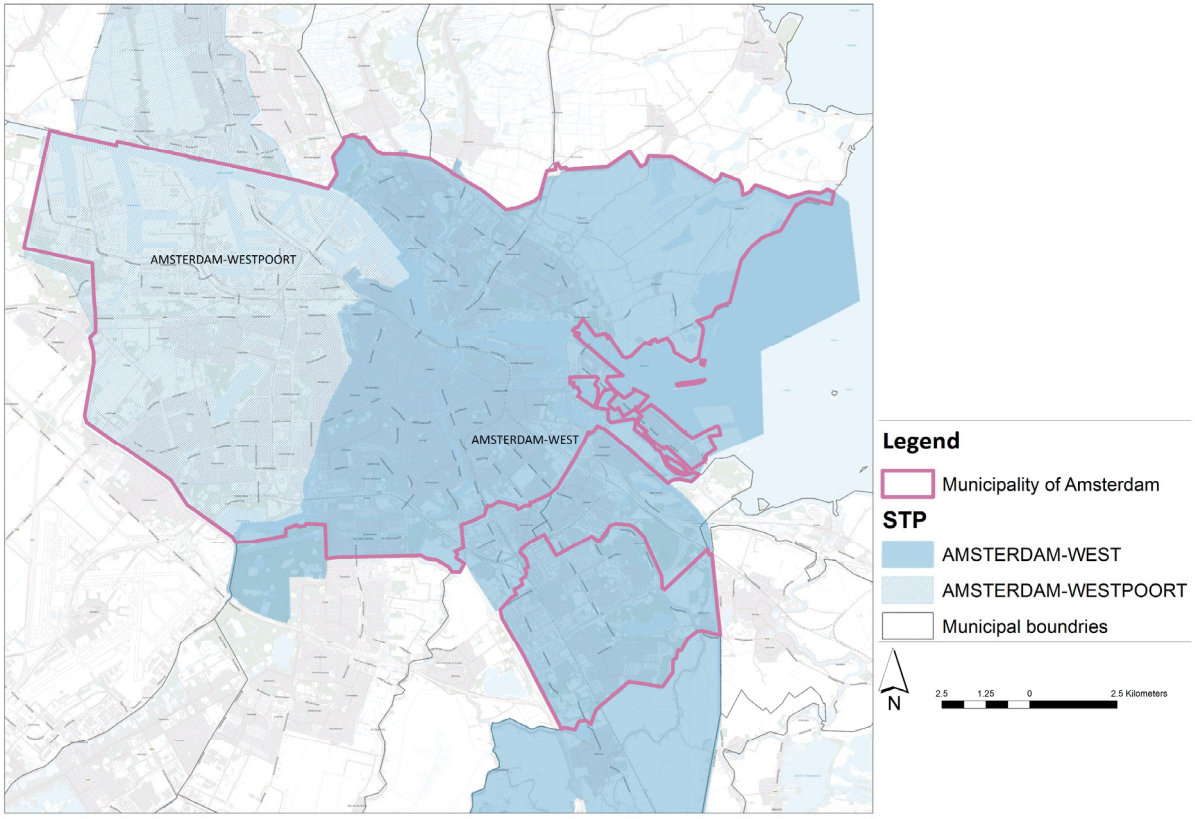
- Number of persons connected to each STP ranges from ≈1,000 to >800,000
- Quantitative measurements of SARS-CoV-2 RNA in all STPs in the Netherlands. Plants provide four 24h flow-proportional samples per week since late 2020
- Direct analysis of samples using qPCR. Targets: N1 and N2. Multiple PCRs per sample
- Daily flow rates of all 313-317 sewage plants to translate number of genome copies to virus particles per 100,000 persons per day
- Tables of numbers of persons in each postal code 4 (PC4) area, with referring STP and municipality from Statistics Netherlands
- Daily age-stratified hospital admissions with confirmed COVID-19 for each of 344 municipalities

Wastewater plants and catchment areas



- STP boundaries often do not overlap with boundaries of municipalities and sometimes also not with those of safety regions and provinces
- There are occasional changes in number of STPs and in numbers of municipalities, mostly by mergers
- Data from Statistics Netherlands (https://www.cbs.nl/nl-nl/maatwerk/2021/06/inwoners-per-rioolwaterzuiveringsinstallatie-1-1-2021)

Example: Amsterdam and Weesp



- The municipality of Amsterdam is served by two large STPs, Amsterdam-Westpoort and Amsterdam-West
- Since March 2022 the municipality Weesp has become part of the municipality of Amsterdam (‘stadsgebied Weesp’), as is the STP Weesp

Statistical model – basic idea

- Basis of the analyses is a Bayesian multilevel penalized spline model where the log-transformed latent virus load $c_i(t)$ of sewage plant i at time t is given by

$$c_i(t) = c(t) + d_i(t),$$

  where the population load $c(t)$ and STP-specific deviations $d_i(t)$ are modelled with a p-spline and zero mean b-splines, respectively

- The Bayesian framework enables incorporation of prior information. Regression coefficients (spline weights) of the deviations are given weakly informative normal prior distributions ($N(0,1)$ or $N(0,0.5)$)

- Provides description of the sewage data that can be used for assessment of trends, incorporation of censored data, extension to other organizational levels, and (perhaps) prediction

Statistical model – non-detects and likelihood

- Probability of detection: $p\big(\text{detection} \mid c_i(t)\big) = 1/\big(1 + e^{-k(c_i(t)-c_0)}\big)$
- Detection distribution: $D \sim \text{Bernoulli}\big(p(\text{detection})\big)$
- Prior distributions of detection function: $c_0 \sim N(12, 0.5)$ and k improper uniform
- Log-likelihood contributions:

$$\mathcal{L}_{it} = \begin{cases} f_D\big(1; p(\text{detection})\big) + f_Y(Y_{it}; c_i(t), \sigma) & \text{if RNA was detected at STP } i \text{ and time } j \\ f_D\big(0; p(\text{detection})\big) & \text{if RNA was not detected,} \end{cases}$$

where f_D and f_Y denote the log-Bernoulli and log-normal probability densities

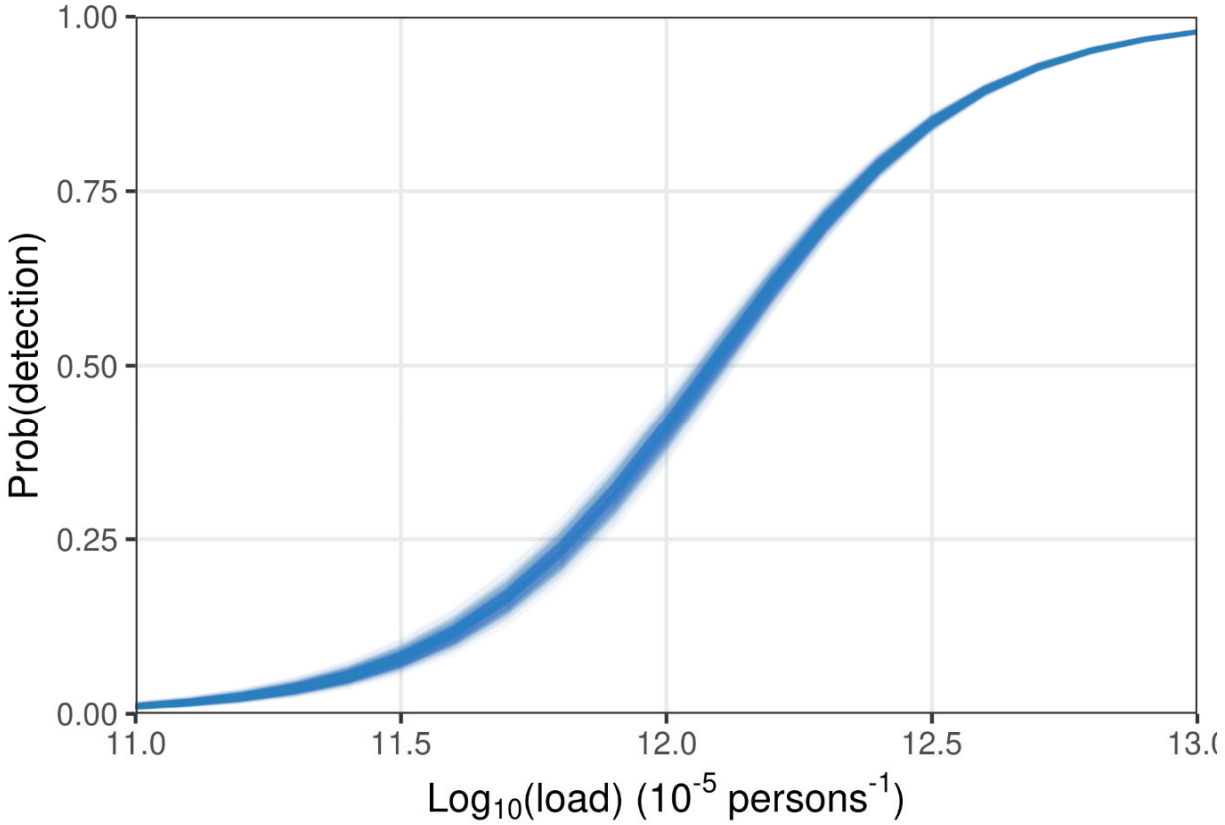
Implementation, cross-validation, model selection

- Estimation is performed in a Bayesian framework using Hamiltonian Monte Carlo
- The model is implemented in Stan with R interface
- Assessment of model fit is based on Bayesian predictive intervals and (visual) assessment of residuals
- Model selection is based on the leave-one-out information criterion to maximise predictive performance

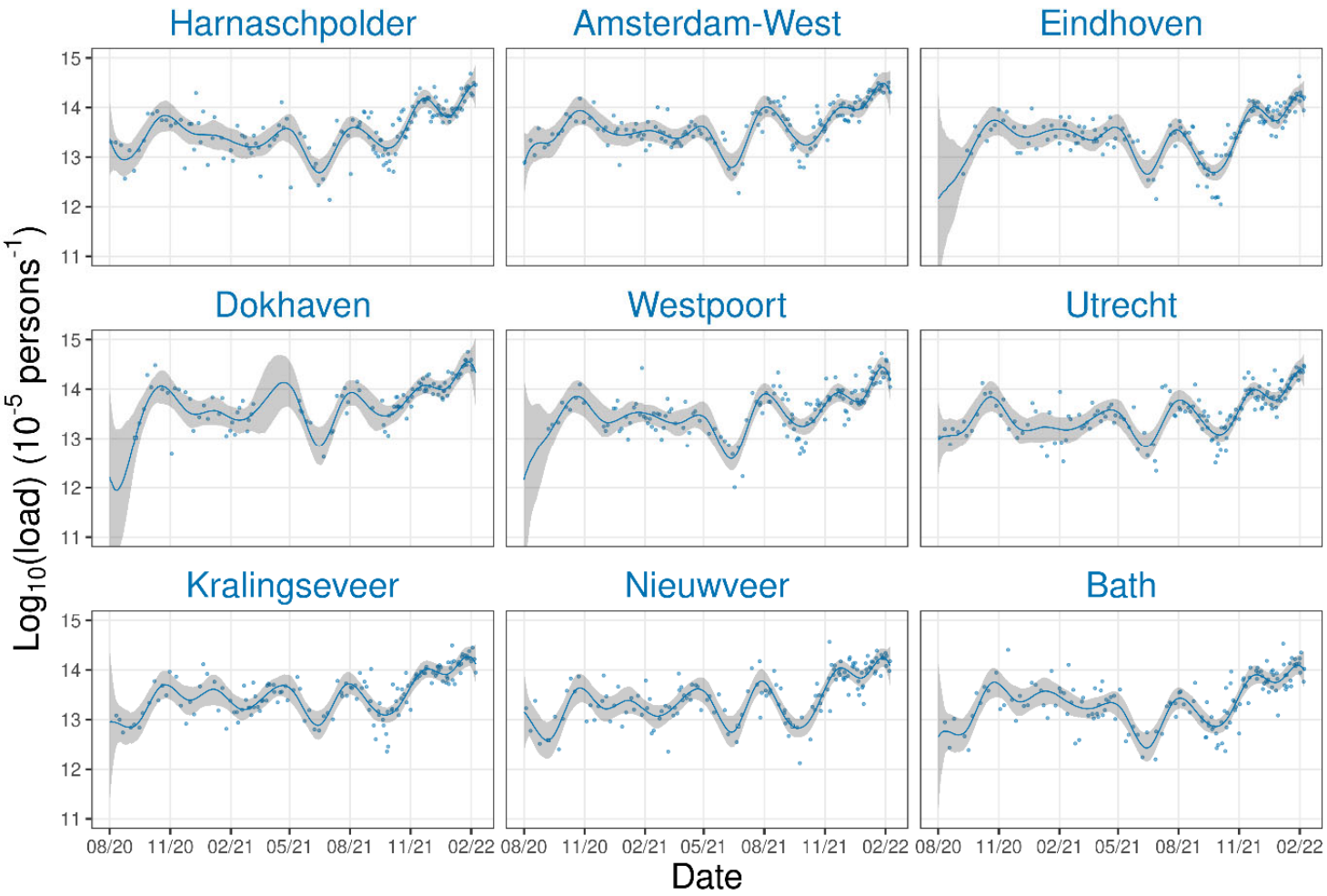
Model selection

- Single national load/include safety regions
- Observation model: single shared standard deviation
- Shared parameters of detection function
- First-order p-spline penalization
- No exhaustive evaluation

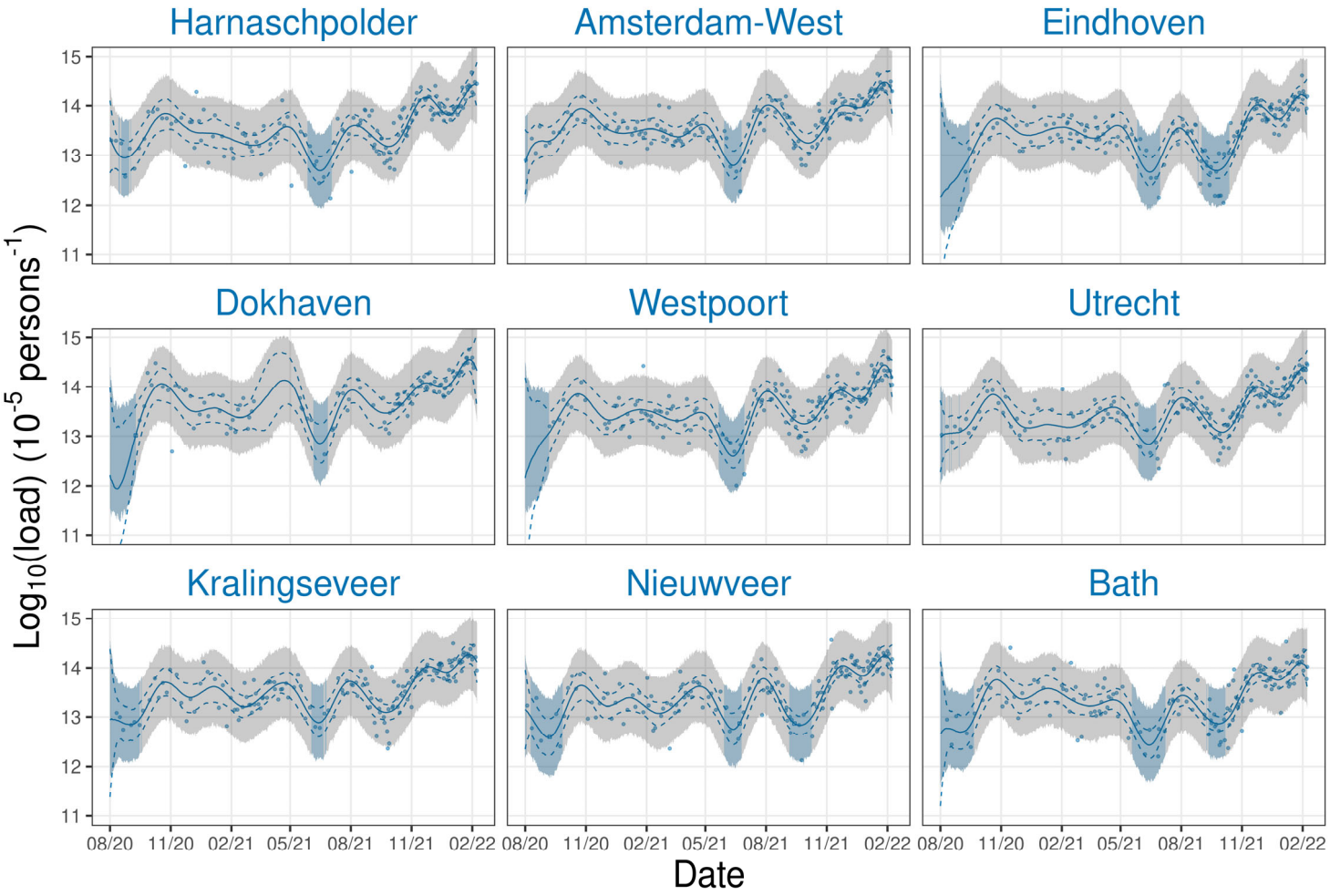
The probability of detection



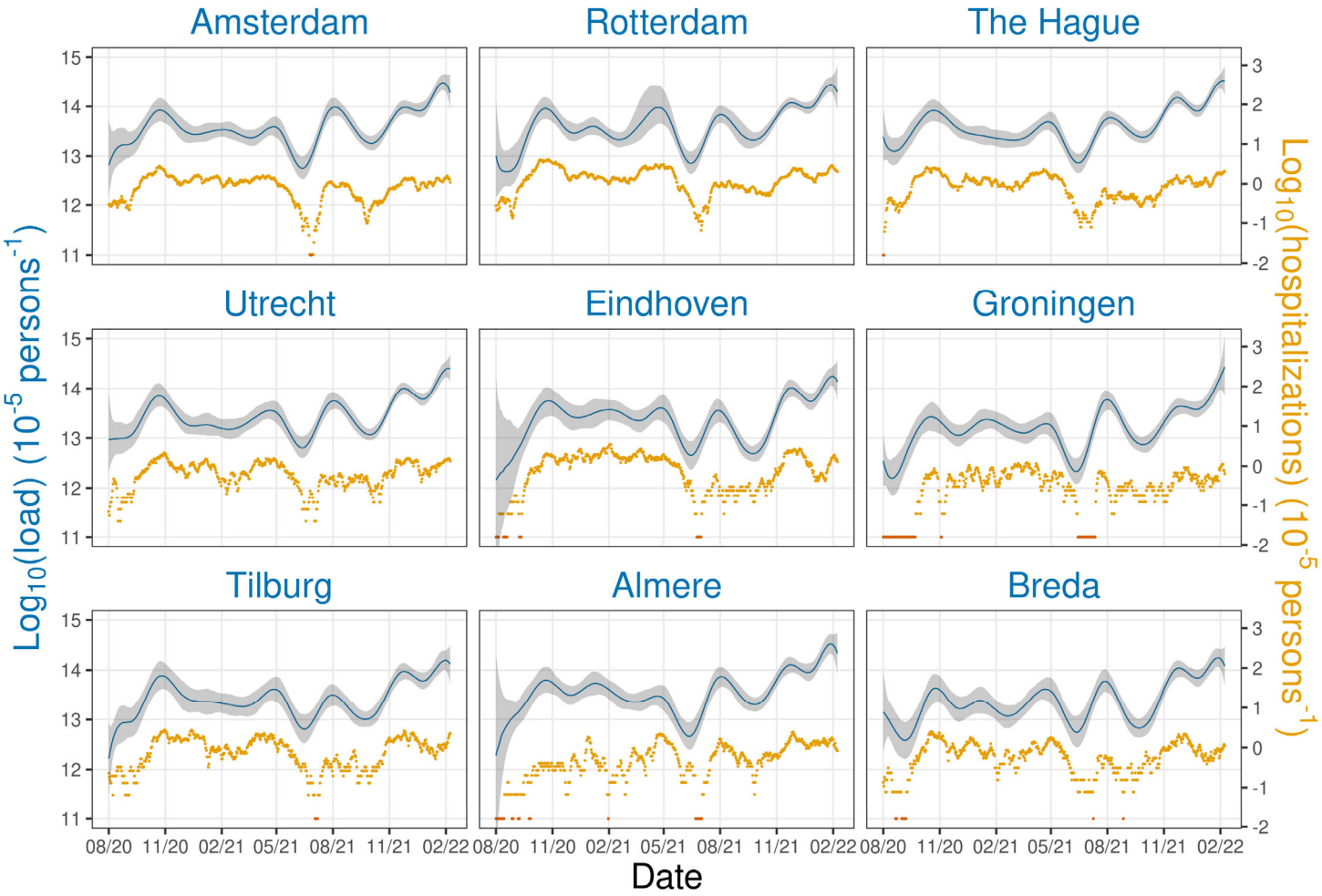
Virus load estimates



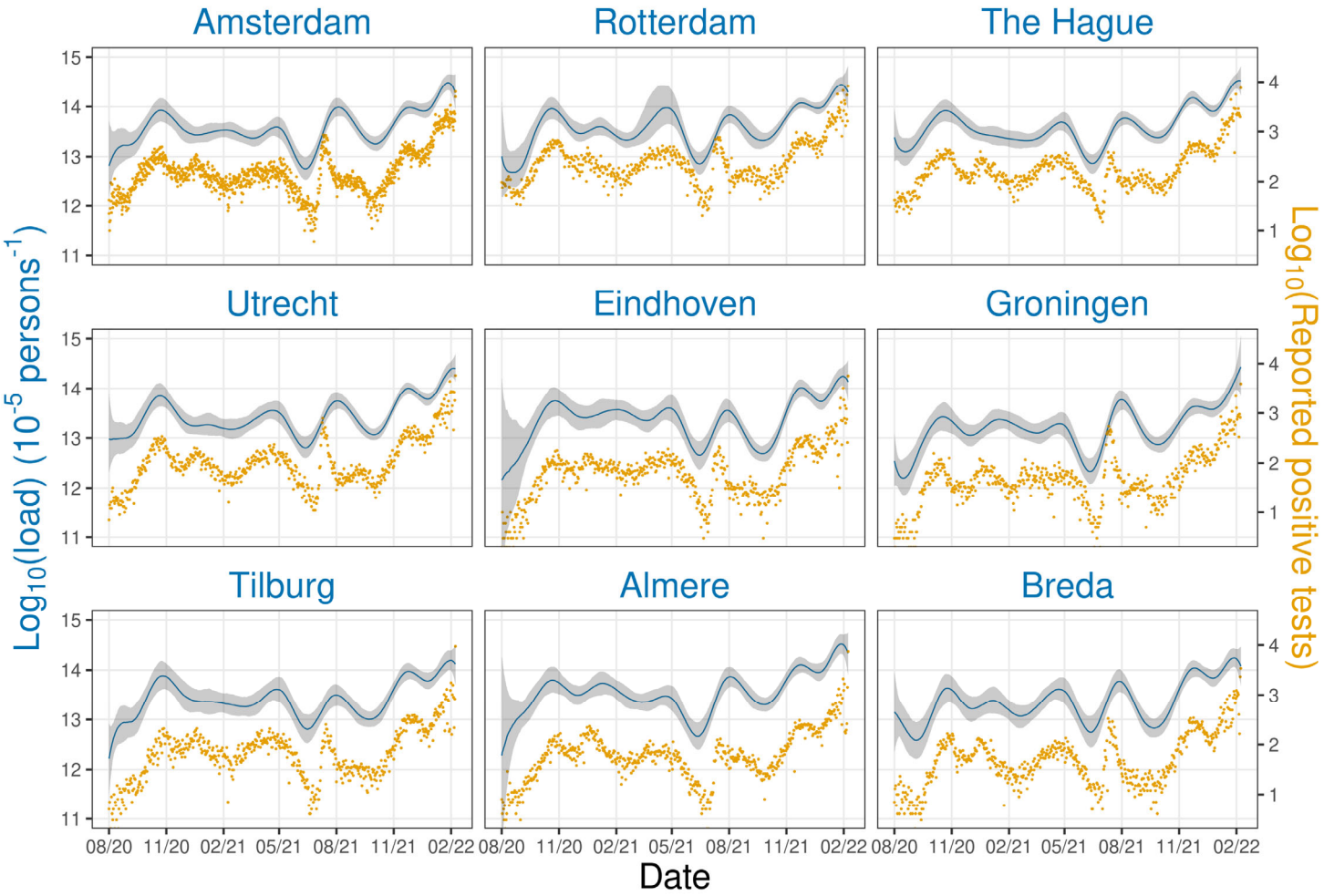
Prediction intervals



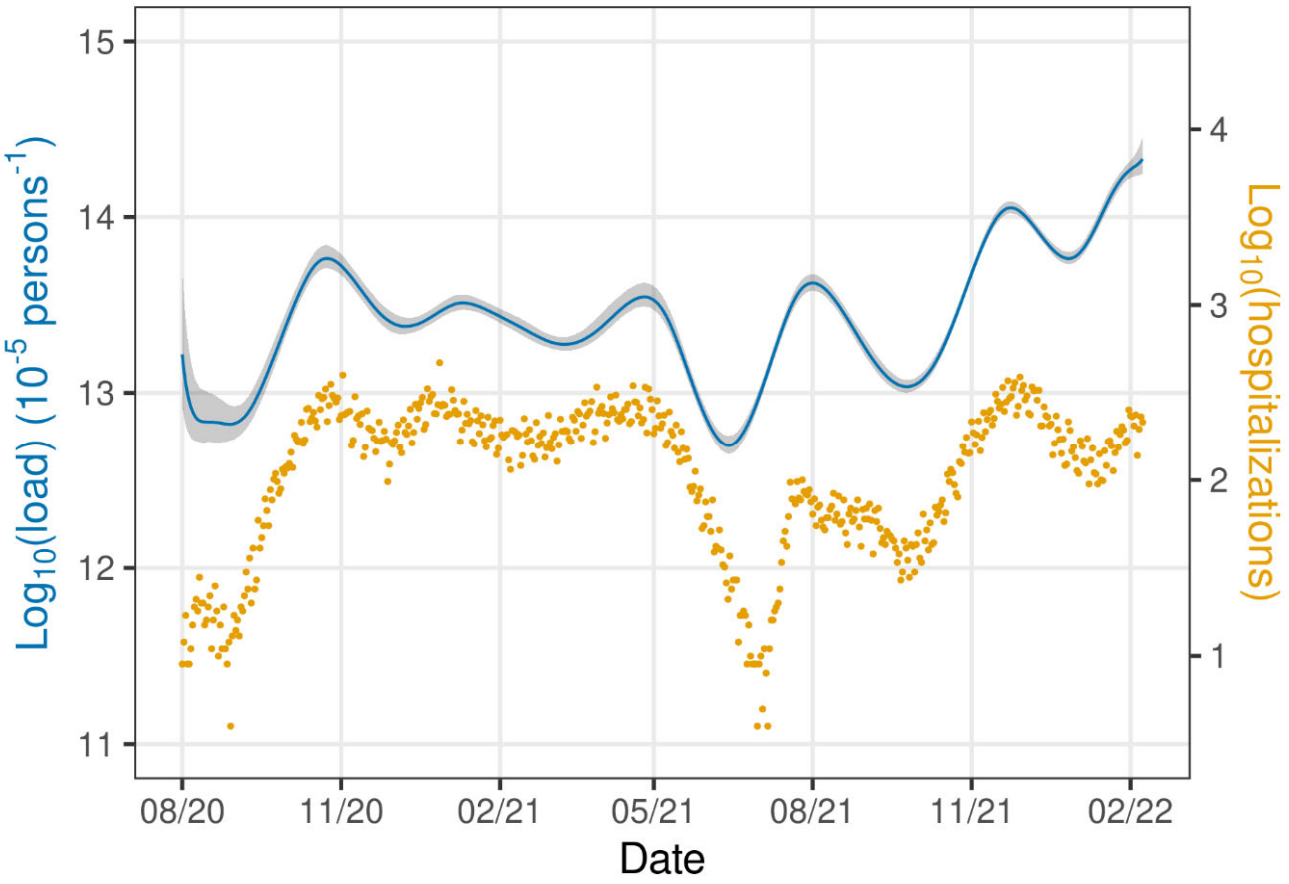
Virus loads and hospitalisations



Virus loads and positive tests



Population average – a word of caution



Discussion

- Epidemic peaks and troughs are not fully synchronised, 1-2 week shifts of peak loads are common
- Intrinsic variations in virus load measurements are estimated to be substantial (σ=0.35), such that measured virus loads can be approximately 5 times higher or lower than the estimated latent virus load ($\approx 10^{2\sigma}$)
- Nevertheless, latent virus loads can be estimated quantitatively with fair precision as increases and decreases during epidemic upswings and downturns are in the order of two orders of magnitude (>100-fold), dwarfing intrinsic variations
- The above argument depends on having sufficiently dense sampling

https://github.com/rivm-syso/SARS-CoV-2_sewage

Discussion

- Model provides an overall adequate description of the data; however, very sudden drops and increases in the data are not always captured well (i.e. October 2021)
- Calculation of virus loads is based on number of inhabitants per PC4, . This may not always work well (working areas, tourist regions). Alternatives, e.g., using Pepper Mild Mottle Virus (PPMoV) are possible
- No age stratification or weighting by age groups
- No local dependencies between STPs/municipalities

https://github.com/rivm-syso/SARS-CoV-2_sewage

Virus loads and hospitalisation admissions

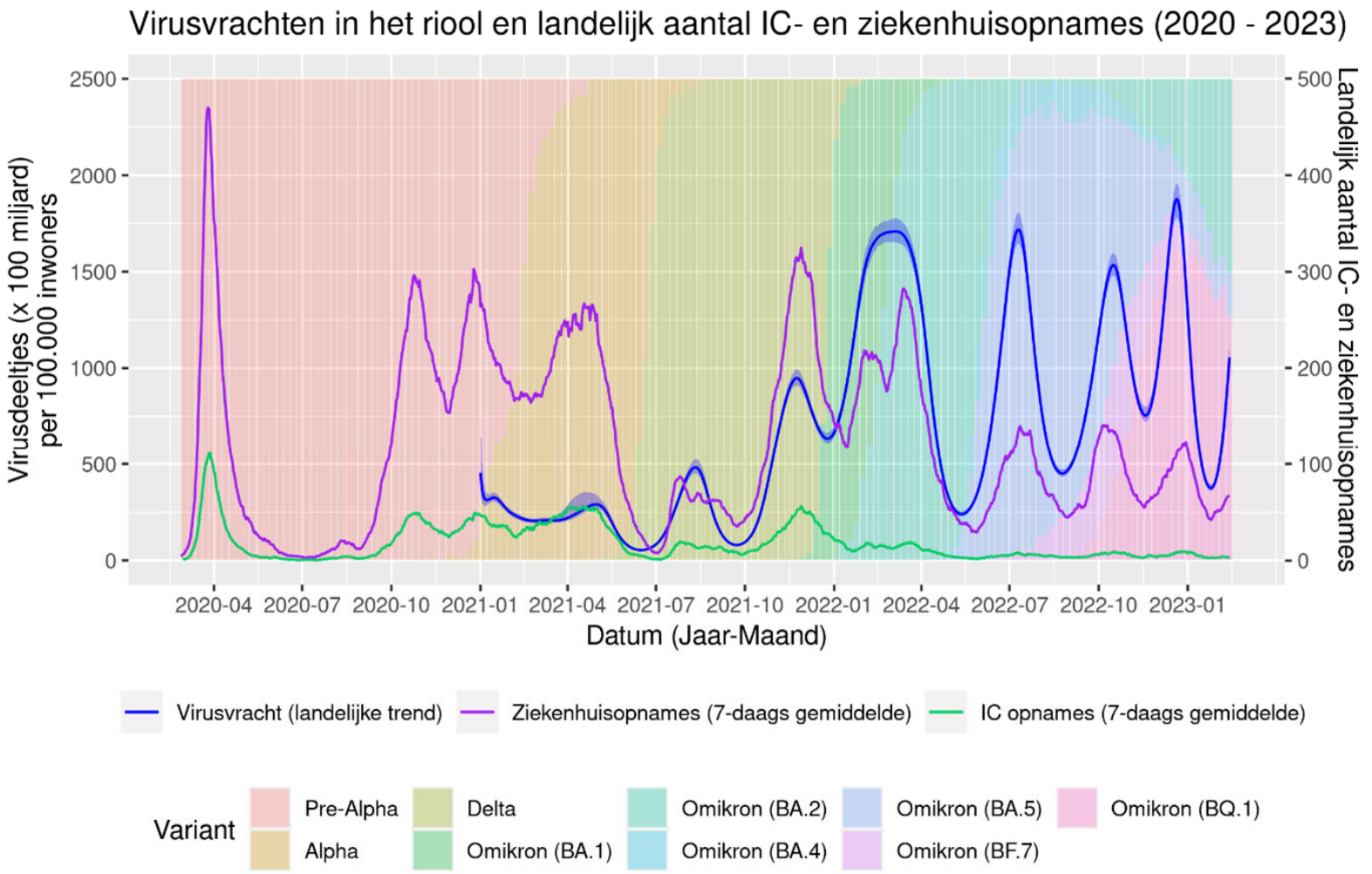
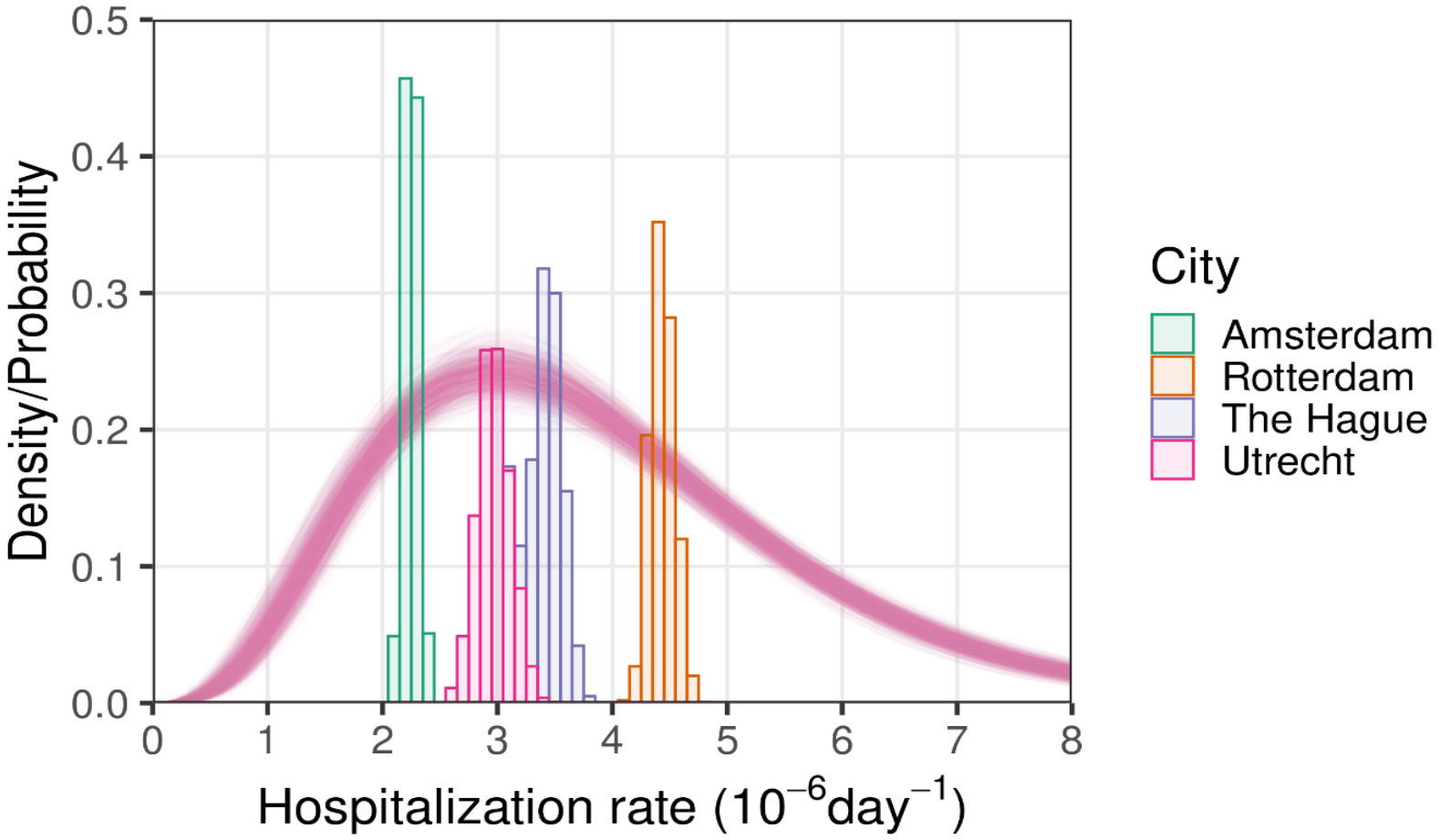


figure courtesy Wouter Hetebrij

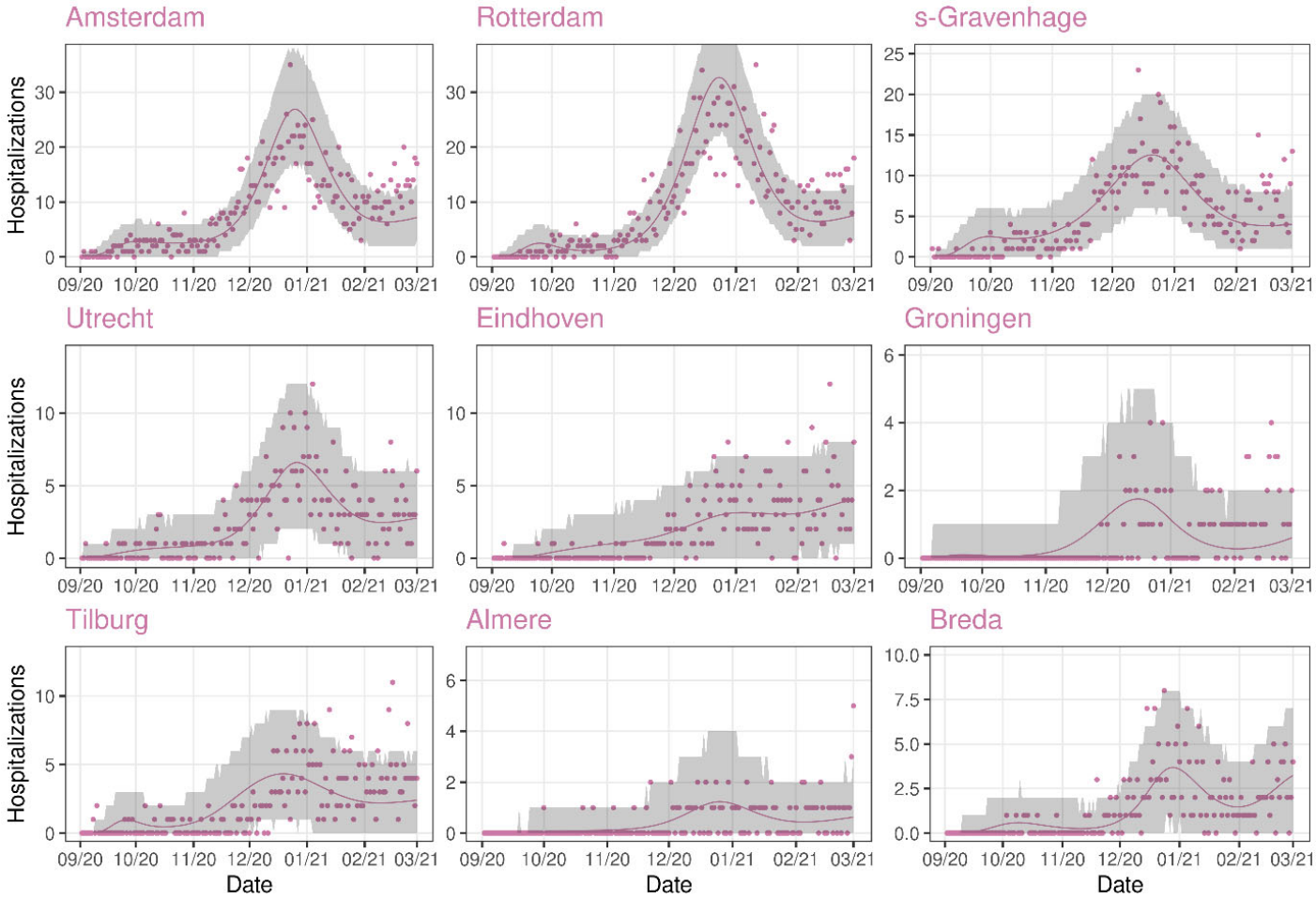
Relating virus loads in sewage to hospitalisations

- Level of analysis: municipality as at this level both hospital admissions and test data are available
- General approach: Bayesian multilevel Poisson regression that takes the inferred virus loads at the municipality level as main explanatory variable
- Per municipality a rate parameter is estimated from a gamma hyperdistribution (Bayesian multilevel/random effects model)
- Up to early 2022: Only age-aggregated hospital admissions
- From early 2022 onwards: age stratified municipal hospital admission and vaccination coverage data available
- Frequencies of variant circulation taken from national virological surveillance

Relating hospital admissions to virus loads in the pre-vaccination era (at load 10^{13})



Relating hospital admissions to virus loads in the pre-vaccination era



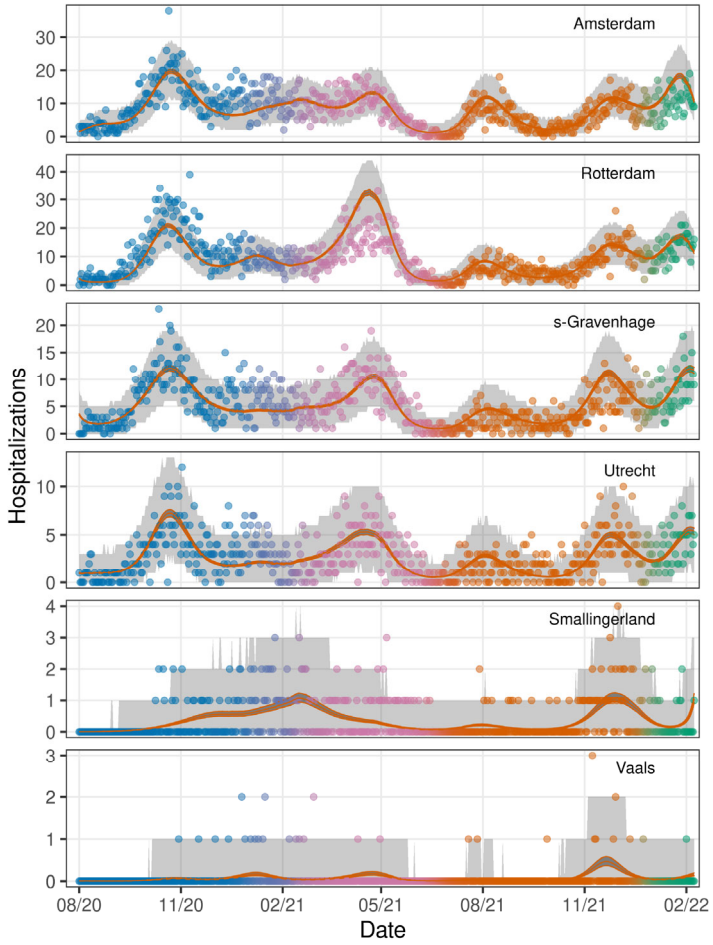
Age-structure, vaccination, variants

- Extended Poisson regression that includes age (0-20, 20-40, 60-80, 80+ years), strain (wildtype, alpha, delta, omicron), and vaccination coverage (first, second, booster) in each of the 344 Dutch municipalities
- Baseline rate parameters in each municipality and age group are taken from uninformative Gamma hyperpriors
- Implicit assumption: fixed distribution of infections in different age groups
- Beta(2,8) prior distributions for reductions in rate parameters after vaccination, implying 0.52-0.97 prior coverage of reductions in shedding (think VE)

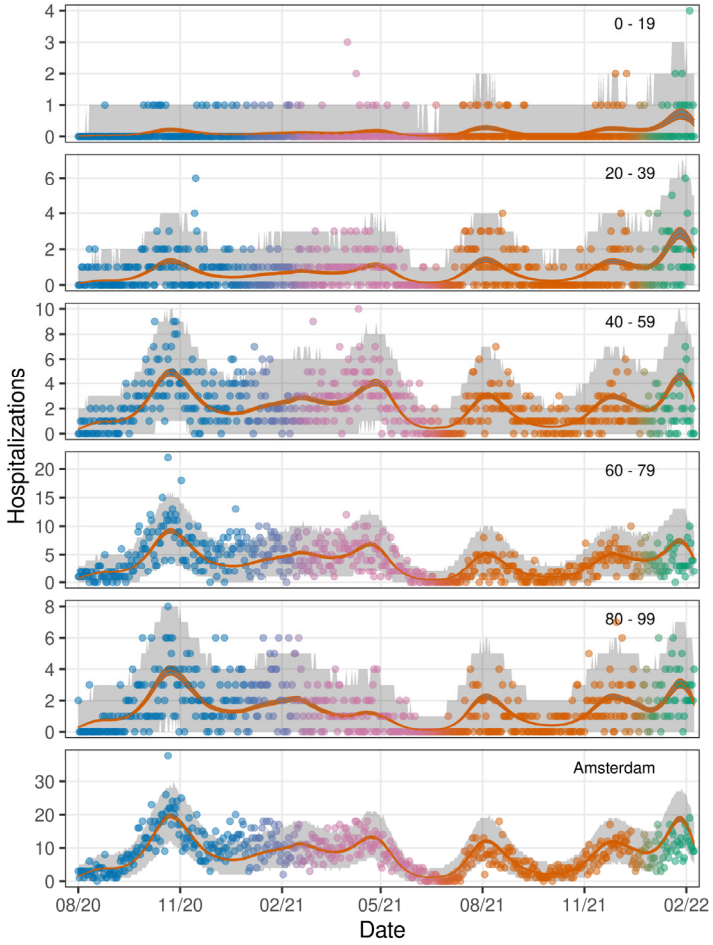
$$H_{m,t,a} \sim \text{Pois}\left(\lambda_{m,a} \times \sum_s \left(\tilde{\lambda}_s \times p_{s,t} \times \sum_v \left(\xi_{a,v,s} \times f_{m,t-14,a,v}\right)\right) \times l_{m,t} \times n_{m,a}\right)$$

Age-structure, variants, and vaccination

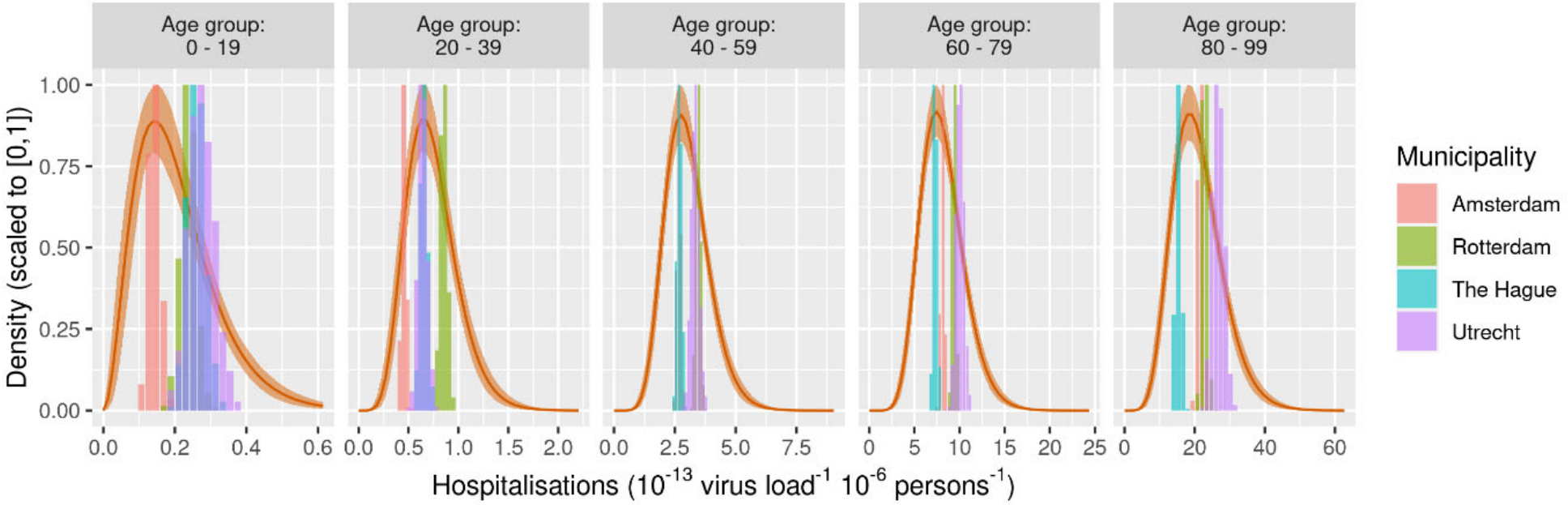
Municipalities, aggregated



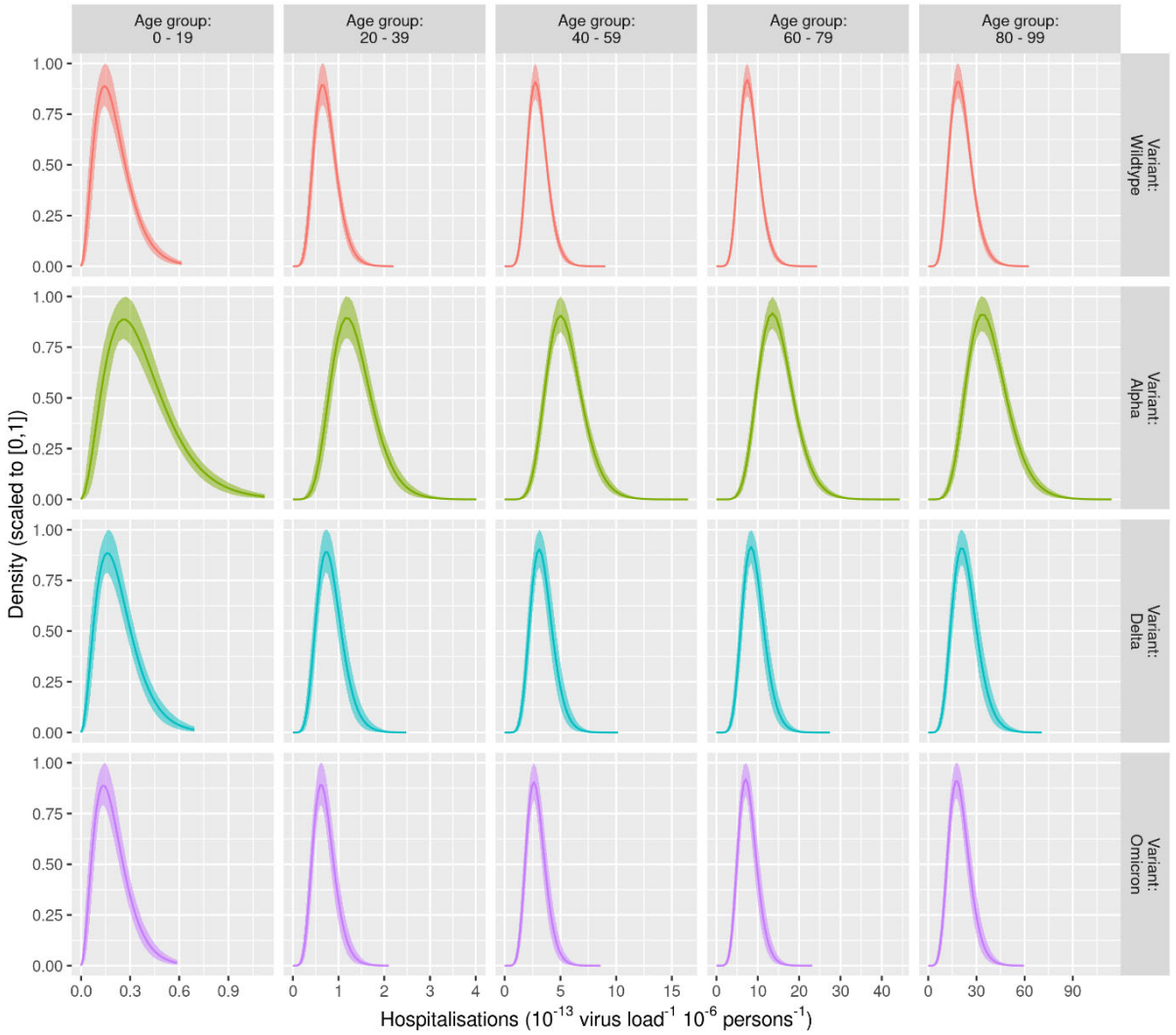
Amsterdam, by age



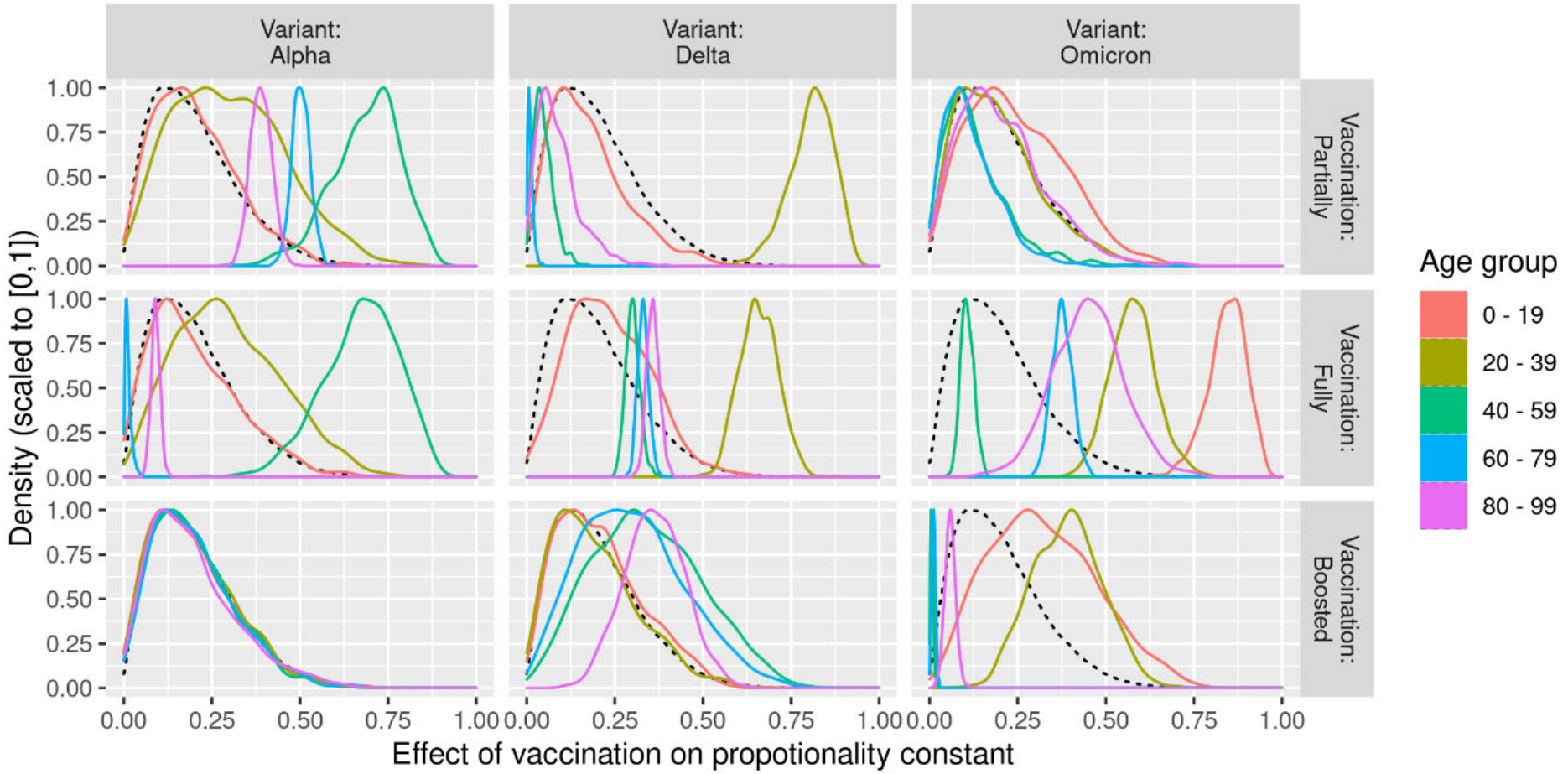
Age-structure, variants, and vaccination



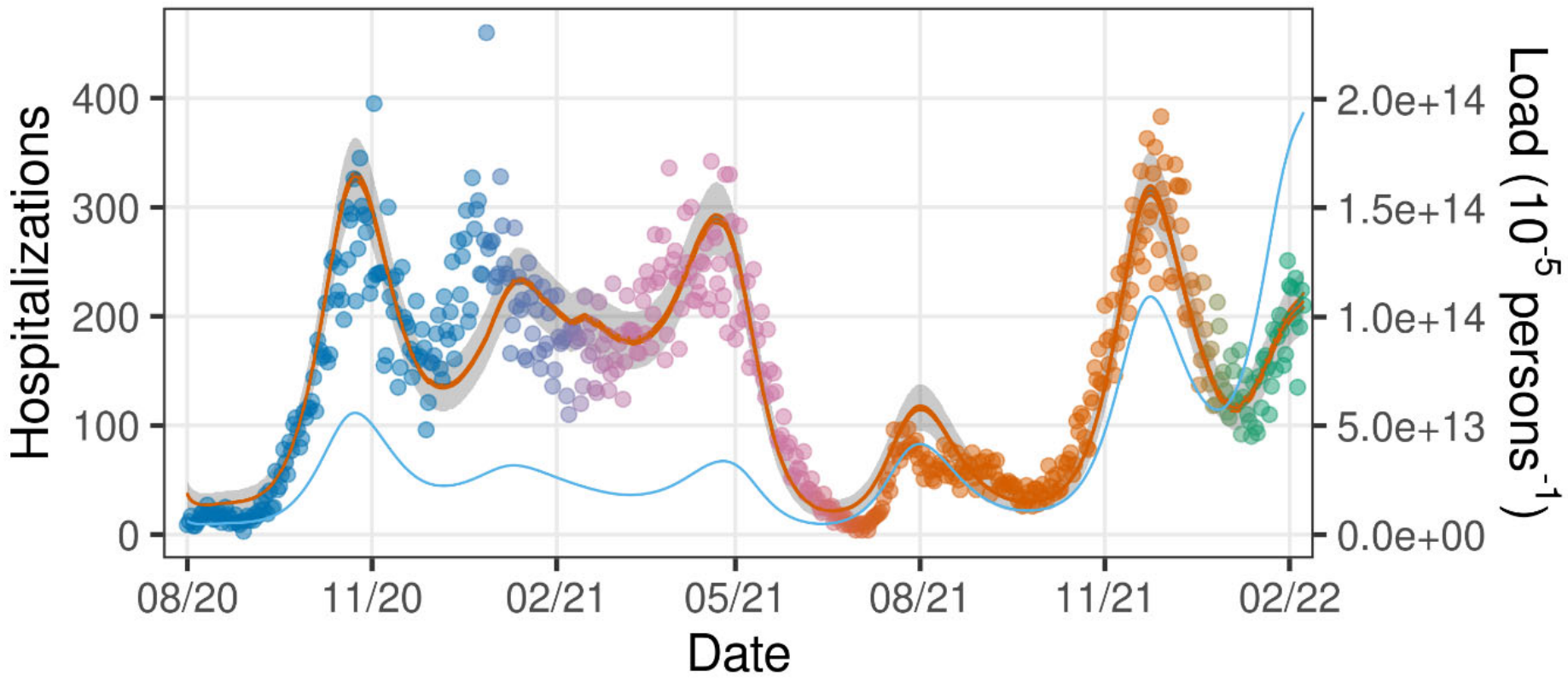
Age-structure, variants, and vaccination



Age-structure, variants, and vaccination



Population average

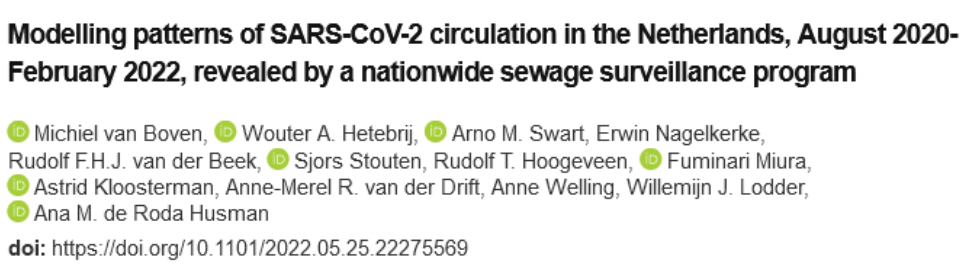


Discussion

- Conceptually simple Poisson regressions can adequately relate estimated virus loads in sewage to hospital admissions when age, vaccination coverage, and strain composition are taken into account
- Move from phenomenological statistical models to fitting mechanistic transmission models using Generalised Profiling
- Predictive ability is probably limited to short periods (1-2 weeks)
- Lags are not taken into account
- Under development: influenza, RSV, enteroviruses, AMR

Acknowledgements

- Wouter Hetebrij, Arno Swart, Ruud van der Beek, Erwin Nagelkerke, Senna van Iersel, Willemijn Lodder, Ana Maria de Roda Husman, and many others
- Ministry of Health, Welfare and Sport
- Union of Waterboards and Regional Waterboards
- Statistics Netherlands



Modelling patterns of SARS-CoV-2 circulation in the Netherlands, August 2020-February 2022, revealed by a nationwide sewage surveillance program

Michiel van Boven, Wouter A. Hetebrij, Arno M. Swart, Erwin Nagelkerke, Rudolf F.H.J. van der Beek, Sjors Stouten, Rudolf T. Hoogeveen, Fuminari Miura, Astrid Kloosterman, Anne-Merel R. van der Drift, Anne Welling, Willemijn J. Lodder, Ana M. de Roda Husman

doi: https://doi.org/10.1101/2022.05.25.22275569

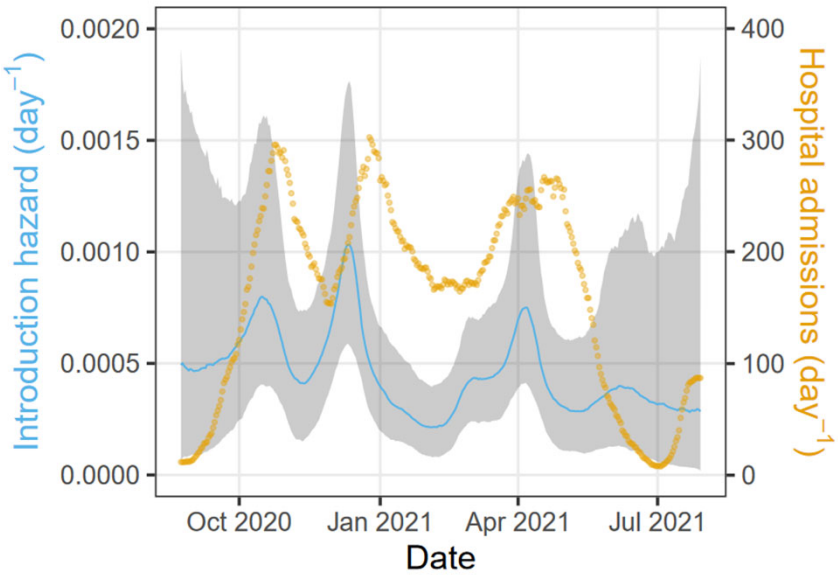
Inferring Hospital Admissions from SARS-CoV-2 Virus Loads in Wastewater in the Netherlands, August 2020 – February 2022

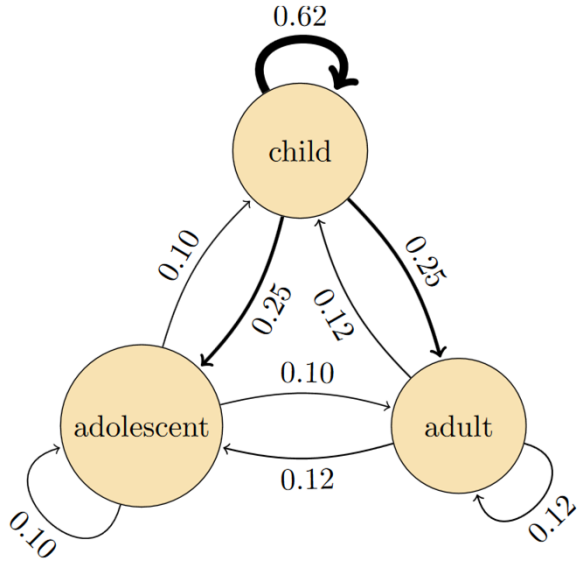
11 Pages • Posted: 31 May 2023

Wouter A. Hetebrij

National Institute for Public Health and the Environment (RIVM) - Centre for Infectious Disease Control

An advertisement





Estimation of introduction and transmission rates of SARS-CoV-2 in a prospective household study

Michiel van Boven, Christiaan H. van Dorp, Ilse Westerhof, Vincent Jaddoe, Valerie Heuvelman, Liesbeth Duijts, Elandri Fourie, Judith Sluiter-Post, Marlies A. van Houten, Paul Badoux, Sjoerd Euser, Bjorn Herpers, Dirk Eggink, Marieke de Hoog, Trisja Boom, Joanne Wildenbeest, Louis Bont, Ganna Rozhnova, Marc J. Bonten, Mirjam E. Kretzschmar, Patricia Bruijning-Verhagen

doi: https://doi.org/10.1101/2023.06.02.23290879

https://github.com/mvboven/sars2-households