Article details: 2021-0085	
Title	Association between new onset anosmia and positive SARS-CoV-2 tests among people accessing outpatient testing in Toronto, Ontario: a retrospective cross-sectional study
Authors	Braden O'Neill MD DPhil, Sumeet Kalia MSc; Peter Gill MD DPhil, Susan Hum MSc, Carla Moran-Venegas BSc, Rebecca Stoller MD; Michelle Greiver MD MSc, Payal Agarwal MD, Abirami Kirubarajan MD MSc, Samuel DeKoven MD, David Eisen MD, Andrew Pinto MD MSc, Sheila Dunn MD MSc
Reviewer 1	Dr. Kate Zinszer
Institution	McGill University, Montréal, Que.
General comments (author response in bold)	The authors present a robust analysis of SARS-CoV-2 testing data to determine the PPV of anosmia. The manuscript is well-written with some comments that should be addressed to strengthen the analyses: Methods - Its unclear, the North York General Hospital has 2 sites? **North York General Hospital has two sites and the COVID-19 assessment centre was located at one site that usually provides outpatient care only. We have updated this as follows in the 'study design': The 'community hospital' setting in this study, North York General Hospital, is a medium-sized hospital with 435 inpatient beds; data in this study are from their COVID-19 Assessment Centre in an outpatient setting in a lower socioeconomic status neighbourhood in northwestern Toronto.** - Why was Sept 2020 chosen as at the cut-off date? **The most recent date was chosen as the cut-off date when we requested the EHR data to be extracted from two hospitals.** - Any information on the data quality (completeness of symptoms, missingness, etc)? How was missing data dealt with? **We believe that the extracted EHR data was of sufficient quality to generate reliable inference around the association of anosmia and positive COVID cases. We reported the completeness (or missingness) of the symptoms in Table 1(b) and also added the appropriate text in the limitation section of the manuscript.** - It was unclear if data include repeated measures on the same individual and how the determinant was made between new infections in the same individual vs viral clearance (like rare but still important to clarify)? **Repeated measures information was available for some individuals in this study. However, in the overall population, the repeated-measures information was very sparse because most of the individuals visited the clinic only once during the study period. Nonetheless, we accounted for the patient-level dependence (i.e. repeated-measures) using robust sandwich estimation of the standard errors in generalized estimating equations.** - Explain how sensitivity/specificity/PPV/NPV were calculated **The diagnostic measures of sensitivity, specificity, PPV, NPV, LR+ and LR- are functions of TP, TN, FP and FN and disease prevalence. We added a citation in the manuscript to refer the interested reader to the definitions of**

	these diagnostic measures: **"We encourage the interested readers to refer to the statistical literature describing the diagnostic measures of medical test for the purpose of classification and prediction."** **And then added the following reference: Pepe, M. S. (2003). The statistical evaluation of medical tests for classification and prediction. Medicine.** - For the GEEs, why was the exchangeable correlation structure chosen? and include a reference, were the results stratified by site (and if not, would be relevant as a sensitivity analysis)? what about the importance of time in terms of an adjustment factor? **There are multiple options available to specify the correlation structure in GEE models. Some examples include exchangeable correlation, auto-regressive correlation, Toeplitz correlation. We chose exchangeable correlation structure to capture inter-dependencies with respect to repeat visits of each patient. It is important to note that the GEE models are robust to the misspecification of the correlation structure when the total number of clusters are large (Zeger et al, 1988). The results for the diagnostic measures were stratified with respect to WCH site (see Figure 2). We chose not to adjust for the time component in the GEE model due to the possibility of adding more instability (i.e. increased standard errors) of the odds ratios.** **Zeger, Scott L., Kung-Yee Liang, and Paul S. Albert. "Models for longitudinal data: a generalized estimating equation approach." Biometrics (1988): 1049-1060.** - Why were other symptoms included in the adjustment? Was multicollinearity assessed? **We adjusted for common symptoms available in both sites. We assessed the multi-collinearity as a source of confounding when comparing the unadjusted (crude/univariate) odds ratios with adjusted odds ratios (see Figure 3). A large change from unadjusted odds to adjusted odds correspond to the presence of confounder (Vach W, 2012).** **Vach, W. (2012). Regression models as a tool in medical research. CRC Press.** - The process of model selection and fit? **We used apriori knowledge about clinical predictors to fit the GEE model. This is in agreement with the conceptual framework literature using the directed acyclic graphs in epidemiology literature (Rothman et al, 2008).** **Rothman, K. J., Greenland, S., & Lash, T. L. (2008). Modern epidemiology (Vol. 3). Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins.** - Include Table S4 in main text but remove unadjusted and indicate what the models were adjusted for **We chose to use the panels in the forest plot as a visualization of the associations with respect to clinical symptoms and COVID positivity. We reported both the crude and adjusted odds ratios to gather more information about a strong confounder in the adjusted covariates. The adjustment covariate set of the GEE model is described in the main text:** **"Generalized estimating equations also adjusted for patient demographics**

	(age, sex, travel history) and common symptoms available at both sites (anosmia, cough and/or shortness of breath, diarrhea and/or abdominal pain).” Results - It would be interesting to see the TPV over time and how that changed with respect to age **Since the positivity rate of COVID in this study was very low (~2%), the estimates of TPV stratified with respect to time period will introduce some instability in the statistical inference. In addition, the sparsity of COVID positive cases coupled with low diagnostic measure of sensitivity (i.e. TPV) will provide temporal estimates with increased imprecision. For these reasons, we chose not to stratify the TPV with respect to study period or age-groups.** Discussion - The study is referred to as a cohort study throughout until the discussion, where its said to be a repeated cross-sectional study? **Conceptually this is an observational cohort study with the study period defined from [Apr 12, 2020] to [Sep 30, 2020]. During this study period, patients may visit the COVID clinic multiple times (i.e. repeated measures) depending on symptomatic flares or other clinical/societal factors etc. In order to avoid confusion, we revised the description of study design as “A retrospective cross-sectional study” throughout the manuscript.** - It would be pertinent to situate your findings of anosmia in the context of variants, what is the literature suggesting? **Since this study was conducted in the early stages of the pandemic (i.e. study period was [Apr 12, 2020] to [Sep 30, 2020]), it will be difficult (if not impossible) to describe the findings using different COVID strains. This is primarily due to not having the COVID variant information captured in the EMR sources at NYGH and WCH during the early stages of the pandemic.**
Reviewer 2	Dr. David Forner
Institution	
General comments (author response in bold)	Thank you for the opportunity to review this paper by O’Neill and colleagues, investigating the association between anosmia and COVID-19 test positivity. In this retrospective cross-sectional study, anosmia data was prospectively collected and logistic regression models under a generalized estimating equation framework were used to analyze the association with COVID-19 test positivity, accounting for repeated measures by individuals with repeat testing over time. While the positive predictive value of anosmia was overall low, it represented the highest PPV amongst other collected symptoms, and had an excellent specificity. Overall, the study is well done with sound statistical methods. The study design is appropriate, but is hampered by multiple different testing criteria over time. The interpretation by the authors is well supported by the data provided. That said, I have several questions for the authors: 1. Is it possible to know exactly how the testing criteria changed over time?

	We have updated details about this under 'study population' as follows: "When COVID-19 assessment centres opened on March 12/20, only symptomatic individuals could be tested; this then changed on May 28/20 when asymptomatic people who were concerned about COVID-19 could be tested (19), and then reverted back to only testing symptomatic individuals as of Sept 24/20.(20)" 2. Is it possible to know how the PPV and NPV of anosmia changed with testing criteria over time? For example, completely asymptomatic screening would presumably have a higher NPV (lower prevalence) compared to symptomatic testing would have a higher PPV (higher prevalence of disease in those being tested) **Technically it is possible to estimate PPV and NPV indexed with respect to study period. However, since the positivity rate of COVID in this study was very low (~2%), the estimates of PPV and NPV stratified with respect to time period will introduce instability in the statistical inference. For this reason, we chose not to stratify the TPV with respect to study period.** 3. Was the anosmia question shown to be reliable and valid? (Page 4) **In relation to other symptoms, Figure 1 and 2 showed that anosmia had reasonably higher specificity and PPV.** 4. Were patients prompted as to what "new" problems in smell meant? ie how recently? (Page 4) **Patients were asked the question as described in the manuscript without additional details. There was no particular guidance given about what "new" meant; it is possible if people asked questions about this, nurses asking the questions provided their own interpretation. Our experience (several authors worked in these clinics although we did not directly patients these questions, they were asked by nurses at registration) is that we were not aware of any requests for clarification about this. The questions were piloted the questions for a few days along with other data elements before they were incorporated into systematic data collection as part of starting up the clinics. It is unclear how this might have affected the diagnostic test characteristics of the questions.** **We have not edited this section of the manuscript to reflect these additional details but would be pleased to do so if the reviewer feels this would help the reader understand the process better.** 5. Please define accuracy – I assume this is TN + TP / TN + TP + FN + FP? (Page 4) **Yes, this is the definition of accuracy we used. We have now included this in the manuscript.** 6. Were model assumptions tested and verified? It seems like some variables have the potential to be collinear e.g. fever and body temperature (Page 4) **We adjusted for the following predictors in the regression model: "patient demographics (age, sex, travel history) and common symptoms available at both sites (anosmia, cough and/or shortness of breath, diarrhea and/or abdominal pain)". Although it is possible that some of these clinical**

	predictors will be correlated, we accounted for this possibility by assessing the magnitude of the variance inflation factor during regression modelling. In the revised draft, we chose not to adjust for both fever and body temperature in regression models (since these predictors are likely to contain similar information). We chose to report fever and body temperature as clinical factors in the descriptive baseline table 1(b). 7. Was there informed written consent or was this waived? (Page 4) **We have added the following sentence to clarify this: "These approvals included permission to waive written informed consent since this study was conducted using routinely collected health information."** 8. In the results, the authors comment that 20-29 year old patients had higher rates of test positivity. While this seems true, those 40-49 also had high rates (Actually higher than 20-29), giving the impression that reporting only for 20-29 may be telling a narrative. (Page 5) **We restructured the sentence and chose to report only the descriptive statistics (without placing any emphasis on comparing age categories). The following sentence has been added in the revised draft:** **"The positivity rate among adults aged 20-29 years was 2.25%, those under age 20 years was 1.46%, and older adults aged 60+ years (1.34%)."** 9. Are there any combination of symptoms, perhaps in post-hoc analysis, that offer a better PPV? For example anosmia AND shortness of breath **We did not report the clustering of symptoms in this manuscript as it will deviate from the main focus of this article (i.e. associations between anosmia and positive COVID test). However, we did some post-hoc analysis using the Bayesian latent class models in which we constructed symptomatic profiles of patients based on the severity of multiple symptoms. Although this is not the focus of this manuscript, the Baysian latent class analysis identified ~30% of the COVID-19 positive cases to be asymptomatic (i.e. low prevalence of symptoms). This finding was consistent with the earlier reports published in New York times:** **Many 'Long Covid' Patients Had No Symptoms From Their Initial Infection. https://www.nytimes.com/2021/03/08/health/long-covid-asymptomatic.html** 10. Is it possible to know the false positive and false negative rates of the tests? This is listed as a limitation, so I assume not? **We considered the results of the "swab test" as the gold standard for the diagnosis of COVID. Hence it is not possible to assess the diagnostic measures of swab test itself. However, we used the diagnosis based on swab test to estimate diagnostic measures using the clinical symptoms.** 11. Did length of symptoms play a role? This was collected for WCH but I do not see results for this in the supplement or Figure 2 (where other WCH symptoms are listed) **Unfortunately, the information on length of stay was not available in the extracted data cut from WCH. We removed the text around "length of stay" in the main text.**
Reviewer 3	Dr. Peter George Jaminal Tian

Institution	University of Alberta, Edmonton, Alta.
General comments (author response in bold)	General comment. This study is clinically relevant; it assesses the importance of anosmia in increasing the index of suspicion for COVID-19. Specific comments. 1. I suggest not classifying this study as a cohort study, as indicated in the title. Diagnostic accuracy studies are basically snapshots in time, i.e., cross-sectional studies (not cohort studies). However, in the methods section, this study is specified as a cross-sectional study. **We revised the description of study design as "A retrospective cross-sectional study" throughout the manuscript.** 2. The research question (as reflected in the title) referred to the positive predictive value (PPV) of anosmia for positivity in COVID-19. The results showed a PPV of 12%, i.e., of those who had anosmia, only 12% tested positive. This is not much. I suggest the conclusion be revised to reflect this low PPV. **We agree with this and appreciate the opportunity to reflect that although there is an association between anosmia and SARS-CoV-2 positivity, it is not particularly strong. We have revised the conclusion to be less prescriptive about SARS-CoV-2 testing for people with anosmia, as follows: "In this study of people attending two community-based COVID assessment centres, presence of anosmia did not reliably identify people with COVID-19. However, anosmia's high specificity and positive predictive value of 12% in this community population with low prevalence of SARS-CoV-2 positivity suggests a moderate clinical suspicion of infection in individuals with this symptom. This supports the recommendation that people with new onset anosmia should consider being tested for SARS-CoV-2."** 3. The usefulness of this study's results is in its applicability to the population tested, i.e., a population that has a low prevalence of COVID-19. If the PPV is low, then it is not very useful in predicting COVID-19 positivity, in this low-prevalent population. I suggest toning down the discussion/conclusion to reflect this. **We hope our edits to the 'conclusion' as described above are sufficient to address this comment. We have reviewed the rest of the 'interpretation' section and believe that it is appropriately reflective of the moderate association between anosmia and SARS-CoV-2 positivity and the fact that the 12% PPV does not make it a particularly 'useful' symptom to strongly recommend testing or to recommend self-isolation.** 4. Kindly check the numbers in the following statement: "Other symptoms that were more common in patients testing positive included fever (6.99% vs. 1.59%), cough and/or shortness of breath (5.11% vs. 1.49%), and diarrhea and/or abdominal pain (2.89% vs. 1.90%)." My manual calculations yielded far different numbers: fever (24.9% vs 6.6%), cough/SOB (34.4% vs. 12.8%), and diarrhea (10.1% vs 6.8%). **Our apologies for this confusion. Our intention was to report the row percentages in Table 1(b). We restructured the sentence to reflect the correct interpretation of row percentages in the main text:** **"The prevalence of positive COVID test was more common among patients with clinical symptoms including fever (7.0% vs. 1.6%), cough and/or**

	shortness of breath (5.1% vs. 1.5%), and diarrhea and/or abdominal pain (2.9% vs. 1.9%).”