Contents lists available at ScienceDirect

Epidemics

journal homepage: www.elsevier.com/locate/epidemics

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

Michiel van Boven [a,b,*], Jantien A. Backer [a], Irene Veldhuijzen [a], Justin Gomme [c,d], Rob van Binnendijk [a], Patricia Kaaijk [a]

[a] *Centre for Infectious Disease Control, National Institute for Public Health and the Environment, Bilthoven, The Netherlands*
[b] *Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands*
[c] *Department of Epidemiology and Social Medicine, University of Antwerp, Antwerp, Belgium*
[d] *NHS Scotland, Edinburgh, Scotland, United Kingdom*

ARTICLE INFO

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 an outbreak of mumps virus genotype G among college students in the Netherlands over the period 2009–2012 using paired serological data. 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 in this study 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 of samples 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.

1. Introduction

Mass vaccination campaigns have been highly successful in reducing transmission and associated morbidity of infectious diseases of childhood. Case in point is the Measles–Mumps–Rubella vaccine, which in the Netherlands is administered at 14 months and 9 years, and has a coverage of 90%–95%. It is known, however, that the antibody response to the mumps component in the vaccine wanes over time, and that IgG antibodies induced by the vaccine have relatively low avidity (Davidkin et al., 2008; Cortese et al., 2011; Kontio et al., 2012; Antia et al., 2018). Perhaps owing to this, mumps outbreaks have occurred worldwide over the past decades, mainly in vaccinated adolescents and young adults. These outbreaks often occurred in close contact settings (schools, households, parties), and were mostly caused by genotypes that are different from the vaccine genotype (van Boven et al., 2013; Willocks et al., 2017; Patel et al., 2017; Veneti et al., 2018; Shah et al., 2018; Westphal et al., 2019; Ferenczi et al., 2020; Perez Duque et al.,

2021; Moncla et al., 2021). For instance, the current vaccine used in the Netherlands contains the Jeryl-Lynn strain (genotype A), and is genetically distant from the recent outbreak strains (genotypes D and G) (Rubin et al., 2012; Gouma et al., 2018). This vaccine is also widely used in the United States and other European countries.

Many but not all infections with mumps virus are asymptomatic or are associated with mild symptoms only, especially in vaccinated individuals (Dittrich et al., 2011). This makes it difficult to assess the true extent of virus circulation in an outbreak. In principle, reliable infection attack rates can be obtained from measuring mumps-specific IgG antibody concentrations, because these generally increase after mumps virus infection (Borgmann et al., 2014). However, a challenge is that paired pre- and post-outbreak samples are often not available, and that there are no generally agreed antibody concentrations that define recent infection in a single sample. Two studies have shown that pre-outbreak mumps neutralisation antibody concentrations in patients

with mumps were generally lower than in persons who were not infected with mumps virus during the outbreak (Cortese et al., 2011; Kaaijk et al., 2019, 2022). However, it proved not possible to define reliable cutoffs separating infected from uninfected persons, and separating patients with clinical symptoms from person with asymptomatic or mild infection (Cortese et al., 2011). Hence, it is not straightforward to deduce who has been infected either clinically or sub-clinically, especially when only a single serum sample is available, and it is not clear how the probability of infection depends on pre-existing antibody concentrations.

In an earlier study, we measured mumps-specific IgG antibody concentrations in paired pre- and post-outbreak samples from university students in the Netherlands, using a fluorescent bead-based multiplex immunoassay (Smits et al., 2012; Gouma et al., 2014b; Kaaijk et al., 2019). In these studies, we calculated the proportions of symptomatic and asymptomatic infections, and determined infection attack rates and risk factors for mumps virus infection using predefined criteria for infection. Specifically, participants would be classified as infected if there was a fourfold increase of antibody concentrations from first to second sample, or a high antibody concentration in the second sample. Further, to identify a correlate of protection, mumps-specific IgG concentrations in pre-outbreak samples were compared between infected and non-infected persons (Sane et al., 2014; Gouma et al., 2014b,a). However, as antibody concentrations decay over time, especially in the first months after an infection (Antia et al., 2018; Kaaijk et al., 2020, 2019, 2022), and as the second sample in our study may have been taken several years after infection, we need analyses with less rigid criteria for infection. Therefore, in this study we use methods that do not rely on predefined criteria for infection, but that estimate infection probabilities directly from the available serological data.

Our aim is to provide estimates of the infection attack rates using methods that make optimal use of the data, and that do not use predefined cutoffs or fixed titer increases for classification. Specifically, we probabilistically classify participants with a two-component binary mixture model in which the component distributions represent infected and uninfected persons (Steens et al., 2011; te Beest et al., 2014; Vink et al., 2015). Usually, the mixing parameter in such analyses represents the prevalence or probability that a person is infected. Instead of using a fixed population-level mixing parameter, we here link the pre- and post-outbreak samples by making the biologically plausible assumption that the probability of infection increases monotonically with the ratio of post- to pre-outbreak antibody concentration. This enables estimation of the infection probability and associated uncertainty for each participant in a manner that is optimally informed by the data.

2. Methods

2.1. Study population and sample collection

All pre-outbreak sera in this retrospective study are taken from first-year medical students from Leiden University and Utrecht University. Recruitment from this population was carried out from 2007–2010, and resulted in a study population of 746 students with both a pre- and post-outbreak sample. The study has been approved by a medical ethical committee (NL38042.041.11), and has been described in detail earlier (Kaaijk et al., 2019). Briefly, a self-sampled dried blot spot sample, a questionnaire concerning vaccination history, risk factors, and mumps symptoms has been obtained from each student. The serological criteria for mumps virus infection in the previous study were a fourfold increase in IgG concentration combined with at least an post-outbreak antibody concentration of 300 RU/ml (Kaaijk et al., 2019). A small subset of 16 participants show a strong (more than fourfold) decrease in the antibody concentration from pre- to post-outbreak sample. These could represent recent infections, as pre-outbreak samples had been taken from 2007–2010, while the outbreak lasted from 2009–2012. We therefore perform analyses of the full data in the body of the text, but also indicate how the results are affected if those 16 participants are left out.

2.2. Mumps-specific IgG multiplex immunoassay

Mumps virus – specific serum immunoglobulin G (hereafter called IgG) antibody concentrations have been determined by a fluorescent bead – based multiplex immunoassay as described before, using purified Jeryl-Lynn mumps vaccine strain as antigen (Smits et al., 2012, 2013). For analysis, all antibody concentrations are $\log_2$-transformed, and in the following the log-transformed concentrations are referred to as (antibody) titers.

2.3. Mixture model for paired data

Participants are assumed to be uninfected in the first sample, and either infected or uninfected in the second sample. Here, we assume that these two classes are characterised by probability distributions for the antibody titers, and are characterised by density functions f^{uninf} and f^{inf}. Based on preliminary analyses, we assume that the titer distributions of the two classes are normally 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 pre-outbreak samples provide information on the uninfected component probability distribution, and the post-outbreak samples inform both the uninfected and infected component probability distributions. In mixture models for cross-sectional data the probability of infection is usually given by the prevalence weighted density of the infected component divided by the sum of the weighted infected and uninfected components (e.g., Vink et al., 2015; Bouman et al., 2021, 2022). Here we take an alternative approach that makes use of the fact that data are paired by postulating that the probability that a post-outbreak sample belongs to the infected component (the mixing parameter) increases monotonically with the ratio of the post- versus pre-outbreak titers. Specifically, we propose a two-parameter logistic infection function such that the mixing parameter for the ith participant ($i = 1, \ldots, 746$), q_i^{inf}, is given by

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

In Eq. (1), Y_i^{pre} and Y_i^{post} denote the antibody titers in the pre- and post-outbreak samples, and k and x_0 represent the steepness and 50% infection probability of the logistic function. Throughout, we use a transformation of the two-parameter logistic function (k and x_0) in terms of the probability that a sample belongs to the infected component if antibody concentrations remain unchanged (q_0), and if a fourfold increase is observed (q_2). This yields parameters that can be interpreted more easily in biological terms, and facilitates making informed choices for the parameter prior distributions. A straightforward calculation shows that

$$k = \frac{1}{2}\left(\log\left(q_0^{-1} - 1\right) - \log\left(q_2^{-1} - 1\right)\right)$$
$$x_0 = -2\frac{\log\left(q_0^{-1} - 1\right)}{\log\left(q_2^{-1} - 1\right) - \log\left(q_0^{-1} - 1\right)} \ .$$

Biological reasoning implies that one would expect $q_0 \approx 0$ and $q_2 \approx 1$ (Smits et al., 2012; Gouma et al., 2014b). Notice further that q_i^{inf} only depends on the difference of the two measurements but not on the individual values of Y_i^{pre} and Y_i^{post}. Also notice that a difference of 1 between antibody titers represents a twofold increase, a difference of 2 a fourfold increase, etcetera.

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

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

where $\mu_{\text{pop}}^{\text{uninf}}$ and $\sigma_{\text{pop}}^{\text{uninf}}$ are the population mean and variance of the distribution of latent antibody concentrations in uninfected persons.

Each measurement of uninfected persons provides a possibly imperfect representation of the latent antibody concentration, such that, for instance, for pre-outbreak samples Y_i^{pre} ($i = 1, \ldots, 746$) we have

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

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

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

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

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

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

where we have suppressed the dependence of p_i^{inf} and q_i^{inf} on the antibody differences $Y_i^{\text{post}} - Y_i^{\text{pre}}$, and the dependence of f_i^{uninf} and f_i^{inf} on post-outbreak antibody concentrations Y_i^{post}. Hence, the 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

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

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

2.4. Prior distributions and inference

Parameters are estimated in a Bayesian framework using Hamiltonian Monte Carlo, implemented in Stan (Carpenter et al., 2017a). The main parameters to be estimated are the mean and standard deviation of the hyper distribution of antibody titers in uninfected persons, $\mu_{\text{pop}}^{\text{uninf}}$ and $\sigma_{\text{pop}}^{\text{uninf}}$, the random noise $\sigma_{\text{noise}}^{\text{uninf}}$, the mean and standard deviation of antibody titers in infected persons, μ^{inf} and σ^{inf}, and the parameters defining the probability of infection when there is no increase in antibody titer from pre- to post-outbreak samples, and when there is a fourfold increase, q_0 and q_2.

Prior distributions for the means and standard deviations of the uninfected component distribution are based on pre-outbreak samples as these are by definition classified as uninfected. Specifically, we take $\mu_{\text{pop}}^{\text{uninf}} \sim \mathcal{N}\left(\overline{\mu}, 0.1\right)$ for the mean of the uninfected hyperdistribution and where $\overline{\mu} = 7.3$ denotes the mean of the pre-outbreak antibody titers. Likewise, we take $\sigma_{\text{pop}}^{\text{uninf}} \sim \mathcal{N}\left(\overline{\sigma}, 0.25\right)$ for the standard deviation of the uninfected hyperdistribution, where $\overline{\sigma} = 1.3$ represents the standard deviation of the pre-outbreak data. Further, we take $q_0 \sim \mathcal{B}(1, 29)$ and $q_2 \sim \mathcal{B}(29, 1)$, such that the one-sided 95% prior ranges of these parameters are approximately $[0, 0.1]$ and $[0.9 - 1]$, respectively. Other parameters ($\sigma_{\text{noise}}^{\text{uninf}}$, μ^{inf}, and σ^{inf}) are not equipped with explicit prior distributions, implying that all values on their domains are a priori equally likely.

All analyses are performed using R (version 4.3.1) and Stan using the RStan interface (version 2.21.8) (Carpenter et al., 2017b). We run 10 Hamiltonian Monte Carlo (HMC) chains in parallel and base the
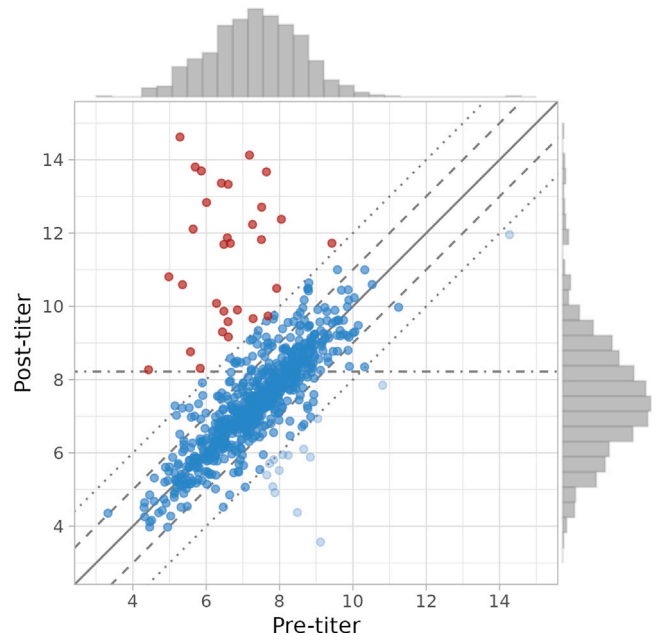

Fig. 1. Overview of the data. Shown are the paired antibody titers (i.e. $\log_2$-transformed antibody concentrations) for each of the 746 participants in the post- versus pre-outbreak survey. Titers range from 3 to 15, so that mumps-specific IgG antibody concentrations range from $2^3 = 8$ RU/mL to $2^{15} = 32{,}768$ RU/mL (Kaaijk et al., 2019). 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 (Kaaijk et al., 2019). Distributions on the top and right represent marginal titer distributions in the pre- and post-outbreak surveys, respectively.

analyses on 10,000 thinned samples from 10 well-mixed chains. Data, scripts, and figures are available in the online repository at https://github.com/rivm-syso/mumps_serology.

3. Results

3.1. Rule-based classification

The data have been described and analysed earlier (Sane et al., 2014; Gouma et al., 2014b,a; Kaaijk et al., 2019), and we here put the earlier findings into a perspective relevant to our analyses. Fig. 1 shows the antibody titer measurements and distributions of the pre- and post-outbreak samples. There is substantial variation in the pre- and post-outbreak data, with mumps-specific IgG antibody titer measurements generally ranging from 4–14 (16–16,384 RU/mL). Hence, antibody concentrations can vary more than 1,000-fold. The figure also shows that there is a strong correlation between the paired pre- and post-outbreak samples, such that for the majority of participants the post-outbreak titer is roughly equal to the pre-outbreak titer. In fact, for most paired samples titer differences are within a twofold change in concentration. Only a small fraction of samples has a more than fourfold difference between pre- and post-outbreak samples, and most of these participants show a more than fourfold increase. In fact, $33/746 = 0.044$ show a fourfold or more antibody increase. Of these, 31 also have a post-outbreak antibody concentration of 300 RU/mL, so that the rule-based cumulative infection attack rate would be $31/746 = 0.042$. On the other hand, only a small fraction of participants ($16/746 = 0.021$) show a more than fourfold antibody decrease. Interestingly, the number of participants with a modest titer increase between two- and fourfold is substantially larger than the number with a similar modest
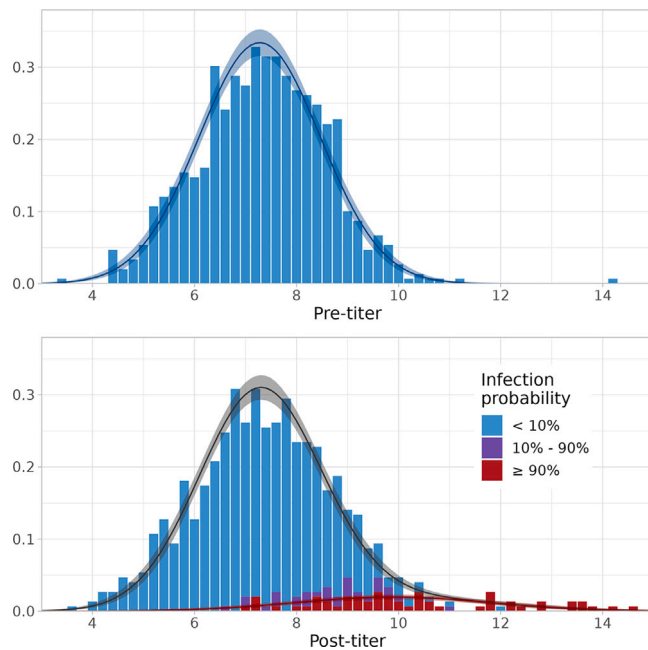
Fig. 2. Overview of the data and fitted mixture distribution. Shown are the data (histograms) with fitted distributions weighted by estimated prevalence (lines). Shaded areas represent 95% credible envelopes. Top panel: pre-outbreak data with fitted distribution of uninfected participants. Bottom panel: 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 ≤ 10%) are represented in blue, and those that are likely infected (posterior probability of infection ≥ 90%) are represented in red. Samples with intermediate posterior probabilities of infection are coloured in purple. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

decrease (56 versus 31). This suggests that some of the participants with modest titer increase may actually have been infected.

In an earlier analyses a slightly different and more lenient classification rule for serological infection has been used (Gouma et al., 2014b). Here, the authors opted for either a fourfold increase of antibody concentrations, or at least a post-outbreak titer of 10.6 (antibody concentration 1500 RU/mL). In this case, the number of infections and infection attack rates are 37 and $37/746 = 0.050$, respectively.

3.2. Parameter estimation

Next, we fit our mixture model (Eqs. (1)–(4)) to the data. Parameter estimates and derived quantities are presented in Table 1, and the data and model fit are presented in Fig. 2. Table 1 shows that the posterior mean of antibody concentrations in infected persons (9.89) is approximately six times higher than the posterior mean in uninfected persons (7.28, hence $2^{9.89-7.28} = 6.1$). However, the estimated antibody distributions are quite broad, especially for samples from uninfected persons. Overall, Fig. 2 illustrates that there is a good correspondence between the model fit and data both for the pre- and post-outbreak surveys. Interestingly, even though there is considerable overlap between the mixing distributions of uninfected and infected persons, for most participants there is little doubt whether they had been infected or not. This is due to the high estimated intraclass correlation (0.85), and random variation in antibody concentrations of uninfected persons is estimated to be at most twofold ($2\widehat{\sigma_{\text{noise}}^{\text{uninf}}} = 1.02$). This, in turn, implies that already a twofold increase in antibody concentration provides evidence that the persons may have been infected. In fact, the estimated probability of infection is essentially 0 if the post-outbreak titer equals the pre-outbreak titer (0.00034, 95%CrI: $0.000014-0.0025$), is close to 1 in case of a fourfold antibody increase (0.994, 95%CrI: $0.963-1.0$), and

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.

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)
q_0	Probability of infection at no antibody increase	0.00034	(0.000014–0.0025)
q_1	Probability of infection at twofold antibody increase	0.19	(0.064–0.38)
q_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)

is almost 20% in case of a twofold increase (0.19, 95%CrI: $0.064-0.38$). The overall estimated cumulative incidence of infection is estimated at 0.095 (95%CrI: $0.082 - 0.11$) (Table 1).

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

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

3.3. Probability of infection as function of pre-outbreak antibody titer

We have plotted pre-titers as a function of the infection probability category (<10%, 10%–90%, and ≥90%)(4), to investigate to what extent
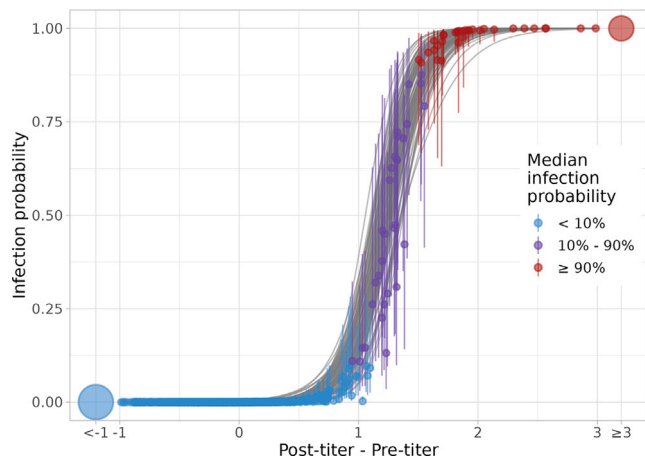
Fig. 3. Estimated infection probabilities as function of differences in antibody concentrations. Shown are 100 samples from the posterior distribution of the logistic function determining the probability that a sample belongs to the infected component, as function of the difference in the antibody titers (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. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 4. Pre-titer as function of the estimated probability of infection, stratified by infection probability. Shown are the pre-titers of all participants as function of the posterior median estimated infection probabilities (dots) together with corresponding boxplots of the distributions (interquartile and overall ranges).

the pre-titer can be gauged from the estimated infection probabilities, and to show the numbers in the various infection groups. The figure shows that the distribution of pre-titers in participants with low estimated probability of infection (<10%) is broad and covers not only high pre-titer values (>8, say) but includes low pre-titers as well, indicating that only a small fraction of participants with low titers have been infected. Conversely, the distribution of pre-titers in participants that are likely infected (probability ≥90%) is also broad, and includes participants with a pre-titer >8. Hence, a high pre-titer does not necessarily protect against infection. Importantly, the figure shows that for the majority of participants there is little uncertainty as to whether they have been infected or not. In fact, the estimated probability of infection is uncertain (i.e. between 10% and 90%) for only a small fraction of the study population ($\frac{31}{746} < 5\%$).

Next, we turn attention to the relation between pre-outbreak antibody titers and cumulative incidence of infection. Fig. 5 shows for all 746 participants the estimated infection probability with associated uncertainty as function of the pre-outbreak titer. In general, there is limited uncertainty for most participants, especially when the estimated probability of infection is close to 0 or close to 1. Only for a small subset of 31 participants with estimated posterior median probability of infection ≥10% and <90% there is substantial uncertainty.

To generalise these results for groups of persons and estimate the probability of infection as function of the pre-outbreak antibody titer we have stratified the study population by pre-outbreak antibody titers and calculated the probabilities of infection in these groups (black lines in 5). Specifically, we have grouped pre-outbreak titers in four equal sized groups, with titer ranges [2, 6.48), [6.48, 7.35), [7.35, 8.18), and [8.18, 15), corresponding to antibody concentrations of 4–89 RU/mL, 89–163 RU/mL, 163–290 RU/mL, and 290–32,768 RU/mL. The estimated probabilities of infection generally decrease with increasing pre-outbreak antibody titer. Specifically, the estimated infection probability is 0.12 (95%CrI: 0.10 − 0.13) for the lowest two quartiles, and 0.056 (95%CrI: 0.044 − 0.068) for the highest quartile. Hence, while there is no antibody titer that provides protection against infection, the probability of infection is more than double in the groups with low pre-outbreak antibody titers as compared to the group with highest antibody titers. The estimated overall probability of infection is 0.95
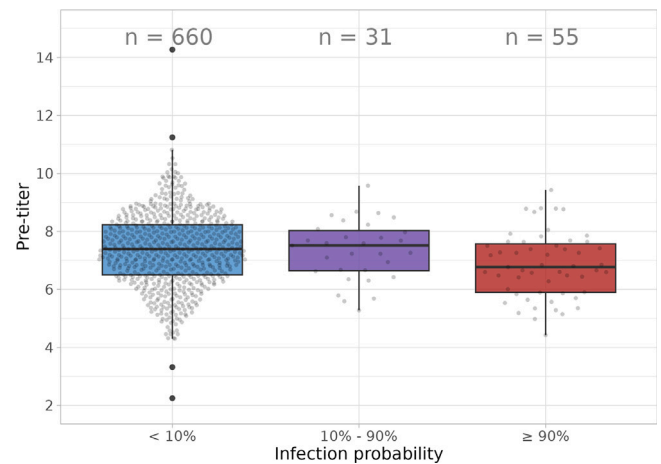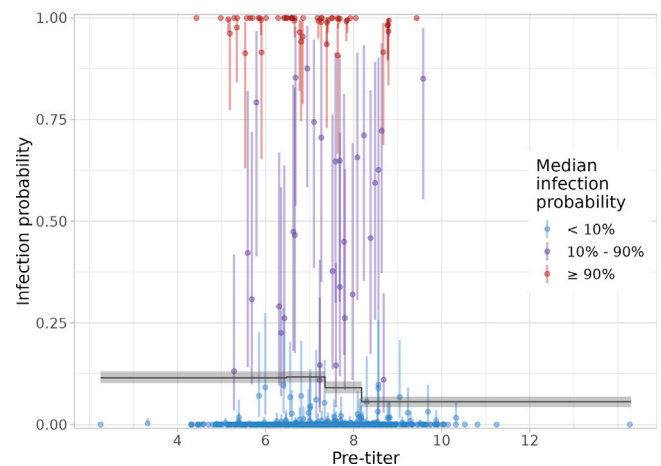

Fig. 5. Estimated infection probabilities as function of the pre-outbreak antibody titer. Shown are the estimated infection probabilities for all participants (dots: posterior medians; lines: 95% credible interval). Colours represent three infection probability classes (cf. Fig. 2). Also shown are estimated infection probabilities (median and 95% credible intervals), stratified by quartile of pre-outbreak antibody concentration.

(95%CrI: 0.082 − 0.11) 1. Thus, our results confirm earlier results that the probability of infection increases with decreasing antibody titers (Kaaijk et al., 2019), but yield substantially higher infection attack rates than reported earlier.

4. Discussion

The past two decades have witnessed a distinct increase in the number of outbreaks of mumps in highly vaccinated populations, mostly by non-vaccine genotypes, with an over-representation of adolescents and young adults, and predominantly in close-contact settings (households, schools and universities, parties) (Lam et al., 2020). This has prompted suggestions that a booster vaccination in adolescence could be beneficial (Cardemil et al., 2017; Lam et al., 2020; Kaaijk et al., 2020, 2019). Our analyses add to these results by showing that the infection attack rate has been high in Dutch universities (9.5%), and that the infection attack rate has been more than twice as high in students with low pre-existing mumps-specific IgG antibody concentrations than in students with high pre-existing antibodies (12% vs 5.6%). However, we also show there is no fixed pre-outbreak antibody concentration that provides full protection against infection. This indicates that not only IgG

5

antibodies but also other compartments of the immune system, such as cellular immunity, play a role in protection against mumps (Jokinen et al., 2007; de Wit et al., 2018). Nevertheless, young adults with low mumps-specific IgG antibody concentrations are at higher risk for infection and may therefore benefit most from a third dose of the MMR vaccine. Interestingly in this context, our previous study showed that individuals with lowest pre-vaccination IgG concentrations also showed the strongest increase in IgG concentrations after vaccination (Kaaijk et al., 2020).

Our analyses extend earlier analyses for the same outbreak (Gouma et al., 2014b; Kaaijk et al., 2019). A strength of the current analyses is that they are fully self-contained, and that all parameters, distributions, and infection rates are estimated from the data. This is a considerable advantage over previous analyses, as there is no fixed predefined level of antibody concentrations that is indicative of infection or protection against infection (Teunis et al., 2016; Teunis and van Eijkeren, 2020). Moreover, antibody concentrations in the population depend in a complex manner on time since last vaccination, potential previous infection(s), and outbreak genotype(s). Therefore, it is unlikely that proper sets of validation samples with confirmed uninfected and infected persons can be obtained. This is true not only for our population of university students but probably holds more generally. Thus, ad hoc choices have previously been made for the infection criterion (a four-fold antibody increase with additional requirements depending on the study). Comparing our results with the earlier analyses of Gouma et al. (2014b) and Kaaijk et al. (2019) we find a similar pattern of infection attack rates increase with decreasing prior antibody concentrations. Our analyses also indicate, however, that infection attack rates are substantially higher (9.5%) than reported earlier (4%–6%). This is due to the fact that many participants with modest increases of antibody titers (two- to threefold) may actually have been infected. In all, our results paint a more dynamic picture of the antibody dynamics induced by infection and waning than hitherto considered.

Previous analyses for bivariate serological cross-sectional studies (te Beest et al., 2014, 2015; Vink et al., 2016) have modelled the serological response with bivariate mixture distributions. Here, by virtue of the fact that we have paired data at our disposal we have been able to couple pre- and post-outbreak data using an infection function (Eq. (1)). This has enabled precise estimation of infection probabilities for the majority of participants even though the distributions of uninfected and infected persons show considerable overlap. In fact, we would not have been able to obtain precise estimates of infection probabilities if the data had been treated as two separate cross-sectional surveys (J. Gomme, unpublished).

In our analyses we have included 16 participants with a strong (at least fourfold) antibody decrease. These person may have had experienced a recent mumps infection before the first sampling, and may therefore not be representative for antibody concentrations in populations with no recent exposure. To study how the results are affected by this choice, we have performed a sensitivity analysis in which these samples are excluded. The results of these analyses are very similar to the main analyses. This is due to the fact that these 16 persons represent only a small minority of the study population (2%), and the fact that the estimated probability of infection for those participants is close to zero.

Two main limitations deserve scrutiny. First, we have throughout focused on probabilistic classification of participants and estimation of the serological infection attack rate. It would be desirable if the analyses could be extended to include not only serological data but also information on the severity of infection. It would be particularly interesting to study how the severity of infection depends on both the pre- and post-outbreak antibody concentrations. Second, our analyses apply to the specific situation of medical students in the Netherlands. This is a specific population and setting. For instance, in the Netherlands the vast majority of students have been vaccinated, and contact intensities in this group are probably substantially higher than in other strata of the population of similar age. It remains therefore an open question to which extent the results still apply in other settings. In particular, one could envisage that our estimated infection attack rates in medical students represent an upper bound for this age group in general.

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

CRediT authorship contribution statement

Michiel van Boven: Conceptualization, Formal analysis, Investigation, Methodology, Software, Writing – original draft. **Jantien A. Backer:** Conceptualization, Software, Writing – review & editing. **Irene Veldhuijzen:** Conceptualization, Writing – review & editing. **Justin Gomme:** Conceptualization, Software, Writing – review & editing. **Rob van Binnendijk:** Conceptualization, Data curation, Resources, Writing – review & editing. **Patricia Kaaijk:** Conceptualization, Data curation, Resources, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

References

Antia, A., Ahmed, H., Handel, A., Carlson, N.E., Amanna, I.J., Antia, R., Slifka, M., 2018. Heterogeneity and longevity of antibody memory to viruses and vaccines. PLoS Biol. 16 (8), e2006601.

te Beest, D.E., Birrell, P.J., Wallinga, J., De Angelis, D., van Boven, M., 2015. Joint modelling of serological and hospitalization data reveals that high levels of pre-existing immunity and school holidays shaped the influenza A pandemic of 2009 in the Netherlands. J. R. Soc. Interface 12 (103).

te Beest, D., de Bruin, E., Imholz, S., Wallinga, J., Teunis, P., Koopmans, M., van Boven, M., 2014. Discrimination of influenza infection (A/2009 H1N1) from prior exposure by antibody protein microarray analysis. PLoS One 9 (11), e113021.

Borgmann, S., Schwab, F., Santibanez, S., Mankertz, A., 2014. Mumps virus infection in vaccinated patients can be detected by an increase in specific IgG antibodies to high titres: A retrospective study. Epidemiol. Infect. 142 (11), 2388–2396.

Bouman, J.A., Kadelka, S., Stringhini, S., Pennacchio, F., Meyer, B., Yerly, S., Kaiser, L., Guessous, I., Azman, A.S., Bonhoeffer, S., Regoes, R.R., 2022. Applying mixture model methods to SARS-CoV-2 serosurvey data from Geneva. Epidemics 39, 100572.

Bouman, J.A., Riou, J., Bonhoeffer, S., Regoes, R.R., 2021. Estimating the cumulative incidence of SARS-CoV-2 with imperfect serological tests: Exploiting cutoff-free approaches. PLoS Comput. Biol. 17 (2), e1008728.

Cardemil, C.V., Dahl, R.M., James, L., Wannemuehler, K., Gary, H.E., Shah, M., Marin, M., Riley, J., Feikin, D.R., Patel, M., Quinlisk, P., 2017. Effectiveness of a third dose of MMR vaccine for mumps outbreak control. N. Engl. J. Med. 377 (10), 947–956.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., Riddell, A., 2017a. Stan: A probabilistic programming language. J. Stat. Softw. Articles 76 (1), 1–32.

Carpenter, B., Gelman, A., Hoffman, M.D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., Riddell, A., 2017b. Stan: A probabilistic programming language. J. Stat. Softw. 76 (1), 1–32.

Cortese, M.M., Barskey, A.E., Tegtmeier, G.E., Zhang, C., Ngo, L., Kyaw, M.H., Baughman, A.L., Menitove, J.E., Hickman, C.J., Bellini, W.J., Dayan, G.H., Hansen, G.R., Rubin, S., 2011. Mumps antibody levels among students before a mumps outbreak: In search of a correlate of immunity. J. Infect. Dis. 204 (9), 1413–1422.

Davidkin, I., Jokinen, S., Broman, M., Leinikki, P., Peltola, H., 2008. Persistence of measles, mumps, and Rubella antibodies in an MMR-vaccinated cohort: A 20-year follow-up. J. Infect. Dis. 197 (7), 950–956.

de Wit, J., Emmelot, M.E., Poelen, M.C.M., van Binnendijk, R.S., van der Lee, S., van Baarle, D., Han, W.G.H., van Els, C.A.C.M., Kaaijk, P., 2018. T-cell memory. J. Allergy Clin. Immunol. 141 (5), 1908–1911.

Dittrich, S., é, S., van Lier, A., Kohl, R., Boot, H., Koopmans, M., van Binnendijk, R., 2011. Assessment of serological evidence for mumps virus infection in vaccinated children. Vaccine 29 (49), 9271–9275.

Ferenczi, A., Gee, S., Cotter, S., Kelleher, K., 2020. Ongoing mumps outbreak among adolescents and young adults, Ireland, August 2018 to January 2020. Euro. Surveill. 25 (4).

Gouma, S., Sane, J., Gijselaar, D., Cremer, J., é, S., Koopmans, M., van Binnendijk, R., 2014a. Two major mumps genotype G variants dominated recent mumps outbreaks in the Netherlands (2009–2012). J. Gen. Virol. 95 (Pt 5), 1074–1082.

Gouma, S., Schurink-Van't Klooster, T.M., de Melker, H.E., Kerkhof, J., Smits, G.P., é, S.J., van Els, C.A., Boland, G.J., Vossen, A.C., Goswami, P.R., Koopmans, M.P., van Binnendijk, R.S., 2014b. Mumps serum antibody levels before and after an outbreak to assess infection and immunity in vaccinated students. Open Forum Infect. Dis. 1 (3), ofu101.

Gouma, S., Vermeire, T., Van Gucht, S., Martens, L., Hutse, V., Cremer, J., Rota, P.A., Leroux-Roels, G., Koopmans, M., van Binnendijk, R., Vandermarliere, E., 2018. Differences in antigenic sites and other functional regions between genotype A and G mumps virus surface proteins. Sci. Rep. 8 (1), 13337.

Jokinen, S., Osterlund, P., Julkunen, I., Davidkin, I., 2007. Cellular immunity to mumps virus in young adults 21 years after measles-mumps-rubella vaccination. J. Infect. Dis. 196 (6), 861–867.

Kaaijk, P., Nicolaie, M.A., van Rooijen, D., van Houten, M.A., van der Klis, F.R., Buisman, A.-M., van Binnendijk, R.S., 2020. Dynamics of the antibody response after a third dose of measles-mumps-rubella vaccine indicate a slower decline compared with a second dose. Open Forum Infect. Dis. 7 (11), ofaa505.

Kaaijk, P., Wijmenga-Monsuur, A.J., van Houten, M.A., Veldhuijzen, I.K., ten Hulscher, H.I., Kerkhof, J., van der Klis, F.R., van Binnendijk, R.S., 2019. A third dose of measles-mumps-rubella vaccine to improve immunity against mumps in Young adults. J. Infect. Dis. 221 (6), 902–909.

Kaaijk, P., Wijmenga-Monsuur, A.J., ten Hulscher, H.I., Kerkhof, J., Smits, G., Nicolaie, M.A., van Houten, M.A., van Binnendijk, R.S., 2022. Antibody levels at 3-years follow-up of a third dose of measles-mumps-rubella vaccine in Young adults. Vaccines 10 (1).

Kontio, M., Jokinen, S., Paunio, M., Peltola, H., Davidkin, I., 2012. Waning antibody levels and avidity: Implications for MMR vaccine-induced protection. J. Infect. Dis. 206 (10), 1542–1548.

Lam, E., Rosen, J.B., Zucker, J.R., 2020. Mumps: An update on outbreaks, vaccine efficacy, and genomic diversity. Clin. Microbiol. Rev. 33 (2).

Moncla, L.H., Black, A., DeBolt, C., Lang, M., Graff, N.R., rez Osorio, A.C., ller, N.F., Haselow, D., Lindquist, S., Bedford, T., 2021. Repeated introductions and intensive community transmission fueled a mumps virus outbreak in Washington State. Elife 10.

Patel, L.N., Arciuolo, R.J., Fu, J., Giancotti, F.R., Zucker, J.R., Rakeman, J.L., Rosen, J.B., 2017. Mumps outbreak among a highly vaccinated University Community-New York City, January-April 2014. Clin. Infect. Dis. 64 (4), 408–412.

Perez Duque, M., San-Bento, A., on, L., dio, P., a, M.A., Albuquerque, M.J., Nascimento, M., Balasegaram, S., Machado, R., 2021. Mumps outbreak among fully vaccinated school-age children and young adults, Portugal 2019/2020. Epidemiol. Infect. 149, e205.

Rubin, S.A., Link, M.A., Sauder, C.J., Zhang, C., Ngo, L., Rima, B.K., Duprex, W.P., 2012. Recent mumps outbreaks in vaccinated populations: No evidence of immune escape. J. Virol. 86 (1), 615–620.

Sane, J., Gouma, S., Koopmans, M., de Melker, H., Swaan, C., van Binnendijk, R., é, S., 2014. Epidemic of mumps among vaccinated persons, The Netherlands, 2009–2012. Emerg. Infect. Dis. 20 (4), 643–648.

Shah, M., Quinlisk, P., Weigel, A., Riley, J., James, L., Patterson, J., Hickman, C., Rota, P.A., Stewart, R., Clemmons, N., Kalas, N., Cardemil, C., Patel, M., Donahue, M., Schneider, A., Ukegbu, U., Wittich, K., Kellogg, J., Beardsley, D., Tran, N., Callaghan, D., Pyatt, A., Kitzmann, T., Kintigh, B., 2018. Mumps outbreak in a highly vaccinated university-affiliated setting before and after a Measles-Mumps-Rubella Vaccination Campaign-Iowa, July 2015-May 2016. Clin. Infect. Dis. 66 (1), 81–88.

Smits, G.P., van Gageldonk, P.G., Schouls, L.M., van der Klis, F.R., Berbers, G.A., 2012. Development of a bead-based multiplex immunoassay for simultaneous quantitative detection of IgG serum antibodies against measles, mumps, Rubella, and varicella-zoster virus. Clin. Vaccine Immunol. 19 (3), 396–400.

Smits, G., Mollema, L., é, S., de Melker, H., Tcherniaeva, I., Waaijenborg, S., van Binnendijk, R., van der Klis, F., Berbers, G., 2013. Seroprevalence of mumps in The Netherlands: Dynamics over a decade with high vaccination coverage and recent outbreaks. PLoS One 8 (3), e58234.

Steens, A., Waaijenborg, S., Teunis, P.F., Reimerink, J.H., Meijer, A., van der Lubben, M., Koopmans, M., van der Sande, M.A., Wallinga, J., van Boven, M., 2011. Age-dependent patterns of infection and severity explaining the low impact of 2009 influenza A (H1N1): Evidence from serial serologic surveys in the Netherlands. Am. J. Epidemiol. 174 (11), 1307–1315.

Teunis, P.F.M., van Eijkeren, J.C.H., 2020. Estimation of seroconversion rates for infectious diseases: Effects of age and noise. Stat. Med. 39 (21), 2799–2814.

Teunis, P.F., van Eijkeren, J.C., de Graaf, W.F., ć, A.B., Kretzschmar, M.E., 2016. Linking the seroresponse to infection to within-host heterogeneity in antibody production. Epidemics 16, 33–39.

van Boven, M., Ruijs, W.L., Wallinga, J., O'Neill, P.D., é, S., 2013. Estimation of vaccine efficacy and critical vaccination coverage in partially observed outbreaks. PLoS Comput. Biol. 9 (5), e1003061.

Veneti, L., Borgen, K., Borge, K.S., Danis, K., Greve-Isdahl, M., Konsmo, K., lstad, G., ø, S.A., ystese, K.S., Rykkvin, R., Sagvik, E., Riise, R., 2018. Large outbreak of mumps virus genotype G among vaccinated students in Norway, 2015 to 2016. Euro. Surveill. 23 (38).

Vink, M.A., Berkhof, J., van de Kassteele, J., van Boven, M., Bogaards, J.A., 2016. A bivariate mixture model for natural antibody levels to human papillomavirus types 16 and 18: Baseline estimates for monitoring the herd effects of immunization. PLoS One 11 (8), e0161109.

Vink, M.A., van de Kassteele, J., Wallinga, J., Teunis, P.F., Bogaards, J.A., 2015. Estimating seroprevalence of human papillomavirus type 16 using a mixture model with smoothed age-dependent mixing proportions. Epidemiology 26 (1), 8–16.

Westphal, D.W., Eastwood, A., Levy, A., Davies, J., Huppatz, C., Gilles, M., Lyttle, H., Williams, S.A., Dowse, G.K., 2019. A protracted mumps outbreak in Western Australia despite high vaccine coverage: A population-based surveillance study. Lancet Infect. Dis. 19 (2), 177–184.

Willocks, L.J., Guerendiain, D., Austin, H.I., Morrison, K.E., Cameron, R.L., Templeton, K.E., DE Lima, V.R.F., Ewing, R., Donovan, W., Pollock, K.G.J., 2017. An outbreak of mumps with genetic strain variation in a highly vaccinated student population in Scotland. Epidemiol. Infect. 145 (15), 3219–3225.