Neighborhood and individual-level predictors of COVID-19 seropositivity among Canadian blood donors across three pandemic waves

Yuan Yu1, Matthew J. Knight1, Sheila F. O’Brien2, David L. Buckeridge1,3, W. Alton Russell1,3

1School of Population and Global Health, McGill University, Montreal, Canada

2Canadian Blood Services, Ottawa, Canada

3COVID-19 Immunity Task Force, Montreal, Canada

**Corresponding author:** W. Alton Russell [alton.russell@mcgill.ca](mailto:alton.russell@mcgill.ca), +1-514-396-1476

Suite 1200, 2001 McGill College Avenue, Montreal, QC, Canada H3A 1G1

**Key words:** Blood Donor; COVID-19; Surveillance; Bayesian multilevel regression; CAR

**Running title:**

# Abstract (~250)

Canadian Blood Services (CBS) is monitoring COVID-19 immunity levels in in the largest continuous seroprevalence study in Canada through the testing of randomly selected blood donations. While blood donors are a convenient and cost-effective group for public health surveillance, their characteristics may not reflect those of the general population. Therefore, it is important to evaluate and adjust for any potential biases. In this study, we assessed the correlation between demographic covariates and SARS-CoV-2 seropositivity across three pandemic waves (Pre-Delta, Delta, and Omicron) in four regions (Ontario, British Columbia, Prairies region and Atlantic region). We also compared multilevel modeling approaches with or without spatial effects. We found that Bayesian multilevel models with conditional autoregressive (CAR) priors consistently achieved better fitting by having the lower posterior predictive error in SARS-CoV-2 seropositivity when using individual factors and spatial factors. Our findings highlight how Bayesian multilevel modeling using antibody test data can provide valuable insights into population COVID-19 immunity levels, informing policy making.

# Introduction

The ability to accurately identify and document positive cases is an essential component of any successful surveillance system. Since most cases of SARS-CoV-2 do not require active care, governments and public health organizations have been forced to modify existing surveillance measures to accurately monitor SARS-CoV-2 in the general population. Surveillance programs have since provided vital information to policymakers regarding infection prevalence, vaccine coverage, and antibody waning [1,2]. Testing of blood samples collected for donation or other healthcare purposes have contributed significantly towards SARS-CoV-2 surveillance within Canada and abroad [3–5]. Blood sample surveillance can model temporal patterns, infection prevalence, and highlight differences in disease dynamics across subgroups [3].

Before the SARS-CoV-2 pandemic, multiple countries had already used blood donor samples for public health surveillance. Denmark established a national blood donor cohort across all donation centers to monitor changes in health status among a relatively healthy population [6]. Blood donor samples have previously been used to estimate the incidence of Chikungunya virus in Puerto Rico and are currently being explored for use in surveillance of other infectious diseases [7,8]. Blood donors present several advantages for population health surveillance, including access to repeated donations over time, low cost, and ability to rapidly increase testing capacity. However, factors such as the use of different measurement assays and low positive case counts in some regions can obscure estimation of some parameters such as seroprevalence [9].

Population seroprevalence of natural infection to SARS-CoV-2 is typically characterized using assays testing for antibody response to the nucleocapsid (anti-N) of the SARS-CoV-2 virus. Previous studies have found that population seroprevalence is sensitive to factors including sampling frame, location, occupation, and other demographic characteristics [10,11]. However, identifying significant predictors of seroprevalence is an equally important surveillance task as it aids resource allocation. Within Canada, anti-N seropositivity has been associated with belonging to a visible minority group and residing within regions experiencing social and/or material deprivation [12]. Geographic variation in predictors of anti-N seroprevalence has also been observed. Using a cohort of blood donors across nine Canadian provinces, O’Brien et al. identified belonging to a visible minority was significantly correlated with higher rates of seropositivity in only Ontario, British Columbia, and the Atlantic provinces [13]. However, population-based seroprevalence studies face several challenges, including small sampling area and use of multiple different detection assays, which may skew estimates [9,14,15]. In addition, blood donors differ from the general population on characteristics such as overall health status, which may reduce generalizability [16]. Application of Bayesian multi-level modeling methods which adjust for these biases may provide more accurate seroprevalence estimates when using blood donor data.

A considerate amount of samples were collected by Canadian blood services and being utilized for the SARS-CoV-2 antibody testing for disease surveillance since the very beginning of the pandemic. Covid-19 pandemic lasts over two years during which several variants of the SARS-CoV-2 dominants the infections of the community and different provinces have conducted different prevention strategies like suggesting wearing mask, lock downs and vaccination distribution etc. in response to the spread of the SARS-CoV-2 in a shorter period. While the possibility of an individual being infected is also quite related to whether remote working opportunities are available, the neighborhood he/she lives in, education level and social and material deprivation status etc. The surveillance for areas are challenging, it is worth to borrow information across neighboring counties to get more accurate county level seropositivity estimation. On the other hand, the corresponding covariates may not always available, utilizing the spatial model may account for the unknown spatial confounding compared to the general mixed effects model and facilitates a better fitting, which may result in different understanding of the overall social demographic features for the policy makers in general.

The current study aimed to estimate individual- and neighborhood-levels predictors of anti-N seroprevalence using a large cohort of Canadian blood donors from nine Canadian provinces. Several fixed and random effects multilevel regression models were compared to identify significant predictors of seropositivity tested by Roche assay during the pre-Delta (December 14 2020 – August 1 2021), Delta (August 1 2021 – December 15 2021), and Omicron (December 15 2021 – December 1 2022) pandemic waves.

# Methods

Generalized linear model with both fixed effects, regular random effects (unstructured) and structured random effects are used to compare the seropositivity estimation across counties as well as the model fitting in Widely applicable information criterion (WAIC) score. Conditional autoregressive prior to construct the relationship between the geographical areas. Estimation results from shorter time range break is also calculated to evaluate the coefficient change according to the progression of the pandemic. Table One of the population distribution according to different individual or areal level covariates for all the targeted regions for each wave are provided.

## Data

We used the donation data across major stages (pre-delta, delta, omicron) of the pandemic from pre-delta period started from December 2020 to the late omicron which is December 2022 of four regions across Canada excludes Quebec. They are Ontario, British Columbia, the prairies region which includes Alberta, Saskatchewan, Manitoba, and the Atlantic region consists of New Brunswick, Newfoundland and Labrador, Nova Scotia, and Prince Edward Island. Since the seropositivity at the FSA level is too fine to arouse interest, it is more meaningful to obtain the county level seropositivity estimation. Knowing the fact that we only have the geographic information of the fsa level, while for some FSA area could be mapped to multiple census divisions, thus we develop a strategy to map the FSA to the census division proportional to the the census population at the donor level. We drop the donations with the missing demographic or geographic information. Since there are also situations that people fill out their hometown address while living in the areas they work or attend schools, we also drop the census divisions that has less than 5 donations in total. In the end, a total number of 420,320 donations from 239346 donors during the 24 months of the pandemic period were utilized into the analysis.

## Modeling method

We are particularly interested at investigating how the social demographic features influence the SARS-CoV-2 across different regions of Canada for each wave. Besides the demographic or the geographic covariates, time trend is also a significant factor to be taken into account, so we also incorporate the sample month into the regression model. We mainly focus on the comparison of the models utilizing the spatial information. Multilevel Regression with the regular fixed effects, as well as different unstructured or structured random effects (ICAR prior) are evaluated to fit the CBS data across four major regions (Ontario, British Columbia, the Parries region and the Atlantic Region) during three pandemic waves (Pre-delta, Delta & Omicron). We conducted our analysis through brms which allows comparing varies GLM.

# Results

Based on the binary outcome from the anti-N tested results, we fit with generalized linear model with the logit link function. The linear portion of the GLM will incorporate both individual and the areal level covariates, together with sampling month to take care of the increasing trend as the virus spread. It is always challenging to include all the significant factors into the regression model to achieve the best fit, thus a more accurate insight of the coefficient estimation, however spatial correlation among the neighboring counties could be utilized to capture the spatial confounding to obtain a more reliable understanding of each demographic feature. The relationship between the adjacent counties could be incorporated through a conditional autoregressive prior which the probability of observing a positive sample in one area is associated with the SARS-CoV-2 seroprevalance from the areas that share borders. We also fitted with adding the regional random effects which is unstructured just in order to compare with the model with CAR prior. The convergence diagnostic is carefully checked for all the corresponding parameters involved in the Bayesian models, with most of the largest r-hat equal to 1 and several cases reach to 1.01 which rarely observed.

## Predictors of seropositivity by region and wave

Overall, material deprivation seems to be positively associated with seropositivity except for the Atlantic region where a few samples collected, while the social deprivation tends to be negatively associated with the seropositivity. This result in general aligns with the common sense that people living in areas that are least lack of material resources associated with low education, insecure job situation and insufficient incomes, and most socially separated with other people will be less likely to be tested positive. (Figure 1)

## Predicitons by census division

Taking the county level estimation from Ontario for example, in general the generalized linear model accounting for the spatial correlation provides more accurate estimation compared to the regular GLM without incorporating any geographic information at all, seeing from the fact that the posterior mean is more closed to the observed value of seropositivity for most of the counties and it always falls within the 95% of the HPD interval from the spatial model. However, the regular GLM fails to capture the true value for certain counties even though regardless of the sample size. Similar results apply for other regions as well (Figure 2).

# Discussion

Overall the material deprivation seems to positively associated with the anti-N seropositivity and social deprivation index seems to be negatively associated with the seropositivity for major regions of Canada, though the effect of material and social deprivation is not significant for Ontario, British Columbia (BC) and the Prairies. While for Atlantic region, the sign of the estimation coefficient of the material deprivation is not consistent with the other regions through the pandemic. Typically during the Delta period, it shows a positive effect for the social deprivation and the seropositivity unlike the other regions. This is caused by the fact that fewer data available from the Atlantic region until the omicron started, so the coefficient estimation from the regression model will not be reliable based on small sample. Again except for the Atlantic region which lacks enough sample to provide reliable understanding of the coefficient, regarding the individual characteristics for first two waves of the pandemic, age has significant negative effect on the seropositivity, with the estimation centered around -0.02 for the delta period and -0.03 for the omicron. The coefficients are small due to applying the original scale of the age instead of grouping it into the age categories. Belonging to white racial group tends to be less likely to be tested positive, while the effect is not significant for the prairies region during pre-delta and delta period. Male group are more likely to have significant positive effect on test positive result except for the samples from the prairies region during the delta wave, which presents negative effect in general, being inconsistent with other regions. In conclusion, the younger non-white male tends to have higher seropositivity across all waves.

In general, the coefficient estimation from the general linear model with/without the structured spatial random effect is much closed and for some circumstances GLM with spatial effect will break the significance of the effect of the material and social deprivation especially for Ontario and BC. However for the individual covariates age, race and sex, the significance of the estimation from different models always remain consistent. Since the material and social deprivation index quantile are all mapped from the census dissemination area level, and urban/rural indicator is calculated from the second digit of the postal code, it makes sense that the areal level covariates are more sensitive to adding the spatial random effect compared to the individual covariates age, race and sex. It is also interesting to find that only for the Prairies living in urban tends to have significant negative association with seropositivity across all the pandemic period. However, except for the estimation for the Atlantic region omicron wave from the glm with spatial effect, living in urban has no significant effect on the seropositivity. Breaking the time period into a three months period across the 24 months during the pandemic allows us to take a further look of how the areal or individual level effect on the seropositivity change overtime as the pandemic develops, combining in three-month is chosen to reflect the estimation trend overtime while assuring a enough sample size for obtaining the reliable estimation from the regression model. Living in urban significantly negatively affects the seropositivity for the prairies region (Figure S1), while for the estimation of BC and Ontario, there is a trend of increasing to the positive side at mid delta period and then slowly towards to the other side at the end of the omicron study at the November 2022. The trend is expected as in the earlier stage, people living in urban areas may suffer from more exposure of the virus till the gradually opening of the community during late omicron period, when the coefficient estimation tends to vanish as the people living in urban rural areas may interact more. We can observe similar diminishing effects for the material deprivation, however for social deprivation the effect is not pronounced at the latter half of the pandemic. Similar tendency is observed for the sex character, for the prairies and BC the race differences seem to decrease with time till the earlier period of Omicron and then started to increase but still remain at the negative side as the pandemic progresses. On the other hand, the age effect remains significant at around -0.025 and become more stable at the later stage of the pandemic (Figure S2). The county level seropositivity estimation from the the glm model with the spatial random effect is more closed to the observed seropositivity for most of the counties especially given enough sample size, as is shown by Figure S2 of Ontario and BC throughout all three waves (Figure 2).

Generalized linear model (GLM) framework with logit link function is applied for modeling the relationship between the binary response variable and various types of covariates, which serves as a base comparison model to hierarchical models with utilized the geographical information of census division identifier or further the neighboring information between the divisions. Widely Applicable Information Criterion (WAIC) is used to compare the Bayesian model fitting result, lower values indicate better model performance in terms of balancing goodness-of-fit and model complexity. It shows that except for the Atlantic region during the Delta wave when there is few positive cases, the GLM with spatial random effect (CAR prior) consistently fitted better compared to the GLM, a little bit better than the model with only the cluster random effect being incorporated. Since we do not have strong preferences or prior beliefs of the parameters involved in the hierarchical models, non-informative flat priors for coefficient parameters and weakly informed prior t(3, 0, 2.5) for the hyperparameter for intercept and standard deviation are being used. That might be the reason why the structured spatial effect only has a slight edge over the GLM with the regular grouping random effect. (Table 1)

We can also get the similar understanding for the individual level covariates from both the modeling results (Figure 1, panel B) and the one comes with aggregated results (Figure S4). There exists a weird performance of the prairies region that the rural seropositivity is comparatively higher than the urban area. The result showing in Figure 1, panel A and Figure S1 is consistent with the aggregated result grouped by category in Figure S3, demonstrating the comparatively huge discrepancy between the urban and rural areas of the prairies region, and the plots look similar withholding one or two among Alberta, Saskatchewan and Manitoba that together make up the prairies. The extremely lower proportion of the reported vaccination within three months recorded by the Canadian Blood Services might be the potential cause for the gap, especially started from the Delta wave (Figure S5). While the donor reported vaccination rates varies little regarding the individual characteristics (Figure S6). Considering the fact that the reported vaccination rate in the Pre-Delta for both BC and the Atlantic region is very similar, another interpretation of this disparity especially happens in the prairies region is the outbreak happened in April 2020 at a meat-packing plant in High River, being located in rural area of Alberta, which result in the leading seropositivity of the rural over the urban area. However, as there is unbalanced donation population between rural and urban areas, other potential risk factor effect on a certain smaller population in prairies rural area could also be another interesting aspect to be investigated.

# Conclusion

The Canadian Blood Services (CBS) plays a major role in monitoring COVID-19 immunity levels in Canada through the testing of randomly selected blood donations. In this study, we assessed the correlation between demographic covariates COVID-19 seropositivity across three pandemic waves (Pre-Delta, Delta, and Omicron) in four regions (Ontario, British Columbia, Prairies region and Atlantic region). We also compared multilevel modeling approaches with or without spatial effects. We found that Bayesian multilevel models with conditional autoregressive priors explained most of the variation in COVID-19 seropositivity when using individual factors and spatial factors. Our findings highlight how Bayesian multilevel modeling using antibody test data can provide valuable insights into COVID-19 immunity levels in the population, informing policy making. It is crucial to control the increasing seropositivity by adding the month covariate, especially during Omicron period as the significant trend involves. Though it is counterfactual from our initial guess that people from urban areas are more likely to be infected by living under larger population density, it indicates that we might need to incorporate other covariates like vaccination status and more general self-precaution being carried out. It is also showed from other studies that the high seropositivity may cause by less precautions from rural areas.

We used individual covariates (age, sex, race) and neighborhood covariates (urbanicity, material and social deprivation) to predict COVID seropositivity. Multilevel modeling with structured random effects consistently provides a better model fit across the regions during each of the three waves, compared with generalized linear model with just fixed effects or the unstructured random effect only. However, the coefficient estimation from the posterior distribution shows that donation from a rural area tends to result in a higher seropositivity, especially for the Prairies region.

# Declarations

**Funding:** This work is founded by Canadian Blood Services (CBS)

**Conflicts:** none declared

**Ethics/Consent:** This research is approved by IRB of McGill university

**Data and materials:** Data is provided by COVID-19 Immunity Task Force (CITF)

**Code availability:** A

**Authors’ contributions:**

# References

1. Manuel DG, Delatolla R, Fisman DN, et al. *The Role of Wastewater Testing for SARS-CoV-2 Surveillance*. Ontario COVID-19 Science Advisory Table; 2021. doi:[10.47326/ocsat.2021.02.40.1.0](https://doi.org/10.47326/ocsat.2021.02.40.1.0)

2. Ward H, Whitaker M, Flower B, et al. Population antibody responses following COVID-19 vaccination in 212,102 individuals. *Nat Commun*. 2022;13(1):907. doi:[10.1038/s41467-022-28527-x](https://doi.org/10.1038/s41467-022-28527-x)

3. Saeed S, Drews SJ, Pambrun C, Yi Q-L, Osmond L, O’Brien SF. SARS-CoV-2 seroprevalence among blood donors after the first COVID-19 wave in Canada. *Transfusion*. 2021;61(3):862-872. doi:[10.1111/trf.16296](https://doi.org/10.1111/trf.16296)

4. Charlton CL, Nguyen LT, Bailey A, et al. Pre-Vaccine Positivity of SARS-CoV-2 Antibodies in Alberta, Canada during the First Two Waves of the COVID-19 Pandemic. *Microbiology Spectrum*. 2021;9(1):e00291-21. doi:[10.1128/Spectrum.00291-21](https://doi.org/10.1128/Spectrum.00291-21)

5. O’Brien SF, Goldman M, Drews SJ. An expanded role for blood donor emerging pathogens surveillance. *CMAJ*. 2023;195(1):E16-E16. doi:[10.1503/cmaj.147635-l](https://doi.org/10.1503/cmaj.147635-l)

6. Erikstrup C, Sørensen E, Nielsen KR, et al. Cohort Profile: The Danish Blood Donor Study. *International Journal of Epidemiology*. October 2022:dyac194. doi:[10.1093/ije/dyac194](https://doi.org/10.1093/ije/dyac194)

7. Simmons G, Brès V, Lu K, et al. High Incidence of Chikungunya Virus and Frequency of Viremic Blood Donations during Epidemic, Puerto Rico, USA, 2014 - Volume 22, Number 7—July 2016 - Emerging Infectious Diseases journal - CDC. doi:[10.3201/eid2207.160116](https://doi.org/10.3201/eid2207.160116)

8. Custer B, Grebe E, Buccheri R, et al. Surveillance for Zika, Chikungunya, and Dengue Virus Incidence and RNAemia in Blood Donors at 4 Brazilian Blood Centers During 2016–2019. *The Journal of Infectious Diseases*. 2023;227(5):696-707. doi:[10.1093/infdis/jiac173](https://doi.org/10.1093/infdis/jiac173)

9. Tuite AR, Fisman D, Abe KT, et al. Estimating SARS-CoV-2 Seroprevalence in Canadian Blood Donors, April 2020 to March 2021: Improving Accuracy with Multiple Assays. *Microbiology Spectrum*. 2022;10(1):e02563-21. doi:[10.1128/spectrum.02563-21](https://doi.org/10.1128/spectrum.02563-21)

10. Bobrovitz N, Arora RK, Cao C, et al. Global seroprevalence of SARS-CoV-2 antibodies: A systematic review and meta-analysis. *PLOS ONE*. 2021;16(6):e0252617. doi:[10.1371/journal.pone.0252617](https://doi.org/10.1371/journal.pone.0252617)

11. Skowronski DM, Kaweski SE, Irvine MA, et al. Serial cross-sectional estimation of vaccine-and infection-induced SARS-CoV-2 seroprevalence in British Columbia, Canada. *CMAJ*. 2022;194(47):E1599-E1609. doi:[10.1503/cmaj.221335](https://doi.org/10.1503/cmaj.221335)

12. Reedman CN, Drews SJ, Yi Q-L, Pambrun C, O’Brien SF. Changing Patterns of SARS-CoV-2 Seroprevalence among Canadian Blood Donors during the Vaccine Era. *Microbiology Spectrum*. 2022;10(2):e00339-22. doi:[10.1128/spectrum.00339-22](https://doi.org/10.1128/spectrum.00339-22)

13. O’Brien SF, Caffrey N, Yi Q-L, et al. Cross-Canada Variability in Blood Donor SARS-CoV-2 Seroprevalence by Social Determinants of Health. *Microbiology Spectrum*. 2023;11(1):e03356-22. doi:[10.1128/spectrum.03356-22](https://doi.org/10.1128/spectrum.03356-22)

14. Goldfarb DM, Mâsse LC, Watts AW, et al. SARS-CoV-2 seroprevalence among Vancouver public school staff in British Columbia, Canada: A cross-sectional study. *BMJ Open*. 2022;12(4):e057846. doi:[10.1136/bmjopen-2021-057846](https://doi.org/10.1136/bmjopen-2021-057846)

15. Lewin A, Therrien R, De Serres G, et al. SARS-CoV-2 seroprevalence among blood donors in Québec, and analysis of symptoms associated with seropositivity: A nested case-control study. *Can J Public Health*. 2021;112(4):576-586. doi:[10.17269/s41997-021-00531-6](https://doi.org/10.17269/s41997-021-00531-6)

16. Atsma F, Veldhuizen I, Verbeek A, Kort W de, Vegt F de. Healthy donor effect: Its magnitude in health research among blood donors. *Transfusion*. 2011;51(8):1820-1828. doi:[10.1111/j.1537-2995.2010.03055.x](https://doi.org/10.1111/j.1537-2995.2010.03055.x)

# Tables

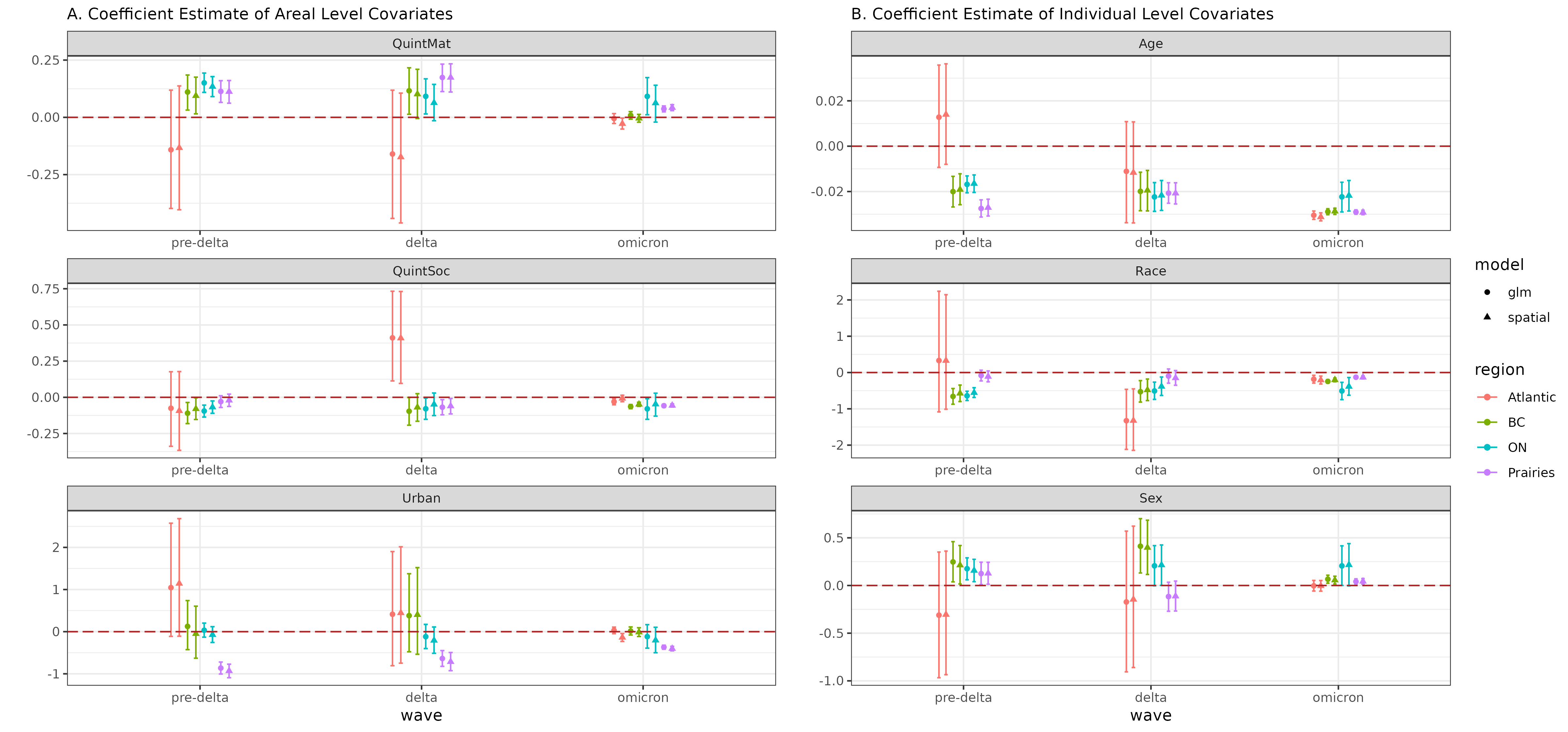
## WAIC by model type, region and wave

**Table** **:** Waic score from different models per region/wave.

| **Region.Wave** | **GLM** | **GLM.w.nonspatial** | **GLM.w.spatial** |
| --- | --- | --- | --- |
| Ontario-PreDelta | 10,953.04 | 10,792.09 | 10,778.91 |
| Ontario-Delta | 3,262.97 | 3,212.39 | 3,211.86 |
| Ontario-Omicron | 161,243.10 | 160,681.99 | 160,674.40 |
| BC-PreDelta | 3,539.64 | 3,494.69 | 3,493.32 |
| BC-Delta | 1,816.50 | 1,787.33 | 1,787.34 |
| BC-Omicron | 56,952.86 | 56,580.27 | 56,578.95 |
| Prairies-PreDelta | 9,967.58 | 9,885.03 | 9,881.98 |
| Prairies-Delta | 5,432.21 | 5,357.38 | 5,354.64 |
| Prairies-Omicron | 99,922.69 | 99,358.87 | 99,351.41 |
| Atlantic-PreDelta | 501.87 | 500.22 | 500.19 |
| Atlantic-Delta | 354.82 | 356.90 | 356.11 |
| Atlantic-Omicron | 31,353.12 | 31,233.64 | 31,223.49 |

# Figures

## Coefficients for individual and neighborhood covariates by COVID-19 wave and region



**Figure :** Coefficient estimate of areal and individual level covariates through different models across three waves.

## Posterior predicted seropositivity vs. observed seropositivy by census division: Multilevel model with spatial effects

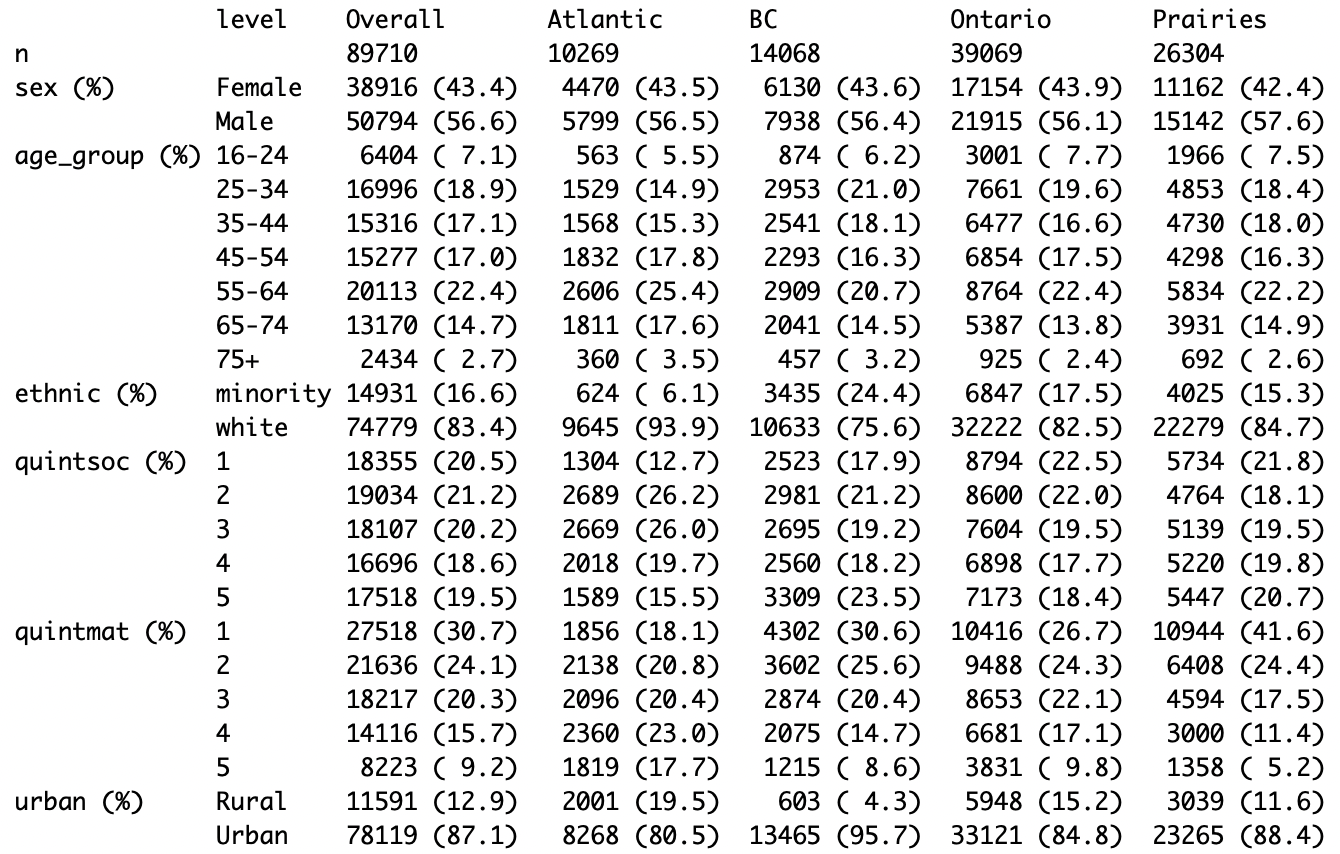


**Figure :** Seropositivity estimation of Ontario and British Columbia across three waves through model with/without spatial random effect.

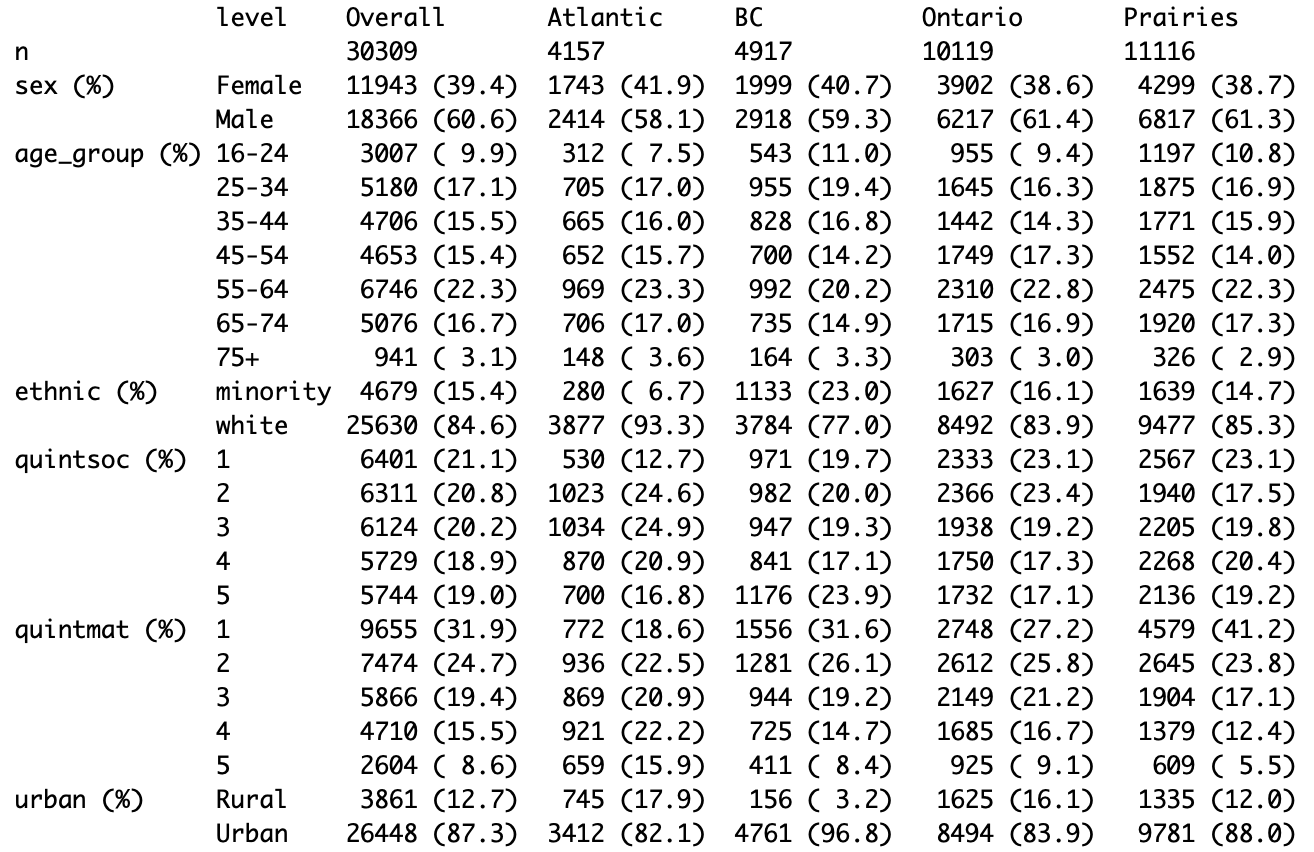
# Supplemental materials

# A. Supplement section

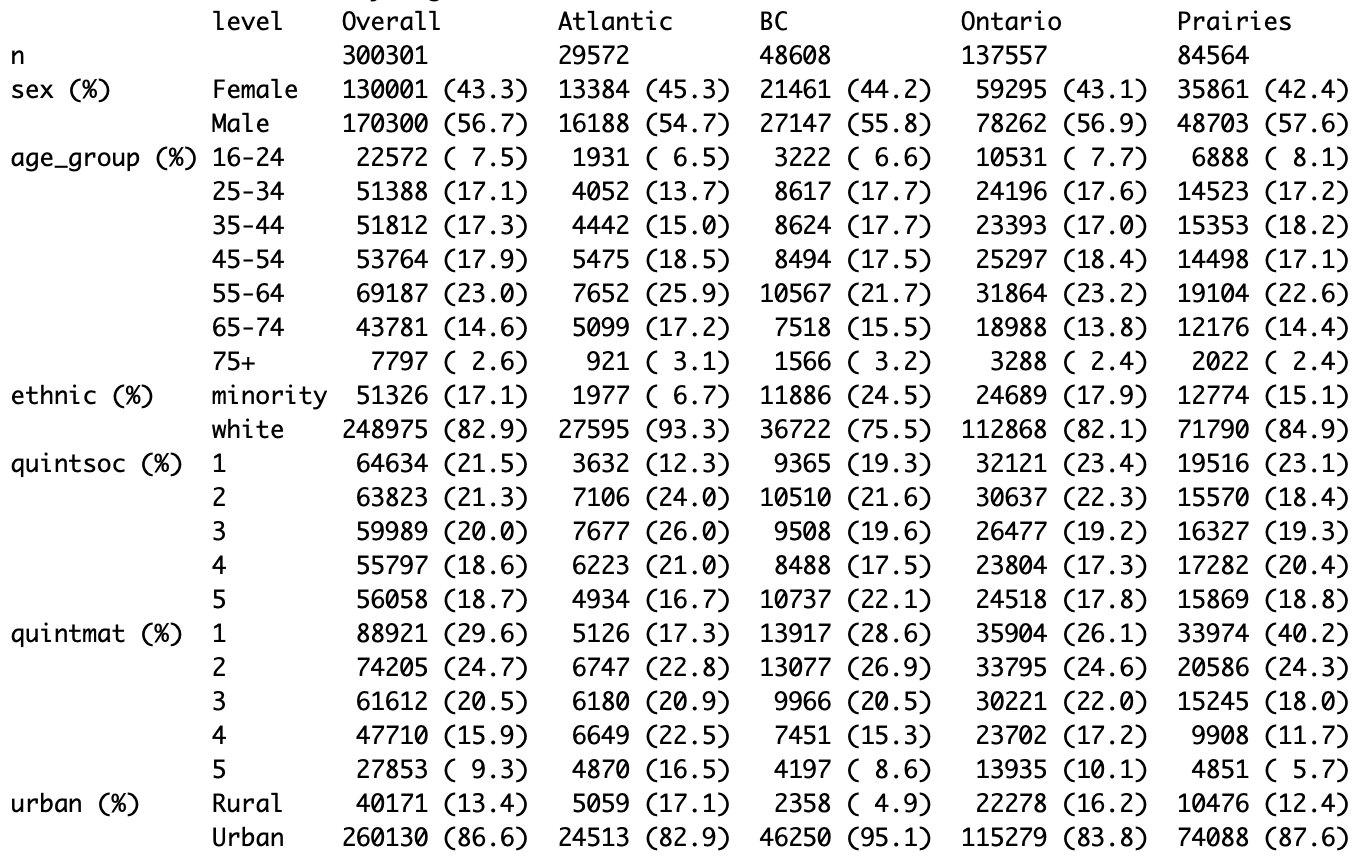
# Supplemental tables



**Table S:** Tableone of Pre-Delta



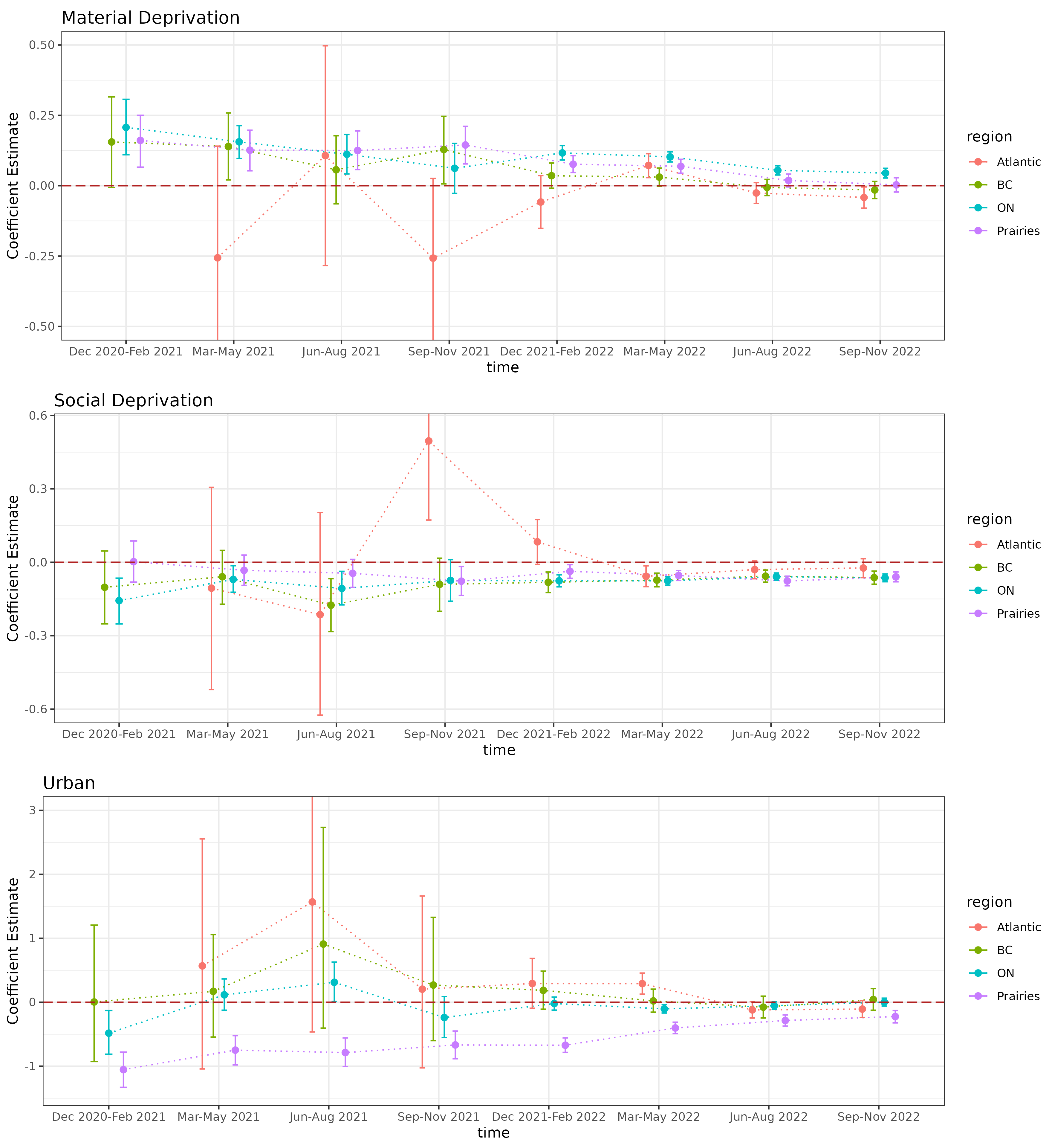
**Table S:** Tableone of Delta



**Table S:** Tableone of Omicron

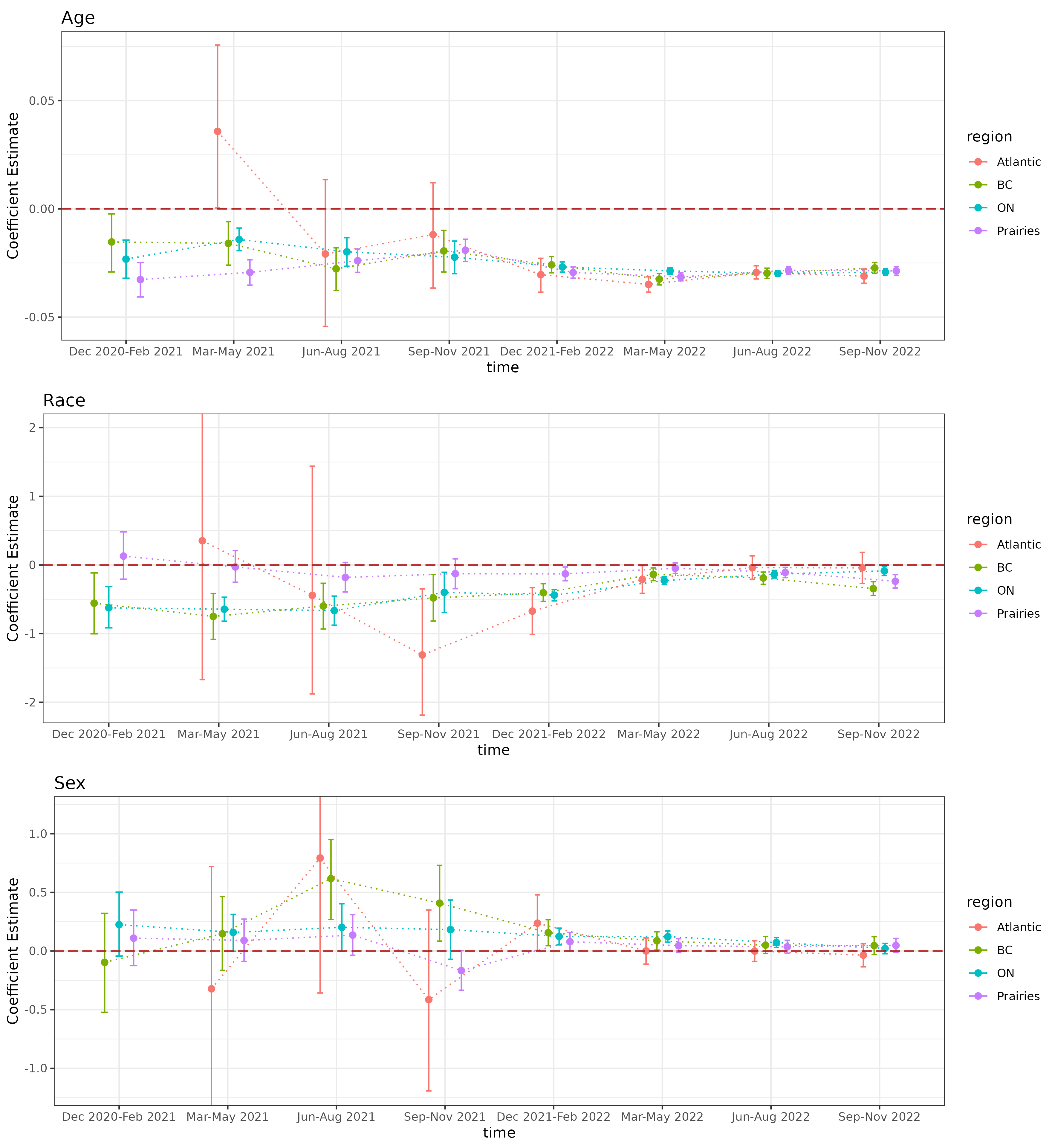
# Supplemental figures

## Coefficient estimation trend of areal covariates



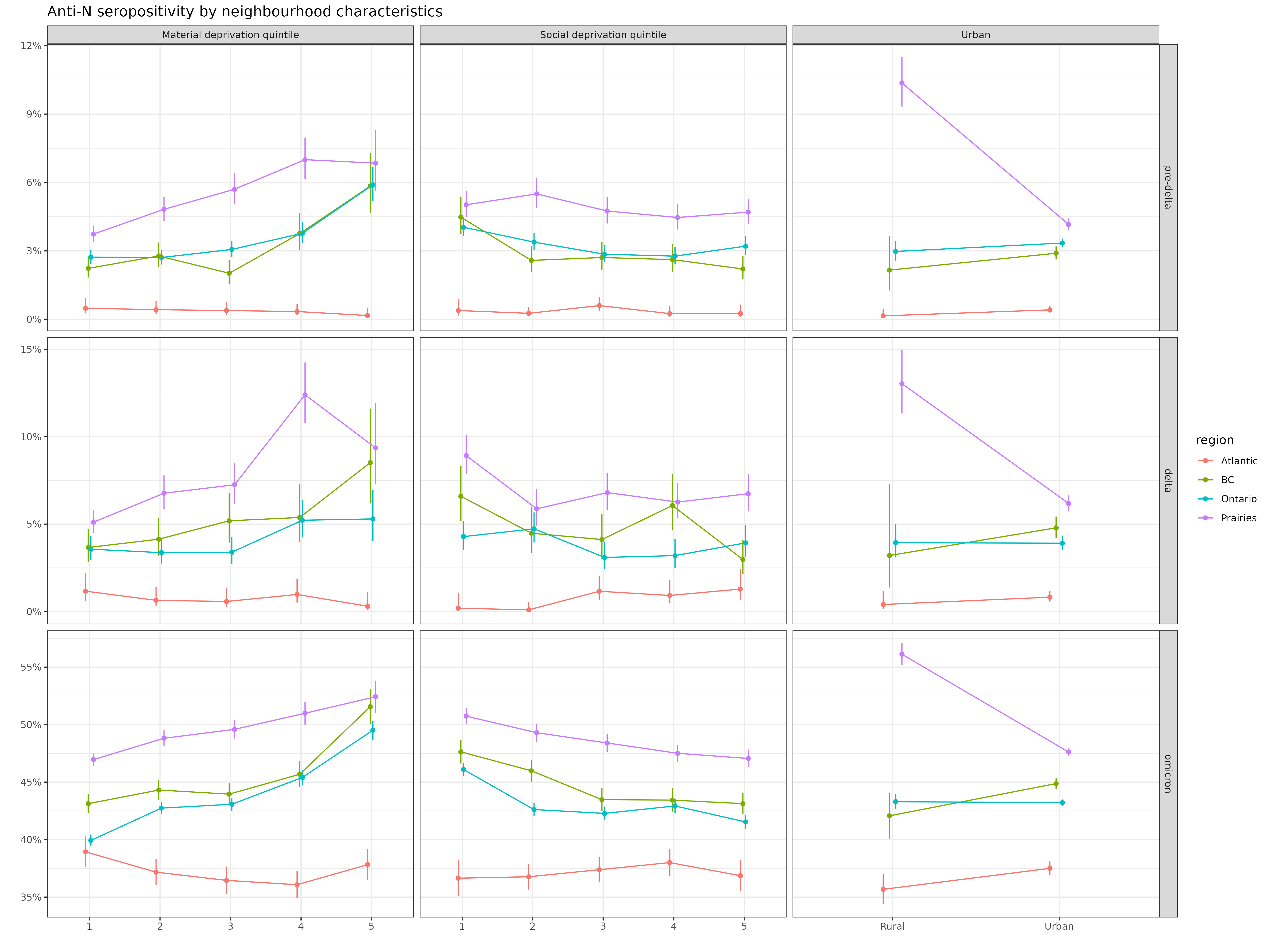
**Figure S:** Coefficient estimation trend of areal covariates

## Coefficient estimation trend of individual covariates



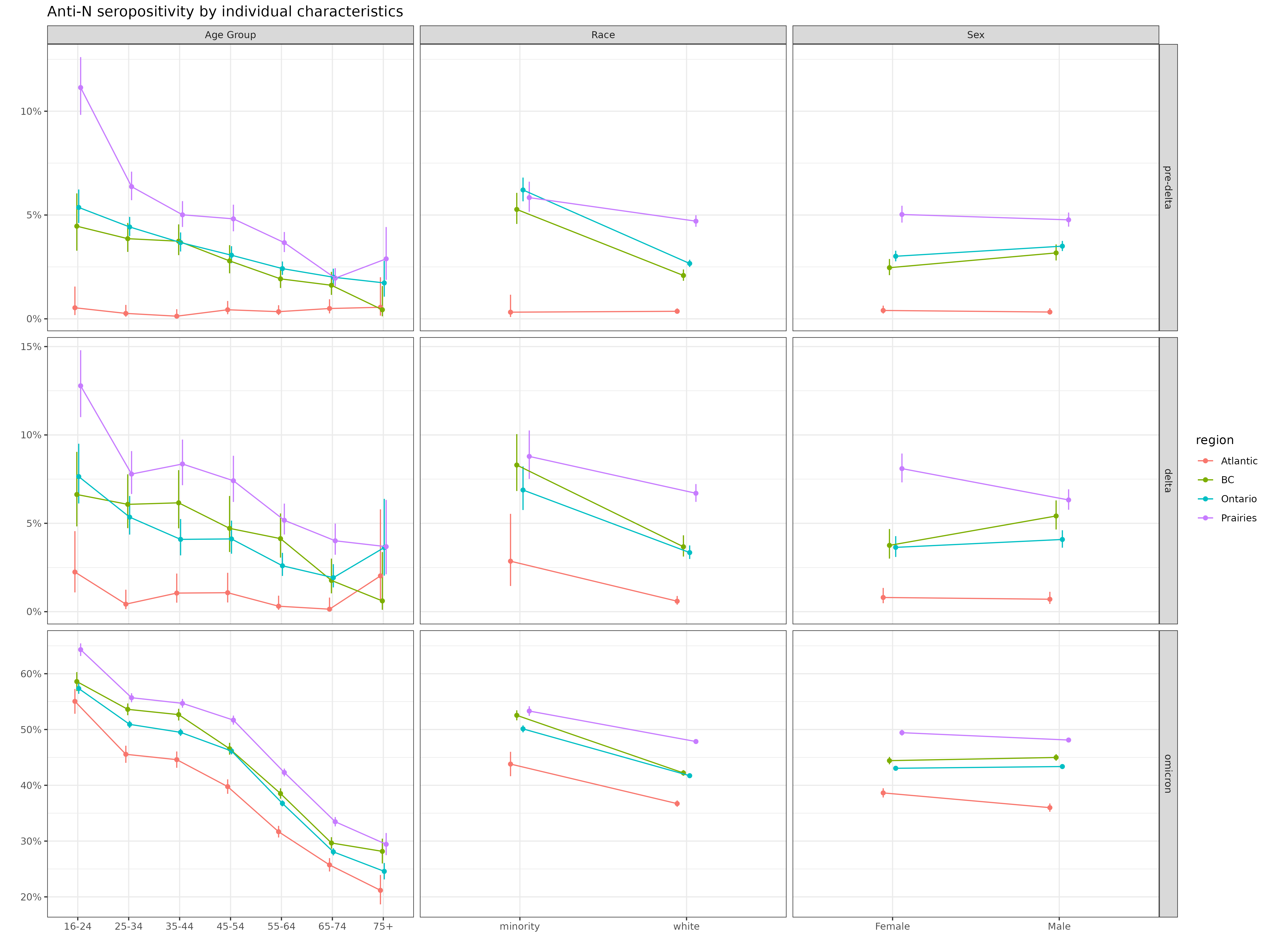
**Figure S:** Coefficient estimation trend of individual level covariates

##Anti-N seropositivity aggregated by areal level covariates, region and wave



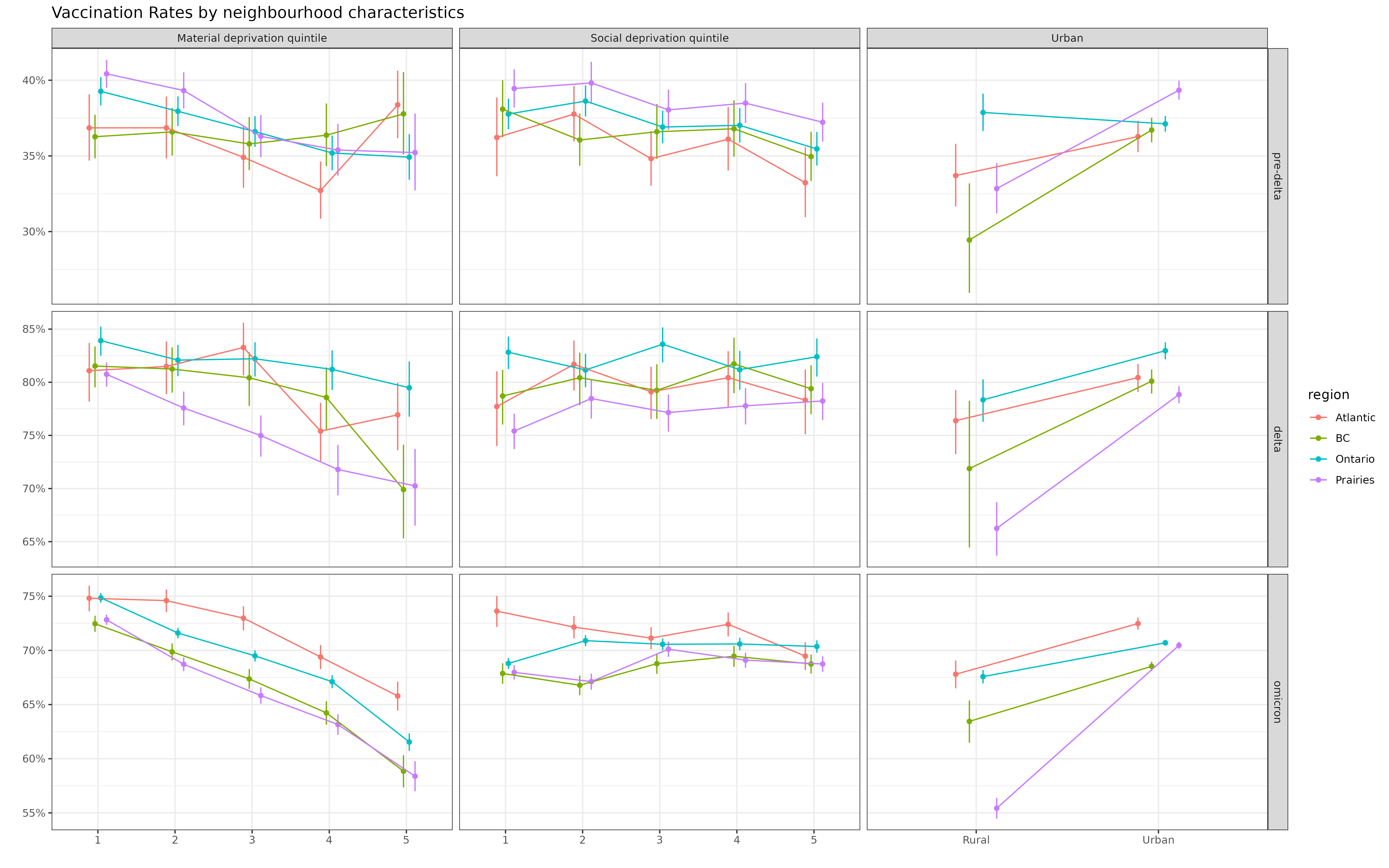
**Figure S:** Anti-N seropositivity of areal level covariates by region/wave

##Anti-N seropositivity aggregated by individual level covariates, region and wave



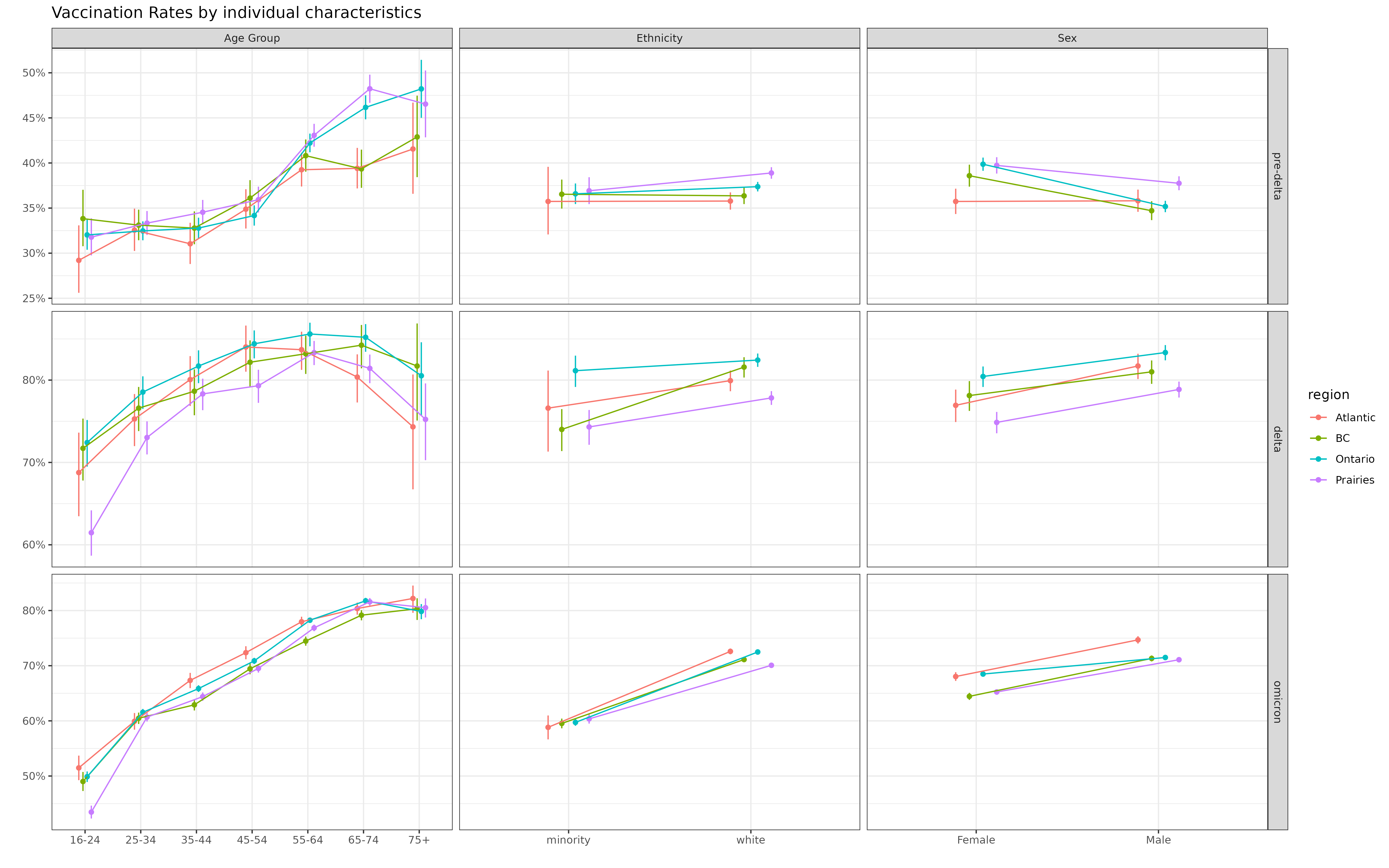
**Figure S:** Anti-N seropositivity of individual level covariates by region/wave

## Reported vaccination proportion aggregated by areal level covariates, region and wave



**Figure S:** Reported vaccination proportion of areal level covariates by region/wave

## Reported vaccination proportion aggregated by individual level covariates, region and wave



**Figure S:** Reported vaccination proportion of individual level covariates by region/wave

## Timing of COVID serological testing by Canadian Blood Services (Might aks Sheila if it is good to include)