Associations between measures of social distancing and SARS-CoV-2 seropositivity: a nationwide population-based study in the Netherlands

Supplementary Information

Eric R.A. Vos, Michiel van Boven, [redacted],
Jantien A. Backer, [redacted], Cheyenne C.E. van Hagen,
[redacted], Rob S. van Binnendijk, [redacted],
Fiona R.M. van der Klis, Hester E. de Melker

December 2020

Contents

1 Introduction

In this supplement we detail our sampling strategy, provide information on non-response rates, and explain how we have included post-stratification weighting in the analyses. The risk factor analyses in the main text use logistic regression based on binary classification of the data (seronegative versus seropositive). Here, we also provide an underpinning of this classification using a two-component mixture model. In this model, samples are not rigidly classified as either seronegative or seropositive, but belong to either the negative or positive component with certain probability [1, 2]. As the probability of seropositivity may depend on age, we model the mixing parameter (i.e. the probability of seropositivity, or seroprevalence) with an age-dependent penalized spline [3]. We fit the model to antibody concentration measurements from the population sample described in the main text while incorporating information from a test panel of proven negative and positive samples [4]. Subsequently, we derive test characteristics (sensitivity, specificity) for various cut-offs, showing that the binary classification used in the main text performs well. In a final step we present additional results for the weighted seropositivity estimates by municipal health services, and for the age-specific odds of seropositivity.

2 Sampling

The current sample includes persons who had participated in an earlier study (sample 1), and (sample 2). Details on the first sample have been described previously [5, 6]. In this previous study, 2,634 participants had been included. Anticipating a 10% drop-out rate from the first study, and given the low estimated seroprevalence in the first study (2.8%), we aimed to increase the overall power of the current study. Hence, the initial cohort was supplemented with an additional sample of randomly sampled participants from the Dutch population registry as of May, 2020.

For this second sample, persons were randomly drawn from five regions with roughly similar population size (North: provinces of Groningen, Friesland, Drenthe and Overijssel; Mid-West: provinces of Flevoland and Noord-Holland; Mid-East: provinces of Gelderland and Utrecht; South-West: provinces of Zuid-Holland and Zeeland; South-East: provinces of Noord-Brabant and Limburg), and from 17 pre-defined age groups (1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-89 years).

A total sample size of $6,400$ participants, i.e. with an average of 380 participants per age group, would enable us to estimate an overall and age-specific seroprevalence with a precision of 1.25% and 5%, respectively. Following previous experience, we anticipate a response rate of at least 15%. Hence, for the additional sample, we randomly selected $27,200$ persons from the population registry, of which $26,854$ remained eligible for participation after an initial screening.

3 Non-response and weighting

All randomly-selected persons who were invited in the first serological study were also invited for the current study. Of these, $2,317$ participated in the current study. This cohort was subsequently supplemented with an additional sample (as described above). Specifically, we invited $26,854$ randomly-selected persons of which $4,496$ participated, resulting in an overall number of $6,813$ participants. Table S1 shows the number of participants and response rates, stratified by sex, age group, region, and ethnic background. See main text for details and discussion.

Table S1. Overview of the study data. Ethnicity is missing for 16 invited persons.

		Non-responder		Responder		Total
		N	%	N	%	31,780
Total		24,967	78.6	6,813	21.4	
Sex	Man	12,609	50.5	3,042	44.7	15,561
	Women	12,358	49.5	3,771	55.4	16,129
Age	1-4	1,740	7.0	220	3.2	1,960
	5-9	1,637	6.6	285	4.2	1,922
	10-14	1,567	6.3	319	4.7	1,886
	15-19	1,591	3.4	304	4.5	1,895
	20-24	1,542	6.2	300	4.4	1,842
	25-29	1,779	7.1	398	5.8	2,177
	30-34	1,293	5.2	369	5.4	1,662
	35-39	1,519	6.1	408	6.0	1,927
	40-44	1,439	5.8	448	6.6	1,887
	45-49	1,423	5.7	457	6.7	1,880
	50-54	1,365	5.5	548	8.0	1,913
	55-59	1,323	5.3	544	8.0	1,867
	60-64	1,226	4.9	591	8.7	1,817
	65-69	1,250	5.0	626	9.2	1,876
	70-74	1,326	5.3	501	7.4	1,827
	75-79	1,410	5.7	134	2.0	1,752
	80-90	1,537	6.2	153	2.3	1,690
Region	North	5,029	20.1	1,357	19.9	6,386
	Mid-West	4,957	19.9	1,211	17.8	6,168
	Mid-East	4,825	19.3	1,469	21.6	6,294
	South-West	5,060	20.3	1,248	18.3	6,308
	South-East	5,096	20.4	1,528	22.4	6,624
Urbanization degree	High	6,038	24.2	1,319	19.4	7,357
	Middle	7,670	30.7	2,101	30.8	9,771
	Low	11,259	45.1	3,393	49.8	14,652
Ethnic background	Dutch	18,598	74.5	5,996	88.0	24,594
	Non-Dutch Western	2,389	9.6	512	7.5	2,901
	Non-Western	3,964	15.9	305	4.5	4,269

Post-stratification weights are assigned to each participant to standardize seroprevalence estimates, using census data of the Netherlands from January 2020. Since our cohort consists of two samples, weights are calculated for each sample separately. Per study sample, weights are assigned to each participant based on their membership to specific census strata (in total 112): for Dutch ethnic background, strata are designed for age group (1-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-90 years), urbanization level (high, middle, low), and sex; and for other ethnicity groups strata were based on age group (1-9, 10-34, 35-59, 60-90 years) and sex.

Subsequently, post-stratification weights are defined as the proportion of each stratum represented in the Dutch population divided by the analogous proportion in the study sample. Specifically, participant weights w_{ij} for participants in stratum i and study j are calculated as

$$w_{ij} = \frac{\frac{X_i}{N}}{\frac{x_{ij}}{n_j}} ,$$

where X_i is the total number of persons in stratum i, N is the total population size (i.e. the Netherlands), x_{ij} is the number of participants in stratum i in study sample j, and n_j is number of participants in sample j.

4 Data

Figure S1 shows the regional distribution of samples, and Figure S2 shows the antibody concentration measurements by age. For the analyses, we also include a validation panel that has been used for validation of the assay [4]. Specifically, we take a set of 384 samples from uninfected persons that had been drawn from the Dutch population before the pandemic, and a set of 115 proven SARS-CoV-2 infections with mild to severe disease [4]. Mean and standard deviation of the (log-transformed) measurements are $\mu_{\text{uninfected}} = -2.3$ (arbitrary units) and $\sigma_{\text{uninfected}} = 1.0$ for the uninfected group, and $\mu_{\text{infected}} = 3.0$ and $\sigma_{\text{infected}} = 2.1$ for the infected group.

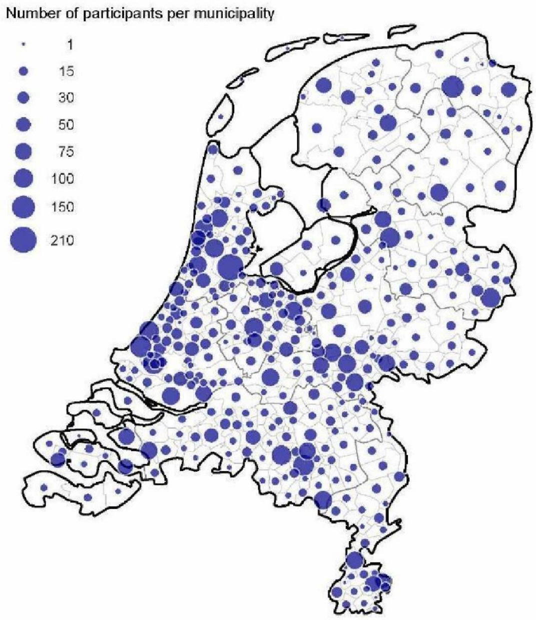

Figure S1. Regional distribution of samples. Notice that the western part of the Netherlands is the most densely populated area and also has large number of samples, thus attaining good population coverage.

5 Mixture model

Survey participants are assumed to be either seropositive or seronegative. These two classes are characterized by distributions for antibody measurements, denoted by f_{neg} and f_{pos} and specified by parameters θ_{neg} and θ_{pos}. Further, the mixing parameter (probability of seropositivity) depends on age and is denoted by $p(a)$. For $n = 6,813$ participants, the set of participant ages and observed measurements are given by $\mathbf{a} = (a_k)$ and $\mathbf{x} = (x_k)$ $(k = 1, \ldots, n)$, respectively. Throughout we use normal distributions for the components of the mixture of

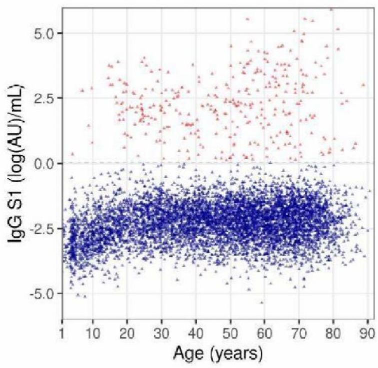

Figure S2. Overview of the data. Shown are (log-transformed) antibody concentrations of all 6,813 samples in the national sample as function of age. Here, samples are classified as seronegative below the cut-off of 0.04 *(log(AU)/mL)(blue) and as seropositive above the cut-off (red).*

the log-transformed data, so that $\theta_{\text{neg}} = (\mu_{\text{neg}}, \sigma_{\text{neg}})$ and $\theta_{\text{pos}} = (\mu_{\text{pos}}, \sigma_{\text{pos}})$, while the mixing parameter is modelled with a Bayesian penalized-spline using cubic basis functions and first order penalization [7, 8]. Throughout, we consider the age range $[0, 100]$ years, placing knots at 10-year intervals (11 knots in total), so that the total number of basis functions is 13 [7, 8].

6 Estimation

Parameters are estimated in a Bayesian framework using Hamiltonian Monte Carlo, implemented in Stan [9]. To improve performance at low prevalence, we employ a logistic transformation for the age-specific prevalence.

Prior distributions for the means and standard deviations of the seronegative and seropositive components are based on the uninfected and infected samples from the validation set. As the uninfected set is obtained from random samples from the Dutch population in 2006/2007 and 2016/2017 as well as a panel comprising cases with influenza-like illness, and the seropositive set contains cases with symptoms of disease and may be less representative of cases in the population, we take informative prior distributions for the parameters of the seronegative component, a weakly informative prior distribution for the mean of the seropositive component, and provide no explicit prior distribution for the standard deviation of the seropositive component. Specifically, we take

$$\mu_{\text{neg}} \sim \mathcal{N}\left(\mu_{\text{uninfected}}, 0.01\right)$$
$$\sigma_{\text{neg}} \sim \mathcal{N}\left(\sigma_{\text{uninfected}}, 0.1\right)$$
$$\mu_{\text{pos}} \sim \mathcal{N}\left(\mu_{\text{infected}}, 0.5\right) \ .$$

For the spline smoothing parameter (RWvar) we take an inverse gamma distribution [8],

$$\text{RWvar} \sim \text{inverse gamma}\,(1, 0.0005) \ ,$$

and for the weights of the spline base functions w_i $(i = 1 \ldots 13)$, we take

$$w_i \sim \mathcal{N}\left(0, 4\right) \ ,$$

where it should be noted that the prior weights are defined on the logistic scale.

Table S2. Parameter estimates (selected posterior quantiles) with selected convergence diagnostics.

Parameter	$\widehat{R}$	n_{eff}	2.5%	50%	97.5%
μ_{neg}	0.997	1071	-2.311	-2.297	-2.284
σ_{neg}	0.997	964	0.742	0.756	0.770
μ_{pos}	0.996	1066	1.967	2.168	2.336
σ_{pos}	1.003	1126	1.216	1.339	1.501
RWvar	1.000	1030	0.008	0.042	0.169

Results from properly converged chains are obtained within hours (using 10 cores on our servers). Estimates for the parameters defining the mixing distribution and the spline smoothing parameter are given in Table S2, together with

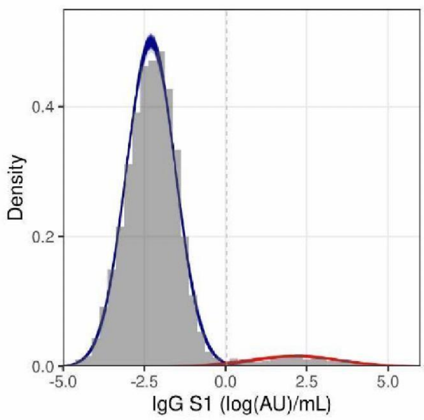

Figure S3. Data and model fit. Shown are the data (gray histograms) and fit of the mixture model (blue: seronegative component; red: seropositive component). The age-specific prevalence is modelled with a penalized spline, and the mixing distributions are weighted with the overall posterior probability of infection. Shown are $1,000$ *samples from the posterior distribution.*

convergence diagnostics $\widehat{R}$ and n_{eff} [9]. In a sensitivity analysis we have re-run the fitting procedure with uninformative prior distributions (only assuming that $\mu_{\text{pos}} > \mu_{\text{neg}}$). These analyses yield virtually identical results (not shown).

Figure S3 gives a visualisation of the data (gray histograms) and model fit (colored lines), suggesting good agreement between the two. Notice also that overlap between the negative and positive component is small which bodes well for efforts to distinguish seronegative from seropositive samples. To further investigate the implications of the analyses, Figure S4 shows the estimated probability of infection as function of antibody concentration. Here, the probability of infection calculated as the estimated positive density (at a certain concentration) divided by the sum of the positive and negative densities (at that concentration) [1]. The figure shows that, in the absence of information on age-specific prevalence, the estimated probability of infection is close to 0 for concentrations of -1 (log(AU)/mL) and lower, and close to 1 at concentrations of 0 (log(AU)/mL) and higher.

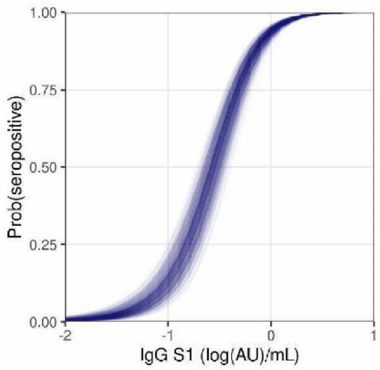

Figure S4. Estimated probability of seropositivity. Shown are estimated probabilities of seropositivity as function of the (log-transformed) antibody concentration. No weighting for prevalence is applied. Shown are 1,000 samples from the posterior distribution.

In a next step we estimated the probability of seropositivity for each of the $n = 6,813$ samples. Here we weighted the posterior seropositive density by the posterior prevalence, and the posterior seronegative density by 1 minus the posterior prevalence, and applied the same procedure as in Figure S4. The figure shows that for the majority of samples (6,722), the posterior median for the probability of infection is either low (< 0.05, 6,437 samples) or high (> 0.95, 285 samples), indicating that only for a small minority of samples (< 100) classification would not be straightforward. This is a robust result that also holds when using less informative priors or when including a random effect at the municipality level (not shown). It is due to the clear separation of the negative and positive components in the analyses (Figure S3).

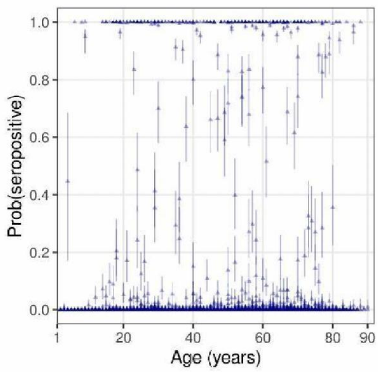

Figure S5. Estimated probability of seropositivity. Shown are estimated probabilities of seropositivity for each of the 6,813 samples as function of age. Estimates are weighted with age-specific prevalence. Dots and whiskers represent posterior medians and 95% credible intervals, respectively. Notice that the posterior probability of seropositivity (i.e. posterior median) is either very low (< 0.05) or very high (> 0.95) for the majority of samples (> 98%).

7 Binary classification

The above results show that for the majority of samples there is limited uncertainty as to whether they should be classified as seronegative or seropositive. Therefore, we feel confident that reliable binary classification of the samples is feasible. Here, we investigate the optimal cut-off value for such binary classification, and associated test characteristics (sensitivity and specificity).

For a given cut-off, the proportion of the negative distribution with concentrations higher than the cut-off defines specificity of the test (high proportion implies low specificity), while the proportion of the positive distribution with concentrations lower than the cut-off defines sensitivity of the test. Technically, both sensitivity and specificity are calculated using cumulative density functions of the negative (specificity) and positive distributions (sensitivity) [1]. Figure S6

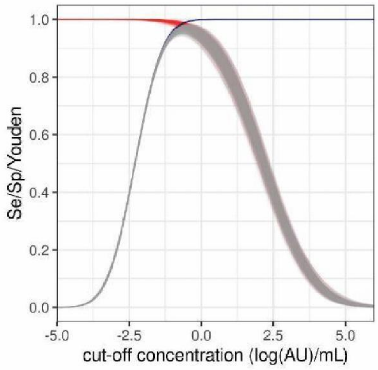

Figure S6. Sensitivity, specificity, and Youden index. Shown are the estimated sensitivity (red), specificity (blue), and Youden index (gray, superposed on top of sensitivity and specificity) as function of the cut-off concentration for seropositivity. Shown are 1,000 samples from the posterior distribution.

shows the test characteristics and the Youden index ($Se + Sp - 1$) as function of the cut-off. For low values of the cut-off, sensitivity of the test is high, at the price of a low specificity. Conversely, at high values of the cut-off, specificity of the test is high, at the price of low sensitivity. At intermediate values both sensitivity and specificity are reasonably high, and the Youden index is maximal.

Table S3. Test characteristics for cut-off that maximizes the Youden index or that selects for high test specificity ($Sp = 0.999$). Shown are posterior medians with 95% credible intervals.

Scenario	cut-off (95%CrI)	Se (95%CrI)	Sp (95%CrI)	Youden (95%CrI)
Youden	-0.56 (-0.67, -0.44)	0.979 (0.965, 0.987)	0.989 (0.985, 0.993)	0.97 (0.95, 0.98)
Sp	0.04 (0.0, 0.08)	0.943 (0.910, 0.966)	0.999	0.94 (0.91, 0.97)

In Table S3 we show test characteristics for two specific scenarios. The first takes cut-offs that maximize the Youden index. Here, the estimated optimal cut-off is -0.56 (95%CrI: -0.67- -0.44) and the estimated maximal Youden index is 0.97 (94%CrI: 0.95-0.98). This cut-off, however, is not useful in practice as expected seroprevalence is low ($< 10\%$), and control of the false positive rate is more important than control of the false negative rate. Therefore, in a second scenario we aimed at a specificity of 0.999. Such specificity can be reached with the test, at a cut-off of 0.04 and a sensitivity of 0.943 (which is really good for such specificity!). In the following and in the main text we have opted for a cut-off of 0.04.

Figure S7 presents the results as a Receiver Operating Characteristic (ROC) diagram (blue lines), together with true and false positive rates at the cut-off of 0.04 (red dots). Variation in the false positive rate is minimal ($\widehat{Sp} = 0.9990$, $95\%CrI : 0.9987 - 0.9992$), while estimated sensitivity is still high ($\widehat{Sp} = 0.944$, $95\%CrI : 0.910 - 0.967$). Estimated Youden index is 0.94 ($95\%CrI : 0.91 - 0.97$).

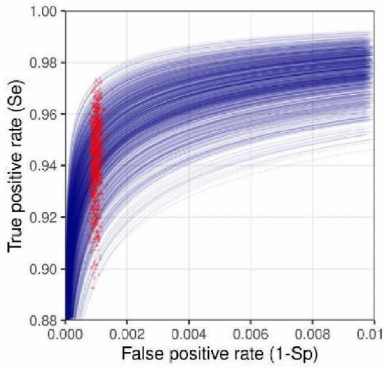

Figure S7. Receiver operator characteristic (ROC) diagram. Shown are the false positive rates ($1 - Sp$) and true positive rates (Se) for 1,000 samples from the posterior distribution (blue). Also shown are the false and true positive rates for cut-off of 0.04 (log(AU)/mL) (1.04 AU/mL)(red).

Finally, Figure S8 shows the posterior distribution of test sensitivity at a cut-off of 0.04 (log(AU)/mL). Mean and standard deviation of the distribution are 0.942 and 0.0151, respectively. These values can be incorporated in Rogan-Gladen-type corrections for estimating true prevalence from observed apparent prevalence in binary classification [10, 11].

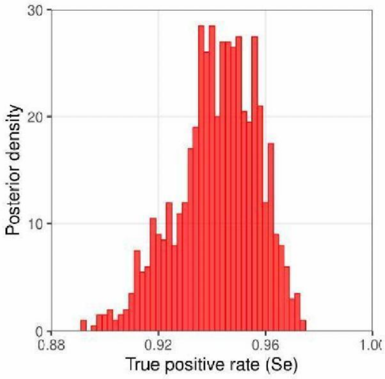

Figure S8. Posterior distribution of the true positive rate (sensitivity) when the cut-off is set at 0.04 *(log(AU)/mL)* (1.04 *AU/mL). Shown is a histogram of* $1,000$ *samples from the posterior distribution. Mean and standard deviation of the distribution are* 0.942 *and* 0.0151, *respectively.*

8 Logistic regression

The main text provides main results and interpretation of the analyses with logistic regression using the binary classification described in the above. Below we provide additional results on the regional estimates of seroprevalence (Figure S9), as well as the age-specific estimates of the unadjusted odd ratios for seropositivity derived from the univariable model (Figure S10).

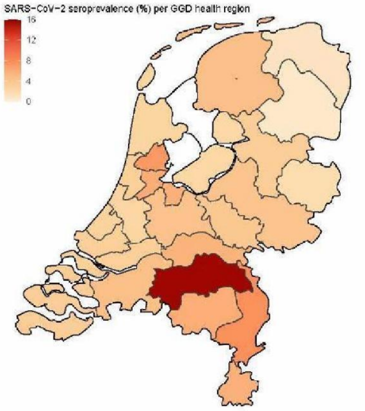

Figure S9. Estimates of seroprevalence by municipal health service area. Shown are estimates of (random effect) logistic regression based on the above binary classification, including sample weights and Rogan-Gladen bias correction [10].

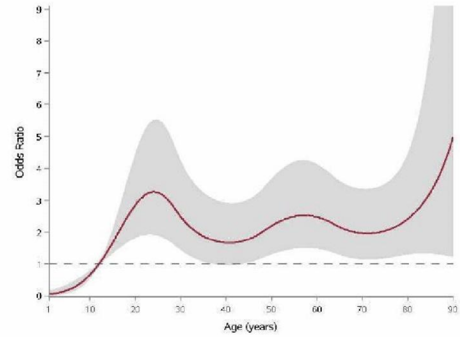

Figure S10. Estimates of the unadjusted odd ratios for seropositivity as function of age (see main text for adjusted odds ratios). The estimate is based on (random-effects) univariable logistic regression. Also shown is the 95% confidence envelope. Reference age is 12 years (odds ratio = 1).

References

[1] P. F. Teunis, M. T. Fonville, D. D. Döpfer, I. A. Eijck, V. Molina, E. Guarnera, and J. W. van der Giessen. Usefulness of sero-surveillance for Trichinella infections in animal populations. *Vet Parasitol*, 159(3-4):345–349, Feb 2009.

[2] A. Steens, S. Waaijenborg, P. F. Teunis, J. H. Reimerink, [redacted], M. van der Lubben, [redacted], M. A. van der Sande, J. Wallinga, and M. van Boven. 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, Dec 2011.

[3] M. van Boven, J. van de Kassteele, M. J. Korndewal, C. H. van Dorp, M. Kretzschmar, F. van der Klis, H. E. de Melker, A. C. Vossen, and D. van Baarle. Infectious reactivation of cytomegalovirus explaining age- and sex-specific patterns of seroprevalence. *PLoS Comput Biol*, 13(9):e1005719, Sep 2017.

[4] G. den Hartog, R. M. Schepp, M. Kuijer, C. GeurtsvanKessel, J. van Beek, N. Rots, [redacted], F. R. M. van der Klis, and R. S. van Binnendijk. SARS-CoV-2-Specific Antibody Detection for Seroepidemiology: A Multiplex Analysis Approach Accounting for Accurate Seroprevalence. *J Infect Dis*, 222(9):1452–1461, 10 2020.

[5] J. D. M. Verberk, R. A. Vos, L. Mollema, J. van Vliet, J. W. M. van Weert, H. E. de Melker, and F. R. M. van der Klis. Third national biobank for population-based seroprevalence studies in the Netherlands, including the Caribbean Netherlands. *BMC Infect Dis*, 19(1):470, May 2019.

[6] Eric R A Vos, [redacted], Rutger M Schepp, [redacted], [redacted]-[redacted], [redacted], [redacted], [redacted], Janneke D M Verberk, Michiel van Boven, Rob S van Binnendijk, Hester E

de Melker, [redacted] 5.1.2e, and Fiona R M van der Klis. Nationwide seroprevalence of sars-cov-2 and identification of risk factors in the general population of the netherlands during the first epidemic wave. *Journal of Epidemiology & Community Health*, 2020.

[7] P. H. Eilers and B. D. Marx. Flexible smoothing with b -splines and penalties. *Statist. Sci.*, 11(2):89–121, 05 1996.

[8] S. Lang and A. Brezger. Bayesian p-splines. *Journal of Computational and Graphical Statistics*, 13, 03 2004.

[9] B. Carpenter, A. Gelman, M. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell. Stan: A probabilistic programming language. *Journal of Statistical Software, Articles*, 76(1):1–32, 2017.

[10] W. J. Rogan and B. Gladen. Estimating prevalence from the results of a screening test. *Am J Epidemiol*, 107(1):71–76, Jan 1978.

[11] Z. Lang and J. Reiczigel. Confidence limits for prevalence of disease adjusted for estimated sensitivity and specificity. *Prev Vet Med*, 113(1):13–22, Jan 2014.