American Journal of Epidemiology **Submitted Manuscript**

Title: Correcting for Antibody Waning in Cumulative Incidence Estimation from Sequential Serosurveys

Authors: Sarah Kadelka, Judith A Bouman, Peter Ashcroft, Roland R Regoes

Correspondence Address: sarah.kadelka@env.ethz.ch, roland.regoes@env.ethz.ch

Affiliations: Institute of Integrative Biology, ETH Zurich, Zurich, Switzerland (Sarah Kadelka, Judith A Bouman, Peter Ashcroft, Roland R Regoes);

Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland (Judith A Bouman);

Novartis Institutes for BioMedical Research, Basel 4056, Switzerland (Peter Ashcroft)

Funding: Roland Regoes gratefully acknowledges funding from the Botnar Research Centre for Child Health, Switzerland (grant number 2020-FS-354).

Data Availability Statement: The data are provided in the supporting R-package.

Thanks: We would like to emphasize that this work was only possible because the authors of "Three-quarters attack rate of SARS-CoV-2 in the Brazilian Amazon during a largely unmitigated epidemic" [Science, 2021] generously and comprehensively shared their data and we would like to thank Ester Sabino and her group for their critical feedback. We thank Sonja Lehtinen for valuable comments on the manuscript.

Conference presentation: N/A

Preprint Information: N/A

Disclaimer: The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of Interest: The authors report no conflict of interest.

Running Head: Cumulative Incidence Estimation from Sequential Serosurveys

Key words: serosurveys, cumulative incidence, SARS-CoV-2, infectious disease

Abbreviations: anti-N, anti-nucleocapsid protein; CI, confidence interval; CR, confidence region; IgG/M, immunoglobulin G/M; SARS-CoV-2 , severe acute respiratory syndrome *coronavirus* type 2; S/C ratio, signal-to-cutoff ratio

Correcting for antibody waning in cumulative incidence estimation from sequential serosurveys

Sarah Kadelka[1,*], Judith A Bouman[1,2], Peter Ashcroft[1,3], and Roland R Regoes[1,*]

[1]Institute of Integrative Biology, ETH Zurich, Zurich, Switzerland
[2]Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland
[3]Novartis Institutes for BioMedical Research, Basel 4056, Switzerland
[*]Corresponding authors: Sarah Kadelka and Roland R. Regoes, (sarah.kadelka@env.ethz.ch, roland.regoes@env.ethz.ch).

Abstract

Serosurveys are a widely used tool to estimate the cumulative incidence, i.e. the fraction of a population that have been infected by a given pathogen. These surveys rely on serological assays that measure the level of pathogen-specific antibodies. Because antibody levels are waning, the fraction of previously infected individuals that have sero-reverted increases with time past infection. To avoid underestimating the true cumulative incidence, it is therefore essential to correct for waning antibody levels. We present an empirically-supported approach for sero-reversion correction in cumulative incidence estimation when sequential serosurveys are conducted in the context of a newly emerging infectious disease. The correction is based on the observed dynamics of antibody titers in sero-positive cases and validated using several *in silico* test scenarios. Furthermore, through this approach we revise a previous cumulative incidence estimate, which relies on the assumption of an exponentially-declining probability of sero-reversion over time, of SARS-CoV-2 of 76% in Manaus, Brazil, by October 2020 to 47.6% (43.5% - 53.5%). This estimate has implications e.g. for the proximity to herd immunity in Manaus in late 2020.

Introduction

Understanding how much an emerging infectious disease has spread in a population or geographical region is critical for the estimation of epidemiological parameters such as the case ascertainment rate or infection fatality rate which consequently influence decision making on measures undertaken to contain the disease (1). One important tool for assessing the cumulative incidence of an infectious disease are serosurveys, in particular when a large fraction of infections pass symptom-free and hence remains undetected (2, 3, 4).

In serosurveys, depending on the purpose and context of the study, the presence of antibodies against a certain antigen is used as a marker for past infection, vaccination or immunity(5). Serosurveys have been conducted for a wide range of antigens, such as for example the measles and rubella viruses (5) and the Zika virus (6, 7). A large body of serosurveys have also been performed to estimate the seroprevalence of SARS-CoV-2 in various regions of the world (8).

Antibody persistence ranges from life-long for some viral infections, such as measles and yellow fever (9) to a relatively fast decay for other viral infections such as West Nile (10), seasonal coronaviruses (11) and SARS-CoV-2 (12, 13, 14). Rapidly decaying antibody levels result in an increasing likelihood of sero-negativity over time despite prior antigen exposure. This transition of antibody levels from above a positivity threshold (sero-positive) to below the threshold (sero-negative) is called sero-reversion and should be corrected for when estimating cumulative incidence from serological surveys. This correction gains in importance as the time from first antigen exposure in the population to the serosurvey grows. The lack of such correction can result in significant underestimates of cumulative incidence. Prior to the current SARS-CoV-2 pandemic, one study considering measles-mumps-rubella vaccine coverage in Australia corrected for sero-reversion by assuming fixed sero-reversion rates for all three antibodies. These rates, in addition to the vaccine coverage, were fitted using serial serosurvey data (15). Another study estimated the temporal IgM antibody profile in West Nile virus infected individuals to derive time-dependent estimates for the probability of IgM-seropositivity (16). These, in combination with the temporal profile of reported cases, were used to estimate the cumulative incidence of the West Nile virus infection in the North Texas region during the 2012 epidemic from serological survey data.

Many serological studies have been conducted in the current SARS-CoV-2 epidemic (8). However, despite the wide-spread acknowledgement of the importance of sero-reversion correction (17, 18, 19), only few studies have actually done so (19, 20, 21, 22, 23, 24). In one such study Buss *et al.* estimated a cumulative incidence of 76% (95% CI, 66.6% to 97.9%) in Manaus, Brazil, by October 2020 compared to an uncorrected 25.8% (20.9%-31.3%) (23) using data from monthly serosurveys between March and October 2020. While this cumulative incidence should

have conferred herd immunity and curbed the epidemic ([25], [26], [27]), Manaus was hit by a very strong second wave in January 2021 ([28], [29]). Multiple explanations for these puzzling patterns have been proposed: methodological issues relating to cumulative incidence estimation, waning of immunity, and new viral variants that evade immunity from previous infection or have increased transmissibility compared to the initially circulating variant ([28]).

Here, we describe a cutoff-based approach for cumulative incidence estimation of an emerging infectious disease that combines elements from several of the studies mentioned above. It requires repeated serosurveys (as in ([23])) and makes use of antibody kinetic data (as in ([16])) to correct the estimates for sero-reversion . In contrast to this empirical derivation of the distribution of times from seroconversion to sero-reversion (similar to ([24])), previous methods by Buss *et al.* and Shioda *et al.* assumed exponentially or Weibull-distributed sero-reversion times ([19], [23]). We validate the method for several *in silico* test cases and investigate the impact of various assumptions on the performance of the proposed method. We then apply the method to the Manaus data from Buss *et al.* and suggest that the cumulative incidence estimate of 76% in Manaus represents an over-estimation.

Methods

Antibody waning is commonly observed after recovery from acute infections and can lead to sero-reversion from a sero-positive to a sero-negative state over time ([30], [31]). To estimate cumulative incidence with correction for such sero-reversion we propose a maximum-likelihood method that incorporates empirically-derived probabilities of sero-reversion. These can be derived from the dynamics of antibody decay.

Distribution of sero-reversion times

To derive empirically supported distributions of the times to sero-reversion we need to make assumptions about the functional form of the antibody kinetics within infected individuals. Specifically, we assume no delay between sero-conversion (transition from sero-negative to sero-positive state) and peak antibody levels, both occurring a fixed duration t_{rec} after time of infection, and define an individual to be recovered at the time of sero-conversion t_{rec}. Furthermore, the term *uninfected individual* refers to any individual who has either not been infected or infected for less than time t_{rec}, while the terms *infected* and *recovered individuals* are used interchangeably for those who were infected longer than time t_{rec} ago.

The distribution of sero-reversion times is derived from the distributions of quantitative antibody measures of positive controls at the time of peak antibody level, negative (pre-pandemic) controls and the distribution of antibody decay rates in positive controls under the assumption that antibodies decay exponentially after their peak. A schematic representation is shown in Figure 1 and details are given in the Web Appendix 1.

Likelihood function for dichotomized data from sequential serosurveys

Maximizing the likelihood function that explains dichotomized (sero-positive or sero-negative) antibody data observed in a serosurvey is a common approach for seroprevalence estimation ([32], [33], [34], [35]). The form of the likelihood and the number of optimized parameters depend on the available data and exact goal of the study. We derive a likelihood function that aims at explaining not only the results from a single serosurvey, but a sequence of serosurveys conducted at regular intervals of one unit of time (e.g. one month or one week) starting shortly after the first recoveries in the population. We assume that the individuals sampled at each time point represent independent draws from the population. We refer to such a sequence of serosurveys as one study. By integrating our control-data-derived knowledge on sero-reversion time distribution, and consequently the probability to sero-revert within t units of time after recovery, $p_{rev}^{|pos,con}(t)$, and our knowledge on test sensitivity ($sens$) and specificity ($spec$), we correct for sero-reversion of individuals due to antibody waning and for test accuracy. Let $\boldsymbol{Z}$ be a $\{0,1\}^{\sum_{j=0}^{m} n_j}$ valued random variable describing the test results of all study participants in a study consisting of $m+1$ surveys with n_j survey participants in survey j. Here, 0 represents a negative and 1 a positive serological test result. Then, the log-likelihood for the unknown per capita new infections between two surveys, $r_0, \ldots, r_m$, given the data $\boldsymbol{Z} = \boldsymbol{z}$, is given by

$$\begin{aligned} &\log L(r_0, \ldots, r_m | \boldsymbol{Z} = \boldsymbol{z}) \\ &\quad = \sum_{j=0}^{m} \Big\{ N_j \times \log(p_j(\boldsymbol{\Theta}_j)) + (n_j - N_j) \times \log(1 - p_j(\boldsymbol{\Theta}_j)) \Big\}, \end{aligned} \tag{1}$$

where the total number of survey participants, n_j, and the number of survey participants with positive test results, N_j, are extracted from the the data $\boldsymbol{z}$, $\boldsymbol{\Theta}_j$ is defined by $\left(sens, spec, p_{rev}^{|pos,con}(0{+}1/2), \ldots, p_{rev}^{|pos,con}(j{+}1/2), r_0, \ldots, r_j \right)$

and $p_j(\boldsymbol{\Theta}_j)$ is the probability that a randomly drawn individual in the population has a positive test result at the time of survey j. For a detailed derivation of the log-likelihood function and $p_j(\boldsymbol{\Theta}_j)$ see the Web Appendix 1. The arguments r_i^* which maximize the log-likelihood function in equation 1 yield the cumulative incidence estimates $c_j^* = \sum_{i=0}^{j} r_i^*$. For a detailed description of the optimization routine and on estimating confidence regions for the cumulative incidence estimates that account for uncertainties arising from both validation and study data, see Web Appendix 1.

Simulations

Creating *in silico* studies – Test cases

Our method is tested using several *in silico* studies. Each *in silico* study consists of both validation and study data, with study data comprising data from a sequence of serosurveys. In brief, these data are sampled from the assumed true underlying distributions of peak level, background level and decay rate of antibodies (see Web Figure 1) , as described in more detail in the Web Appendix 1.

Performance analysis

We define the statistical power of this method as the proportion of *in silico* studies that estimate the cumulative incidence successfully at at least 75% of all survey times within the study and semi-successfully at the time of the latest survey within the study. Success is defined as a relative difference of less than 10% or an absolute difference of less than 2% between estimated and true cumulative incidence in addition to a difference of less than 2 standard deviations between estimated and true cumulative incidence. A semi-success is reached if the relative difference between estimated and true cumulative incidence is less than 20% or the absolute difference is less than 3%. It is important to note that what we call power is different from the classical statistical power for hypothesis tests.

Results

Theoretical results – Method validation via *in silico* studies

To validate the described method for cumulative incidence estimation from sequential serosurveys, we consider seven test cases where an epidemic is observed for 9 months (in 5 test cases) and 15 and 17 months in one test case each. In these test cases the true epidemic size is ranging from consistently low, with monthly per capita incidences below 2%, to strong waves with up to 30% of the population infected within a single month. An overview summarizing the seven test cases is given in Table 1. (For a detailed description see Web Appendix 1 and equation 8 therein). *In silico* studies are created under the simplifying assumption that recoveries are uniformly distributed between any two consecutive surveys (alternative assumptions are discussed in the Web Appendix 2 and Web Figures 2-5). Furthermore, the number of surveyed individuals in any survey of any *in silico* study is a randomly drawn integer between 800 and 900, the number of peak and background antibody levels in the validation data of any *in silico* study is set to 900 and the number of antibody decay rates in the validation data is fixed to 100. For each test case $N_{sim} = 10000$ *in silico* studies are conducted and cumulative incidences are estimated by maximizing the log-likelihood function in equation 1 with fixed parameters (sensitivity, specificity and seroreversion probabilities) derived from the validation data of the respective *in silico* study. Figure 2 shows good agreement between true cumulative incidences and cumulative incidences fitted to data from the *in silico* studies, with powers ranging from 87.4% for test case 4 to 96.1% for test case 2 (see Web Table 1, row 1). The impact of variations in the delay between two consecutive surveys, in the correlation between peak antibody levels and antibody decay rates in the number of participants per survey and in the assumed shape of the distribution of sero-reversion times on the method's performance are discussed in the Web Appendix 2 and Web Figures 6-15. These analyses include several cases of mismatches between the model used for simulating and the one assumed when fitting the data, e.g. different correlations or different distributions of sero-reversion times (empirically derived for simulation vs various others when fitting). Additionally, we compare the impact of a constant population followed throughout the various surveys within one study instead of disjoint survey populations.

A real world example – Estimating the spread of SARS-CoV-2 in Manaus from March to October 2020

In their paper on the attack rate of SARS-CoV-2, which corresponds to what we call cumulative incidence, in the Brazilian Amazon (23), Buss *et al.* used data on anti-N IgG levels (signal/cutoff (S/C) ratios) from repetitive (monthly from March-October 2020) but independent serosurveys in Manaus to estimate the monthly cumulative

incidence. Here, independence means that different cohorts comprising 800-900 individuals each, sampled from the same population (blood donors in Manaus), were surveyed at each time point. In addition to the serosurvey data, Buss *et al.* reported anti-N IgG levels of convalescents 20-50 days past symptom onset (positve controls), anti-N IgG gradients in convalescents and anti-N IgG levels of pre-pandemic controls (negative controls). A detailed description of these validation data can be found in the Web Appendix 1.

As anti-N IgG is estimated to peak approximately $3-4$ weeks after symptom onset, we assume that the levels from the positive control group are a good representation of peak IgG levels (12). Consequently, when deriving the distribution of sero-reversion times as described in i)–iii) (Web Appendix 1), and schematically represented in Figure 1, we assume that peak and background antibody levels are distributed according to the empirical distributions of the positive and the negative control groups, respectively (see Figure 3A). Meanwhile, since we assume antibody levels are decreasing, we ignore positive gradients (7/88) and fit a gamma distribution to the absolute values of the negative gradients (81/88; see Figure 3B). The derived distribution of sero-reversion times is shown in Figure 3C. Since serosurveys are performed monthly, we extract the probabilities to sero-revert within the first 0.5–8.5 months after peak antibody titer (see Figure 3D, black curve). We estimate that within 6 months past seroconversion approximately 46% (38-53%) sero-revert. This is roughly in accordance with the results reported in (17) but lower than the 81% found in (23) (see Web Figure 16A for a comparison of sero-reversion times derived using the empirical approach described in this article (red curve) and those derived by Buss *et al.* (green curve)). Buss *et al.* did not fit the sero-reversion rate to longitudinal antibody data but rather derived it as a side product when estimating monthly cumulative incidence from serosurvey data.

When correcting the cumulative incidence estimate for sero-reversion using the empirically-supported profile, we obtain a cumulative incidence estimate of 47.6% (bootstrap 95% CR, 43.5 to 53.5%) in October 2020 — roughly 30 percentage points lower than Buss *et al.'s* estimate of 76.0% and outside the 95% confidence interval (66.6% to 97.9%), i.e. significantly lower (see Figure 4, solid black curve and Table 2). The cumulative incidence increased from 0.8% (95% CR: (0,1.7)%) in March 2020 to 47% (95% CR: (42.3,50)%) in May 2020 and has been almost constant from May through October 2020 with minimal non-significant increases in June. These results are consistent with the estimated cumulative incidence of 41.53%-44.82% in Manaus 6 months after the start of the epidemic determined in the DETECTCoV-19 cohort study (36). Our cumulative incidence estimate in June is significantly lower than Buss *et al.'s* age, sex, sensitivity and specificity adjusted estimate. This suggests that the raw seroprevalence estimates observed in June may be high, based on some bias introduced by the convenience sample of blood donors. The selection bias hypothesis is further supported by re-estimating cumulative incidences when dropping the data from the June survey. Compared to using data from all surveys (including June) this reduced the cumulative incidence estimates from May through October (reductions not significant, see Web Figure 17). By contrast, dropping the data from either the May or July surveys rendered the estimates almost unchanged (except for wider confidence intervals due to reduced number of samples).

Discussion

In this paper we have proposed an empirically-supported method for sero-reversion correction in serosurveys. The method can be used in any situation where a disease enters a previously naive population, individuals are protected from reinfection for the duration of the study and antibodies against the new antigen wane exponentially in recovered individuals. It also requires the availability of a quantitative serological assay for antibodies against the new antigen, validation data from individuals with known past infection and from individuals prior to the introduction of the disease in the population. This empirical approach to sero-reversion correction is based on the distribution of sero-reversion times after recovery, the estimation of which requires peak antibody levels and decay rates, i.e. longitudinal data, of positive controls.

Considering seven different test scenarios, simulating studies consisting of repeated sero-surveys under various assumed incidence curves, has shown that, in general, the method successfully approximates the cumulative incidences at the times of the surveys. At times when the disease incidence increases, the method over-corrects for sero-reversion while it under-corrects when disease incidence decreases (see Web Appendix 2). The strength of under- or over-correction depends on the rate of antibody waning and the delay between consecutive surveys. Thus, if antibody waning is fast, or disease incidence changes rapidly, frequent sampling is required (see Web Appendix 2). The method can be improved in the future by allowing the integration of information on the general shape of the incidence curve during the time of the study.

For all test scenarios maximal power has been reached when sampling $10^{3.5}$ or more individuals per survey (see Web Appendix 2). Lower sample sizes resulted in reduced powers while sampling more individuals did not increase the method's power. These results were based on the assumption that the cohorts tested in the different surveys within one study are disjoint. If instead a constant cohort was followed over time, the proposed method in general still succeeded at estimating the monthly cumulative incidences for large enough cohort sizes. However, the size of a constant cohort required to reach the same power as when using disjoint cohorts needed to be greater than the

sizes of the disjoint cohorts (see Web Figure 13). The reason for this is that once infected future antibody levels are pre-determined (by peak and decay rate of the individual's antibody level). Hence the number of survey participants that yield new information reduces with every survey.

Often sero-reversion is not corrected for in serosurveys, which in the context of antibodies that decay relatively fast, can lead to significant underestimation of cumulative incidences. If we had ignored to correct for sero-reversion by setting the corresponding probabilities to zero (see Web Appendix 2), we would have failed to correctly estimate cumulative incidence in all test cases with the exception of test case 6 (see Web Table 2). Two previous methods for sero-reversion correction in the context of sequential serosurveys, introduced by Buss *et al.* (23) and Shioda *et al.* (19), assumed exponentially distributed sero-reversion times (times from seroconversion to sero-reversion) (23) or Weibull-dsitributed sero-reversion times (19). Due to identifiability reasons, Shioda *et al.* assumed a fixed standard deviation of 50 days for this Weibull-distribution. By simulating study data using empirical data on antibody kinetics and using sero-reversion probabilities derived under the assumptions that sero-reversion times are exponentially or Weibull-distributed in the model used for fitting the simulated data (model mismatch), we have shown that these previous assumptions are in conflict with empirical data on antibody kinetics (see Web Appendix 2).

Previous studies have reported a correlation between peak antibody levels and antibody decay rates after e.g. SARS-CoV-2 infection. We have shown that if the validation data approximately mirrors the true underlying correlation, our method, in general, performs well at estimating the cumulative incidence irrespective of the strength of the correlation (see Web Appendix 2). In some cases the method even predicted the cumulative incidence accurately if the correlation in the validation data and the true correlation did not match. However this was not true for all test scenarios.

The presented method can be applied to stratified data (e.g. age-stratified) by estimating cumulative incidences for each subpopulation individually and if needed combining the stratified estimates into a weighted population average. If not only the cumulative incidences but also the antibody dynamics vary between subpopulation (as might be the case for young vs elderly), then stratified validation data is required and sero-reversion probabilities need to be estimated separately for each subpopulation.

Many infectious diseases do not confer perfect immunity after infection. In situations where (i) the aim of a study is to - at the time of each survey - estimate the fraction of individuals who have been infected at least once and (ii) antibody dynamics after reinfection resembles antibody dynamics after primary infection (in terms of peak, decay and background level), the method can be adapted to account for reinfections. To this end, the fractions of individuals who had recovered for a given amount of time (say a units) at the time of a given survey needs to be replaced by the fractions of individuals whose latest recovery had been a units of time before the given survey. The precise definition of this depends on if, for how long and at what level infection confers immunity and is outside the scope of this article.

Recently, it has been shown that if quantitative antibody measurements are available, cutoff-free methods which avoid dichotomizing study participants into antibody positive or negative are beneficial compared to cutoff-based methods when estimating cumulative incidence from a single serological survey (34, 37). In the future, we plan to adapt the approach presented in this article and introduce a similar cutoff-free approach for cumulative incidence estimation from sequential cross-sectional serological surveys.

Applying our method to serosurvey data from Manaus (23) suggested that the previously reported cumulative incidence estimate of 76% by October 2020 is a significant over-estimation. We predicted an approximately 30 percentage points lower cumulative incidence of 47.6% (43.5% to 53.5%) which is in line with (36). Similar to Buss *et al.*, we estimated the cumulative incidence under the assumption that incidence can only increase. The observed seroprevalences show large drops from May to June to July. To explain this, Buss *et al* predicted large parts of seroconverted individuals to serorevert within the first or second month past seroconversion. This however, is not in accordance with the anti-N IgG dynamics observed in the convalescent control group (see Web Figure 16A, green vs red curves). By contrast, our method based the sero-reversion correction on the antibody dynamics observed in convalescent plasma donors and predicted cumulative incidence estimates in June significantly below the respective sensitivity and specificity adjusted seroprevalence estimates, suggesting a possible selection bias in the data. This hypothesis of selection bias in the June serosurvey was also supported by re-estimation of cumulative incidences when ignoring data from a single serosurvey. While ignoring the data from May or July left the estimates almost unchanged, they were reduced when ignoring the June serosurvey. The data from convalescent individuals showed some evidence for a correlation between anti-N IgG peak and decay rate. Accounting for this in the sero-reversion probabilities (see Web Figure 18A), however, did not result in significantly different cumulative incidence estimates (see Web Appendix 2 and Web Figure 18B). The distributions of positive control peak antibody levels and decay rates and the distribution of negative control antibody levels used to derive the *in silico* studies in the test scenarios were chosen very similar to those observed in the validation data from Manaus. When each survey consisted of 800-900 individuals, our method displayed a relatively high power of more than 75% for all test scenarios under the assumption of disjoint cohorts at each survey within one study. It is not clear to us whether this is guaranteed in the Manaus data set or whether there is some overlap of the cohorts. However, since the predicted powers of the proposed method still ranged above 67% even if a constant cohort was followed through time (see Web Figure 13),

these results justify a certain degree of trust in our cumulative incidence estimates for Manaus.

As a byproduct, this method returned the distribution of sero-reversion times estimated from positive controls' dynamics. From this distribution one could derive the fraction of individuals that have sero-reverted at a given time past recovery. We have predicted from the Manaus data set that for anti-N using the Abbott Architect SARS-CoV-2 IgG assay 46% of seroconverters sero-revert within the first 6 months past seroconversion, lower than the 81% found by Buss *et al.* (23) but roughly in accordance with Krutikov *et al.* (17).

The positive control data from Manaus that was used to derive the sero-reversion probability is only representative of symptomatic, non-hospitalised COVID-19 cases. However, antibody levels and therefore time to sero-reversion vary with disease severity (38). In the Web Appendix 2 we compared the fitted cumulative incidences with those obtained when using a set of positive controls that is closer to the survey data in terms of disease severity. We found that, while estimated cumulative incidences are slightly larger, due to faster sero-reversion derived from the alternative positive control group, the estimated cumulative incidence in October 2020 is still significantly below that estimated by Buss *et al.*.

Even though the alternative, empirically-supported sero-reversion correction does not provide a full explanation for the unexpected resurgence of the SARS-CoV-2 epidemic in Manaus, it contributes to solving the puzzle by providing a lower estimate of the cumulative incidence. Beyond its relevance for the Brazilian sero-survey, the approach to adjusting for sero-reversion presented here provides an important, empirically-supported method that could be used in any sero-survey in which the antibody levels wane over time.

Acknowledgments

We would like to emphasize that this work was only possible because the authors of (23) generously and comprehensively shared their data and we would like to thank Ester Sabino and her group for their critical feedback.
We thank Sonja Lehtinen for valuable comments on the manuscript.
Roland Regoes gratefully acknowledges funding from the Botnar Research Centre for Child Health, Switzerland (grant number 2020-FS-354).

References

[1] Fontanet A, Tondeur L, Grant R, et al. SARS-CoV-2 infection in schools in a northern French city: a retrospective serological cohort study in an area of high transmission, France, January to April 2020. *Eurosurveillance*. 2021, 26(15):2001695.

[2] Buitrago-Garcia D, Egli-Gany D, Counotte MJ, et al. Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: A living systematic review and meta-analysis. *PLoS Med*. 2020, 17(9):e1003346.

[3] Byambasuren O, Cardona M, Bell K, et al. Estimating the extent of asymptomatic COVID-19 and its potential for community transmission: systematic review and meta-analysis. *J Assoc Med Microbiol Infect Dis Can*. 2020, 5(4):223–234.

[4] Metcalf CJE, Farrar J, Cutts FT, et al. Use of serological surveys to generate key insights into the changing global landscape of infectious disease. *Lancet*. 2016, 388(10045):728–730.

[5] Dimech W and Mulders MN. A 16-year review of seroprevalence studies on measles and rubella. *Vaccine*. 2016, 34(35):4110–4118.

[6] Saba Villarroel PM, Nurtop E, Pastorino B, et al. Zika virus epidemiology in Bolivia: A seroprevalence study in volunteer blood donors. *PLoS Negl Trop Dis*. 2018, 12(3):e0006239.

[7] Flamand C, Bailly S, Fritzell C, et al. Impact of Zika virus emergence in French Guiana: A large general population seroprevalence survey. *J Infect Dis*. 2019, 220(12):1915–1925.

[8] Chen X, Chen Z, Azman AS, et al. Serological evidence of human infection with SARS-CoV-2: a systematic review and meta-analysis. *Lancet Glob Health*. 2021.

[9] Paul WE, ed. *Fundamental Immunology*. 5th edition. Philadelphia, PA. Lippincott Williams & Wilkins, 2003.

[10] Busch MP, Kleinman SH, Tobler LH, et al. Virus and antibody dynamics in acute West Nile virus infection. *J Infect Dis*. 2008, 198(7):984–993.

[11] Callow K, Parry H, Sergeant M, et al. The time course of the immune response to experimental coronavirus infection of man. *Epidemiol Infect*. 1990, 105(2):435–446.

[12] Röltgen K, Powell AE, Wirz OF, et al. Defining the features and duration of antibody responses to SARS-CoV-2 infection associated with disease severity and outcome. *Sci Immunol.* 2020, 5(54).

[13] Dan JM, Mateus J, Kato Y, et al. Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection. *Science.* 2021, 371(6529):eabf4063.

[14] Antia A, Ahmed H, Handel A, et al. Heterogeneity and longevity of antibody memory to viruses and vaccines. *PLoS Biol.* 2018, 16(8):e2006601.

[15] Wood JG, Goeyvaerts N, MacIntyre CR, et al. Estimating vaccine coverage from serial trivariate serologic data in the presence of waning immunity. *Epidemiology.* 2015, 26(3):381–389.

[16] Williamson P, Custer B, Biggerstaff B, et al. Incidence of West Nile virus infection in the Dallas–Fort Worth metropolitan area during the 2012 epidemic. *Epidemiol Infect.* 2017, 145(12):2536–2544.

[17] Krutikov M, Palmer T, Tut G, et al. Prevalence and duration of detectable SARS-CoV-2 nucleocapsid antibodies in staff and residents of long-term care facilities over the first year of the pandemic (VIVALDI study): prospective cohort study in England. *Lancet Healthy Longev.* 2022, 3(1):e13–e21.

[18] Angulo FJ, Finelli L, and Swerdlow DL. Estimation of US SARS-CoV-2 infections, symptomatic infections, hospitalizations, and deaths using seroprevalence surveys. *JAMA Netw Open.* 2021, 4(1):e2033706–e2033706.

[19] Shioda K, Lau MS, Kraay AN, et al. Estimating the cumulative incidence of SARS-CoV-2 infection and the infection fatality ratio in light of waning antibodies. *Epidemiology.* 2021, 32(4):518–524.

[20] Sullivan PS, Siegler AJ, Shioda K, et al. Severe acute respiratory syndrome coronavirus 2 cumulative incidence, United States, August 2020–December 2020. *Clin Infect Dis.* 2022, 74(7):1141–1150.

[21] Lamba K, Bradley H, Shioda K, et al. SARS-CoV-2 cumulative incidence and period seroprevalence: Results from a statewide population-based serosurvey in California. *Open Forum Infect Dis.* 2021, 8(8):ofab379.

[22] Chamberlain AT, Toomey KE, Bradley H, et al. Cumulative incidence of SARS-CoV-2 infections among adults in Georgia, United States, August to December 2020. *J Infect Dis.* 2022, 225(3):396–403.

[23] Buss LF, Prete CA, Abrahim CM, et al. Three-quarters attack rate of SARS-CoV-2 in the Brazilian Amazon during a largely unmitigated epidemic. *Science.* 2021, 371(6526):288–292.

[24] Prete Jr CA, Buss LF, Whittaker C, et al. SARS-CoV-2 antibody dynamics in blood donors and COVID-19 epidemiology in eight Brazilian state capitals: A serial cross-sectional study. *Elife.* 2022, 11:e78233.

[25] Anderson RM, Heesterbeek H, Klinkenberg D, et al. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *Lancet.* 2020, 395(10228):931–934.

[26] Fontanet A and Cauchemez S. COVID-19 herd immunity: where are we? *Nat Rev Immunol.* 2020, 20(10):583–584.

[27] Britton T, Ball F, and Trapman P. A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2. *Science.* 2020, 369(6505):846–849.

[28] Sabino EC, Buss LF, Carvalho MP, et al. Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. *Lancet.* 2021, 397(10273):452–455.

[29] MS Windows NT kernel description. https://covid19.manaus.am.gov.br/monitoramento/#. Accessed: 2021-05-24.

[30] De Graaf W, Kretzschmar M, Teunis P, et al. A two-phase within-host model for immune response and its application to serological profiles of pertussis. *Epidemics.* 2014, 9:1–7.

[31] Hamady A, Lee J, and Loboda ZA. Waning antibody responses in COVID-19: what can we learn from the analysis of other coronaviruses? *Infection.* 2022, 50:11–25.

[32] Larremore DB, Fosdick BK, Zhang S, et al. Jointly modeling prevalence, sensitivity and specificity for optimal sample allocation [preprint]. *BioRxiv.* 2020. (doi: https://doi.org/10.1101/2020.05.23.112649). Accessed September 17, 2020.

[33] Larremore DB, Fosdick BK, Bubar KM, et al. Estimating SARS-CoV-2 seroprevalence and epidemiological parameters with uncertainty from serological surveys. *Elife.* 2021, 10:e64206.

[34] Bouman JA, Riou J, Bonhoeffer S, et al. Estimating the cumulative incidence of SARS-CoV-2 with imperfect serological tests: Exploiting cutoff-free approaches. *PLoS Comput Biol.* 2021, 17(2):e1008728.

[35] Bouman JA, Kadelka S, Stringhini S, et al. Applying mixture model methods to sars-cov-2 serosurvey data from geneva. *Epidemics*. 2022, 39:100572.

[36] Lalwani P, Salgado BB, Pereira Filho IV, et al. SARS-CoV-2 seroprevalence and associated factors in Manaus, Brazil: baseline results from the DETECTCoV-19 cohort study. *Int J Infect Dis*. 2021, 110:141–150.

[37] Vos ER, van Boven M, den Hartog G, et al. Associations between measures of social distancing and severe acute respiratory syndrome coronavirus 2 seropositivity: A nationwide population-based study in the Netherlands. *Clin Infect Dis*. 2021, 73(12):2318–2321.

[38] Hashem AM, Algaissi A, Almahboub SA, et al. Early humoral response correlates with disease severity and outcomes in COVID-19 patients. *Viruses*. 2020, 12(12):1390.

Table 1: Description of test cases.

Test case	Number of monthly surveys[a]	Number of epidemic waves[b]	Time of wave(s)[b]	Total fraction infected during each wave (in %)
1	9	1	3-4	40
2	9	1	6-7	40
3	9	2	3-4, 7-8	11, 22
4	9	2	3-4, 7-8	22, 11
5	15	2	3-5, 12-14	41, 25
6	9	0[c]		
7	17	0[c]		

[a] In each survey 800-900 individuals are sampled from a population that is assumed to be large enough for these samples to have little overlap.

[b] Any month in which more than 3% of the entire population is infected is considered to be (part of) an epidemic wave. Consecutive months each with more than 3% infected are defined as a single wave. Incidences in months that are not part of any epidemic wave range from 0-3%. An exact definition of the seven test cases is given in equation (8) in the Web Appendix 1.

[c] Test cases 6 and 7 are characterized by consistently low monthly incidences of less than 2% of the population infected per month.

Table 2: Cumulative incidence estimates using the empirically derived sero-reversion times and age and sex adjusted seroprevalence estimates[a].

Month (2020)	Age and sex adjusted seroprevalence estimate (in %)[a]	Cumulative incidence (in %)	95% confidence region (in %)
March	0.72	0.77	0.00, 1.69
April	4.10	5.42	3.18, 7.10
May	37.40	46.95	42.29, 50.01
June	44.13	47.55	43.31, 51.86
July	33.91	47.56	43.39, 51.90
August	25.54	47.56	43.50, 51.91
September	24.42	47.57	43.50, 52.38
October	21.66	47.57	43.52, 53.46

[a] Reported by Buss *et al.* (23).

Figure 1: Schematic representation of how the distribution of sero-reversion times is derived. A) Distributions of the decimal logarithms of antibody (AB) levels for positive controls (at peak, blue) and negative controls (red). The vertical line marks the cutoff for sero-positivity. B) Distribution of antibody level decay rates for positive controls. C) Antibody dynamics exemplified for two individuals described by the tuples (A_0^1, r_1, A_{neg}^1) and (A_0^2, r_2, A_{neg}^2), where $A_0^i = A(0)^i$ and A_{neg}^i are peak and background antibody levels drawn from the distributions in A), and r_i are antibody level decay rates drawn from the distribution in B). The horizontal line marks the cutoff for sero-positivity. D) Density (black) and cumulative distribution function (green) of sero-reversion times approximated using n tuples $A(0), r, A_{neg}$ as described in (i)–(iii) (in Web Appendix 1).

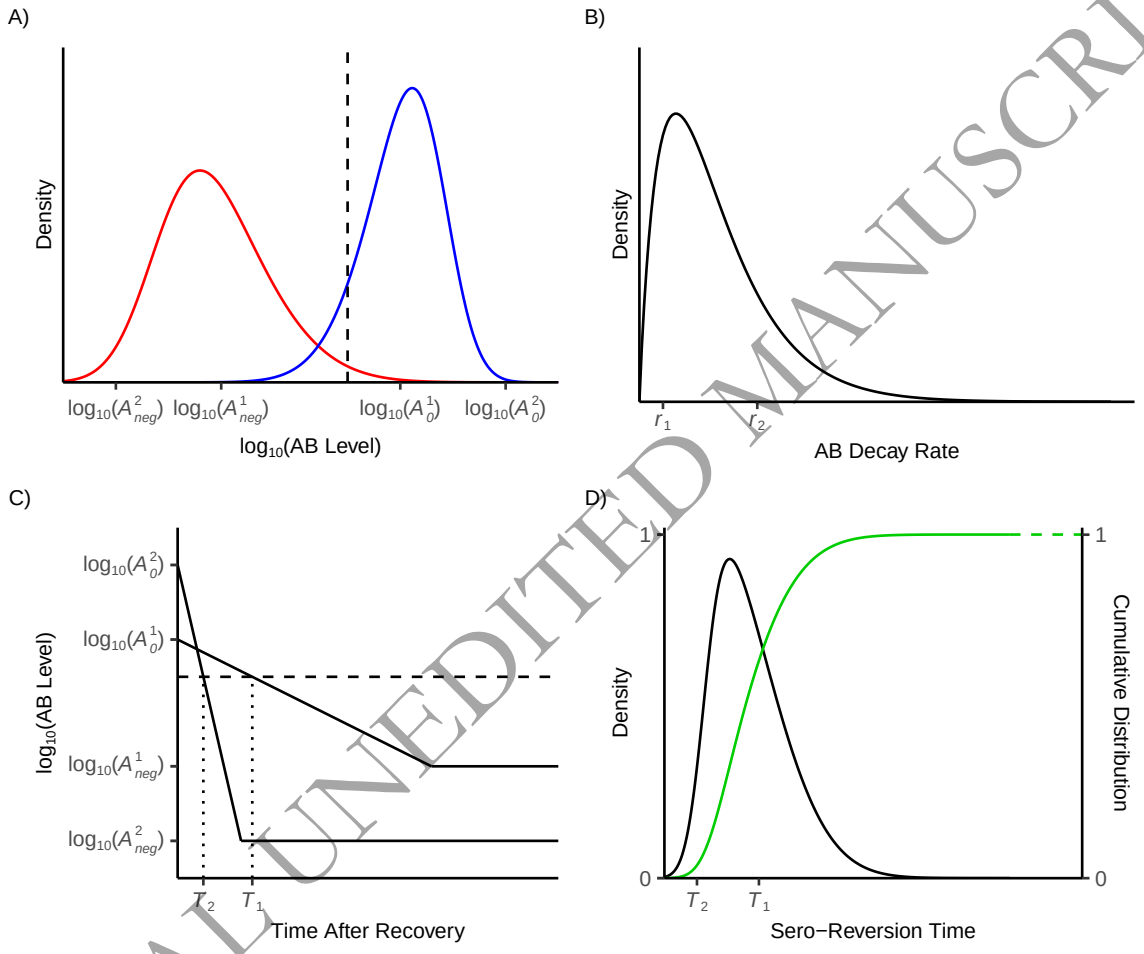

Figure 2: True cumulative incidences (red) and median of fitted cumulative incidences from $N_{sim} = 10000$ *in silico* studies (black) with uniformly distributed recovery times between sequential surveys and a delay of one month between surveys for test cases 1-7 (A)-G)). The shaded gray region is bounded by the 2.5% and 97.5% quantiles of the estimated cumulative incidences.

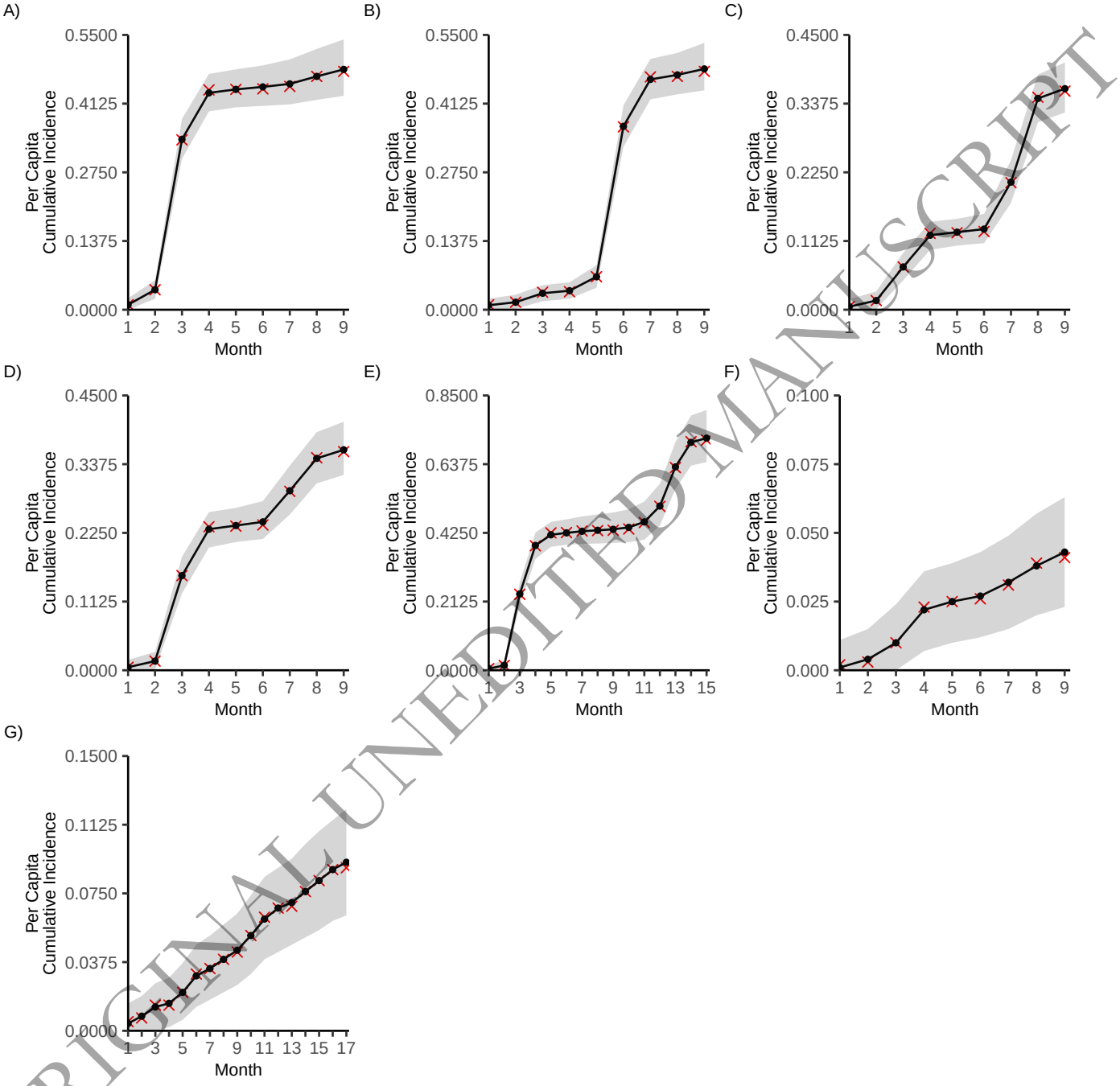

Figure 3: A) Histogram of decimal logarithms of anti-N IgG S/C ratios for case and control data (from (23)) with the densities of the respective smoothed empirical distributions. The vertical line marks the cutoff for sero-positivity. B) Histogram of the anti-N IgG decay rates. More specifically, the histogram of the gradients of the decimal logarithm of anti-N IgG S/C ratios in convalescent controls (from (23)) with the density of the gamma distribution that was fitted to the absolute values of the negative gradients. C) Histogram, probability density function f_{T_Θ} (blue) and cumulative distribution function F_{T_Θ} (red) for sero-reversion times T_Θ of initially sero-converted individuals. For better visibility we combined all sero-reversion times larger than 5 years and display them at 5 years plus 90 days. Furthermore, we combined all individuals that never sero-revert and display them at 5 years plus 150 days. D) Probabilities to sero-revert within a given time past recovery. Shaded region is bounded by the 2.5 and 97.5% quantiles obtained when resampling the validation data and recalculating the sero-reversion probabilities 100 times.

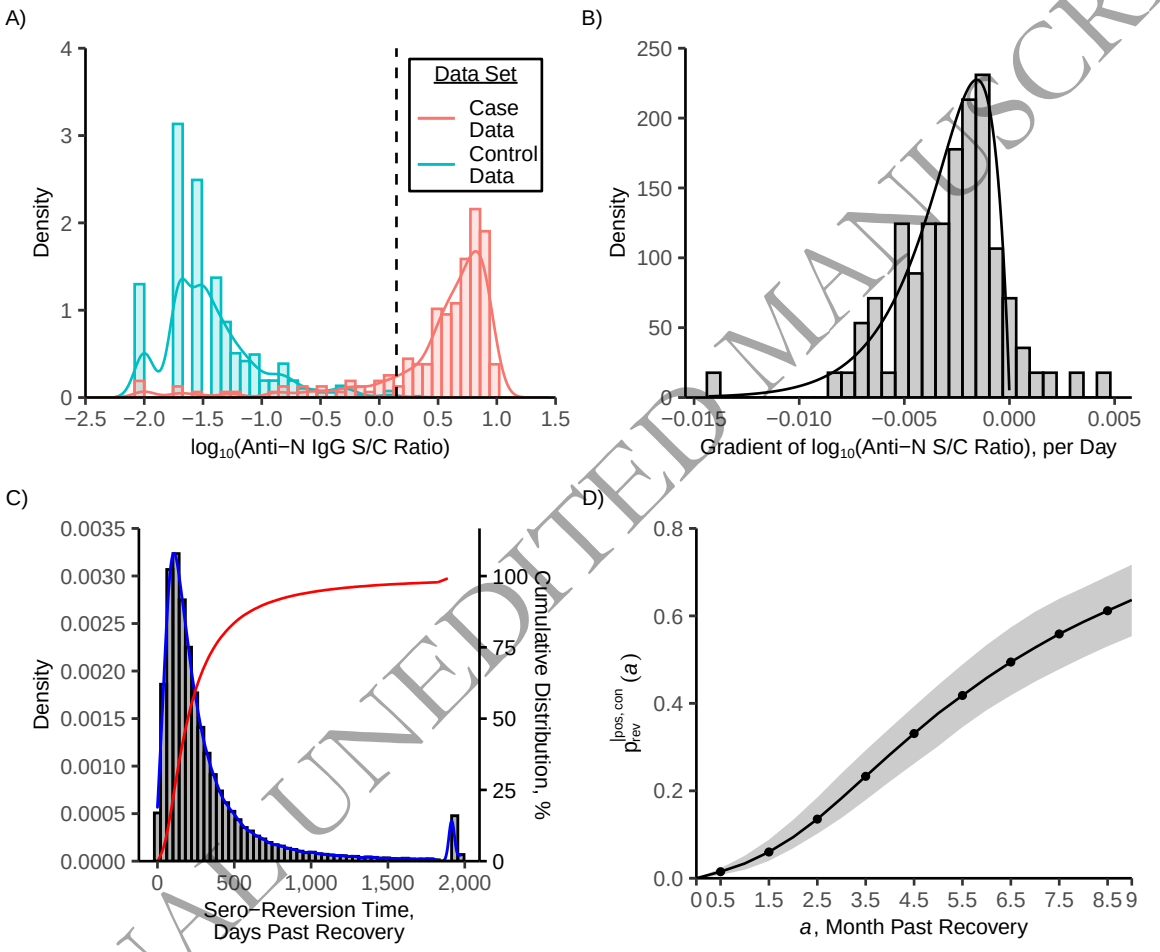

Figure 4: Cumulative incidence estimates applying our method (black dots) to age and sex adjusted observed seroprevalences reported by Buss *et al.* (23) (blue crosses). The shaded region is the bootstrapped 95% confidence region using 1000 bootstrap samples. For comparison seroprevalence observations additionally adjusted for test sensitivity and specificity (blue triangles) and Buss *et al.*'s cumulative incidence estimates (black squares) are shown (23).