Inferring the Role of School Children in Bringing Infection into Households

Thomas House

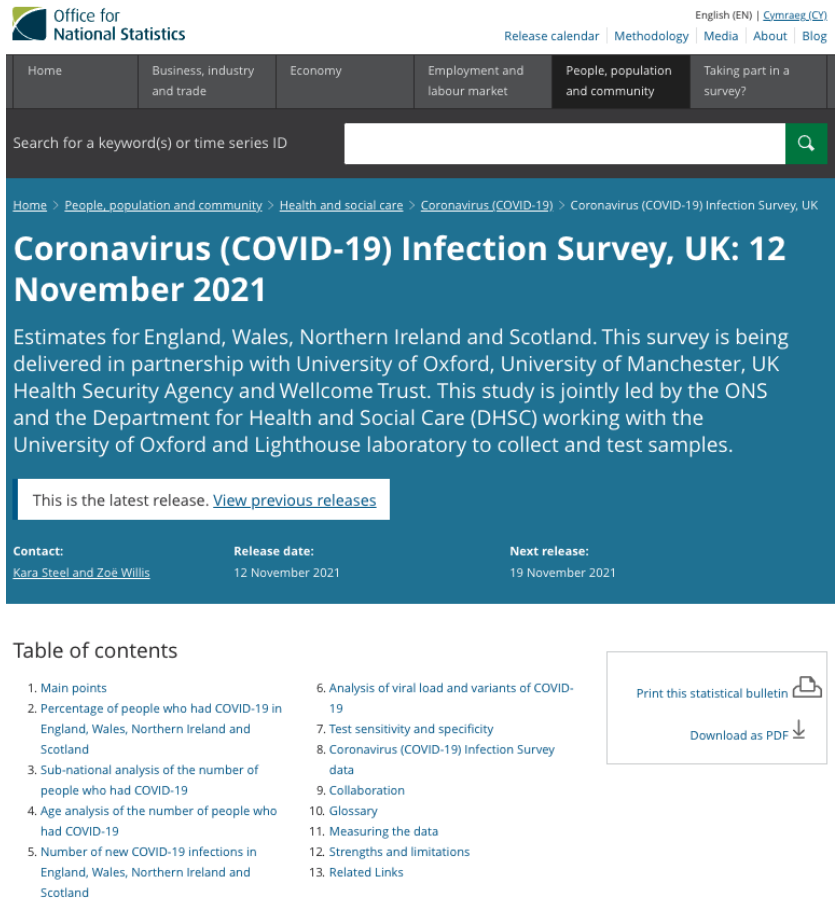
University of Manchester

Controlling COVID-19 in Schools: Lessons Learned and Open Questions
INI, 27 April 2022, 10.35am

Community Surveillance

- The UK Government funded lots of excellent pandemic science, including two very large community surveillance studies based on (approximately) uniformly random sampling from the population.
- One of these, REACT, has an individual-based cross-sectional design.
- The UK Office for National Statistics' COVID-19 Infection Survey (ONS CIS) has a household-based longitudinal design.
- There is also case-finding, death, and hospitalisation data, as well as a mobile-based symptom study (CSS or 'Zoe') and the official NHS App.
- Of these, cases, REACT has now stopped and Zoe has lost Government funding so we have to adapt modelling to diminished data.

The ONS CIS has a regular reporting role: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveypilot/previousReleases



Office for National Statistics

English (EN) | Cymraeg (CY)

Release calendar | Methodology | Media | About | Blog

Home | Business, industry and trade | Economy | Employment and labour market | People, population and community | Taking part in a survey?

Search for a keyword(s) or time series ID

Home > People, population and community > Health and social care > Coronavirus (COVID-19) > Coronavirus (COVID-19) Infection Survey, UK

Coronavirus (COVID-19) Infection Survey, UK: 12 November 2021

Estimates for England, Wales, Northern Ireland and Scotland. This survey is being delivered in partnership with University of Oxford, University of Manchester, UK Health Security Agency and Wellcome Trust. This study is jointly led by the ONS and the Department for Health and Social Care (DHSC) working with the University of Oxford and Lighthouse laboratory to collect and test samples.

This is the latest release. View previous releases

Contact: Kara Steel and Zoë Willis

Release date: 12 November 2021

Next release: 19 November 2021

Table of contents

1. Main points
2. Percentage of people who had COVID-19 in England, Wales, Northern Ireland and Scotland
3. Sub-national analysis of the number of people who had COVID-19
4. Age analysis of the number of people who had COVID-19
5. Number of new COVID-19 infections in England, Wales, Northern Ireland and Scotland
6. Analysis of viral load and variants of COVID-19
7. Test sensitivity and specificity
8. Coronavirus (COVID-19) Infection Survey data
9. Collaboration
10. Glossary
11. Measuring the data
12. Strengths and limitations
13. Related Links

Print this statistical bulletin

Download as PDF

Why households?

- One of the reasons that the ONS study has a household design is that people live in households.
- Another is that there is the possibility of using this to infer something about **transmission**, which is something special about infections.
- In the past, the CIS has tended to play a more confirmatory role of results obtained through case finding, which is intrinsically faster due to study design (more below).
- Due to the changes in data streams, though, we are expecting to be in the spotlight more during the next phase of the pandemic.

Households in Disease Dynamics

Household models are an integral part of the history of infectious disease epidemiology, alongside the better known whole-population models like the SIR equations. Households are important for various reasons:

- The close, repeated nature of contact within the household means that within-household transmission of infectious disease is common.
- Most of the population lives in relatively small, stable households and so these are a natural unit for data collection.
- We can design interventions at the household level – this pandemic, the emphasis has been on whole-household isolation, and school LFD testing has a strong household element, for example.

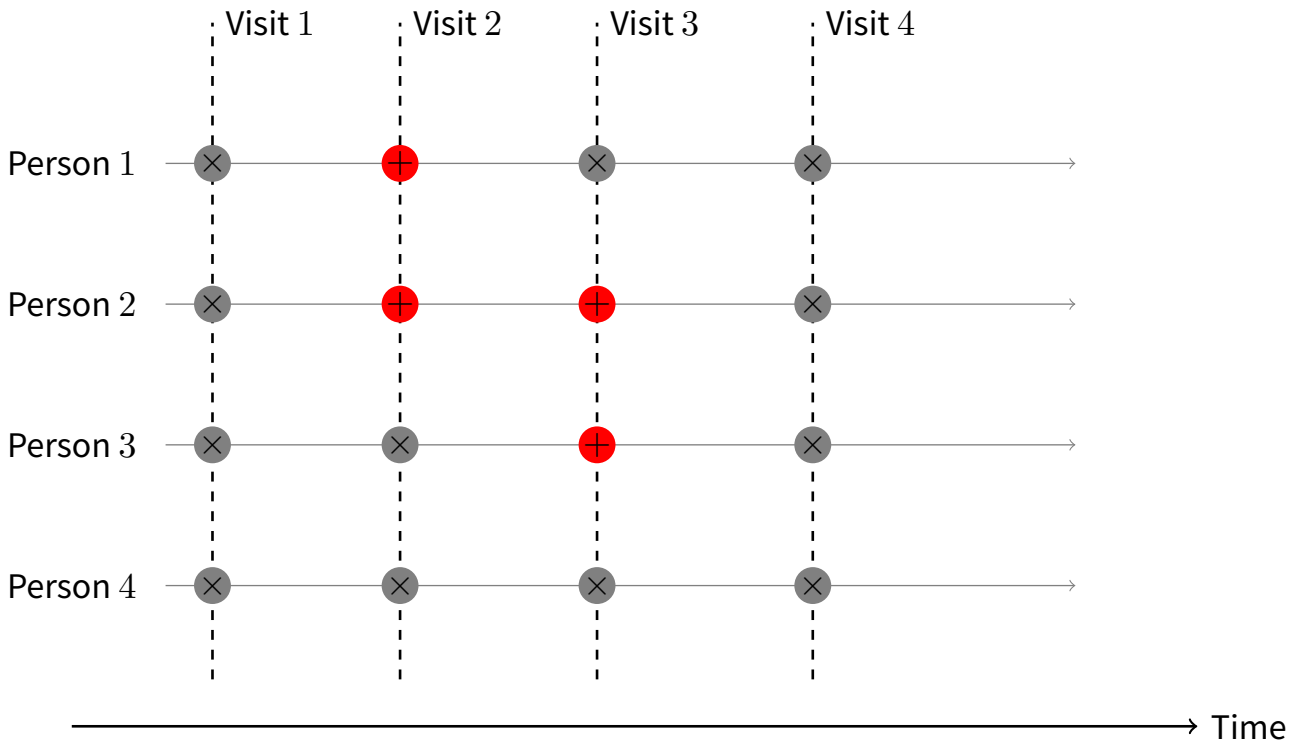
History

Personal view – there have been three ‘eras’:

1. Early-mid 20th century: Reed and Frost’s unpublished work in the 1920s on the first stochastic epidemic model (simulated using a modified roulette table). Theoretical developments by e.g. Bailey and symptom-based empirical observations by e.g. Hope Simpson.
2. Late 20th century: General final-size formula from Ball, Statistical work using this by e.g. Addy, Longini, Halloran on e.g. Tecumseh study based on viral culture.
3. 21st century: Modern computational methods (e.g. MCMC – Demiris and O’Neill) available as well as modern molecular techniques such as PCR for empirical work.

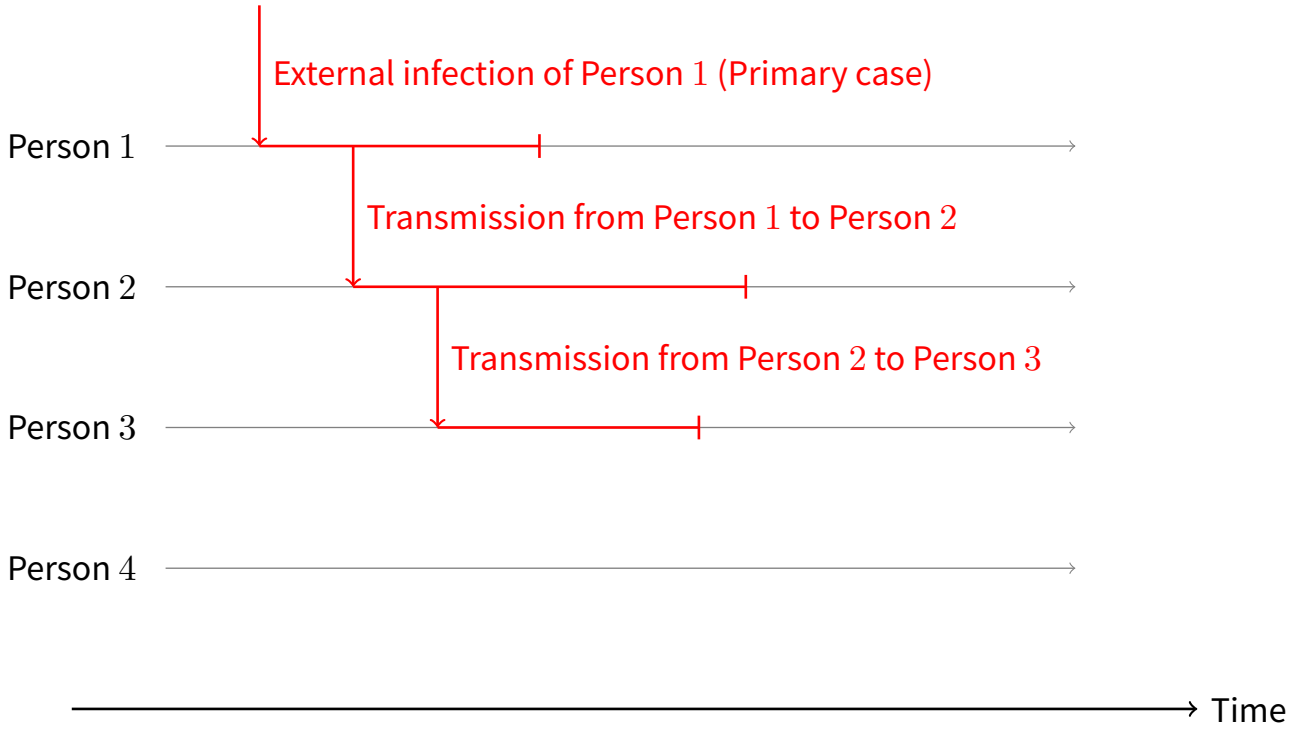
As is often the case, in an emergency, we will use the last era’s methods to get a timely answer!

ONS CIS study design



The study design involves weekly then monthly household visits for a year, with PCR testing, WGS and serology. This involves less sampling bias than data obtained through case ascertainment, in the public health system, but at the cost that the actual transmission routes are not observed, only point positivity (red +).

Compare to Test and Trace



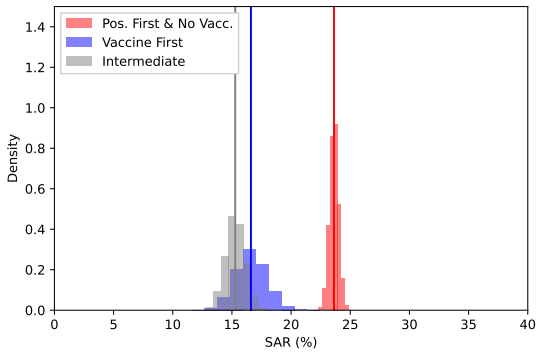
In an idealised version of test and trace, the actual routes of transmission are observed, and so the primary, secondary, tertiary etc. cases are known, although often it is more complex, and the first case ascertained is assumed primary, with all others secondary.

Secondary attack rates

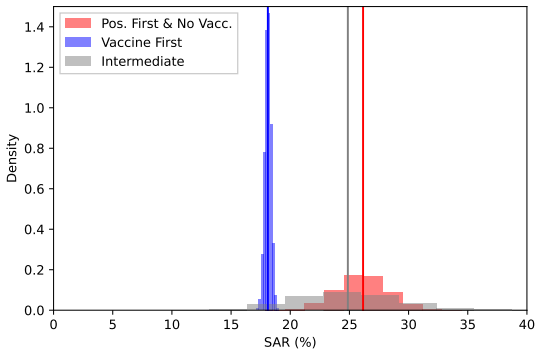
- Based on : **T. House**, H. Riley, L. Pellis, A. Eidukas, C. Ferguson, Z. Janes, E. Pritchard, A. R. McLean, A. S. Walker, “Total Effect Analysis of Vaccination on Household Transmission in the Office for National Statistics COVID-19 Infection Survey” [arXiv:2107.06545].
- Much updated since then!
- The basic idea is to stratify our sample **by household**, vaccine status and dominant strain / time period, and then calculate a SAR within each, using bootstrapping to work out CIs.
- Formally If we have m households and the i-th household has size n_i and y_i positives, then let the set of households with at least one infection be $\mathcal{I} = \{i \in [m] | y_i \geq 1\}$, then the SAR is

$$\mathrm{SAR} = \frac{\sum_{i \in \mathcal{I}} (y_i - 1)}{\sum_{i \in \mathcal{I}} (n_i - 1)} \,. \tag{1}$$

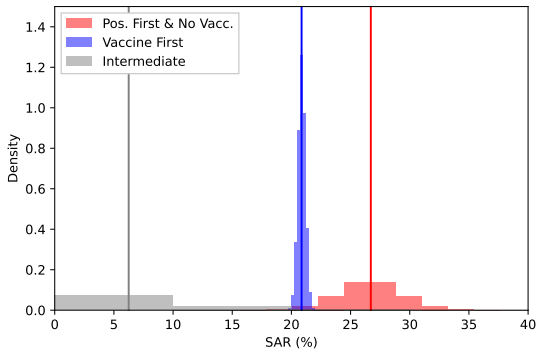
SAR results – Graphs



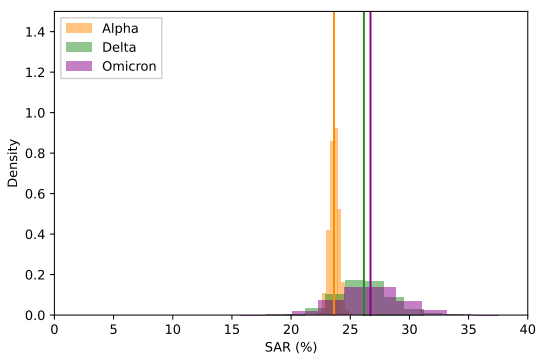
(a) Alpha period (1 December 2020 to 16 May 2021)



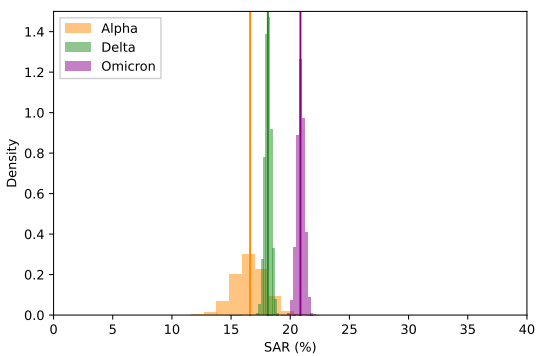
(b) Delta period (17 May 2021 to 14 December 2021)



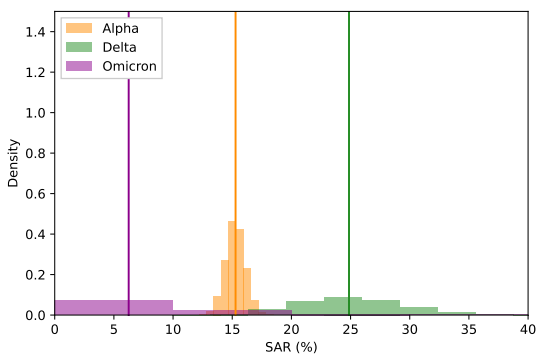
(c) Omicron period (15 December 2021 to 24 January 2022)



(d) Variants for positive first & no vaccine households



(e) Variants for vaccine first households



(f) Variants for intermediate vaccine households

SAR results – Tables

Household vaccine status	SAR (95% CI)	p-value for column $>$ row		
		PF&NV	Inter	VF
Alpha period (1 December 2020 to 16 May 2021)				
Positive First & No Vaccine	23.6 (22.8, 24.4) %	·	$< \epsilon$	$< \epsilon$
Intermediate	16.6 (14.1, 19.2) %	$> 1 - \epsilon$	·	0.81
Vaccine first	15.3 (13.7, 16.9) %	$> 1 - \epsilon$	0.19	·
Delta period (17 May 2021 to 14 December 2021)				
Positive First & No Vaccine	26.2 (22.1, 30.4) %	·	0.39	$< \epsilon$
Intermediate	18.1 (17.6, 18.6) %	0.61	·	0.053
Vaccine first	24.9 (16.8, 33.5) %	$> 1 - \epsilon$	0.947	·
Omicron period (15 December 2021 to 24 January 2022)				
Positive First & No Vaccine	26.7 (21.7, 32.1) %	·	0.02	0.01
Intermediate	20.9 (20.3, 21.5) %	0.98	·	0.953
Vaccine first	6.3 (0.000, 25.0) %	0.99	0.047	·

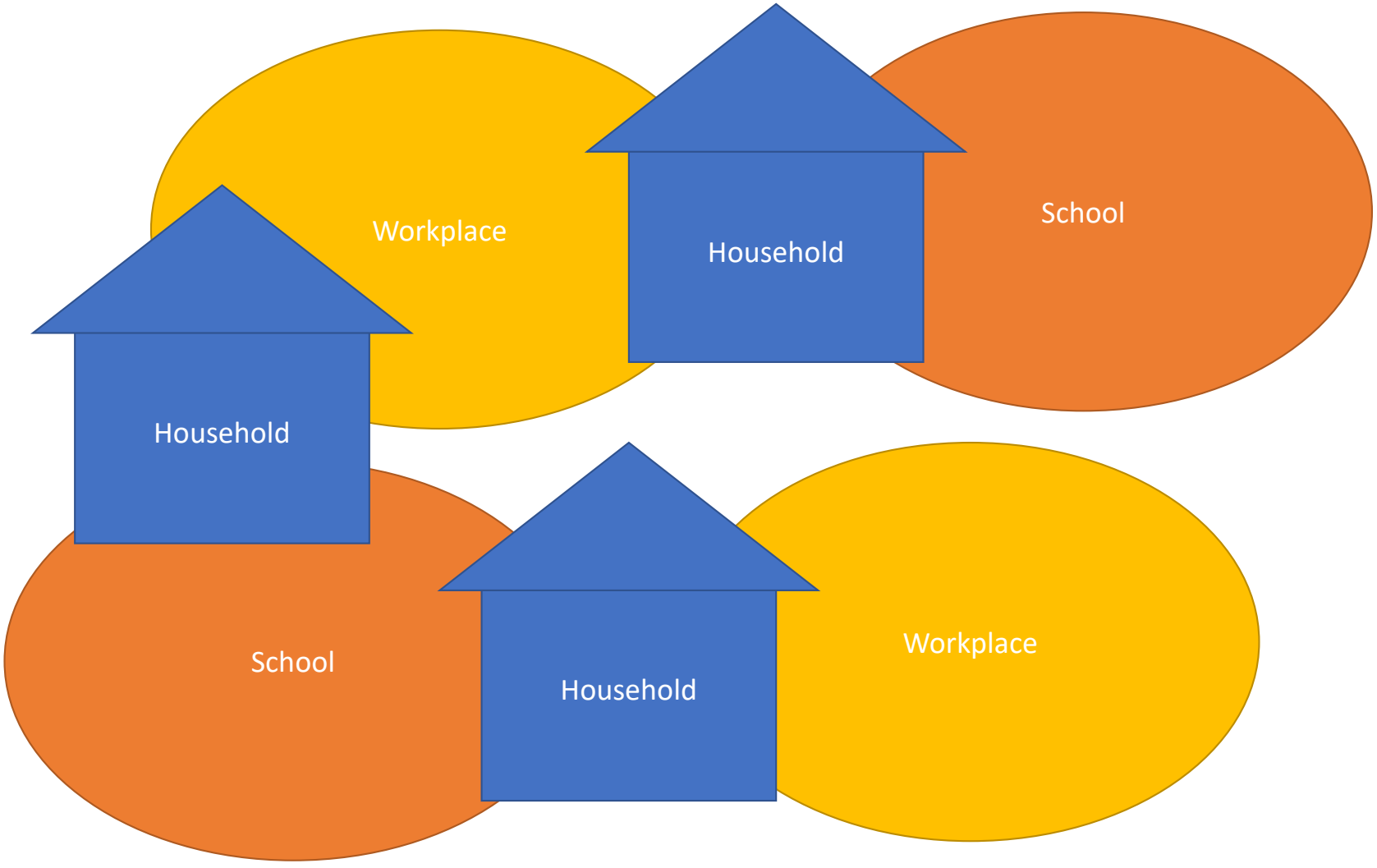
Table: Numerical values for the SAR with confidence invervals and p-values, stratified by time period. ϵ represents the minimum value that can be reliably estimated from 2×10^4 bootstrap samples, i.e. below 0.001.

SAR results – Tables

Dominant variant	SAR (95% CI)	p-value for column $>$ row		
		Alpha	Delta	Omicron
Positive First & No Vaccine				
Alpha	23.6 (22.8, 24.4) %	·	0.88	0.87
Delta	26.2 (22.1, 30.4) %	0.12	·	0.56
Omicron	26.7 (21.7, 32.1) %	0.13	0.44	·
Intermediate				
Alpha	16.6 (14.1, 19.2) %	·	0.87	0.999
Delta	18.1 (17.6, 18.6) %	0.13	·	$> 1 - \epsilon$
Omicron	20.9 (20.3, 21.5) %	0.001	$< \epsilon$	·
Vaccine first				
Alpha	15.3 (13.7, 16.9) %	·	0.02	0.01
Delta	24.9 (16.8, 33.5) %	0.98	·	0.953
Omicron	6.3 (0.000, 25.0) %	0.99	0.047	·

Table: Numerical values for the SAR with confidence invervals and p-values, stratified by vaccine status. ϵ represents the minimum value that can be reliably estimated from 2×10^4 bootstrap samples, i.e. below 0.001.

Role of the household in wider transmission



‘Tranche’ work

- Presentation is mainly on: **T. House**, H. Riley, L. Pellis, K. B. Pouwels, S. Bacon, A. Eidukas, K. Jahanshahi, R. M. Eggo, A. S. Walker, “Inferring Risks of Coronavirus Transmission from Community Household Data.” To appear in Statistical Methods in Medical Research [arXiv:2104.04605].
- Methodology developed in arXiv:1911.12115: T. M. Kinyanjui and **T. House**, “Generalised Linear Models for Dependent Binary Outcomes with Applications to Household Stratified Pandemic Influenza Data.”
- Based on application of Ball/Addy-type final size equations, justified by splitting the data into the following tranches, with associated time periods and notable events. Code and updates at: https://github.com/thomasallanhouse/covid19-housefs
- This allows us to deal with the study design in a reasonably principled way, but there is room for improvement!

- **Tranche 1:** 26 April 2020 to 31 August 2020; low prevalence; schools closed; Alpha and Delta variants not emerged yet; no vaccine available.
- **Tranche 2:** 1 September 2020 to 14 November 2020; high prevalence; schools open; negligible Alpha variant; Delta variant not emerged yet; no vaccine available.
- **Tranche 3:** 15 November 2020 to 31 December 2020; high prevalence; schools open; Alpha variant becomes dominant; Delta variant not emerged yet; negligible vaccine coverage.
- **Tranche 4:** 1 January 2021 to 14 February 2021; high prevalence; schools closed (except for pre-school); Alpha variant dominant; Delta variant not emerged yet; over 10 million first vaccine doses by end of time period.
- **Tranche 5:** 15 February 2021 to 29 April 2021; low prevalence; schools open; Delta variant negligible; over 35 million first and 15 million second vaccine doses by end of time period.
- **Tranche 6:** 30 April 2021 to 15 July 2021; high prevalence; schools open; Delta variant becomes dominant; over 45 million first and 35 million second doses distributed by end of time period.

Explanatory Variables Considered

The point of the split into tranches is to leave the variables of interest well approximated as fixed during the time period in question:

- Swab positivity
- Household size
- The Alpha and Delta variant (identified via S-gene target failure and time)
- Age of participant
- Work in patient-facing roles

These variables are distributed in the sample as in the table below.

‘Table 1’ – Sample properties (1/2)

	Tranche 1	Tranche 2	Tranche 3				
Number of participants	89624	293570	315187				
Number of households	43300	144904	157432				
Number of positive individuals	242	5625	6078				
Households with $1+$ positive	206	4074	4433				
Children $<$12	7483	23257	24045				
Children 12–16	4814	15503	16790				
OR+N+S positives	124	4051	2263				
OR+N positives	12	547	2535				
Patient-facing participants	3335	9464	10046				

‘Table 1’ – Sample properties (2/2)

	Tranche 4	Tranche 5	Tranche 6	Overall
Number of participants	329532	343821	351879	408278
Number of households	165238	171809	178955	200876
Number of positive individuals	6925	1440	1890	23392
Households with $1+$ positive	5123	1071	1506	17180
Children $<$12	24686	25408	25050	32307
Children 12–16	18098	19012	19294	22250
OR+N+S positives	695	33	1382	9543
OR+N positives	4353	1036	244	8842
Patient-facing participants	10069	11103	11437	15213

Setup

Suppose we have a set of n individuals (participants), indexed $i, j, \ldots \in [n]$, where we use the notation $[k]$ to stand for the set of integers from 1 to k inclusive. These individuals are members of m households, and we represent the a-th household using a set of individual indices H_a. These are specified such that each individual is in exactly one household, so formally,

$$H_a \subseteq [n], \forall a \in [m]\,, \quad H_a \cap H_b = \varnothing, \forall a \in [m], b \in [m] \setminus \{a\}\,,$$

$$\bigcup_{a=1}^{m} H_a = [n]\,.$$

The size of the a-th household is then $n_a = |H_a|$. We let $\mathbf{x}_i$ be the length-p feature vector (also called covariates) associated with the i-th individual, and y_i be the test result so that $y_i = 1$ if the swab is positive and $y_i = 0$ if not.

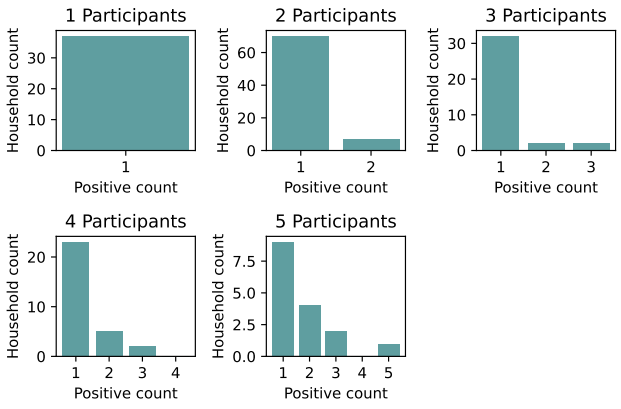
Exploratory analysis – histograms

- Before jumping in to modelling (as I teach MSc students!) we should do an exploratory analysis of the data.
- The heights of the histogram bars are given by

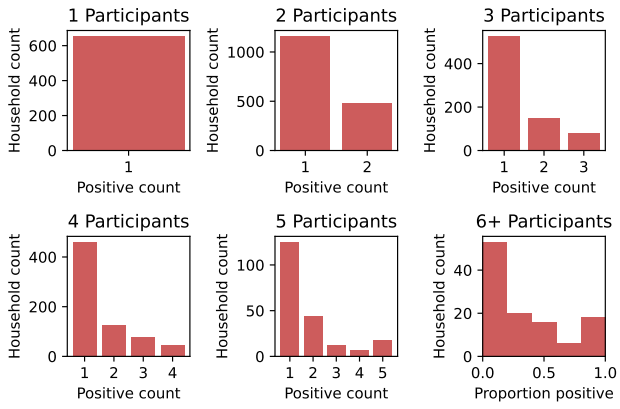
$$Z_{k,\ell} = \sum_{a=1}^{m} \mathbb{1}_{\{n_a=\ell\}} \mathbb{1}_{\{\sum_{i\in H_a} y_i=k\}} ,$$

$$k \in \{2,3,4,5,6\} , \quad \ell \in \{0,\ldots,k\} ,$$

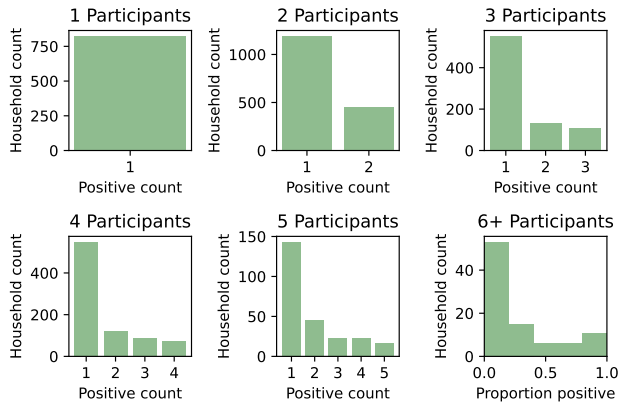
  where $\mathbb{1}$ stands for the indicator function.
- Verbally, $Z_{k,\ell}$ is the count of households of size ℓ with k participants testing positive.



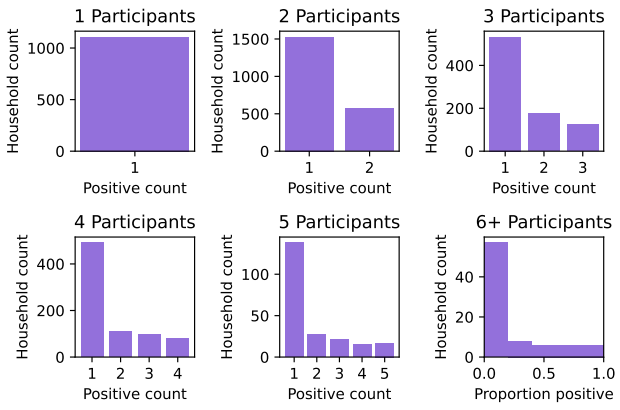
(g) Tranche 1



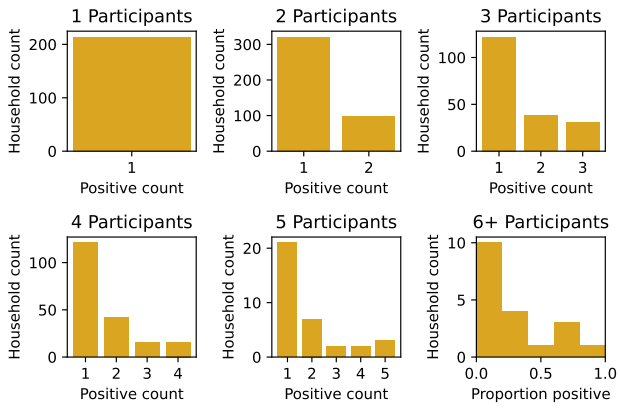
(h) Tranche 2



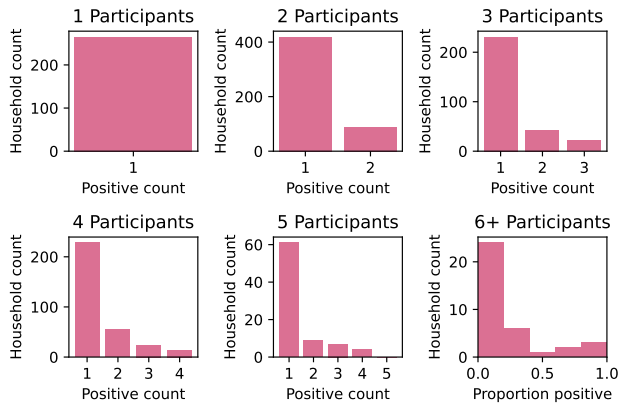
(i) Tranche 3



(j) Tranche 4



(k) Tranche 5



(l) Tranche 6

Figure: Histograms of household attack rates

Exploratory analysis – Density plots

- The density plots are obtained by considering some feature (in this case, age) that takes values 0 or 1. We then form a point $\mathbf{r}_a \in [0,1]^2$ for each household H_a such that

$$\sum_{i \in H_a} \mathbb{1}_{\{y_i=1\}} > 0\,, \quad \sum_{i \in H_a} \mathbb{1}_{\{x_i=1\}} > 0\,, \quad \sum_{i \in H_a} \mathbb{1}_{\{x_i=0\}} > 0\,,$$

  through the definition

$$\mathbf{r}_a = \left(\frac{\sum_{i \in H_a} \mathbb{1}_{\{y_i=1 \& x_i=1\}}}{\sum_{i \in H_a} \mathbb{1}_{\{x_i=1\}}}, \frac{\sum_{i \in H_a} \mathbb{1}_{\{y_i=1 \& x_i=0\}}}{\sum_{i \in H_a} \mathbb{1}_{\{x_i=0\}}} \right) .$$

- Then we can construct a kernel density estimate in the usual way by summing then normalising kernel functions around the points, in particular the width-w square kernel function

$$\mathcal{K}(\mathbf{r}, \mathbf{r}_a) = \mathbb{1}_{\{||\mathbf{r}-\mathbf{r}_a||_\infty < w\}} \,.$$

- We use age (16 years old and under versus over 16 years old) as the feature in making the density plots below.

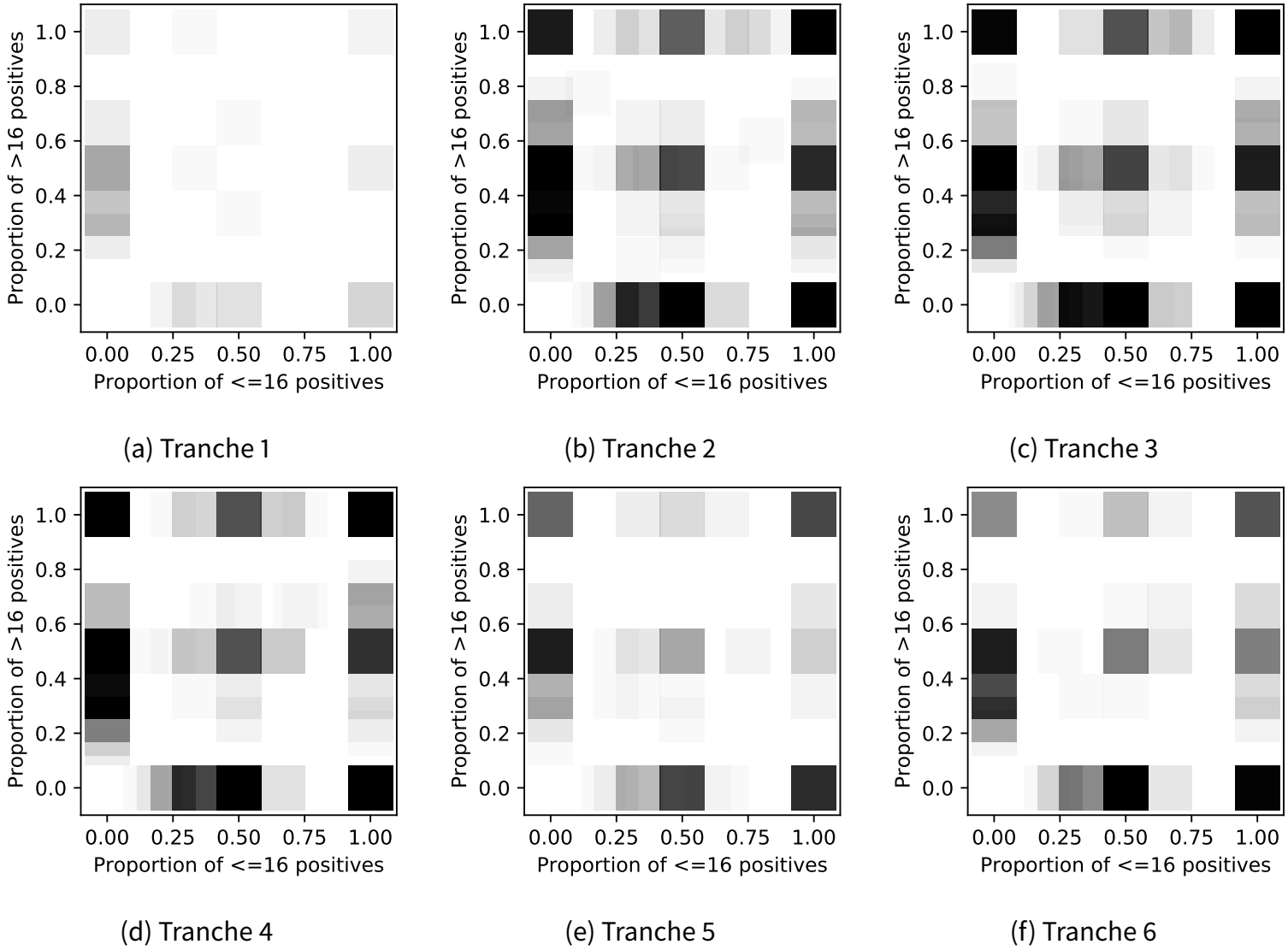


Figure: Density plots

Residual analysis

- Pearson residuals let us tabulate features and positives in households in a manner that allows their clustering to be assessed.
- Let x_i be the feature for individual i that takes values with generic labels $A, B, \ldots$ in particular PCR gene patterns.
- We are then interested in the table of pairs of individuals in households in the set $\mathcal{H} \subseteq [m]$ with certain properties,

$$Y_{AB} = \sum_{a \in \mathcal{H}, i \in H_a, j \in H_a \setminus \{i\}} \mathbb{1}_{\{x_i = A\}} \mathbb{1}_{\{x_j = B\}} \cdot$$

- Verbally, Y_{AB} is the count in the sample of A-B pairs in the set of households.

Residual analysis

- The null hypothesis to compare to is independent state probabilities $\boldsymbol{\pi} = (\pi_A)$ with MLE

$$\hat{\pi}_A = \frac{1}{n} \sum_{i \in [n]} \mathbb{1}_{\{x_i = A\}} \, .$$

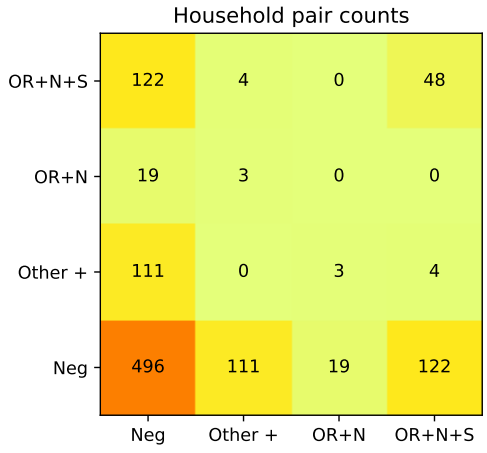
- The expected table under the null is

$$M_{AB} = \pi_A \pi_B \sum_{a \in \mathcal{H}} n_a (n_a - 1) \, .$$

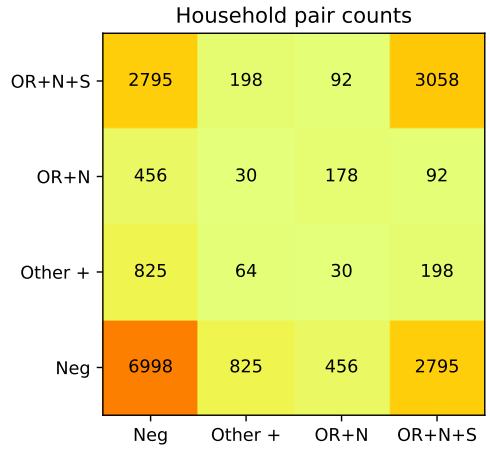
- And the Pearson residual associated with the (A, B)-th table entry is

$$R_{AB} = \frac{Y_{AB} - M_{AB}}{\sqrt{M_{AB}}} \, .$$

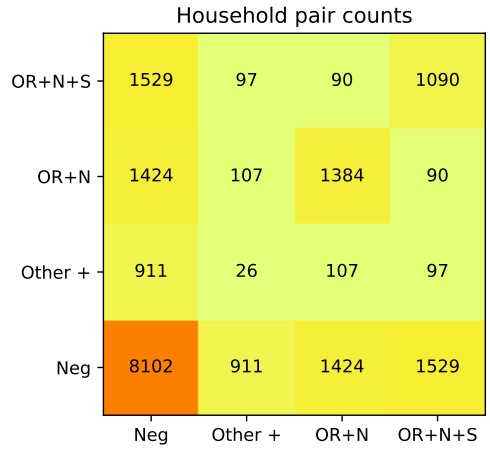
- While these residuals will be zero at the null and have values indicative of the corellations compared to the null, we do not have results for their asymptotic distribution yet.



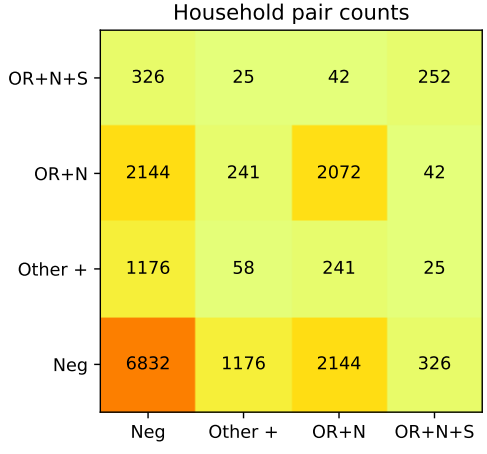
(g) Tranche 1



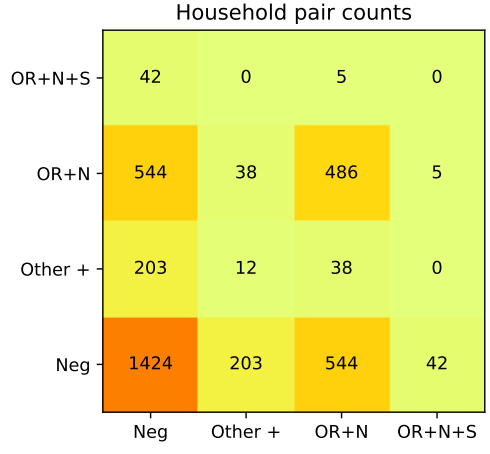
(h) Tranche 2



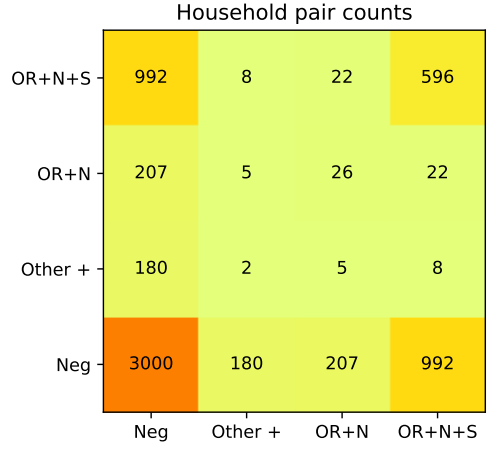
(i) Tranche 3



(j) Tranche 4

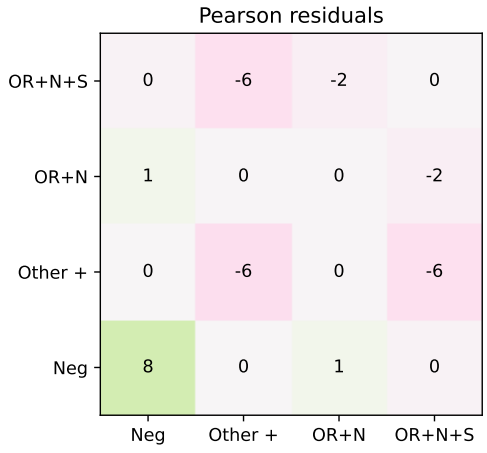


(k) Tranche 5

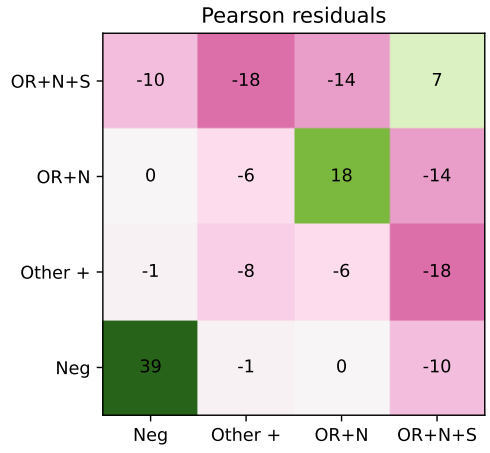


(l) Tranche 6

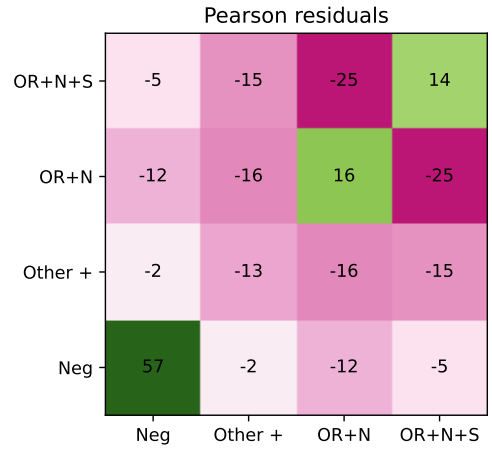
Figure: Pair counts



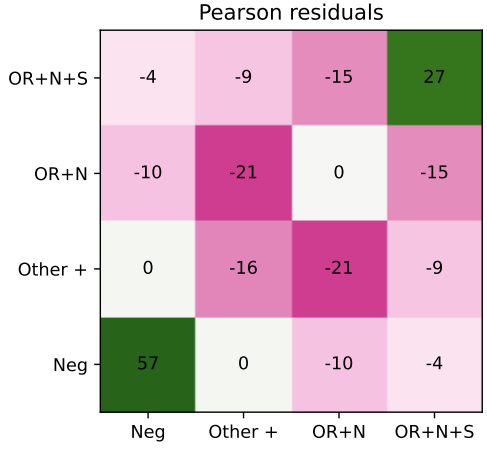
(a) Tranche 1



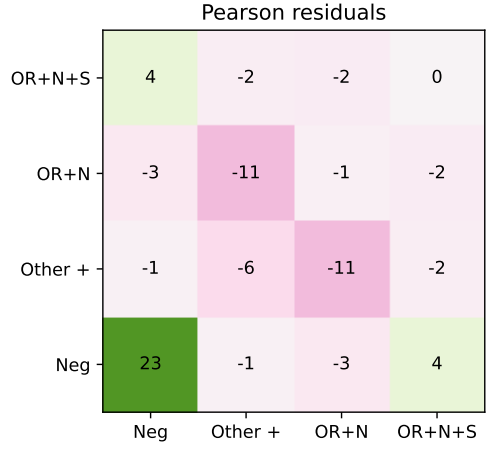
(b) Tranche 2



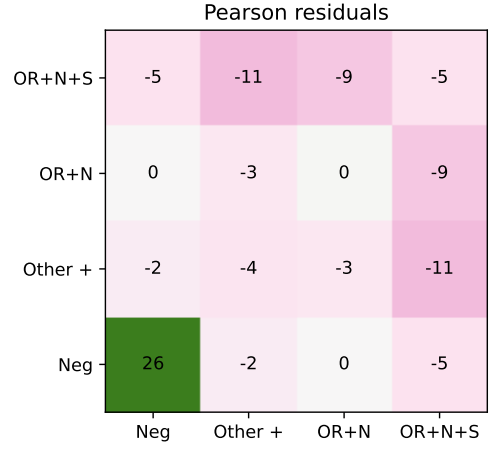
(c) Tranche 3



(d) Tranche 4



(e) Tranche 5



(f) Tranche 6

Figure: Residual Plots

Sellke construction

We are now going to think about how to model the within-household epidemic, which starts with the Sellke construction.

- We suppose that each individual i has a stochastic variable T_i for its infectious period, picked from the infectious period distribution, and that susceptible individuals have a random threshold $\Xi_i \sim \mathrm{Exp}(1)$.
- The individual then becomes infectious when their threshold is exceeded by the total force of infection they have experienced. To see why this is equivalent to the standard definition, consider

$$\begin{aligned}\Pr(\Xi > \Lambda(t+\delta t) | \Xi > \Lambda(t)) &= \frac{\int_0^{\Lambda(t+\delta t)} \exp(-\xi)\mathrm{d}\xi}{\int_0^{\Lambda(t)} \exp(-\xi)\mathrm{d}\xi} \\ &= 1 - \Lambda(t)\delta t + o(\delta t)\,.\end{aligned}$$

Final size equations

We will now write down the relevant equations for a household H of size n with outcome vector $\mathbf{y}$ and feature matrix $\boldsymbol{X}$ (i.e. suppressing the household index a to simplify notation). In particular, given a map $\iota : \{0,1\}^n \to \{1,\ldots,2^n\}$, we will be able to form the vector $\mathbf{P} = (P_{\iota(\mathbf{y})})_{\mathbf{y}\in\{0,1\}^n}$ of probabilities of different outcomes in the household. This will be a solution to the set of linear equations

$$\boldsymbol{B}(\boldsymbol{\theta})\mathbf{P} = \mathbf{1}\,, \tag{2}$$

where $\mathbf{1}$ is a length-2^n vector of all ones, and $\boldsymbol{B} = [B_{\iota(\boldsymbol{\nu}),\iota(\boldsymbol{\omega})}]_{\boldsymbol{\nu},\boldsymbol{\omega}\in\{0,1\}^n}$, which has

$$B_{\iota(\boldsymbol{\nu}),\iota(\boldsymbol{\omega})} = \mathcal{B}_{\boldsymbol{\nu},\boldsymbol{\omega}} = \frac{1}{\prod_{j\in H}\Phi\left(\sum_{i\in H}(1-\nu_i)\lambda_{ij}\right)^{\omega_j} Q_j^{1-\nu_j}}\,,$$

$\boldsymbol{\nu} \leq \boldsymbol{\omega} \in \{0,1\}^n$, and other elements equal to zero, where we write $\leq$ between vectors to stand for the statement that each element on the left-hand side is less than or equal to the corresponding element on the right-hand side.

Final size equations

The first model component is the probability of avoiding infection from outside; for the i-th individual this is

$$Q_i = \mathrm{e}^{-\Lambda_i} \,, \qquad \Lambda_i = \Lambda \mathrm{e}^{\boldsymbol{\alpha}\cdot\mathbf{x}_i} = \mathrm{e}^{\alpha_0 + \boldsymbol{\alpha}\cdot\mathbf{x}_i} \,.$$

In the language of infectious disease modelling, Λ_i is the cumulative force of infection experienced by the i-th individual. Then $\exp(\alpha_k)$ is the relative external exposure associated with the k-th feature / covariate, meaning that it is the multiplier in front of the baseline force of infection, which is that for an individual whose feature vector is all zeros, $\mathbf{0}$. This baseline probability of avoiding infection from outside is then

$$q = \exp(-\Lambda) = \exp(-\exp(\alpha_0)) \,, \tag{3}$$

and we will report $(1 - q)$ in tables, alongside the relative external exposures that are elements of the vector $\boldsymbol{\alpha}$, although it would also be possible to use (3) to relate this to the baseline force of infection Λ or intercept of the linear predictor, α_0.

Final size equations

The second component of the model is variability in the infectiousness at the individual level, usually interpreted as arising from the distribution of infectious periods. We assume that each individual picks from a unit-mean Gamma distribution since this provides a natural one-parameter distribution with appropriate support. The Laplace transform of this is used and is

$$\Phi(s) = (1 + \vartheta s)^{-1/\vartheta} .$$

The parameter ϑ is the variance of the Gamma distribution, i.e. it is larger for more individual variability. To see why the Laplace transformation is appropriate, consider the Sellke construction and assume a baseline rate of infection, λ, to be multiplied by infectious duration T to give total force of infection $\Lambda = \lambda T$, so

$$\Pr(\Xi > \Lambda) = \int_0^\infty F_\Xi(\lambda t) f_T(t) \mathrm{d}t = \int_0^\infty \mathrm{e}^{-\lambda t} f_T(t) \mathrm{d}t = \mathcal{L}[f_T](\lambda) .$$

Final size equations

The third component of the model is the infection rate from individual j to individual i,

$$\lambda_{ij} = n^{\eta}\lambda\sigma_i\tau_j = n^{\eta}\lambda \mathrm{e}^{\boldsymbol{\beta}\cdot\mathbf{x}_i}\mathrm{e}^{\boldsymbol{\gamma}\cdot\mathbf{x}_j} = \mathrm{e}^{\boldsymbol{\beta}\cdot\mathbf{x}_i}\mathrm{e}^{\gamma_0+\eta\log(n)+\boldsymbol{\gamma}\cdot\mathbf{x}_j}\,. \quad (4)$$

In this equation: λ is the baseline rate of infection; $\sigma_i = \mathrm{e}^{\boldsymbol{\beta}\cdot\mathbf{x}_i}$ is the relative susceptibility of the i-th participant, and $\exp(\beta_k)$ is the relative susceptibility associated with the k-th feature; $\tau_j = \mathrm{e}^{\boldsymbol{\gamma}\cdot\mathbf{x}_j}$ is the relative transmissibility of the j-th participant, and $\exp(\gamma_k)$ is the relative transmissibility associated with the k-th feature / covariate. As can be seen from (4), we can interpret $\log(\lambda)$ as the intercept of the linear predictor for transmissibility. The term n^{η} is a modelling approach to the effect of household size usually attributed to Cauchemez; as can be seen from (4), this is equivalent to taking $\log(n)$ as a covariate for transmissibility.

Likelihood function and fitting

- We can then produce a likelihood for the data from the probability model.
- This will take the form of a product of probabilities derived from solving the Ball equations (2).
- Actually fitting this model to $> 3 \times 10^6$ observations on a secure environment is non-trivial, and involves a significant numerical linear algebra computational effort.
- For the results here, NumPy was sufficient, but we are experimenting with implementation in Numba.
- We carried out approximate Bayesian inference.
- This was done using the Laplace approximation and a standard normal prior on each parameter.
- Multi-restart numerical optimisation using a Quasi-Newton method was used.

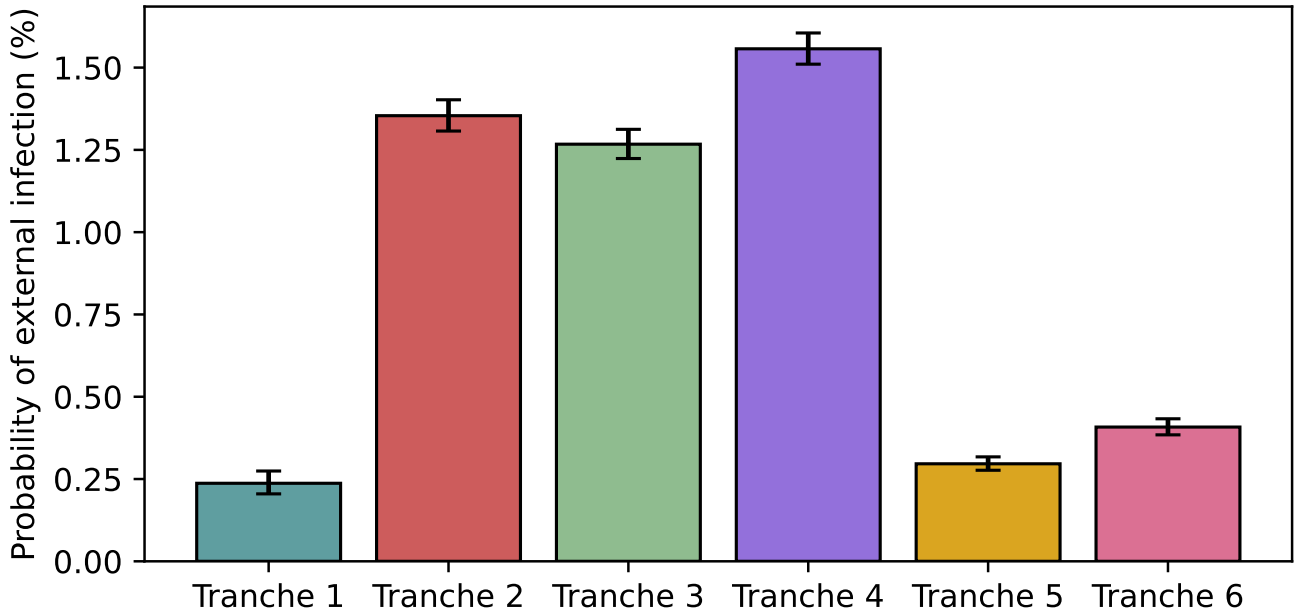
Results – ‘Table 2’; Point estimates and CrIs (1/2)

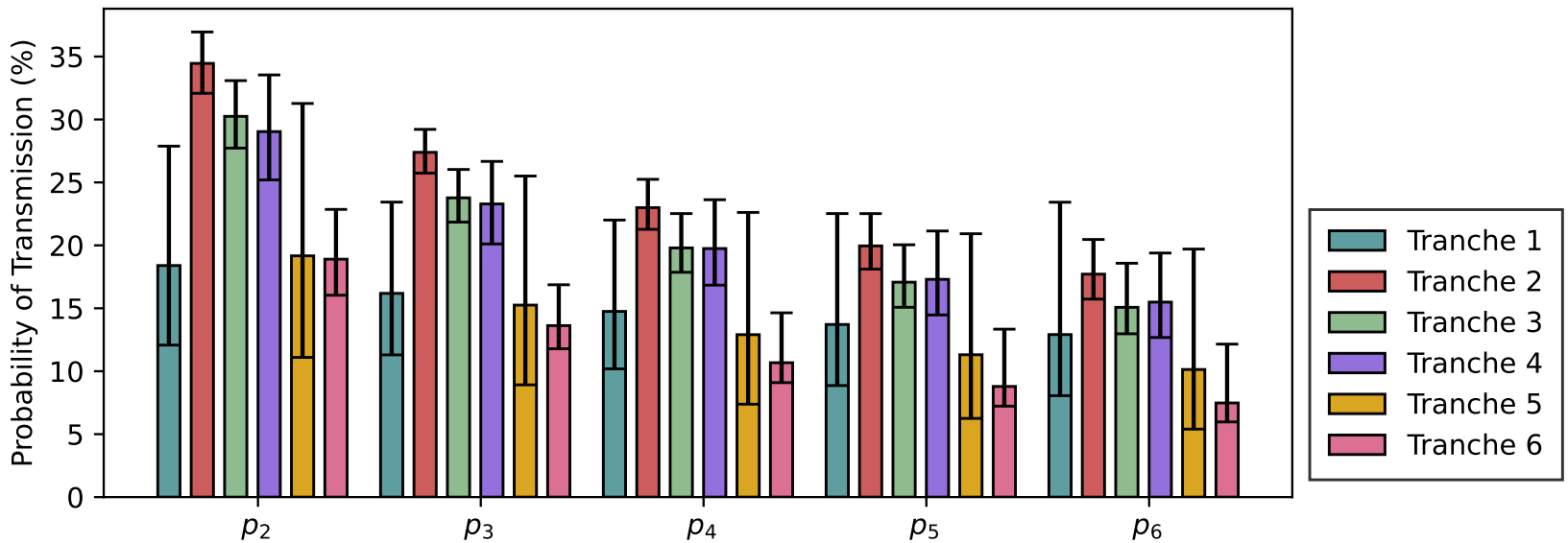
	Tranche 1	Tranche 2	Tranche 3			
$1-q$	0.237 (0.205,0.274) %	1.35 (1.31,1.4) %	1.27 (1.22,1.31) %			
p_2	18.4 (12.1,27.9) %	34.5 (32.1,37.0) %	30.2 (27.7,33.1) %			
p_3	16.2 (11.3,23.4) %	27.4 (25.7,29.2) %	23.8 (21.8,26.0) %			
p_4	14.8 (10.2,22.0) %	23.0 (21.3,25.2) %	19.8 (17.9,22.5) %			
p_5	13.7 (8.86,22.5) %	20.0 (18.1,22.5) %	17.1 (15.1,20.0) %			
p_6	12.9 (8.06,23.4) %	17.7 (15.7,20.5) %	15.1 (13.0,18.6) %			
$\exp(\alpha_{\text{2-11}})$	0.883 (0.525,1.49)	0.845 (0.723,0.987)	1.39 (1.23,1.56)			
$\exp(\alpha_{\text{12-16}})$	0.546 (0.26,1.15)	1.64 (1.44,1.87)	2.35 (2.1,2.63)			
$\exp(\alpha_{\text{PF}})$	2.93 (1.91,4.49)	1.26 (1.06,1.49)	1.61 (1.38,1.87)			
$\exp(\beta_{\text{2-11}})$	0.984 (0.393,2.46)	0.824 (0.636,1.07)	0.865 (0.7,1.07)			
$\exp(\beta_{\text{12-16}})$	0.786 (0.298,2.07)	0.778 (0.578,1.05)	0.872 (0.68,1.12)			
$\exp(\gamma_{\text{2-11}})$	0.922 (0.266,3.2)	0.715 (0.476,1.07)	0.824 (0.593,1.15)			
$\exp(\gamma_{\text{12-16}})$	0.815 (0.237,2.8)	0.771 (0.542,1.1)	0.662 (0.488,0.899)			
$\exp(\gamma_{\text{OR+N}})$	0.576 (0.199,1.67)	0.572 (0.447,0.731)	1.52 (1.33,1.75)			
$\exp(\gamma_{\text{CT-oth}})$	0.157 (0.062,0.398)	0.097 (0.0626,0.15)	0.0926 (0.0607,0.141)			

Results – ‘Table 2’; Point estimates and CrIs (2/2)

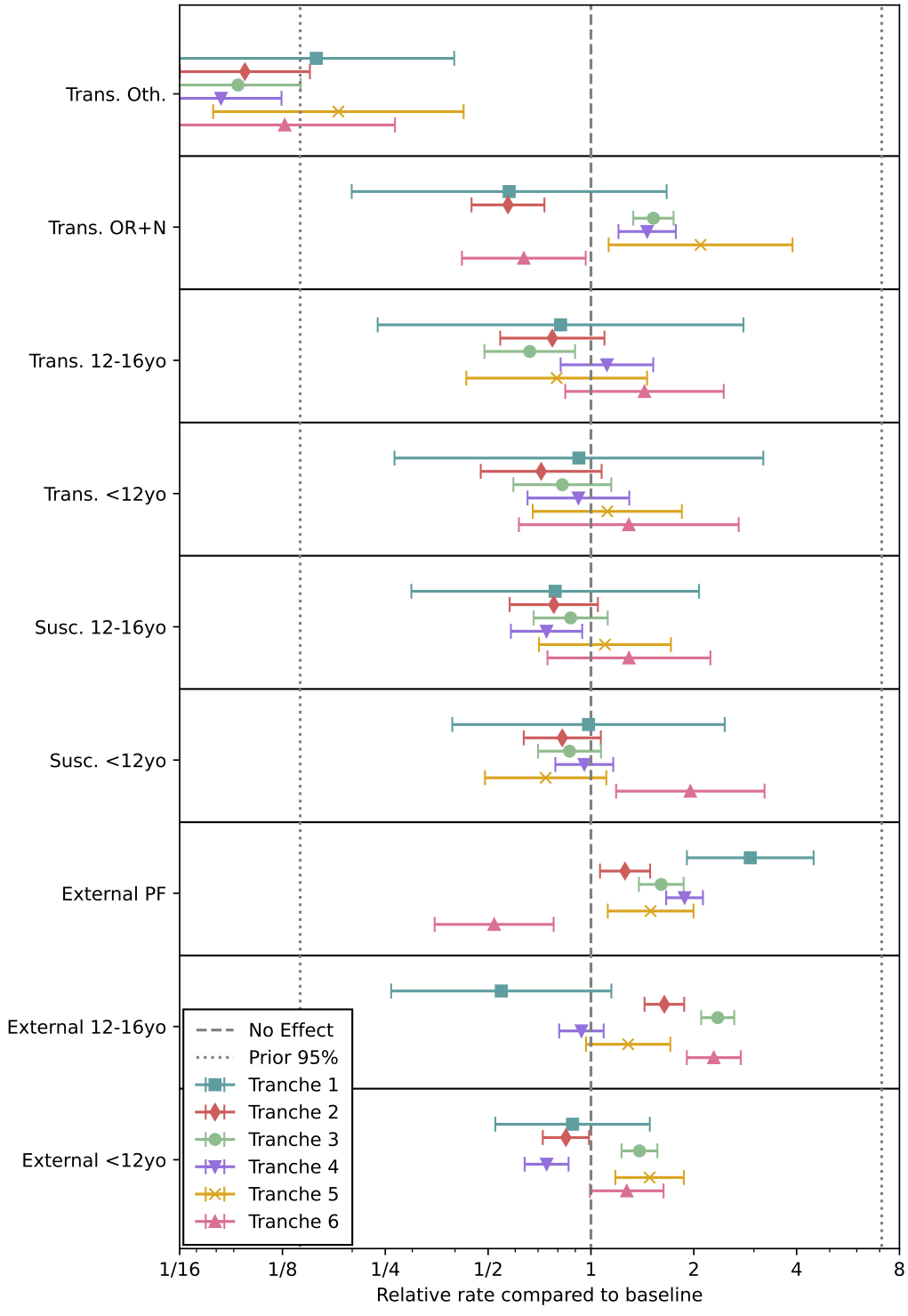
	Tranche 4	Tranche 5	Tranche 6
$1-q$	1.56 (1.51,1.61) %	0.296 (0.277,0.317) %	0.408 (0.384,0.433) %
p_2	29.0 (25.2,33.5) %	19.2 (11.1,31.3) %	18.9 (16.0,22.9) %
p_3	23.3 (20.1,26.7) %	15.3 (8.92,25.5) %	13.6 (11.8,16.9) %
p_4	19.7 (16.8,23.6) %	12.9 (7.38,22.6) %	10.7 (9.09,14.6) %
p_5	17.3 (14.5,21.1) %	11.3 (6.25,20.9) %	8.79 (7.22,13.3) %
p_6	15.5 (12.7,19.4) %	10.1 (5.4,19.7) %	7.48 (5.97,12.2) %
$\exp(\alpha_{\text{2-11}})$	0.742 (0.64,0.86)	1.48 (1.18,1.87)	1.27 (0.993,1.63)
$\exp(\alpha_{\text{12-16}})$	0.938 (0.807,1.09)	1.29 (0.967,1.71)	2.29 (1.91,2.74)
$\exp(\alpha_{\text{PF}})$	1.88 (1.66,2.13)	1.5 (1.12,2.0)	0.521 (0.349,0.778)
$\exp(\beta_{\text{2-11}})$	0.956 (0.787,1.16)	0.737 (0.49,1.11)	1.95 (1.18,3.22)
$\exp(\beta_{\text{12-16}})$	0.741 (0.583,0.943)	1.1 (0.704,1.71)	1.29 (0.746,2.24)
$\exp(\gamma_{\text{2-11}})$	0.919 (0.652,1.29)	1.12 (0.676,1.85)	1.29 (0.615,2.71)
$\exp(\gamma_{\text{12-16}})$	1.11 (0.815,1.52)	0.794 (0.432,1.46)	1.43 (0.841,2.45)
$\exp(\gamma_{\text{OR+N}})$	1.46 (1.2,1.77)	2.09 (1.13,3.89)	0.636 (0.419,0.965)
$\exp(\gamma_{\text{CT-oth}})$	0.0826 (0.055,0.124)	0.182 (0.0783,0.424)	0.127 (0.0604,0.267)

Results – Overall infection from outside & Within-household infection



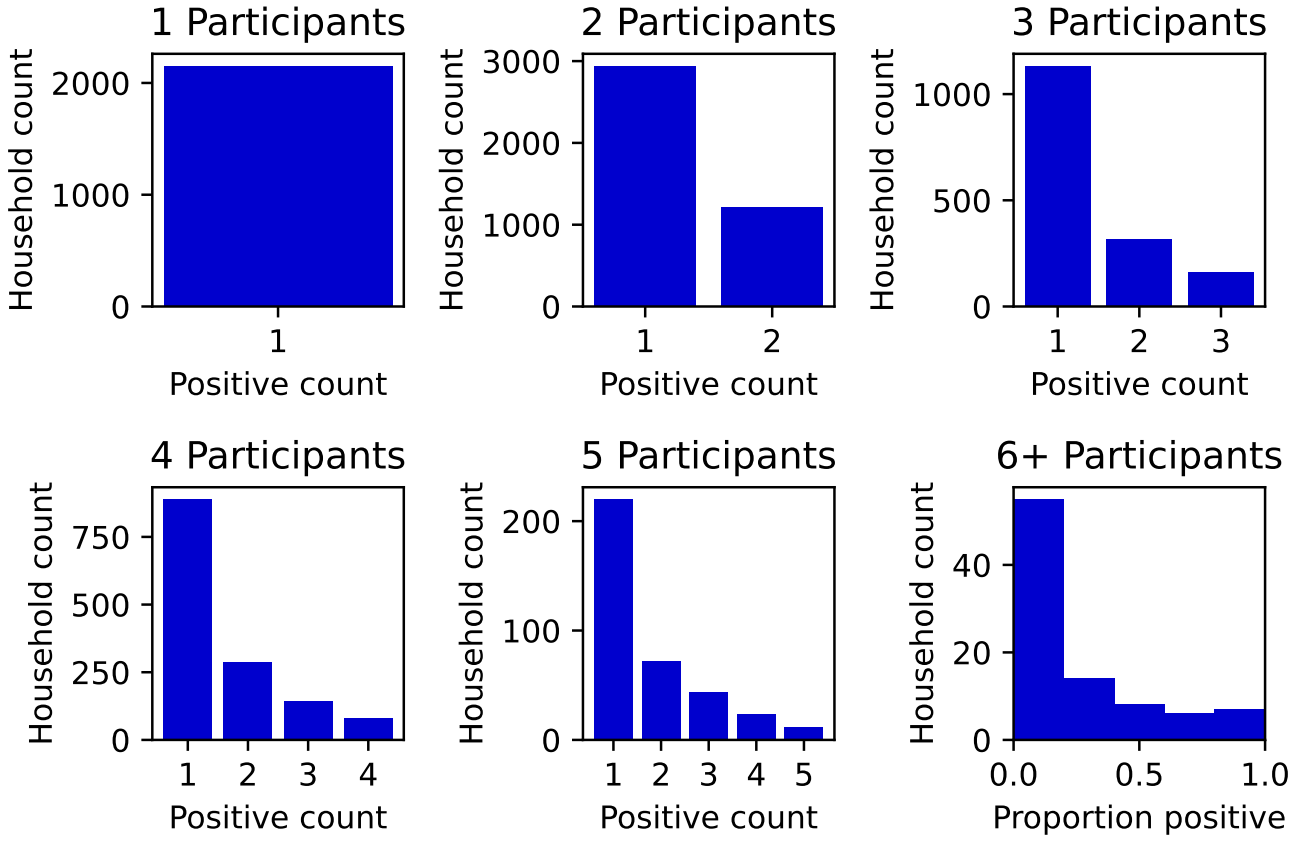


Results – Effect sizes

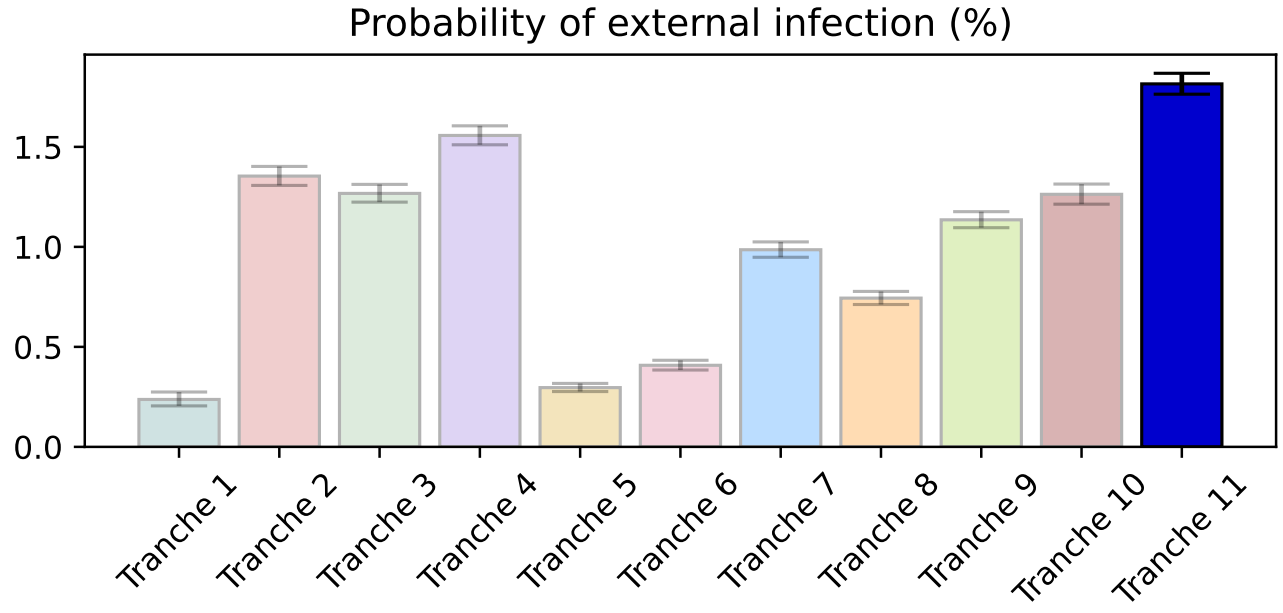


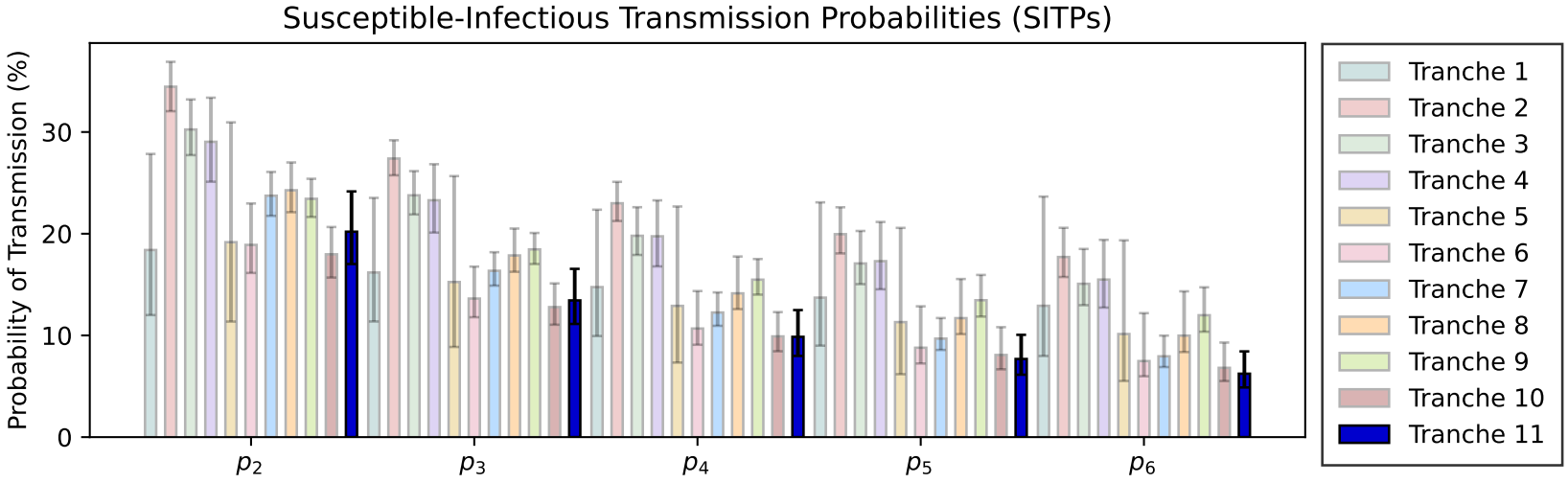
Latest Update – Tranche 11

- **Tranche 11:** 7 January 2022 to 6 February 2022; high prevalence; schools open; Omicron BA.1 lineage dominant and Omicron BA.2 lineage emerging; over 52 million first, 48 million second and 37 million third doses distributed by end of time period.



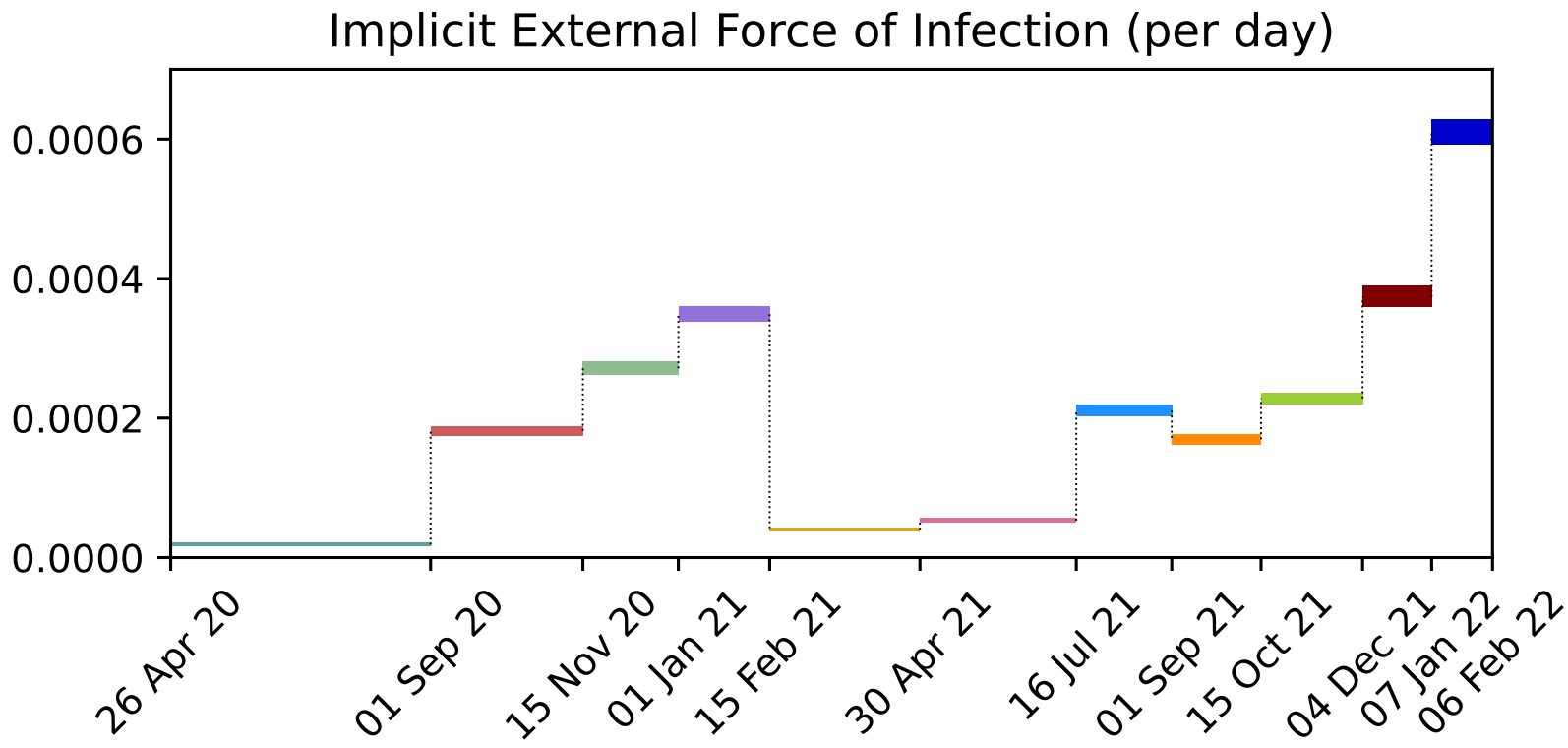
Update for Tranche 11 – Overall infection from outside & Within-household infection



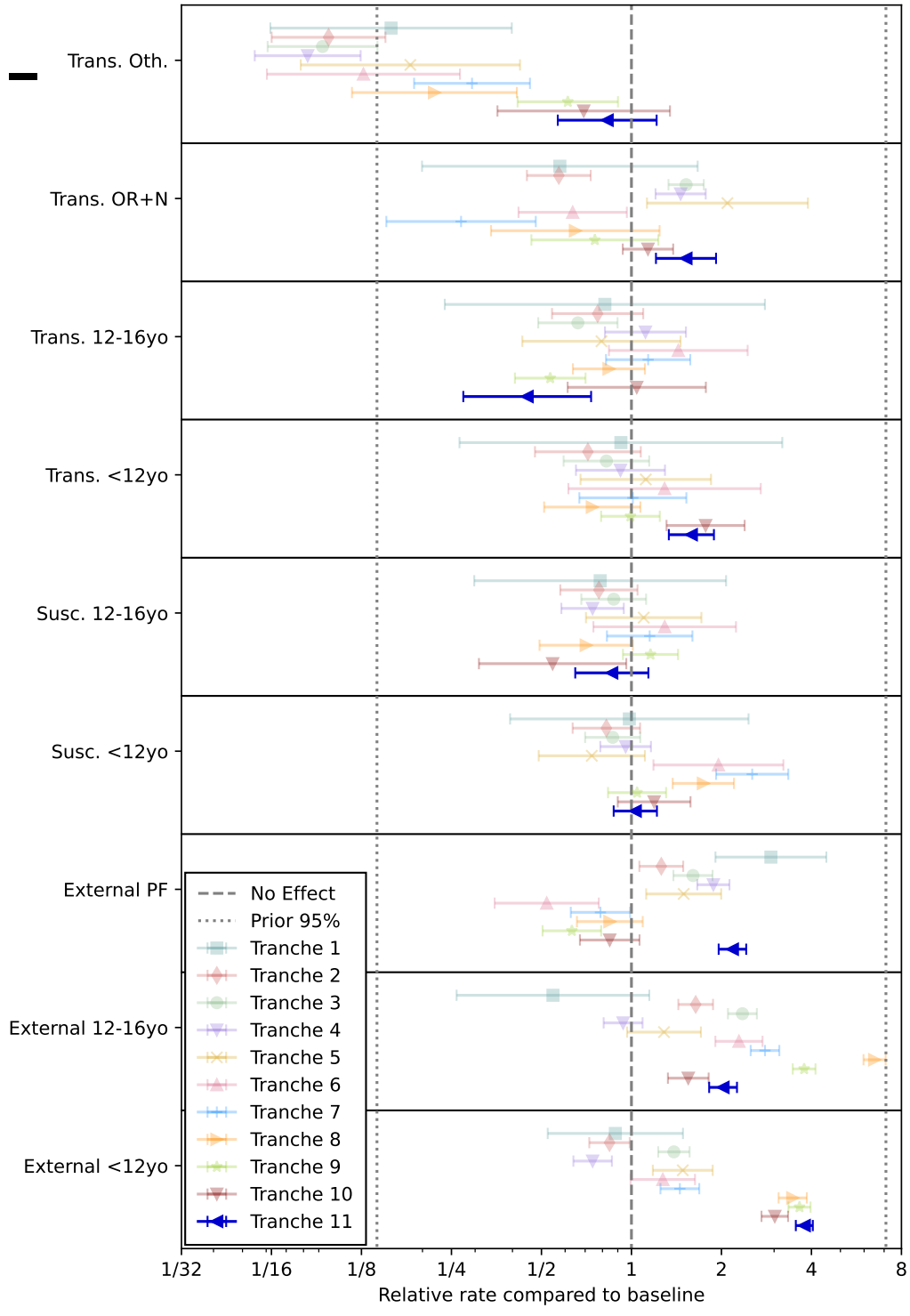


New for Tranche 11 – Infection from outside by time

Previously, we considered the baseline probability of avoiding infection from outside the household, $q = \mathrm{e}^{-\Lambda}$, where Λ is a quantity known as the *cumulative force of infection* in infectious disease epidemiology. Formally, if the force of infection λ is constant over a time period from time 0 to time τ, then $\Lambda = \int_{t=0}^{\tau} \lambda \, \mathrm{d}t = \tau\lambda$. We can then obtain $\lambda = \frac{-\log(q)}{\tau}$. By quoting this value, at low numerical values approximately equal to the daily probability of infection from outside the household, we can adjust for the duration of different time periods considered.



Update for Tranche 11 – Effect sizes



Acknowledgements

- The ONS CIS team
- Manchester Mathematical Epidemiology group
- JUNIPER consortium
- Funders: Royal Society; Alan Turing Institute; UKRI; DHSC
- SAGE and in particular SPI-M members and secretariats
- Many others!

Thanks for your time!