## Supplementary Material

## Collider bias results in underestimation of an association between smell loss and COVID-19

As others have noted,<sup>35</sup> collider bias, resulting from selection or conditioning on variables involved in the analysis, may result in the distorted association between COVID-19 and candidate symptoms or patient attributes. In the present sample, it is likely that we have selected for both a higher probability of COVID-19 and a higher probability of smell and taste disorders than the population at large. However, rather than leading to an overestimation of the positive correlation between smell loss and COVID-19, collider bias is expected to lead to an underestimation of this correlation (**Figure S1)**. If we consider the hypothetical scenario in which there is no association between smell loss and COVID-19 status in the general population, we would expect a distribution similar to that depicted in **Figure S1A**, where the correlation between the likelihood of smell change and likelihood of COVID-19 is  $r = 0$ . Based on our recruitment method, we expect that the participants who elected to complete the GCCR core questionnaire were likely to have COVID-19, smell loss, or both. We can simulate participant selection to reflect this hypothesis by censoring subjects which do not meet a fixed sum of smell loss and COVID-19 probabilities (i.e., the red dots are excluded from the calculation of the correlation; **Figure S1B**). As a result, the estimated correlation between smell loss and COVID-19 status originating from a population with  $r = 0$  would be negative (**Figure S1B)**. A similar scenario would manifest if the association between smell loss and COVID-19 status in the general population is positive (**Figure S1C**)**.** Again, simulating the removal of participants with low likelihood of having COVID-19 and/or reporting smell loss would result in a bias of the estimated correlation towards more negative values (**Figure S1D**). This collider bias indicates that the positive correlation between smell loss and C19+ is underestimated in the present sample. Indeed, a direct comparison of the binary (y/n) smell loss questions in the two empirical samples yields an C19 odds ratio of 5.96 in the YouGov sample (**Table S1**) but only 4.89 for GCCR. Therefore, our analyses represent a conservative scenario for the prediction of C19+ and C19- based on chemosensory alterations.



Figure S1. Collider bias leads to underestimation of the positive correlation between smell loss and COVID-19-positive status. (A) Hypothetical scenario depicting no relationship between smell change and likelihood of COVID-19 positive status. Black dots indicate individual potential subjects, each of whom has a latent likelihood of COVID-19 and of smell loss. (B) Hypothetical scenario depicting the emergence of a negative correlation between smell change and likelihood of COVID-19 positive status following a baseline lack of correlation, if participants with greater smell loss and/or COVID-19 positive are preferentially included in the sample. Red dots indicate subjects not observed due to this selection bias; subjects observed remain in black. **(C)** Hypothetical scenario depicting a positive relationship between smell change and likelihood of COVID-19 positive status. **(D)** Hypothetical scenario depicting the emergence of a negative correlation between smell change and likelihood of CO COVID-19 positive status following a positive baseline correlation, if participants with greater smell loss and/or COVID-19 positive are preferentially included in the sample.





Column 1 includes the list of countries available in the YouGov database. Column 2 indicates the number of participants over the history of their survey up<br>to July 3, 2020. Minimum N=1000 per time point. Column 3 reports t of participants who reported C19 Columns 5-7 report the odds ratios using Smell Loss for either C19+ vs. C19 (Column 5, and as reported in Figure 5 of participants who reported C19-. Columns 5-7 report the odds ratios using Smell Loss for either C19+ vs. C19- (Column 5, and as reported in Figure 5 based on the data of the GCCR survey), for C19+ vs. untested individuals (Column 6) and C19- vs. untested individuals (Column 7). Untested individuals in YouGov's survey are those who did not report to be sick but were contacted as representative participants of a country. Column 8 reports the probability of smell loss in the C19+ group. The first row indicates the Global average (across countries) weighted by sample size. The global odds ratio for Smell loss<br>calculated from a binary question for the group C19+ vs. C19- is calculated from a binary question for the group C19+ vs. C19- is 6.72, which is greater than what we identify in the GCCR survey (OR for changes in smell (binary question) = 4.89). The OR for C19+ vs. untested individuals is 58 and lowers to 11 for C19-. This confirms that smell loss is also associated with other etiologies, but is not nearly as prevalent as in participants with C19+

## How representative is the GCCR sample?

As with most COVID-19 studies,<sup>19</sup> the sample studied here is not representative of the general population. To better understand the extent to which this is the case, we computed a cross-correlation between GCCR and YouGov data.<sup>36</sup> These data were aligned by weighting YouGov samples to achieve an identical survey date distribution to the GCCR samples. Specifically, GCCR survey dates were converted to a YouGov "week number" because YouGov surveys only weekly. The distribution of week numbers was computed for each country in the GCCR data. The YouGov data for the same country was then weighted by week number to match the corresponding GCCR distribution for that country. So, for example, if a country had 10 GCCR survey responses in week 1 of the YouGov survey period, and 30 in week 2 of that period, then the YouGov data in week 1 would be weighted at 25% and in the YouGov data in week 2 at 75%. This procedure was applied independently for each country, and the weights were used to compute a weighted mean COVID-19-positive rate for each country from the YouGov data. This was then directly compared against the raw COVID-19-

positive rate for each country in the GCCR data. A lag (x-axis value in Figure S2) of 0 exactly reflects the above description. Other values of the lag indicate that the alignment was shifted: for example, a lag ag of one week means that the hypothetical GCCR responses above would be weighted 25/75 towards weeks 2 and 3, instead of weeks 1 and 2. Under the hypothesis that the COVID-19-positive rates in the two surveys are related, but may have different temporal dynamics, changing the lag allows these dynamics to be estimated. Figure S2 depicts the country-wise correlation in participants with a positive COVID-19 test results (C19+) fraction between the two datasets, as a function of the lag between GCCR survey date and YouGov survey date. The country-wise C19+ fraction is correlated ( $r \sim 0.45$ ) when responses from the same calendar week are aligned, but diminishes outside of that window, showing both surveys capture a similar within-country temporal component of the epidemic.



**Figure S2.** COVID-19 status in the GCCR cohort is correlated with a representative YouGov sample.

## Sample description

Based on responses to question 7 of the GCCR survey ("Have you been diagnosed with COVID-19?", **Appendix 1**), participants can be split into six groups (see **Figure 1**). Participants who responded with Option 2 ("Yes – diagnosed with viral swab") or 3 ("Yes – diagnosed with another lab test") were classified as C19+; participants who responded with option 5 ("No – I had a negative test, but I have symptoms") were classified as symptomatic C19-; participants who responded with option 4 ("No – I was not diagnosed, but I have symptoms") were classified as C19 Unknown; participants who responded with option 6 ("No – I do not have any symptoms"), with option 7 ("Don't know"), or with option 8 ("Other") were classified as undefinable and excluded from the final analyses. To replicate our previous findings,<sup>8</sup> we first compared individuals newly included in the GCCR dataset (responses from 14 May to 2 July, 2020, replication sample in **Figure 1**) with COVID-19 who were lab tested and those who were diagnosed by a clinician based on the self-reported quantitative changes in smell, taste, chemesthesis, and nasal obstruction (**Figure S3)**. Participants with labtest confirmed C19+ did show slightly greater chemosensory deficits than did those diagnosed with C19+ clinically, but the difference was not clinically meaningful (smell: 4.4±28.6, p=2.7e-13) (**Figure S3, Table S2**). We then focused our descriptive and predictive analyses of participants who received a positive (C19+) or a negative (C19-) lab test for COVID-19. We also computed descriptive and predictive analysis for the C19+ subsample who reported partial or full signs of recovery from their recent respiratory illness. Lastly, the unknown group was originally hypothesized as similar to the C19- group. Yet the ratings of smell ability during illness suggest that the majority of these participants has a smell profile closer to C19+ than C19- (**Figure S4**). To maximize the validity of the COVID-19 diagnosis in our sample, we therefore excluded the C19 Unknown group from further analyses.



Figure S3. This figure describes a pre-registered replication of Parma et al, 2020 and includes only new data collected between May 14th and July 3rd 2020 via the GCCR survey. *(*A-D) Changes in smell (A), taste (B), chemesthesis (C) and nasal blockage (D) during versus before in COVID-19-positive individuals (Groups 1, 2 and 3, see Figure 1). All subjects had a COVID-19-positive status either via lab test (darker shades) or via clinical assessment (lighter shades). (E-F) Principal component analysis shows that smell, taste, and chemesthesis changes in both the lab test (E) and clinical assessment (F) groups) were orthogonal to blocked nose changes, i.e., the three chemosensory changes were highly correlated across subjects whereas blocked nose changes were mostly uncorrelated.



**Figure S4. (A)** Self-reported smell change and comparison of smell change between four diagnosis groups: Positive COVID-19 lab-test (C19+), positive COVID-19 clinical assessment (C19+ (Clin)), COVID-19 Unknown (Unkn; lack of clinical and lab test diagnosis, but reported symptoms), and negative COVID-19 lab test (C19-). Solid horizontal lines reflect the median; dashed lines reflect the quartiles. (**B, C**) Differences between groups, in terms of (B) effect size (Cohen's D) and (C) means (on a 0-100 scale).

## Replication of previous analyses

The replication of Parma et al.<sup>8</sup> used the same Bayesian linear regression approach with Cauchy prior [r = sqrt(2)/2]. This approach is appropriate for estimating the strength of the evidence in support of the alternative hypothesis: the clinical assessment and the lab test C19+ groups show similar smell, taste, chemesthesis and nasal obstruction changes before vs. during the illness. The interpretation of the Bayes factors BF follows the classification scheme proposed by Lee and Wagenmakers<sup>37</sup> and adjusted from Jeffreys<sup>38</sup>, which considers BF

> 3 as moderate evidence, BF > 10 as strong evidence, BF > 30 as very strong evidence and BF > 100 as extreme evidence for  $H_0$  or  $H_1$ .



**Table S2**. Differences between lab-tested and clinically-assessed COVID-19-positive participants on changes in smell, taste, chemesthesis and nasal blockage.

Change means the rating "before" illness minus the rating "during" illness on the 0-100 visual-analog scale. Δ indicates the mean difference in change between lab-test and clinically-assessed COVID-<br>19-positive subjects, while σ indicates the standard deviation. D indicates effect size (Cohen's D), p 19-positive subjects, while σ indicates the standard deviation. *D* indicates effect size (Cohen's D). *<sup>p</sup>* indicates p-value from a Mann-Whitney U-test. In contrast to the prediction of the pre-registration, we found statistically significant differences between groups. However, the effect sizes are small and thus unlikely to be of practical importance.

## Chemosensory characterization of C19+ and C19-

We asked how accurately COVID-19 status could be predicted from the survey responses. The data matrix had strictly non-negative values and was normalized (column-wise min=0, max=1) to apply regularization in an equitable fashion across features and give regression coefficients the same interpretation for each feature. Compared with the main text, models with similar AUC values (but with non-zero coefficients for additional, likely spurious features) were obtained for smaller values of α, and inferior results for larger ones (which contained fewer or no non-zero coefficients). Quantitatively similar AUC values were obtained for other models predicting COVID-19 status using multiple features including ridge regression and random forest, but L1 regularized logistic regression consistently produced sparser models with comparable cross-validation accuracy. Each logistic regression model included an intercept term and one or more normalized features. Each model attempted to predict, using the value of the response to a single question (and an additive constant), whether a subject reported a C19+ or C19- status. Coefficients in a logistic regression model can be interpreted as changes in odds, or as odds ratios when two values are compared. Each ROC curve - constructed using predictions on holdout test sets and concatenated over these test sets -- summarizes the tradeoff between sensitivity (fraction of C19+ cases correctly identified) and specificity (fraction of C19- cases correctly identified) as the threshold value for the predictor is varied.

## Value of using a scale rather than a binary response to detect C19+

We quantified the information entropy for each survey question used the following standard equation: *I*  $\sum_{i=1}^{n}(-p_i * log_2(p_i))$  evaluated over the *n* response options. Re-binning to mimic new scales was achieved by dividing response values by a constant and rounding to the nearest integer. Relative mutual information was calculated by computing the mutual information between survey response and COVID status based on the following standard equation:  $MI = \sum_{i}^{n} \sum_{j}^{2} (p_{ij} * log_2(\frac{p_{ij}}{p_{i}p_{i}}))$  $\frac{n}{i}\sum_{j}^{n}(p_{ij} * log_2(\frac{p_{ij}}{p_{i}p_{j}}))$  where survey response options are indexed with *i* and the C19+/C19- status (two possible values) are indexed with *j*, and then dividing by the entropy available from that same C19 status distribution, calculated using the first equation. Results indicate that soliciting responses

on either a continuous 100-point scale or a downsampled 10-point numeric version of the scale is more informative about symptoms themselves and about COVID-19 status (given the symptoms) than soliciting binary responses (**Figure S5**).



**Figure S5. (A)** Relative information available from the distribution of responses to the two primary "Smell" survey questions. Binary refers to the yes/no question about symptomatic smell loss. A relative information of 1 would correspond to a question whose response is perfectly informative about COVID-19 status. By contrast, a similar question asked on a numeric scale (0-100, the original scale; or a hypothetical 10-point scale obtained by rounding responses) contains ins substantially more information due to the resolution of the scale. A 10-point scale may be familiar from clinical self lf-reports of pain. (**B)** The relative mutual information about COVID status contained in the survey response is also higher for the full numeric scale or the hypothetical 10-point scale than for the binary question.

## Prediction of recovery from COVID-19-associated smell loss

We applied the same predictive modeling framework used in Figure 4 to try to predict smell recovery in C19+ participants. In other words, we asked which survey responses predicted that a subject would fall into the Recovered Smell rather than the Persistent Smell Loss cluster, given both smell loss during the disease and C19+ status. The only predictive feature of any practical significance was "Days Since Onset" of respiratory symptoms (AUC=0.62), indicating that those who experienced their first respiratory symptoms less recently are more likely to have Recovered Smell (Figure S6A). Adding additional features to the model provided modest improvement (AUC=0.65 for the optimal model), but overall it was difficult to predict whether a C19+ participant would exhibit Recovered Smell or Persistent Smell Loss based on the data available (**Figure S6B**). **Ta Table S3**  includes the means and SD by recovery group for C19+ and C19-participants.



**Figure S6**. COVID-19 recovery. Similar to Figure 1, but self-reported smell (**A,B**), taste (**C,D**), chemesth thesis (E,F), and nasal blockage(G,H) during and after respiratory illness in C19+ (darker) versus C19- (lighter). (A,C,E,G) mean values during and after respiratory illness, respectively. (B,D,F,H) Change (after minus during) as a distribution over subjects.



### Table S3. Main chemosensory and relevant demographic features in the three clusters of recovering C19+ and C19- participants.

# APPENDIX 1

## GCCR core questionnaire

The core questionnaire of the Global Consortium for Chemosensory Research (GCCR) has been deployed in Compusense Cloud in 32 languages. The questionnaire was published previously<sup>8</sup> and also appears in the NIH Office of Behavioral and Social Sciences Research (OBSSR) research tools for COVID-19.<sup>39</sup> Responses to the GCCR core questionnaire in 23 languages were collected between April 7 and July 2, 2020 and included in the final dataset, on which we conducted the analyses reported in this paper.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.22.20157263.this version posted July 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted med



### **Welcome and Consent**



 $\bigcirc$  [1] Yes  $\bigcirc$  [0] No

Question Type: Choose only 1<br>Branching logic: if [0]=checked, then go to Section 4) End of test

### About your illness

2. In which year (YYYY) were you born? Question Type: Numeric<br>Branching logic: if year of birth greater than 2001 then go to Section 4) End of test

3. What is your current country of residence?

Question Type: Comment



Global Consortium for Chemosensory Research

GĆ



medRxiv preprint doi: https://doi.org/10.1101/2020.07.22.20157263.this version posted July 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted med



The next section of this survey is focused on your experience of smell, taste, and food flavor during your recent respiratory illness or diagnosis.

### These questions relate to your sense of smell (for example, sniffing flowers or soap, or smelling garbage) but not the flavor of food in your mouth.



16. Optional: Please describe any changes in smell



L,



### The following questions are related to your sense of taste. For example sweetness, sourness, saltiness, bitterness experienced in the mouth



medRxiv preprint doi: https://doi.org/10.1101/2020.07.22.20157263.this version posted July 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted med



31. Rate your ability to taste AFTER your recovery

No sense Excellent sense of taste of taste of taste<br><del>طلب المسافر المسافر المسافر المسافر المسافر المسافر المسافر المسافر</del> المسافر المسافر المسافر المسافر المسافر ال

(Place a mark on the scale above)

Question Type: Line Scale

Ī



medRxiv preprint doi: https://doi.org/10.1101/2020.07.22.20157263.this version posted July 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.<br>It is made available under a CC-BY-NC 4.0 International license .

Question Type: Comment

41. Optional: Is there anything we didn't ask about that you would like to share with us?

Question Type: Comment

### Re-contact

42. We may want to re-contact you for follow up research on this topic. Is it okay if our team or other researchers re-contact you to participate in future research? By saying yes, you agree that we can share your email address with other researchers for this purpose.

 $\bigcirc$  [1] Yes  $\bigcirc$  [0] No

Question Type: Choose only 1

Branching logic: if [0]=checked, then go to: Section 5)End of test

43. Please provide your full email address, so you can be contacted for future studies by our team or other researchers.

Question Type: Comment

### **End of Test**

You have now completed the survey and may close your browser

Thank you for your time!

#### **Notes**

"In which year (YYYY) were you born?

- value must be 1900 or greater

'What date did you first notice symptoms of your recent respiratory illness? Provide your best guess or leave blank if you do not remember. Click the box below to display a calendar"

-- format (mm/dd/yyyy)

'Have you been diagnosed with COVID-19"

-- if [8] Other was selected, a comment is required

"Were you diagnosed with any other respiratory illnesses (not COVID-19) in the last two weeks? (Select all that apply)" -- if [6] None was selected, no other options can be selected.

'Have you had any of the following symptoms with your recent respiratory illness or diagnosis? (Select all that apply)" -- if [17] No symptoms was selected, no other options can be selected.

'Rate your ability to smell BEFORE your recent respiratory illness or diagnosis"

-- Line Scale Range 0-100, intervals of 1. All following line scales formatted similarly

All line scales had anchors indented at ~10% and 90% of scale range.

'OPTIONAL: During the past 30 days, on how many days did you smoke combustible cigarettes or cigars?" and "OPTIONAL: During the past -- value must be between 0-30

Did you have any of the following in the 6 months prior to your recent respiratory illness or diagnosis? (Select all that apply) -- if [12] None was selected, no other options can be selected.