Supplementary materials 2 (Murphy et al. 2023 CMAJ)

Model specification of main model: multivariate beta-binomial regression of anti-N and anti-S on time splines

Beta-binomial regression is suitable for modelling proportions where the underlying sample size for each seroprevalence measurement varies (Young-Xu, 2008). The logit of the expected anti-N seroprevalence was predicted by independent variables for time, modelled as natural cubic splines (Perperoglou, 2019), with province-specific intercepts and spline coefficients, with partial pooling using normally distributed hyperpriors. Seroprevalence from vaccination or infection, measured as anti-S seroprevalence, was jointly modelled: the logit of anti-S seropositivity was predicted from the logit of expected anti-N seroprevalence and province-specific natural spline time terms. In a separate model, anti-N seroprevalence by age group was estimated by adding age terms to the anti-N model described above. The mid-point of the sample age range for contributed age-stratified estimates was modelled as a continuous variable using cubic polynomial terms. The linear term for age was interacted with the time spline terms to allow the age association to vary over calendar time. Quebec was not included in the age model because age-stratified estimates for Quebec in the Omicron waves were not available. Weekly seroprevalence estimates were obtained from population-averaged predictions using total or age-group provincial population as weights (Kleinman, 2009). To assess the sensitivity of the results to informative prior distributions on between-province variance in the intercept and time spline coefficients, we repeated the analysis with priors having less precision (Supplementary Notes Section 3). For both models (anti-N and anti-S), 95% credible intervals were calculated from the MCMC posterior predictive samples (Gelman, 2017).

Notes

Anti-N model

The intercept is the mean anti-N seroprevalence at the end of the analysis (March 15, 2023) and is allowed to vary by province-specific. Time is coded as reverse days, where analysis end data equals 0. The time trend is represented by fourteen natural basis splines and the coefficients, like the intercept, are allowed to vary by province. The model assumes those province-specific log-

odds are normally distributed around a mean log-odds of seropositivity at that time (controlling for other the other bases). Variance around the expected probability is assumed to follow a beta-binomial distribution.

Anti-S model

The intercept is the logit of the expected anti-N seroprevalence for the same calendar week and province as the anti-S observation. Time is coded as days since the start of vaccine roll-out (December 15, 2020) and is represented by five natural basis splines. The time spline coefficients are allowed to vary by province. The model assumes those province-specific log-odds are normally distributed around a mean log-odds of seropositivity at that time (controlling for other the other bases). Variance around the expected probability is assumed to follow a beta-binomial distribution.

We used the mode and concentration parameterization of beta distribution including the (kappa - 2 + 1) 'trick' to avoid errors by estimates of kappa less than 3 [Kruschke 2014].

The described models are shown in algebraic notation below; for simplicity, the observation-level index is dropped.

Likelihood

$$
\begin{aligned}
posn &\sim Binomial(\theta,\ denomn) \\
\theta &\sim Beta(\mu * (\kappa - 2) + 1,\ (1 - \mu) * (\kappa - 2) + 1) \\
logit(\mu) &= \alpha_{prov} + \Sigma_1^k \beta_{k,prov} * splinen_k + \sigma[prov]
\end{aligned}
$$

$$
\begin{aligned}
poss &\sim Binomial(\psi,\ denoms) \\
\psi &\sim Beta(\eta * (\gamma - 2) + 1,\ (1 - \eta) * (\gamma - 2) + 1) \\
logit(\eta) &= logit(\mu_{prov}) + \Sigma_1^z \delta_{z,prov} * splines_z + \nu[prov]
\end{aligned}
$$

Priors

The intercepts, alpha[prov], are the log-odds of province's anti-N seroprevalence at the end of the analysis, but the mean of the provincial intercepts is not necessarily the expected seroprevalence

for Canada, overall, because the input data are not weighted for provincial population. See the Methods for post-estimation.

As a sensitivity analysis, the precision of the priors for distribution of province-specific intercept and regression coefficients (shown in red) was reduced by half. See supplementary materials for the result.

$$
\begin{aligned}
\kappa &= kappaMinusTwo + 2 \\
kappaMinusTwo &\sim Gamma(shape = 1.5, rate = 0.01) \\
sigma &\sim N(1, 1)[0, inf]
\end{aligned}
$$

$$
\begin{aligned}
\gamma &= gammaMinusTwo + 2 \\
gammaMinusTwo &\sim Gamma(shape = 1.5, rate = 0.01) \\
\nu &\sim N(1, 1)[0, inf]
\end{aligned}
$$

$$
\begin{aligned}
\alpha_{prov} &\sim Student(intn, sdintn, 3) \\
\beta_{k,prov} &\sim N(betapool_k, 1)
\end{aligned}
$$

$$
\delta_{z,prov} \sim N(deltapool_z, 1)
$$

Hyperpriors:

$$
\begin{aligned}
intn &\sim N(0, 3) \\
sdintn &\sim N(1, 1)[0, inf]
\end{aligned}
$$

$$
\begin{aligned}
betapool_{1:8} &\sim Student(mean = -2, sd = 3, df = 5) \\
betapool_{9:13} &\sim Student(mean = -5, sd = 2, df = 5) \\
betapool_{14} &\sim N(mean = -25, sd = 3)
\end{aligned}
$$

$$
deltapool_{1:5} \sim Student(mean = 0, sd = 3, df = 9)
$$

Sensitivity analysis

The model for the main outcome, anti-N seroprevalence overall and by province (all ages combined), was repeated with the magnitude of the standard deviation on the hyperparameter priors doubled (see the red parameters in model specification above). The wider priors generally only had an influence in the third decimal place of model parameters or at the tenth of a percent in the province-specific intercepts on the probability scale. The 'study bias' correlated error term was most affected by the doubling of its prior distribution standard deviation, but the change was small relative to the standard error.

Estimate	**Main model**	**Sensitivity analysis**
intercept mean, *intn*	1.40 (1.23 to 1.59)	1.41 (1.23 to 1.61)
SD of the intercept, *sd_intn*	0.155 (0.020 to 0.434)	0.156 (0.022 to 0.431)
Time spline coefficients		
betapool[1]	-0.1026 (-0.8975 to 0.6747)	-0.0784 (-0.9106 to 0.7561)
betapool[2]	-0.3695 (-1.121 to 0.3821)	-0.3412 (-1.1057 to 0.3832)
betapool[3]	-0.9939 (-1.7532 to -0.2111)	-0.9857 (-1.779 to -0.2217)
betapool[4]	-1.331 (-2.0763 to -0.5994)	-1.321 (-2.068 to -0.5953)
betapool[5]	-1.669 (-2.417 to -0.934)	-1.679 (-2.433 to -0.9385)
betapool[6]	-2.641 (-3.391 to -1.938)	-2.651 (-3.398 to -1.927)
betapool[7]	-4.568 (-5.342 to -3.789)	-4.616 (-5.383 to -3.847)
betapool[8]	-4.431 (-5.239 to -3.671)	-4.473 (-5.258 to -3.682)
betapool[9]	-4.777 (-5.595 to -3.947)	-4.853 (-5.689 to -4.024)
betapool[10]	-5.203 (-6.035 to -4.358)	-5.242 (-6.062 to -4.387)
betapool[11]	-7.168 (-8.391 to -5.967)	-7.160 (-8.461 to -5.898)
betapool[12]	-6.943 (-10.669 to -4.687)	-8.314 (-14.725 to -4.907)
betapool[13]	-5.215 (-9.969 to -0.774)	-5.006(-18.219 to 4.613)
betapool[14]	-30.29 (-34.41 to -26.75)	-34.02 (-40.57 to -28.67)
SD of study bias, *sigma*	0.767 (0.397 to 1.49)	0.789 (0.410 to 1.64)
beta distribution concentration, *kappa*	243 (209 to 281)	243 (209 to 280)

Provincial Anti-N seroprevalences estimate for March 15, 2023, on the probability scale, for CBS study (reference group), *invlogit(alpha[k]))*		
British Columbia	0.7732 (0.7394 to 0.8033)	0.7742 (0.7399 to 0.8052)
Alberta	0.8199 (0.7929 to 0.8477)	0.8210 (0.7936 to 0.8494)
Saskatchewan	0.7907 (0.7560 to 0.8190)	0.7928 (0.7577 to 0.8207)
Manitoba	0.8086 (0.7798 to 0.8383)	0.8097 (0.7815 to 0.8398)
Ontario	0.7906 (0.7642 to 0.8154)	0.7909 (0.7628 to 0.8155)
Quebec	0.8215 (0.7720 to 0.8924)	0.8229 (0.7714 to 0.8942)
Nova Scotia	0.8046 (0.7739 to 0.8343)	0.8066 (0.7770 to 0.8370)
New Brunswick	0.7711 (0.7387 to 0.8021)	0.7723 (0.7386 to 0.8038)
Prince Edward Island	0.7690 (0.7260 to 0.8083)	0.7701 (0.7306 to 0.8092)
Newfoundland	0.7768 (0.7437 to 0.8084)	0.7779 (0.7431 to 0.8095)