

1 Estimation of the infection attack rate of mumps in an outbreak among 2 college students using paired serology

3 Michiel van Boven^{a,b,*}, Jantien A. Backer^a, Irene Veldhuijzen^a, Justin Gomme^{c,d},
4 Rob van Binnendijk^a, Patricia Kaaijk^a

5 ^a*Centre for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven, the*
6 *Netherlands*

7 ^b*Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University,*
8 *Utrecht, the Netherlands*

9 ^c*Department of Epidemiology and Social Medicine, University of Antwerp, Antwerp, Belgium*

10 ^d*NHS Scotland, Edinburgh, Scotland*

11 Abstract

Mumps virus is a highly transmissible pathogen that is effectively controlled in countries with high vaccination coverage. Nevertheless, outbreaks have occurred worldwide over the past decades in vaccinated populations. Here we analyse serological data from outbreaks of mumps virus genotype G among college students in the Netherlands over the period 2009-2012. To identify infections in the presence of preexisting antibodies we compared mumps specific serum IgG concentrations in two consecutive samples ($n = 746$), whereby the first sample was taken when students started their study prior to the outbreaks, and the second sample was taken 2-5 years later. We fit a binary mixture model to the data. The two mixing distributions represent uninfected and infected classes. Throughout we assume that the infection probability increases with the ratio of antibody concentrations of the second to first sample. The estimated infection attack rate is higher than reported earlier (0.095 versus 0.042). The analyses yield probabilistic classifications of participants, which are mostly quite precise owing to the high intraclass correlation in uninfected participants (0.85, 95%CrI: 0.82 – 0.87). The estimated probability of infection increases with decreasing antibody concentration in the pre-outbreak sample, such that the probability of infection is 0.12 (95%CrI: 0.10 – 0.13) for the lowest quartile of the pre-outbreak samples and 0.056 (95%CrI: 0.044 – 0.068) for the highest quartile. We discuss the implications of these insights for the design of booster vaccination strategies.

*Corresponding author

12 Highlights

- 13 • We use paired pre- and post-outbreak serological data to estimate mumps infection rates
14 in college students.
- 15 • We use a two-component mixture model to provide individual estimates of infection for
16 each participant.
- 17 • The estimated population infection attack rate is higher than reported earlier (9.5% vs 4.2%).
- 18 • The estimated individual probability of infection increases with decreasing pre-outbreak an-
19 tibody
20 concentrations, from 12% in the lowest to 5.6% in the highest quartile.

21 Introduction

22 Mass vaccination campaigns have been highly successful in reducing transmission and associated
23 morbidity of infectious diseases of childhood. Case in point is the Measles-Mumps-Rubella vaccine,
24 which in the Netherlands is administered at 14 months and 9 years and has a coverage of 90%–95%.
25 It is known, however, that the antibody response to the mumps component in the vaccine wanes
26 over time, and that IgG antibodies induced by the vaccine have relatively low avidity [1, 2, 3, 4].
27 Perhaps owing to this, mumps outbreaks have occurred worldwide over the past decades, mainly
28 in vaccinated adolescents and young adults. These outbreaks often occurred in close contact
29 settings (schools, households, parties), and were mostly caused by genotypes that are different
30 from the vaccine genotype [5, 6, 7, 8, 9, 10, 11, 12, 13]. For instance, the current vaccine used
31 in the Netherlands contains the Jeryl-Lynn strain (genotype A), and is genetically distant from
32 the recent outbreak strains (genotypes D and G) [14, 15]. This vaccine is also widely used in the
33 United States and other European countries.

34 Many but not all infections with mumps virus are asymptomatic or are associated with mild symp-
35 toms only, especially in vaccinated individuals [16]. This makes it difficult to assess the true extent
36 of virus circulation in an outbreak. In principle, reliable infection attack rates can be obtained
37 from measuring mumps-specific IgG antibody concentrations, because these generally increase after
38 mumps virus infection [17]. However, a challenge is that paired pre- and post-outbreak samples are
39 often not available, and that there are no generally agreed antibody concentrations that define re-
40 cent infection in a single sample. Two studies have shown that pre-outbreak mumps neutralisation

41 antibody concentrations in patients with mumps were generally lower than in persons who were
42 not infected with mumps virus during the outbreak [2, 18, 19]. However, it proved not possible
43 to define reliable cutoffs separating infected from uninfected persons, and separating patients with
44 clinical symptoms from person with asymptomatic or mild infection [2]. Hence, it is not straight-
45 forward to deduce who has been infected either clinically or sub-clinically, especially when only a
46 single serum sample is available, and it is not clear how the probability of infection depends on
47 pre-existing antibody concentrations.

48 In an earlier study, we measured mumps-specific IgG antibody concentrations in paired pre- and
49 post-outbreak samples from university students in the Netherlands, using a fluorescent bead-based
50 multiplex immunoassay [20, 21, 18]. In these studies, we calculated the proportions of symptomatic
51 and asymptomatic infections and determined infection attack rates and risk factors for mumps
52 virus infection using predefined criteria for infection. Specifically, participants would be classified
53 as infected if there was a fourfold increase of antibody concentrations from first to second sample
54 or high antibody concentration in the second sample. Further, to identify a correlate of protection,
55 mumps-specific IgG concentrations in pre-outbreak samples were compared between infected and
56 non-infected persons [22, 21, 23]. However, as antibody concentrations decay over time, especially
57 in the first months after an infection [4, 24, 18, 19], and as the second sample in our study may
58 have been taken several years after infection, we need analyses with less rigid criteria for infection.
59 Therefore, we will in this study use methods that do not rely on predefined criteria for infection,
60 but that derive informed probabilities of infection directly from the available serological data.

61 Specifically, to provide a probabilistic classification of participants we employ a two-component
62 binary mixture model in which the component distributions represent infected and uninfected
63 persons [25, 26, 27]. Usually, the mixing parameter in such analyses represents the prevalence or
64 probability that a person is infected. Instead of using a fixed population-level mixing parameter, we
65 here link the pre- and post-outbreak samples by making the biologically plausible assumption that
66 the probability of infection increases monotonically with the ratio of post- to pre-outbreak antibody
67 concentration. This enables estimation of the infection probability and associated uncertainty for
68 each participant in a manner that is optimally informed by the data.

69 **Methods**

70 *Study population and sample collection*

71 All pre-outbreak sera in this retrospective study are taken from first-year medical students. Re-
72 cruitment from this population was carried out from 2007-2010, and resulted in a study population
73 of 746 students with both a pre- and post-outbreak sample. The study has been approved by a
74 medical ethical committee (NL38042.041.11), and has been described in detail earlier [18]. Briefly,
75 a self-sampled dried blot spot sample, a questionnaire concerning vaccination history, risk factors,
76 and mumps symptoms has been obtained from each student. The serological criteria for mumps
77 virus infection in the previous study were a fourfold increase in IgG concentration combined with
78 at least an post-outbreak antibody concentration of 300 RU/ml [18]. A small subset of 16 par-
79 ticipants show a strong (more than fourfold) decrease in the antibody concentration from pre- to
80 post-outbreak sample. These could represent recent infections, as pre-outbreak samples had been
81 taken from 2007-2010, while the outbreak lasted from 2009-2012. We therefore perform analyses
82 of the full data in the body of the text, but also indicate how the results are affected if those 16
83 participants are left out.

84 *Mumps-specific IgG multiplex immunoassay*

85 Mumps virus-specific serum immunoglobulin G (hereafter called IgG) antibody concentrations
86 have been determined by a fluorescent bead-based multiplex immunoassay as described before,
87 using purified Jeryl-Lynn mumps vaccine strain as antigen [20, 28]. For analysis, all antibody
88 concentrations are log₂-transformed, and in the following the log-transformed concentrations will
89 be referred to as (antibody) titers.

90 *Mixture model for paired data*

91 Participants are assumed to be uninfected in the first sample, and either infected or uninfected
92 in the second sample. Here, we assume that these two classes are characterised by probability
93 distributions for the antibody titers, and are characterised by density functions f^{uninf} and f^{inf} .
94 Based on preliminary analyses, we assume that the titer distributions of the two classes are normally
95 distributed with parameters $\theta^{\text{uninf}} = \{\mu^{\text{uninf}}, \sigma^{\text{uninf}}\}$ and $\theta^{\text{inf}} = \{\mu^{\text{inf}}, \sigma^{\text{inf}}\}$. For our data, the
96 pre-outbreak samples provide information on the uninfected component probability distribution,
97 and the post-outbreak samples inform both the uninfected and infected component probability
98 distributions. In mixture models for cross-sectional data the probability of infection is usually given

99 by the prevalence weighted density of the infected component divided by the sum of the weighted
100 infected and uninfected components (e.g., [27, 29, 30]). Here we take an alternative approach
101 that makes use of the fact that data are paired, and in which we postulate that the probability
102 that a post-outbreak sample belongs to the infected component (the mixing parameter) increases
103 monotonically with the ratio of the post- versus pre-outbreak titers. Specifically, we propose a
104 two-parameter logistic infection function such that the mixing parameter for the i -th participant
105 ($i = 1, \dots, 746$), q_i^{inf} , is given by

$$q_i^{\text{inf}} = \frac{1}{1 + e^{-k(Y_i^{\text{post}} - Y_i^{\text{pre}} - x_0)}} . \quad (1)$$

In equation (1), Y_i^{pre} and Y_i^{post} denote the antibody titers in the pre- and post-outbreak samples, k and x_0 represent the steepness and 50% infection probability of the logistic function. Throughout, we use a transformation of the logistic function in terms of the probability that a sample belongs to the infected component at no increase in antibody concentration (d_0) and fourfold increase in the antibody concentration (d_2), as this yields parameters that can be interpreted more easily in biological terms. This, in turn, facilitates making informed choices for the parameter prior distributions. A straightforward calculation shows that

$$k = \frac{1}{2} (\log(d_0^{-1} - 1) - \log(d_2^{-1} - 1))$$
$$x_0 = -2 \frac{\log(d_0^{-1} - 1)}{\log(d_2^{-1} - 1) - \log(d_0^{-1} - 1)} .$$

106 Biological reasoning implies that one would expect $d_0 \approx 1$ and $d_2 \approx 1$ [20, 21]. Notice further that
107 q_i^{inf} only depends on the difference of the two measurements but not on the individual values of
108 Y_i^{pre} and Y_i^{post} . Also notice that a difference of 1 between antibody titers represents a twofold
109 increase, a difference of 2 a fourfold increase, etcetera.

110 To estimate parameters and take correlations between pre- and post-outbreak samples into account,
111 we estimate the unknown latent antibody titers of uninfected persons. Here we assume that the
112 latent titers, y_i^{uninf} , are drawn from a normal hyper distribution representing the unobserved true
113 titers of uninfected persons. Specifically we take

$$y_i^{\text{uninf}} \sim \mathcal{N}(\mu_{\text{pop}}^{\text{uninf}}, \sigma_{\text{pop}}^{\text{uninf}}) , \quad (2)$$

114 where $\mu_{\text{pop}}^{\text{uninf}}$ and $\sigma_{\text{pop}}^{\text{uninf}}$ are the population mean and variance of the distribution of latent antibody
115 concentrations in uninfected persons. Each measurement of uninfected persons provides a possibly

116 imperfect representation of the latent antibody concentration, such that, for instance, for pre-
117 outbreak samples Y_i^{pre} ($i = 1, \dots, 746$) we have

$$Y_i^{\text{pre}} \sim \mathcal{N}\left(y_i^{\text{uninf}}, \sigma_{\text{noise}}^{\text{uninf}}\right), \quad (3)$$

118 where $\sigma_{\text{noise}}^{\text{uninf}}$ represents the measurement noise. On the other hand, post-outbreak samples can be
119 either uninfected or infected, and therefore these samples are distributed according to the mixture
120 distribution

$$Y_i^{\text{post}} \sim \left(1 - q_i^{\text{inf}}\right) \mathcal{N}\left(y_i^{\text{uninf}}, \sigma_{\text{noise}}^{\text{uninf}}\right) + q_i^{\text{inf}} \mathcal{N}\left(\mu^{\text{inf}}, \sigma^{\text{inf}}\right). \quad (4)$$

121 Together, equations (1) - (4) specify the model. Notice that the pre-outbreak samples inform the
122 uninfected component distribution, and that the post-outbreak samples inform both the uninfected
123 and infected component distributions.

124 With the above model at hand we can calculate for each participant the probability of infection, p_i^{inf} ,
125 as the prevalence weighted density of the infected component divided by the sum of the weighted
126 infected and uninfected components, i.e.

$$p_i^{\text{inf}} = \frac{q_i^{\text{inf}} f_i^{\text{inf}}}{q_i^{\text{inf}} f_i^{\text{inf}} + (1 - q_i^{\text{inf}}) f_i^{\text{uninf}}}. \quad (5)$$

127 where we have suppressed the dependence of p_i^{inf} and q_i^{inf} on the antibody differences $Y_i^{\text{post}} - Y_i^{\text{pre}}$,
128 and the dependence of f_i^{uninf} and f_i^{inf} on post-outbreak antibody concentrations Y_i^{post} . Hence, the
129 probability of infection depends on both the mixing parameter and the mixing distributions.

Finally, to quantify correlations of antibody concentrations in pre- and post-outbreak samples of uninfected persons, we calculate the intraclass correlation r_{ICC} as

$$r_{\text{ICC}} = \frac{(\sigma_{\text{pop}}^{\text{uninf}})^2}{(\sigma_{\text{pop}}^{\text{uninf}})^2 + (\sigma_{\text{noise}}^{\text{uninf}})^2}.$$

130 The intraclass correlation will be low if measurement noise is high and samples in uninfected
131 persons are not strongly correlated, and it will be high if the reverse is true. Of course, whether
132 post-outbreak samples are actually infected is not known, but will be estimated.

133 *Prior distributions and inference*

134 Parameters are estimated in a Bayesian framework using Hamiltonian Monte Carlo, implemented
135 in Stan [31]. The main parameters to be estimated are the mean and standard deviation of the
136 hyper distribution of antibody titers in uninfected persons, $\mu_{\text{pop}}^{\text{uninf}}$ and $\sigma_{\text{pop}}^{\text{uninf}}$, the random noise

137 $\sigma_{\text{noise}}^{\text{uninf}}$, the mean and standard deviation of antibody titers in infected persons, μ^{inf} and σ^{inf} , and
138 the parameters defining the probability of infection when there is no increase in antibody titer
139 from pre- to post-outbreak samples, and when there is a fourfold increase, d_0 and d_2 .

140 Prior distributions for the means and standard deviations of the uninfected component distribution
141 are based on pre-outbreak samples as these are by definition classified as uninfected. Specifically,
142 we take $\mu_{\text{pop}}^{\text{uninf}} \sim \mathcal{N}(\bar{\mu}, 0.1)$ for the mean of the uninfected hyperdistribution and where $\bar{\mu} = 7.3$
143 denotes the mean of the pre-outbreak antibody titers. Likewise, we take $\sigma_{\text{pop}}^{\text{uninf}} \sim \mathcal{N}(\bar{\sigma}, 0.25)$ for
144 the standard deviation of the uninfected hyperdistribution, where $\bar{\sigma} = 1.3$ represents the standard
145 deviation of the pre-outbreak data. Further, we take $d_0 \sim \mathcal{B}(1, 29)$ and $d_2 \sim \mathcal{B}(29, 1)$, such that the
146 one-sided 95% prior ranges of these parameters are approximately $[0, 0.1)$ and $[0.9 - 1]$, respectively.
147 Other parameters ($\sigma_{\text{noise}}^{\text{uninf}}$, μ^{inf} , and σ^{inf}) are not equipped with explicit prior distributions, implying
148 that all values on their domains are a priori equally likely.

149 All analyses are performed using R (version 4.3.1) and Stan using the RStan interface (version
150 2.21.8) [32]. We run 10 MCMC chains in parallel and base the analyses on 10,000 thinned samples
151 from 10 well-mixed chains. Data, scripts, and figures are available in the online repository at
152 https://github.com/rivm-syso/mumps_serology.

153 Results

154 *Rule-based classification*

155 The data have been described and analysed earlier [22, 21, 23, 18], and we here put the earlier
156 findings into a perspective relevant to our analyses. Figure 1 shows the antibody titer measurements
157 and distributions of the pre- and post-outbreak samples. There is substantial variation in the pre-
158 and post-outbreak data, with mumps-specific IgG antibody titer measurements generally ranging
159 from 4 – 14 (16 – 16, 384 RU/mL). Hence, antibody concentrations can vary more than 1,000-fold.
160 The figure also shows that there is a strong correlation between the paired pre- and post-outbreak
161 samples, such that for the majority of participants the post-outbreak titer is roughly equal to the
162 pre-outbreak titer. In fact, for most paired samples titer differences are within a twofold change in
163 concentration. Only a small fraction of samples has a more than fourfold difference between pre-
164 and post-outbreak samples, and most of these participants show a more than fourfold increase.
165 In fact, $33/746 = 0.044$ show a fourfold or more antibody increase. Of these, 31 also have a
166 post-outbreak antibody concentration of 300 RU/mL, so that the rule-based cumulative infection
167 attack rate would be $31/746 = 0.042$. On the other hand, only a small fraction of participants

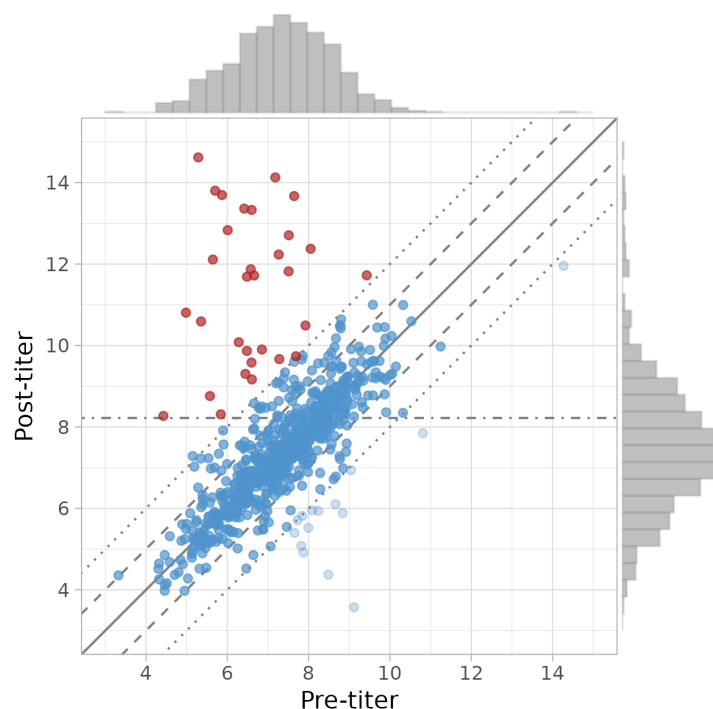


Figure 1: Overview of the data. Shown are the paired antibody titers (i.e. \log_2 -transformed antibody concentrations) for each of the 746 participants in the post- versus pre-outbreak survey. Titers range from 3 to 15, so that mumps-specific IgG antibody concentrations range from $2^3 = 8$ RU/mL to $2^{15} = 32,768$ RU/mL [18]. The solid line represents the identity function with equal antibody concentrations in the pre- and post-outbreak samples. Dashed lines represent a twofold increase and decrease of the antibody concentration, and dotted lines represent a fourfold increase and decrease. The dashed-dotted line represents the minimum titer in the post-outbreak survey for positive classification (8.2, corresponding to $2^{8.2} = 300$ RU/mL). Participants that were previously classified as infected are depicted in red [18]. Distributions on the top and right represent marginal titer distributions in the pre- and post-outbreak surveys, respectively.

168 (16/746 = 0.021) show a more than fourfold antibody decrease. Interestingly, the number of
169 participants with a modest titer increase between two- and fourfold is substantially larger than the
170 number with a similar modest decrease (56 versus 31). This suggests that some of the participants
171 with modest titer increase may actually have been infected.

172 In an earlier analyses a slightly different and more lenient classification rule for serological infection
173 has been used [21]. Here, the authors opted for either a fourfold increase of antibody concentrations,
174 or at least a post-outbreak titer of 10.6 (antibody concentration 1500 RU/mL). In this case, the
175 number of infections and infection attack rates are 37 and $37/746 = 0.050$, respectively.

176 *Parameter estimation*

177 Next, we fit our mixture model (equations (1)-(4)) to the data. Parameter estimates and derived
178 quantities are presented in Table 1, and the data and model fit are presented in Figure 2. Table 1

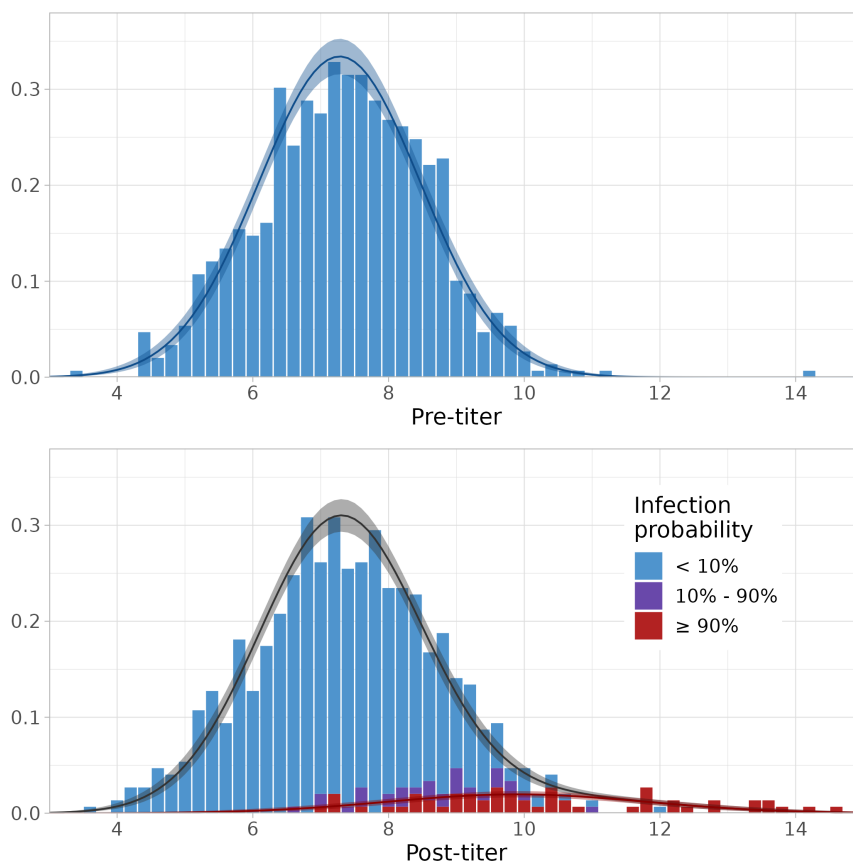


Figure 2: Overview of the data and fitted mixture distribution. Shown are the data (histograms) with fitted distributions weighted by estimated prevalence (lines). Shaded areas represent 95% credible envelopes. Top panel: pre-outbreak data with fitted distribution of uninfected participants. Bottom panel: post-outbreak data with fitted infected mixing distribution (red line) and overall mixture distribution (black line). Participants that are likely uninfected (posterior probability of infection $\leq 10\%$) are represented in blue, and those that are likely infected (posterior probability of infection $\geq 90\%$) are represented in red. Samples with intermediate posterior probabilities of infection are colored in purple.

179 shows that the posterior mean of antibody concentrations in infected persons (9.89) is six times
 180 higher than the posterior mean in uninfected persons (7.28), hence $2^{9.89-7.28} = 6.1$. However, the
 181 estimated antibody distributions are quite broad, especially for samples from uninfected persons.
 182 Overall, Figure 2 illustrates that there is a good correspondence between the model fit and data
 183 both for the pre- and post-outbreak surveys. Interestingly, even though there is considerable
 184 overlap between the mixing distributions of uninfected and infected persons, for most participants
 185 there is little doubt whether they had been infected or not. This is due to the high estimated
 186 intraclass correlation (0.85), and random variation in antibody concentrations of uninfected persons
 187 is estimated to be at most twofold ($2\widehat{\sigma}_{\text{noise}}^{\text{uninf}} = 1.02$). This, in turn, implies that already a twofold
 188 increase in antibody concentration provides evidence that the persons may have been infected.

189 In fact, the estimated probability of infection is essentially 0 if the post-outbreak titer equals
190 the pre-outbreak titer (0.00034, 95%CrI: 0.000014 – 0.0025), is close to 1 in case of a fourfold
191 antibody increase (0.994, 95%CrI: 0.963 – 1.0), and is almost 20% in case of a twofold increase
192 (0.19, 95%CrI: 0.064 – 0.38). The overall estimated cumulative incidence of infection is estimated
193 at 0.095 (95%CrI: 0.082 – 0.11) (Table 1).

194 To gauge the robustness of the above results with respect to the included participants, we rerun
195 all analyses without the 16 participants with strongly decreased antibody concentrations. Overall,
196 the results remain very similar, owing to the fact that these 16 only represent a small fraction of
197 the total study population ($\approx 2\%$) while these participants are with high certainty classified as
198 uninfected. Biggest difference is that the random variation from the latent antibody concentration
199 in uninfected persons is estimated to be even smaller (0.37 versus 0.51), leading to a further
200 increase in the intraclass correlation from 0.85 to 0.91. This in turn, leads to even higher precision
201 with which individual probabilities of infection are estimated, and overall estimated cumulative
202 incidence of infection is slightly increased (0.11, 95%CrI: 0.098 – 0.13).

203 Figure 3 shows 100 samples from the posterior distribution of the logistic infection function (equa-
204 tion (1)) together with estimated individual infections probabilities (equation (5)) as a function
205 of the titer difference between pre- and post-outbreak samples. Most participants either have a
206 post-outbreak titer that is close to the pre-outbreak titer, or have a post-outbreak titer that is sub-
207 stantially higher than the pre-outbreak titer. For these participants there is little doubt whether
208 they have been infected or not. For instance, a titer difference of 0.5 translates to an increase
209 of 41% of the antibody concentration ($2^{0.5} = 1.4$), and in this case it is highly unlikely that the
210 increase has been caused by infection. Conversely, a difference of 1.6 translates to a threefold
211 increase of the antibody concentration, and in this case it is approximately threefold more likely
212 that the participant has been infected rather than that it has not been infected. Between these
213 two extremes is the main uncertainty with respect to the infection status. Overall, there is a good
214 correspondence between the infection function and individual estimated probabilities of infection,
215 and this indicates that the model fits the data well.

216 *Probability of infection as function of pre-outbreak antibody titer*

217 Next, we turn attention to the relation between pre-outbreak antibody titers and subsequent
218 probability of infection. Figure 4 shows for all 746 participants the estimated infection probability
219 with associated uncertainty as function of the pre-outbreak titer. In general, there is limited

Parameter	Description	Estimate	95%CrI
$\mu_{\text{pop}}^{\text{uninf}}$	Mean of antibody concentrations in uninfected participants	7.28	(7.20 – 7.36)
$\sigma_{\text{pop}}^{\text{uninf}}$	Standard deviation of antibody concentrations in uninfected participants	1.19	(1.13 – 1.26)
$\sigma_{\text{noise}}^{\text{uninf}}$	Random variation of antibody concentrations (e.g., measurement noise)	0.51	(0.48 – 0.54)
μ^{inf}	Mean of antibody concentrations in infected participants	9.89	(9.42 – 10.4)
σ^{inf}	Standard deviation of antibody concentrations in infected participants	1.91	(1.63 – 2.27)
d_0	Probability of infection at no antibody increase	0.00034	(0.000014 – 0.0025)
d_1	Probability of infection at twofold antibody increase	0.19	(0.064 – 0.38)
d_2	Probability of infection at fourfold antibody increase	0.994	(0.963 – 1.0)
r_{ICC}	Intraclass correlation in uninfected participants	0.85	(0.82 – 0.87)
p^{inf}	Overall probability of infection (cumulative incidence)	0.095	(0.082 – 0.11)

Table 1: Parameter estimates and selected generated quantities. Estimates and generated quantities are represented by posterior medians and 95% posterior credible intervals. See text for explanation and details.

220 uncertainty for most participants, especially when the estimated probability of infection is close
 221 to 0 or close to 1. Only for a small subset of 31 participants with estimated posterior median
 222 probability of infection $\geq 10\%$ and $< 90\%$ is there substantial uncertainty.

223 To generalise these results for groups of persons and estimate the probability of infection as function
 224 of the pre-outbreak antibody titer we have stratified the study population by pre-outbreak antibody
 225 titers and calculated the probabilities of infection in these groups. Specifically, we have grouped
 226 pre-outbreak titers in four equal sized groups, with titer ranges [2, 6.48), [6.48, 7.35), [7.35, 8.18),
 227 and [8.18, 15), corresponding to antibody concentrations of 4 – 89 RU/mL, 89 – 163 RU/mL,
 228 163 – 290 RU/mL, and 290 – 32,768 RU/mL. As it appears, the estimated probabilities of infection
 229 generally decrease with increasing pre-outbreak antibody titer. Specifically, the estimated infection
 230 probability is 0.12 (95%CrI: 0.10 – 0.13) for the lowest two quartiles, and 0.056 (95%CrI: 0.044 –
 231 0.068) for the highest quartile. Hence, while there does not appear to be a fixed antibody titer

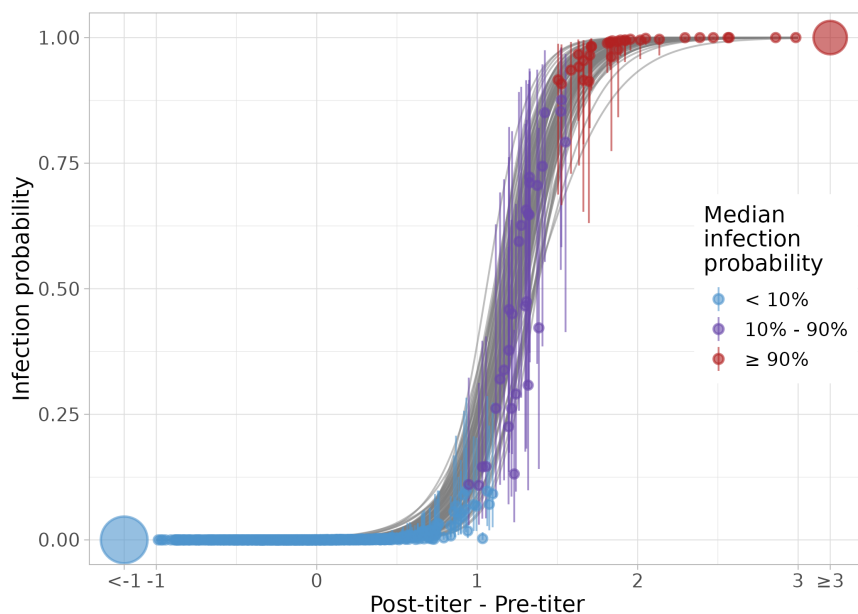


Figure 3: Estimated infection probabilities as function of differences in antibody concentrations. Shown are 100 samples from the posterior distribution of the logistic function determining the probability that a sample belongs to the infected component, as function of the difference in the antibody titers (grey lines). Also shown are the estimated (posterior median) infection probabilities for each of the 746 participants (dots), and uncertainty in the infection probabilities (95%CrI, whiskers). Notice that a difference of 1 between post- and pre-outbreak samples corresponds to a twofold increase, that a value of 2 corresponds to a fourfold increase, etcetera.

232 that provides protection against infection, the probability of infection is more than double in the
233 groups with low pre-outbreak antibody titers as compared to the group with highest antibody
234 titers. The estimated overall probability of infection is 0.95 (95%CrI: 0.082 – 0.11) (1). Thus, our
235 results confirm earlier results that the probability of infection increases with decreasing antibody
236 titers [18], but yield substantially higher infection attack rates than reported earlier.

237 Discussion

238 The past two decades have witnessed a distinct increase in the number of outbreaks of mumps
239 in highly vaccinated populations, mostly by non-vaccine genotypes, with an over-representation of
240 adolescents and young adults, and predominantly in close-contact settings (households, schools and
241 universities, parties) [33]. This has prompted suggestions that a booster vaccination in adolescence
242 could be beneficial ([18, 33] and references therein). Indeed, limited experience with such booster
243 doses has largely been positive [34, 33, 24, 18]. Our analyses have provided additional support for
244 the potential benefit of an extra vaccination in this age group to improve immunity to mumps.
245 Here, we have shown that the infection attack rate can be high in Dutch university university

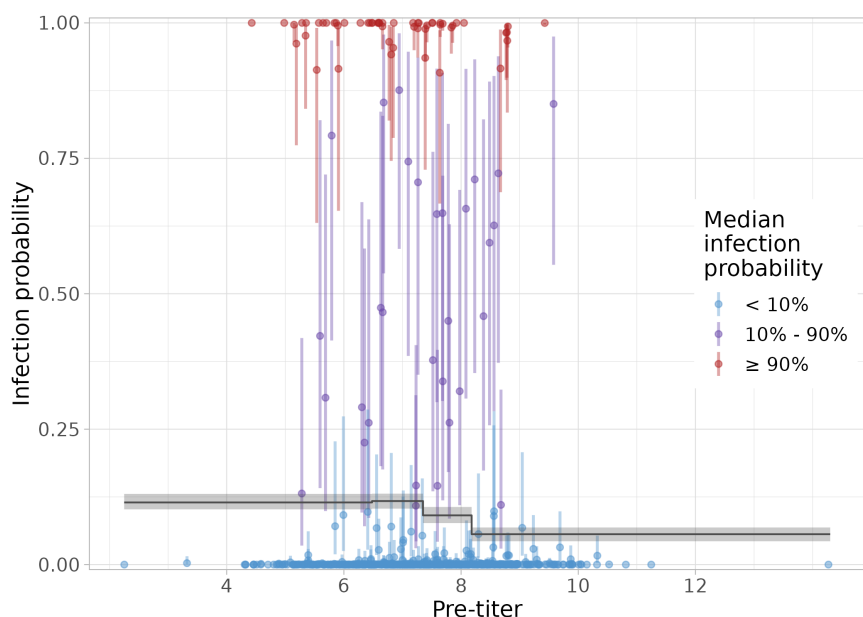


Figure 4: Estimated infection probabilities as function of the pre-outbreak antibody titer. Shown are the estimated infection probabilities for all participants (dots: posterior medians; lines: 95% credible interval). Colors represent three infection probability classes (cf. Figure 2). Also shown are estimated infection probabilities (median and 95% credible intervals), stratified by quartile of pre-outbreak antibody concentration.

246 (9.5%), and that in addition the infection attack rate is more than twice as high in students
247 with low pre-existing mumps-specific IgG antibody concentrations than in students with high pre-
248 existing antibodies (12% vs 5.6%). This suggests that the hazard of infection can be at least twice
249 as high in adolescents and young adults with low antibody concentrations than in those with high
250 antibody concentrations. However, no fixed pre-outbreak antibody concentration could be reliably
251 determined that fully correlates with protection against infection. This indicates that not only IgG
252 antibodies but also other compartments of the immune system, such as cellular immunity, play
253 a role in protection against mumps [35, 36]. Nevertheless, young adults with low mumps-specific
254 IgG antibody concentrations are at higher risk for infection and may therefore benefit most from
255 a third dose of the MMR vaccine. Interestingly in this context, our previous study showed that
256 individuals with lowest pre-vaccination IgG concentrations also showed the strongest increase in
257 IgG concentrations after vaccination [24].

258 Our analyses build on and extend earlier analyses for the same outbreak [21, 18]. An advantage of
259 the current analyses is that they are fully self-contained, and that all parameters, distributions, and
260 infection rates are directly estimated from the data. This is a considerable advantage and makes
261 the analyses quite robust, as there is no fixed predefined level of antibody concentrations that is

262 indicative of infection or protection against infection [37, 38]. Moreover, antibody concentrations
263 in the population depend in a complex manner on time since last vaccination, potential previous
264 infection(s), and outbreak genotype(s). Therefore, it is unlikely that proper sets of validation
265 samples with confirmed uninfected and infected persons can be obtained. This is true not only
266 for our population of university students but probably holds more generally. Thus, ad hoc choices
267 have previously been made for the infection criterion (a fourfold antibody increase with additional
268 requirements depending on the study). Comparing our results with the earlier analyses of Gouma
269 et al. [21] and Kaaijk et al. [18] we find a similar pattern that infection attack rates increase
270 with decreasing prior antibody concentrations. Our analyses also indicate, however, that infection
271 attack rates may have been substantially higher (9.5%) than reported earlier (4% – 6%). This is
272 due to the fact that many participants with modest increases of antibody titers (two- to threefold)
273 may actually have been infected. In all, our results paint a more dynamic picture of the antibody
274 dynamics induced by infection and waning than hitherto considered.

275 Previous analyses for bivariate serological cross-sectional studies [26, 39, 40] have modelled the
276 serological response with bivariate mixture distributions. Here, by virtue of the fact that we have
277 paired data at our disposal we have been able to couple pre- and post-outbreak data using an
278 infection function (equation (1)). This has enabled precise estimation of infection probabilities for
279 the majority of participants even though the distributions of uninfected and infected persons show
280 considerable overlap. In fact, we would not have been able to obtain precise estimates of infection
281 probabilities if the data had been treated as two separate cross-sectional surveys (J. Gomme,
282 unpublished report). In principle, our model and estimates can also be used for prediction of
283 infection probabilities of paired samples in future studies that have been analysed with the test
284 employed here.

285 In our analyses we have included 16 participants with a strong (at least fourfold) antibody decrease.
286 These person may have had experienced a recent mumps infection, and may therefore not be
287 representative for antibody concentrations in populations with no recent exposure. To study how
288 the results are affected by this choice, we have performed a sensitivity analysis in which these
289 samples are excluded. Fortunately, the results are largely unaffected by inclusion of exclusion of
290 those participants. This due to the fact that these 16 persons represent only a small minority
291 of the study population (2%), and the fact that the estimated probability of infection for those
292 participants is essentially 0.

293 Two main limitations deserve scrutiny. First, we have throughout focused on probabilistic classifi-

294 cation of participants and estimation of the serological infection attack rate. It would be desirable,
295 but not possible with the limited information available in our study, if the analyses could be ex-
296 tended to include not only serological data but also information on the severity of infection. It
297 would be particularly interesting to study how the severity of infection depends on both the pre-
298 and post-outbreak antibody concentrations. Second, our analyses apply to the specific situation of
299 medical students in the Netherlands. This is a very specific population and setting. For instance,
300 in the Netherlands the vast majority of students have been vaccinated, and contact intensities
301 in this group are probably substantially higher than in other strata of the population of similar
302 age. It remains therefore an open question to which extent the results still apply in other settings.
303 In particular, one could envisage that our estimated infection attack rates in medical students
304 represent an upper bound for this age group in general.

305 Finally, analyses for paired data such as presented here may have wider applicability, and are
306 not restricted to serological surveys for infectious diseases. Indeed, our methods of analysis apply
307 whenever a latent response is usually stable but highly individual-specific. In such instances, it
308 may be of interest to be able to decide when there are substantial deviations from this stable
309 individual-specific response. Such situations occur often in hospital settings when patients are
310 regularly screened for the onset of disease. Our study has shown one way how such data can be
311 analysed elegantly with mixture model-based methods.

312 **Acknowledgements**

313 We gratefully acknowledge the participants of the study. Part of this study has been performed
314 during an internship of JG from the Department of Epidemiology and Social Medicine at University
315 of Antwerp at the National Institute of Public Health and the Environment (RIVM). This study
316 was supported by the Dutch Ministry of Health, Welfare and Sport.

317 **References**

- 318 [1] I. Davidkin, S. Jokinen, M. Broman, P. Leinikki, and H. Peltola. Persistence of measles, mumps, and rubella
319 antibodies in an MMR-vaccinated cohort: a 20-year follow-up. *J Infect Dis*, 197(7):950–956, Apr 2008.
- 320 [2] M. M. Cortese, A. E. Barskey, G. E. Tegtmeier, C. Zhang, L. Ngo, M. H. Kyaw, A. L. Baughman, J. E. Menitove,
321 C. J. Hickman, W. J. Bellini, G. H. Dayan, G. R. Hansen, and S. Rubin. Mumps antibody levels among students
322 before a mumps outbreak: in search of a correlate of immunity. *J Infect Dis*, 204(9):1413–1422, Nov 2011.
- 323 [3] M. Kontio, S. Jokinen, M. Paunio, H. Peltola, and I. Davidkin. Waning antibody levels and avidity: implications
324 for MMR vaccine-induced protection. *J Infect Dis*, 206(10):1542–1548, Nov 2012.
- 325 [4] A. Antia, H. Ahmed, A. Handel, N. E. Carlson, I. J. Amanna, R. Antia, and M. Slifka. Heterogeneity and
326 longevity of antibody memory to viruses and vaccines. *PLoS Biol*, 16(8):e2006601, Aug 2018.

- 327 [5] M. van Boven, W. L. Ruijs, J. Wallinga, P. D. O'Neill, and S. é. Estimation of vaccine efficacy and critical
328 vaccination coverage in partially observed outbreaks. *PLoS Comput Biol*, 9(5):e1003061, 2013.
- 329 [6] L. J. Willocks, D. Guerendiain, H. I. Austin, K. E. Morrison, R. L. Cameron, K. E. Templeton, V. R. F.
330 DE Lima, R. Ewing, W. Donovan, and K. G. J. Pollock. An outbreak of mumps with genetic strain variation
331 in a highly vaccinated student population in Scotland. *Epidemiol Infect*, 145(15):3219–3225, Nov 2017.
- 332 [7] L. N. Patel, R. J. Arciuolo, J. Fu, F. R. Giancotti, J. R. Zucker, J. L. Rakeman, and J. B. Rosen. Mumps
333 Outbreak Among a Highly Vaccinated University Community-New York City, January-April 2014. *Clin Infect*
334 *Dis*, 64(4):408–412, Feb 2017.
- 335 [8] L. Veneti, K. Borgen, K. S. Borge, K. Danis, M. Greve-Isdahl, K. Konsmo, G. Istad, S. A. ø, K. S. ystese,
336 R. Rykkvin, E. Sagvik, and R. Riise. Large outbreak of mumps virus genotype G among vaccinated students
337 in Norway, 2015 to 2016. *Euro Surveill*, 23(38), Sep 2018.
- 338 [9] M. Shah, P. Quinlisk, A. Weigel, J. Riley, L. James, J. Patterson, C. Hickman, P. A. Rota, R. Stewart,
339 N. Clemmons, N. Kalas, C. Cardemil, M. Patel, M. Donahue, A. Schneider, U. Ukegbu, K. Wittich, J. Kellogg,
340 D. Beardsley, N. Tran, D. Callaghan, A. Pyatt, T. Kitzmann, and B. Kintigh. Mumps Outbreak in a Highly
341 Vaccinated University-Affiliated Setting Before and After a Measles-Mumps-Rubella Vaccination Campaign-
342 Iowa, July 2015-May 2016. *Clin Infect Dis*, 66(1):81–88, Jan 2018.
- 343 [10] D. W. Westphal, A. Eastwood, A. Levy, J. Davies, C. Huppatz, M. Gilles, H. Lyttle, S. A. Williams, and G. K.
344 Dowse. A protracted mumps outbreak in Western Australia despite high vaccine coverage: a population-based
345 surveillance study. *Lancet Infect Dis*, 19(2):177–184, Feb 2019.
- 346 [11] A. Ferenczi, S. Gee, S. Cotter, and K. Kelleher. Ongoing mumps outbreak among adolescents and young adults,
347 Ireland, August 2018 to January 2020. *Euro Surveill*, 25(4), Jan 2020.
- 348 [12] M. Perez Duque, A. San-Bento, L. on, P. dio, M. A. a, M. J. Albuquerque, M. Nascimento, S. Balasegaram,
349 and R. Machado. Mumps outbreak among fully vaccinated school-age children and young adults, Portugal
350 2019/2020. *Epidemiol Infect*, 149:e205, Aug 2021.
- 351 [13] L. H. Moncla, A. Black, C. DeBolt, M. Lang, N. R. Graff, A. C. rez Osorio, N. F. ller, D. Haselow, S. Lindquist,
352 and T. Bedford. Repeated introductions and intensive community transmission fueled a mumps virus outbreak
353 in Washington State. *Elife*, 10, Apr 2021.
- 354 [14] S. A. Rubin, M. A. Link, C. J. Sauder, C. Zhang, L. Ngo, B. K. Rima, and W. P. Duprex. Recent mumps
355 outbreaks in vaccinated populations: no evidence of immune escape. *J Virol*, 86(1):615–620, Jan 2012.
- 356 [15] S. Gouma, T. Vermeire, S. Van Gucht, L. Martens, V. Hutse, J. Cremer, P. A. Rota, G. Leroux-Roels, M. Koop-
357 mans, R. van Binnendijk, and E. Vandermarliere. Differences in antigenic sites and other functional regions
358 between genotype A and G mumps virus surface proteins. *Sci Rep*, 8(1):13337, Sep 2018.
- 359 [16] S. Dittrich, S. é, A. van Lier, R. Kohl, H. Boot, M. Koopmans, and R. van Binnendijk. Assessment of serological
360 evidence for mumps virus infection in vaccinated children. *Vaccine*, 29(49):9271–9275, Nov 2011.
- 361 [17] S. Borgmann, F. Schwab, S. Santibanez, and A. Mankertz. Mumps virus infection in vaccinated patients can
362 be detected by an increase in specific IgG antibodies to high titres: a retrospective study. *Epidemiol Infect*,
363 142(11):2388–2396, Nov 2014.
- 364 [18] Patricia Kaaijk, Alienke J Wijmenga-Monsuur, Marlies A van Houten, Irene K Veldhuijzen, Hinke I ten Hulscher,
365 Jeroen Kerkhof, Fiona R van der Klis, and Rob S van Binnendijk. A Third Dose of Measles-Mumps-Rubella
366 Vaccine to Improve Immunity Against Mumps in Young Adults. *The Journal of Infectious Diseases*, 221(6):902–
367 909, 04 2019.
- 368 [19] Patricia Kaaijk, Alienke J. Wijmenga-Monsuur, Hinke I. ten Hulscher, Jeroen Kerkhof, Gaby Smits,
369 Mioara Alina Nicolaie, Marianne A. van Houten, and Rob S. van Binnendijk. Antibody levels at 3-years
370 follow-up of a third dose of measles-mumps-rubella vaccine in young adults. *Vaccines*, 10(1), 2022.
- 371 [20] G. P. Smits, P. G. van Gageldonk, L. M. Schouls, F. R. van der Klis, and G. A. Berbers. Development of
372 a bead-based multiplex immunoassay for simultaneous quantitative detection of IgG serum antibodies against
373 measles, mumps, rubella, and varicella-zoster virus. *Clin Vaccine Immunol*, 19(3):396–400, Mar 2012.
- 374 [21] S. Gouma, T. M. Schurink-Van't Klooster, H. E. de Melker, J. Kerkhof, G. P. Smits, S. J. é, C. A. van Els, G. J.
375 Boland, A. C. Vossen, P. R. Goswami, M. P. Koopmans, and R. S. van Binnendijk. Mumps serum antibody
376 levels before and after an outbreak to assess infection and immunity in vaccinated students. *Open Forum Infect*
377 *Dis*, 1(3):ofu101, Dec 2014.
- 378 [22] J. Sane, S. Gouma, M. Koopmans, H. de Melker, C. Swaan, R. van Binnendijk, and S. é. Epidemic of mumps
379 among vaccinated persons, The Netherlands, 2009-2012. *Emerg Infect Dis*, 20(4):643–648, Apr 2014.
- 380 [23] S. Gouma, J. Sane, D. Gijselaar, J. Cremer, S. é, M. Koopmans, and R. van Binnendijk. Two major mumps
381 genotype G variants dominated recent mumps outbreaks in the Netherlands (2009-2012). *J Gen Virol*, 95(Pt
382 5):1074–1082, May 2014.

- 383 [24] Patricia Kaaijk, M Alina Nicolaie, Debbie van Rooijen, Marianne A van Houten, Fiona R van der Klis, Anne-
384 Marie Buisman, and Rob S van Binnendijk. Dynamics of the Antibody Response After a Third Dose of Measles-
385 Mumps-Rubella Vaccine Indicate a Slower Decline Compared With a Second Dose. *Open Forum Infectious*
386 *Diseases*, 7(11):ofaa505, 10 2020.
- 387 [25] A. Steens, S. Waaijenborg, P. F. Teunis, J. H. Reimerink, A. Meijer, M. van der Lubben, M. Koopmans, M. A.
388 van der Sande, J. Wallinga, and M. van Boven. Age-dependent patterns of infection and severity explaining
389 the low impact of 2009 influenza A (H1N1): evidence from serial serologic surveys in the Netherlands. *Am J*
390 *Epidemiol*, 174(11):1307–1315, Dec 2011.
- 391 [26] D. te Beest, E. de Bruin, S. Imholz, J. Wallinga, P. Teunis, M. Koopmans, and M. van Boven. Discrimination
392 of influenza infection (A/2009 H1N1) from prior exposure by antibody protein microarray analysis. *PLoS One*,
393 9(11):e113021, 2014.
- 394 [27] M. A. Vink, J. van de Kasstele, J. Wallinga, P. F. Teunis, and J. A. Bogaards. Estimating seroprevalence
395 of human papillomavirus type 16 using a mixture model with smoothed age-dependent mixing proportions.
396 *Epidemiology*, 26(1):8–16, Jan 2015.
- 397 [28] G. Smits, L. Mollema, S. é, H. de Melker, I. Tcherniaeva, S. Waaijenborg, R. van Binnendijk, F. van der Klis,
398 and G. Berbers. Seroprevalence of mumps in The Netherlands: dynamics over a decade with high vaccination
399 coverage and recent outbreaks. *PLoS One*, 8(3):e58234, 2013.
- 400 [29] J. A. Bouman, J. Riou, S. Bonhoeffer, and R. R. Regoes. Estimating the cumulative incidence of SARS-CoV-2
401 with imperfect serological tests: Exploiting cutoff-free approaches. *PLoS Comput Biol*, 17(2):e1008728, Feb
402 2021.
- 403 [30] J. A. Bouman, S. Kadelka, S. Stringhini, F. Pennacchio, B. Meyer, S. Yerly, L. Kaiser, I. Guessous, A. S.
404 Azman, S. Bonhoeffer, and R. R. Regoes. Applying mixture model methods to SARS-CoV-2 serosurvey data
405 from Geneva. *Epidemics*, 39:100572, Jun 2022.
- 406 [31] B. Carpenter, A. Gelman, M. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and
407 A. Riddell. Stan: A probabilistic programming language. *Journal of Statistical Software, Articles*, 76(1):1–32,
408 2017.
- 409 [32] Bob Carpenter, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus
410 Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *Journal of*
411 *Statistical Software*, 76(1):1–32, 2017.
- 412 [33] E. Lam, J. B. Rosen, and J. R. Zucker. Mumps: an Update on Outbreaks, Vaccine Efficacy, and Genomic
413 Diversity. *Clin Microbiol Rev*, 33(2), Mar 2020.
- 414 [34] C. V. Cardemil, R. M. Dahl, L. James, K. Wannemuehler, H. E. Gary, M. Shah, M. Marin, J. Riley, D. R.
415 Feikin, M. Patel, and P. Quinlisk. Effectiveness of a Third Dose of MMR Vaccine for Mumps Outbreak Control.
416 *N Engl J Med*, 377(10):947–956, Sep 2017.
- 417 [35] S. Jokinen, P. Osterlund, I. Julkunen, and I. Davidkin. Cellular immunity to mumps virus in young adults 21
418 years after measles-mumps-rubella vaccination. *J Infect Dis*, 196(6):861–867, Sep 2007.
- 419 [36] J. de Wit, M. E. Emmelot, M. C. M. Poelen, R. S. van Binnendijk, S. van der Lee, D. van Baarle, W. G. H.
420 Han, C. A. C. M. van Els, and P. Kaaijk. T-cell memory. *J Allergy Clin Immunol*, 141(5):1908–1911, May 2018.
- 421 [37] P. F. Teunis, J. C. van Eijkeren, W. F. de Graaf, A. B. é, and M. E. Kretzschmar. Linking the seroresponse to
422 infection to within-host heterogeneity in antibody production. *Epidemics*, 16:33–39, Sep 2016.
- 423 [38] P. F. M. Teunis and J. C. H. van Eijkeren. Estimation of seroconversion rates for infectious diseases: Effects of
424 age and noise. *Stat Med*, 39(21):2799–2814, Sep 2020.
- 425 [39] D. E. Te Beest, P. J. Birrell, J. Wallinga, D. De Angelis, and M. van Boven. Joint modelling of serological and
426 hospitalization data reveals that high levels of pre-existing immunity and school holidays shaped the influenza
427 A pandemic of 2009 in the Netherlands. *J R Soc Interface*, 12(103), Feb 2015.
- 428 [40] M. A. Vink, J. Berkhof, J. van de Kasstele, M. van Boven, and J. A. Bogaards. A Bivariate Mixture Model
429 for Natural Antibody Levels to Human Papillomavirus Types 16 and 18: Baseline Estimates for Monitoring the
430 Herd Effects of Immunization. *PLoS One*, 11(8):e0161109, 2016.