Supplementary methods

Introduction

Our aim was to infer the proportion of the population having any antibody against SARS-CoV-2, as well as the proportion of those who acquired antibodies through natural infection as opposed to vaccination. We do so by modelling jointly the antibody response measured by the Roche-N and S immunoassays together with participant responses to a vaccination questionnaire. We disentangle natural infections from vaccination antibody responses using the fact that available vaccines in Switzerland (Moderna and Pfizer) both elicit a response exclusively to the S protein of SARS-CoV-2, as opposed to natural infections which typically elicit a response to both the N and S virus proteins. We expand previous Bayesian modelling frameworks used for seroprevalence estimates that account for demographic parameters (sex and age), test performance and household infection clustering [1, 2]. The main changes are that we model jointly the response to both tests, and that we account for vaccination-induced antibody response.

Multinomial response model

We model the Roche-S and Roche-N test results for participant i, $\boldsymbol{x}_i$, consisting of one of four possible outcome combinations $[n_{S^+N^+}, n_{S^-N^+}, n_{S^+N^-}, n_{S^-N^-}]$ ($^+$ indicates antibody presence, $^-$ indicates absence) with $\boldsymbol{x}_i \in \{[1,0,0,0],[0,1,0,0],[0,0,1,0],[0,0,0,1]\}$ using a multinomial distribution parameterised by parameter vector $\boldsymbol{\pi}_i = [\pi_i^{++}, \pi_i^{-+}, \pi_i^{+-}, \pi_i^{--}]$, where π^{lm} is the probability of having Roche-S test result l and Roche-N result m, accounting both for the underlying probability of each antibody status $p^{\pm/\pm}$, and test sensitivity, θ^+ and specificity, θ^-:

$$
\begin{aligned}
\boldsymbol{x_i} &\sim \text{Multinomial}(\boldsymbol{\pi}_i),\\
\pi_i^{++} &= \theta_S^+\theta_N^+ p_i^{++} + (1-\theta_S^-)\theta_N^+ p_i^{-+} + \theta_S^+(1-\theta_N^-)p_i^{+-} + (1-\theta_S^-)(1-\theta_N^-)p_i^{--},\\
\pi_i^{-+} &= (1-\theta_S^+)\theta_N^+ p_i^{++} + \theta_S^-\theta_N^+ p_i^{-+} + (1-\theta_S^+)(1-\theta_N^-)p_i^{+-} + \theta_S^-(1-\theta_N^-)p_i^{--},\\
\pi_i^{+-} &= \theta_S^+(1-\theta_N^+)p_i^{++} + (1-\theta_S^-)(1-\theta_N^+)p_i^{-+} + \theta_S^+\theta_N^- p_i^{+-} + (1-\theta_S^-)\theta_N^- p_i^{--},\\
\pi_i^{--} &= (1-\theta_S^+)(1-\theta_N^+)p_i^{++} + \theta_S^-(1-\theta_N^+)p_i^{-+} + (1-\theta_S^+)\theta_N^- p_i^{+-} + \theta_S^-\theta_N^- p_i^{--}.
\end{aligned}
$$

The underlying probability of antibody status accounts both for the probability of natural infection λ_i and vaccinations status, $v_i \in \{0,1\}$. Following [1, 2] we model the probability of natural infection as a function of sex and age category, and accounting for household infection clustering through a random effect, α_h:

$$
\begin{aligned}
\text{logit}(\lambda_i) &= \alpha_h + \mathbf{X}_i\boldsymbol{\beta}\\
\alpha_h &\sim \text{Normal}(0, \sigma_h^2),
\end{aligned}
$$

where $\mathbf{X}_i$ is the matrix of covariates, and $\boldsymbol{\beta}$ the vector of regression coefficients. The probabilities of antibody status are then given by:

$$
\begin{aligned}
p_i^{++} &= \gamma^{++}\lambda_i + \gamma^{-+}\nu_i\\
p_i^{-+} &= \gamma^{-+}\lambda_i(1-\nu_i)\\
p_i^{+-} &= (1-\lambda_i)\nu_i + \gamma^{+-}\lambda_i\\
p_i^{--} &= 1 - \nu_i(1-\lambda_i) - \lambda_i,
\end{aligned}
$$

where $\gamma^{++}, \gamma^{-+}, \gamma^{+-}$ are the conditional probability of having S^+N^+, S^-N^+, S^+N^- responses respectively upon natural infection, ν_i is the probability of having a vaccine-induced S^+ response as a function of the conditional probability of antibody response upon infection η_i, $\nu_i = \eta_i \times v_i$.

Vaccination

To obtain population-level seroprevalence estimates we also model the proportion of vaccinated individuals in each sex/age class following the approach used for natural infection:

$$\begin{aligned} v_i &\sim \text{Bernoulli}(\phi_i) \\ \text{logit}(\phi_i) &= \alpha_{v,h} + \mathbf{X}_i\boldsymbol{\beta_v} \\ \alpha_{v,h} &\sim \text{Normal}(0, \sigma_v^2). \end{aligned}$$

Given vaccination policy recommendations in the canton of Geneva, previously infected individuals were discouraged to be vaccinated in the early phase of the campaign, thus making the probability of vaccination dependent on the infection status of the individual. We account for this dependence by modelling separately the probability of vaccination given the infection status and marginalising out the infection status:

$$\begin{aligned} P(v_i|\boldsymbol{\Theta}) &= \text{Bernoulli}(v_i|\phi_i^I)\lambda_i + \text{Bernoulli}(v_i|\phi_i^{\sim I})(1-\lambda_i), \\ \text{logit}(\phi_i^{\sim I}) &= \alpha_{v,h} + \mathbf{X}_i\boldsymbol{\beta_v}, \\ \text{logit}(\phi_i^I) &= \alpha_{v,h} + \mathbf{X}_i\boldsymbol{\beta_v} + \mathbf{X}_i\boldsymbol{\beta_v^I}, \\ \alpha_{v,h} &\sim \text{Normal}(0, \sigma_v^2), \end{aligned}$$

where $\boldsymbol{\Theta}$ is the vector of all model parameters, $I, \sim I$ indicates infection and non-infection respectively, and $\boldsymbol{\beta_v^I}$ is the vector of regression coefficients giving the difference in probability of vaccination between infected and non-infected individuals.

When estimating the population-level seroprevalence we account for the conditional probability of vaccination given non-infection, $p_{v|\sim I}$, in the probability of a negative S and N response accounting for household vaccination clustering, p^{--}, as:

$$p_{s,k}^{--} = 1 - p_{v|\sim I,s,k} \times (1 - p_{I,s,k}) - p_{I,s,k},$$

where s, k denote the sex and age categories, $p_{v|\sim I,s,k} = \int_0^1 \phi_{s,k}^{\sim I}(t)dt = \int_0^1 \boldsymbol{\beta}_{v,s,k}\boldsymbol{X}_{s,k} + \sigma_v\Phi^{-1}(t)dt$, with $\Phi^{-1}(t)$ being the normal quantile function, and similarly $p_{I,s,k}$ is the probability of infection with $p_{I,s,k} = \int_0^1 \lambda_{s,k}(t)dt = \int_0^1 \boldsymbol{\beta}_{s,k}\boldsymbol{X}_{s,k} + \sigma\Phi^{-1}(t)dt$.

Diagnostic test performance

The individual performance of both N and S tests is incorporated hierarchically following Gelman & Carpenter [3]. The sensitivity, θ^+, is determined using n^+ RT-PCR positive controls from a lab validation study [4], of which x^+ tested positive. The specificity, θ^-, is determined using n^- pre-pandemic negative controls, of which x^- tested positive. For the Roche N test, these values are modulated by data in [5]. For the Roche S test, the lab study data are modulated by those available on the Roche website (last accessed 2021-07-19).

Priors

We follow a similar setting of the priors on the tests' sensitivity and specificity as Gelman & Carpenter [3]. For study j, the specificity θ_j^- and sensitivity θ_j^+ are drawn from normal

distributions on the log odds scale,

$$\text{logit}(\theta_j^-) \sim \text{Normal}(\mu_{\theta^-}, \sigma_{\theta^-}),$$
$$\text{logit}(\theta_j^+) \sim \text{Normal}(\mu_{\theta^+}, \sigma_{\theta^+}).$$

Hyperparameters μ_z and σ_z for $z \in (\theta^-, \theta^+)$ follow, on the logit scale, normal distributions $\mu_z \sim \mathcal{N}(4, 2)$ and positive half-normals $\sigma_z \sim \mathcal{N}^+(0, 1)$ respectively. These priors on test performance were identical for both the Roche S and Roche N tests.

We used standard normal $\mathcal{N}(0, 1)$ priors for the logistic regression coefficients for infection $\boldsymbol{\beta}$. For coefficients of vaccination $\boldsymbol{\beta_v}$ and for coefficients of the difference in probability of vaccination between infected and non-infected individuals $\boldsymbol{\beta_v^I}$ we also used standard normals except for the youngest age groups i.e. 0–5 year olds and 6 – 11 year olds. For these two age groups, $\beta_v \sim \mathcal{N}(-10, 0.01)$ to reflect the fact that there was almost no vaccination in these youngest age groups in Geneva at the time of the study (NB vaccination registration for those aged 12 – 15 opened on 16th June 2021 (https://www.ge.ch/en/getting-vaccinated-against-covid-19/covid-19-vaccination-campaign-geneva last accessed 2021-07-20). Correspondingly, $\beta_v^I \sim \mathcal{N}(0, 0.01)$ for these two age groups.

The priors for the means of the household random effects α_h and $\alpha_{h,v}$, followed standard normals, and for standard deviations of the household random effects were positive half-normals, $\sigma_h \sim \mathcal{N}^+(0, 2)$ and $\sigma_v \sim \mathcal{N}^+(0, 2)$. We use a Dirichlet prior on the conditional probability of having S^+N^+, S^-N^+, S^+N^- responses upon natural infection, $\gamma^{++}, \gamma^{-+}, \gamma^{+-}$, $\boldsymbol{\gamma} \sim \text{Dir}(10, 1, 1)$, to highly favour production of both S and N antibodies upon infection. Finally, we put a strong prior on the conditional probability of antibody response after vaccination $\eta_i \sim \text{Beta}(10, 0.1)$.

Implementation

The model was coded in the probabilistic programming language Stan [6] using the Rstan package [7] as the interface. R [8] version 4.1 was used for data analysis. 4 chains were run with 1500 iterations each, 250 of which were warmup, to give a total of 5000 posterior samples. Convergence was assessed by checking that $\widehat{R} \approx 1$, that the effective sample size was reasonable for all parameters, and visually using shinystan [9] diagnostics checks.

References

[1] S. Stringhini et al. "Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland (SEROCoV-POP): a population-based study". In: *The Lancet* 396.10247 (2020), pp. 313–319.

[2] S. Stringhini et al. "Seroprevalence of anti-SARS-CoV-2 antibodies after the second pandemic peak." In: *The Lancet. Infectious Diseases* 21.5 (2021), pp. 600–601.

[3] A. Gelman and B. Carpenter. "Bayesian analysis of tests with unknown specificity and sensitivity". In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 69.5 (2020), pp. 1269–1283.

[4] P. Muench et al. "Development and validation of the Elecsys Anti-SARS-CoV-2 Immunoassay as a highly specific tool for determining past exposure to SARS-CoV-2". In: *Journal of Clinical Microbiology* 58.10 (2020).

[5] M. Ainsworth et al. "Performance characteristics of five immunoassays for SARS-CoV-2: a head-to-head benchmark comparison". In: *The Lancet Infectious Diseases* 20.12 (2020). https://www.thelancet.com/action/showPdf?pii=S1473-3099%2820%2930634-4, pp. 1390–1400.

[6] Stan Development Team. *Stan Modeling Language Users Guide and Reference Manual*. Version 2.21.0. 2019. URL: https://mc-stan.org.

[7] Stan Development Team. *RStan: the R interface to Stan*. R package version 2.21.2. 2020. URL: http://mc-stan.org/.

[8] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2021. URL: https://www.R-project.org/.

[9] J. Gabry. *shinystan: Interactive Visual and Numerical Diagnostics and Posterior alysis for Bayesian Models*. R package version 2.5.0. 2018. URL: https://CRAN.R-project.org/package=shinystan.