Pre-Registration of Main Questions for 'The effect of social race and socio-economic status on COVID-19 chemosensory loss'

BACKGROUND

SARS-CoV2 attaches to air pollutants ($PM^{2.5}$, PM^{10}, and NO_2) and differentially affects risk of infection [1–9] and mortality in individuals exposed to higher rates of pollution [4–10]. Loss of smell and/or taste in the absence of other symptoms is a key indicator of COVID-19 infection [11,12] but may be useless for communities already suffering from pre-existing smell and taste disorders due to the effects of environmental racism [13–15]—the pollution burden carried by racial minorities and socio-economically disadvantaged groups on local [16–21] and global scales [22] is disproportionate to the general population. US and UK COVID-19 mortality data race indicate that black populations are more than twice as likely to die than other groups [23] but there are no data for other aspects of infection or disease progression. Race is not genetically or biologically discrete but the human social environment is a key predictor of adult morbidity and mortality [24] and demographic differences in allostatic load (e.g., race, gender) predispose marginalized populations to greater sensitivity to chronic disease [25,26]. Pandemic management and planning must include plans for at-risk groups but despite these known differences, European countries ignore social race and clinical studies on COVID-19 in the UK and America detached race data from clinical status. The proposed study would be the first to attempt to understand the global experience of race and socio-economic status (SES) on pandemic dynamics.

Two GCCR papers analyzed data for a large body of case studies and clinical findings regarding taste and smell loss relative to COVID-19 but these papers did not include questions regarding race or SES. While there are limited data in the UK and US on race, these data are not systematically collected in hospitals and there are no representative samples that can inform hypothesis building. Further, these data do not explain the experience prior to death and there are no data on the impact of SES on COVID-19, beyond the general knowledge that lower SES equates to poorer health outcomes. One published study proposes that COVID-19 is a syndemic, interaction and exacerbation of pre-existing health inequities relative to a pandemic [27] and an online report indicates that marginalized groups are linked to lower indices of deprivation and health [28]. Further, given that this will be an unprecedented attempt to understand social race in a global context via self-reported data on individual local perceptions of race and self-reported relative status, an analysis plan will have to be open-ended. Due to the limited data available, the global scale of the study (with various understandings of race and SES), the study is not hypothesis driven other than the US sample and predicting COVID based on smell loss. The survey questions collect a variety of data that can be further explored to develop hypotheses to test in future efforts.

The primary research question is: Is smell loss predictive for COVID-19 when considering race and SES? We aim to identify if there are differences attributable to social race and SES and their interaction relative to COVID-19 and chemosensory loss by country or region (depending on sample sizes). We also ask questions about sexual orientation and gender due to recent reports that LGTBQIA+ populations may be experiencing greater rates of severe COVID-19

disease progression [29]. These questions are not required to answer due to the sensitivity of those topics in some cultures/countries.

DATA COLLECTION METHODS

The only method used for data collection is a survey. The survey considers cross-cultural variation in race and SES. We ask questions about income and resources and whether the participant is a marginalized member of their larger community. We provide definitions for all terms that are not common or may need clarification. We have two survey groups.

1) The first will be a FormR-based survey that will be advertised via a mass blind cc email to participants from a previous survey conducted by the Global Consortium for Chemosensory Research. The email will contain a link to our survey which will not collect their email address. The survey will also be advertised via social media to expand the sample size. We will contact various partners via email or social media to request they boost our signal. The social media text and email are included as part of the advertising/recruitment efforts in this package along with the electronic informed consent without signature and the survey questions themselves. The social media platforms we will use will include Twitter, Facebook, Whatsapp, and Instagram. Other social media platforms may be included for countries where these are not available. The only screening requirement for the survey is that participants must be 18 years of age or older.
2) The second survey group will be via Qualtrics, a professional survey company that we have paid ($3,000) for a guaranteed sample of 520 participants that each have an income below $41,000, with the sample split equally across our demographics of interest: white, black, and Hispanic. The same 18+ age requirement applies to this survey. Qualtrics pays participants a token (~$4.50) in exchange for their time). Because we have additional screening requirements for this survey, the FormR survey will be copied into Qualtrics survey software in two parts: The first part will contain questions directed to Americans (income in three brackets and race) and the second part will be the remainder of the survey questions.

For both forms, participants cannot retrieve answers to past responses. For FormR, participants get an access token for the study, which functions as a strong password. The access token is stored on participant's devices/browsers by default, so the only possible breach of protection is access to the device used to consent to the survey—but once the survey is submitted, the token is no longer active and the survey and answers cannot be retrieved. IP addresses are not automatically collected in either platform.

- The Qualtrics survey is operated by Qualtrics and will run until we get the 520 participants.
- The FormR survey will run until 15 January or until we have 520 participants. We will advertise for the FormR survey once a week. We will share the survey via GCCR Slack channels (key_announcemtns) and the GCCR newsletter.

PARTICIPANT CHARACTERISTICS

The only screening requirement for both surveys is that participants must be 18 years of age or older. The second survey group administered via Qualtrics, a professional survey company, has additional requirements: income below $41,000, race self-identified as white, black, or Hispanic.

1. Inclusion Criteria. All surveys: Adults over the age of 18 years only. Participants must click on a button to accept the terms and confirm they are 18+.
2. Qualtrics Specific inclusion criteria: must be white, black, or Hispanic. Must also have an income under 41,000 for a household of 3 or more.

INFORMED CONSENT

The conditions for consenting to participate in the survey are provided on the initial page of the survey. Once participants agree to continue, they can end their participation at any time and we will exclude partial submissions. Once the submit survey button is clicked, the data are recorded anonymously and we cannot retrieve and delete individuals responses nor can participants edit their responses. A copy of the language is found here: https://github.com/kchoover14/covidSESrace/blob/main/Informed%20Consent.pdf

ETHICS

A copy of the approval letter is in the Github repository for this study and that link is shared in the informed consent/information part of the survey. https://github.com/kchoover14/covidSESrace/blob/main/IRB%20Approval.pdf

ESTIMATE OF REQUIRED SAMPLE SIZE

We cannot estimate effect size for our analysis because there is no precedent for a global analysis of COVID symptoms relative to social race and SES, even if there are some studies in the US and the UK on those specific country demographics. The closest analysis available for estimation of sample size is a global study (also survey-based) on COVID-19 symptoms but that did not consider demographic factors such as race or SES [11]. Our target sample of 520 exceeds two separate calculations of power.

1) If we sample from a population of 100,000 (depending on the celebrities we engage or other social media groups we target) and target a 95% confidence level and 5% confidence interval and only 50% of population participate, the size of the required sample is 384 (https://www.surveysystem.com/sample-size-formula.htm).
2) If we sample from a population of 15,000, using a hypothesized frequency of smell loss among COVID positive people of 88% (based on findings from Gerkin et al., 2021), the required sample size is 280 for 99% confidence level, and 456 for 99.9% confidence level. (OpenEpi - Toolkit Shell for Developing New Applications)

GLOBAL DATA ANALYSIS METHODS (item numbers are from Codebook)

Data Cleaning

We will exclude participants with the following implausible combinations of answers:

- reported smell and taste loss (32) but no change on 10-point scales before and during illness (40-41)

- reported smell loss (32) but no “smell changes” as a symptom during illness (37)
- reported “smell changes” as a symptom during illness (37), but no change on 10-point scales before and during illness (40,41)
- reported smell loss as a factor in seeking diagnosis (38), but no change on 10-point scales before and during illness (40,41)
- reported smell loss as a factor in seeking diagnosis, but no “smell changes” as a symptom during illness (38,37)
- reported more than two generations in household (19), but then gives a number lower than 3 on the question how many (20).
- Checked the box of being over 18 years old on the consent form and answered being under 18 years old (22, formR only)

Global Data Transformations

- Items 9-17: Eight questions on household income and resources are will be used to create two variables
    - Needs Index (69): Household needs relative to income and/or resources
        - Income that meets needs
        - Income with resources that meets needs
        - Income that does not meet needs
        - Income with resources that does not meet needs.
    - SES Index (70): Relative SES ranking (distribution, dimension reduction)
        - Within group with income that meets needs
        - Within country with income and resources that meets needs
        - Within group with income that does not meet needs
        - Within country with income and resources that does not meet needs
- Items 19-21 and 23: Depending on bin size, information on household (two+ generations, how many generations, rural versus urban) will be collapsed into one new variable with four categories.
    - Household Index (71)
        - Urban household with 1-2 generations
        - Urban household with 3+ generations
        - Rural household with 1-2 generations
        - Rural household with 3+ generations
- Item 21: Environmental exposure to agents that impede olfaction will be used to create a new variable
    - Environmental Index (72)
        - Count of options selected (0-4): a higher number indicates a higher risk
- Items 29-30: Depending on bin size, these may be collapsed into new variables. We may have too small a sample of non-cis gendered individuals and we may have too small a sample of non-heterosexual individuals.
- Item 34-36 on smell/taste loss cause and diagnosis will be collapsed into one new variable with three categories.
    - ST Loss Cause (73)
        - Loss attributable to non-viral illness

        - Loss attributable to COVID-19 viral illness
        - Loss attributable to non-COVID-19 viral illness
- Item 37 on viral illness will be split into two variables:
    - Symptom Index (74)
        - Count of symptom boxes ticked (1-13): a higher number indicates greater severity of illness
    - ST Loss (75)
        - smell/taste loss
        - no smell/taste loss
- Items 59-62 on odor awareness will be collapsed into one variable
    - Odor Awareness Index (77)
        - We will perform a factor analysis and use a scree plot to determine the number of factors underlying this set of items.
            - On items 60-62, the answers “not applicable” and “never” will both be coded as 1.
            - The answers “Seldom”, “Sometimes”, “Often”, “Always” will be recoded to 2,3,4,5 respectively.
            - If a single underlying factor is found, we will sum the four recoded 4 odor awareness items 59-62 into a single score from 4-20.
            - If more than one factor is found, we will sum the items that load onto the same factor and proceed with odor awareness subscales in our analyses.
- Items 63-67 on health hazards that impede olfaction will be used to create a new variable
    - Health Index (78)
        - Count of options selected (0-5): a higher number indicates a higher risk

Global Analysis Parameters

- Alpha = 0.05
- Standard Bonferroni correction in the case of multiple dependent tests, within paper.

COVID RACE AND SES

Are there differences in taste and smell loss experience attributable to race and socio-economic status?

Hypothesis 1: Are there baseline differences in chemosensory ability in different SES and race groups as prior reports have found?

- Qualtrics Survey: All participants will report their income in two brackets at or below poverty level in the US and based on a household size of three (26). The three race categories included in data collection are white, black, and Hispanic (27). We predict that there will be a positive correlation between income bracket (26) and self-rated current ability to smell (42, 11-point scale). Because prior studies are not clear on the influence of SES on lower olfactory ability scores and lower SES status and race are tightly bound, we have no prediction for the correlation between race and self-rated olfactory ability.
- FormR Survey: We will use the data reduction procedure above to reduce income and resources into two new variables (70-71). We predict a positive correlation between needs/relative SES and self-reported olfactory ability (11-point scale)--the higher the SES status and needs being met, the higher the self-rating of olfactory ability. We predict that race (a dichotomous nominal variable) will be associated with lower SES status. We predict that environmental factors (72) are associated with lower income and minority race as well as self-rated ability to smell. We predict that health factors (73) increase in groups where needs are not being met (69), lower SES (70), and minority race (27) as well as self-rated ability to smell.

Analysis: If data are normally distributed, multivariate linear regression will be conducted with self-reported olfactory ability as the independent variable with household income, household resources, environmental exposure, and the four health-related questions as independent variables. If data are not normal, generalized linear regression will be used (or Kendall's Tau correlation analysis). Regression on item 7 (independent variable) and item 42 (dependent variable) with interactions from item 21 and other items on dependent variable 42.

Hypothesis 2: Is smell loss in marginalized populations (race, socio-economics) the primary predictive symptom of COVID-19 as has been observed in prior studies (but that failed to collect information on race and SES)? For this hypothesis, only those that responded to having experienced a prior smell loss (whether due to COVID or other respiratory illnesses) are included in the analysis.

- Qualtrics Survey: We expect a negative correlation between seeking and/or obtaining a COVID diagnosis due to smell and taste loss and SES/race.
- FormR Survey: We expect a negative correlation between seeking and/or obtaining a COVID diagnosis due to smell and taste loss and SES/race.

Analysis: We will conduct a multivariate linear regression using diagnosis as the dependent variable and SES and race as the independent variable. Confounding factors will be health related questions. The interaction between race and SES will be significant in determining if these are separate effects or if one has a greater impact on olfactory ability than the other. We

will also conduct a cluster analysis for SES and for race (separately) and the following variables: ability to smell before infection, ability to smell after infection, and ability to smell currently. We predict that the changes in smell and taste will be less variable in lower SES and minority race. We will also conduct a cluster analysis for SES and for race (separately) and the following variables: importance of odors in everyday life, noticing the odor of houses, odors evoking memory and/or emotions, and noticing odors. We predict that lower SES and race will be less likely to agree with those statements. Finally, we will compare the predictive power of smell and taste loss (on infection) in our study to that previously reported [11,12] in studies that did not collect race or SES data. We expect the predictive power to be lower when considering these additional demographic factors. For predictive modelling, we will use Qualtrics data to train the model because those data have equal sample sizes across race categories and contain only lower SES (two brackets). We will then split the FormR data into sample sizes of 60% for testing the model and 40% for validation of the model.

Exploratory questions

Does respiratory illness occur at similar rates across lower socio-economic groups of different races (Black, Hispanic, white) or is there a racial buffer for whites? We can also examine the subset of COVID-19, which has been shown to have a greater infection rate in minority races. Are there differences across races and SES in COVID-19 symptoms? These can be explored through cluster analysis.

ODOR AWARENESS SES AND RACE SCOPE OF WORK

Are there odor awareness differences between locations, household size, number of generations within a household, race, and socio-economics? While not apparently central to the question of smell and taste loss being equally predictive across race and SES categories, this analysis underlies the issue of whether minority races at the interaction of lower SES have pre-existing smell and/or taste loss (or inattention to chemosensory stimuli) that explain why smell and taste loss are not useful predictors (relative to the published literature cited in the background section).

Hypothesis: We predict that lower relative SES (Needs-Index,SES-Index, 69, 70, ordinal categorical variables) , minority race (item 7, dichotomous nominal variable), Household Index (71), and Exposure Index (item 72) all are associated with awareness of odors.

Analyses: After confirming normality, we will perform Spearman's Rho correlation analyses between items 69, 70, 71, 72, and the Odor Awareness Index (77). If normality is violated, we will perform a Kendall's Tau correlation analysis. We will perform an independent t-tests with item 7 as the independent variable and the Odor Awareness Index (77) as the dependent variable.