Variation in SARS-CoV-2 seroprevalence across districts, schools and classes: baseline measurements from a cohort of primary and secondary school children in Switzerland

Appendix 1: Description of Bayesian model

A Ulyte, T Radtke, IA Abela, SR Haile, J Blankenberger, R Jung, C Capelli, C Berger, A Frei, M Huber, M Schanz, M Schwarzmueller, A Trkola, J Fehr, MA Puhan and S Kriemler

Methods

The model formulation is essentially that proposed by "Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland (SEROCoV-POP): a population-based study" Stringhini *et al* 2020 (see also the appendix to that paper), with full description and example code in their github repository `HopkinsIDD/serocovpop`. Both *Ciao Corona* and *SEROCoV-POP* are part of the Corona Immunitas program of studies in Switzerland.

Statistical model

In this paper, our goal is to estimate the true underlying seroprevalence of the population of schoolchildren (in grades 1, 2, 4, 5, 7, or 8) as measured in the of the Canton of Zurich, denoted p^*.

We start by estimating the probability that each person in the survey is seropositive using a Bayesian logistic regression model that accounts for school-level clustering, the sensitivity and specificity of the antibody test, each individual's grade level and geographic district of the school:

$$\begin{aligned} x_i &\sim Bernoulli(p_i\theta^+ + (1-p_i) * (1-\theta^-)) \\ logit(p_i) &= \alpha_h + \mathbf{X}_i\boldsymbol{\beta} \\ \alpha_h &\sim Normal(0, \sigma^2) \\ x^+ &\sim Binomial(n^+, \theta^+) \\ x^- &\sim Binomial(n^-, 1-\theta^-) \end{aligned}$$

where x_i is the result of the antibody test (in primary analyses) for the ith person ($i = 1, \ldots, N$) in the serosurvey. The sensitivity, θ^+, is determined using n^+ RT-PCR positive controls from the lab validation study, of which x^+ tested positive. The specificity, θ^- , is determined using n^- pre-pandemic negative controls, of which x^- tested positive. The model estimates of the sensitivity and specificity are shown below.

ABCORA 2.0	Cohort	N tested	N positive	N negative
SARS-CoV-2 PCT Positive	104	97	7	93.27% sensitivity
pre-pandemic healthy	251	1	250	99.6% specificity

The probability of observing a diagnostic positive is a function of the true positive rate and the false negative rate with regards to the true underlying probability of seropositivity p_i for that person. This probability itself is a function of covariates $\mathbf{X}$, which consists of grade level and geographic district of the school, and their coefficients $\boldsymbol{\beta}$, and a random effect for school, α_h $(h = 1, \ldots, H)$, with variance σ^2. We used naive priors on all parameters to allow for an exploration of the parameter space. The priors on the sensitivity and specificity were flat from 0 to 1, equivalent to $Uniform(0, 1)$ or $Beta(1, 1)$. We used weak $Normal(0, 1)$ priors for the logistic regression coefficients $\boldsymbol{\beta}$. The prior on the standard deviation of the school effect, σ, was flat from 0 to infinity.

We implemented this model in the Stan probabilistic programming language and used the RStan package in R to run the model and analyse outputs. We ran 5,000 iterations (4 chains with 1,500 iterations each with 250 for warm-up) and assessed convergence visually.

Seroprevalence estimate

We estimated the seroprevalence p^* by post-stratifying the posterior samples of our parameter estimates to match the demographics of all schoolchildren in grades 1, 2, 4, 5, 7, or 8 in the Canton of Zurich. For every combination of grade level and the geographic district of the school, we estimated the probability of seropositivity for each posterior draw of β and σ. We can estimate the seroprevalence by taking a weighted average of the p, where weights are determined by the demographic distribution of the Canton of Zurich:

$$p = \int_0^1 logit^{-1}(\mathbf{X}\beta + \sigma\Phi^{-1}(t))dt$$

$$p^* = \sum_{grade} \sum_{district} \frac{pop_{grade,district} * p_{grade,district}}{pop}$$

where $\Phi^{-1}(t)$ is the quantile function of a standard normal distribution, $pop_{grade,district}$ is the population of each demographic "cell" and *pop* is the school age population of the Canton of Zurich. We estimate the average probability of seropositivity for the population in each demographic cell by integrating across all values of a logit-normal distribution with the standard deviation defined by the school random effect σ.