**Working title: Developing and validating an individual-level risk calculator for COVID-19 in the United States**

**Authors:**

Xindi C. Hu, ScD\* [CH]

Mathematica, [chu@mathematica-mpr.com](mailto:chu@mathematica-mpr.com)

Jennifer Starling, PhD [JS]

Mathematica, [jstarling@mathematica-mpr.com](mailto:jstarling@mathematica-mpr.com)

Erin Lipman, MS [EL]

Mathematica, [elipman@mathematica-mpr.com](mailto:elipman@mathematica-mpr.com)

Emma Pendl-Robinson [EPR]

Mathematica, [EPendl-Robinson@mathematica-mpr.com](mailto:EPendl-Robinson@mathematica-mpr.com)

\* = Corresponding author

Word Count - xxx

**Keywords:**

Covid-19

Risk prediction

**Word count:**

**Abstract**

Background

PLACEHOLDER

Methods

PLACEHOLDER

Findings

PLACEHOLDER

Interpretation

PLACEHOLDER

Funding

PLACEHOLDER

**Introduction [CH]**

* Effective pandemic management needs accurate and equitable risk assessment and management.
* Impacts of COVID-19 vary drastically in different population segments. Science is imperfect and evolves daily.
* COVID-19 risk needs to consider geography, demographics, health condition, and health behaviors (adherence to public health guideline). The last element is often overlooked by many risk assessment tools and infectious disease forecasting models.
* In this paper, we developed and validated an individual COVID-19 risk calculator integrating user input with the best available information from the official report and peer-reviewed literature. We provided an open-source, web-based interactive risk calculator for general public consumption, and developed an Application Programming Interface (API) for easy integration with other decision support tools. We compared individual-level estimates with the UK-based Nexoid COVID-19 Survival Calculator and observed similar performance from the two calculators.
* Our risk calculator empowers individuals to understand their risk profiles and encourage health behaviors that reduce the individual and community COVID-19 risk such as social distancing, wearing masks, and getting the vaccine. On a policy level, our calculator allows risk stratification and effective management of population health. The risk score can be used to effectively allocate vaccine and other pandemic management resources.

**Methods**

**Risk calculator**

* Conceptual framework: data, major modules (exposure/susceptibility/risk characterization) [CH]
* Exposure module
  + Community transmission [EL]
    - Under-reporting
  + Symptomatic cases [EPR]
  + Efficacy of handwashing and PPE [CH]
* Susceptibility module [EL]
  + Hospitalization, ICU, deaths
  + Risk factors
* Risk characterization [CH]
  + Normalization to flu
  + Log-transformation
  + Semi-quantitative score

**Validation**

* ~~Nexoid COVID-19 calculator [JS]~~
  + ~~Data overview~~
  + ~~Exposure~~
  + ~~Mortality~~
* OHDSI’s COVER app [EL]
  + Hospitalization/ICU/Death

[JS] Online risk calculators, including 19andMe, aim to educate the public and provide guidance which may reduce risk of transmission and mortality. However, many risk calculators are available, with varying methodology. We compare exposure and mortality risk estimates from 19andMe with Nexoid’s Covid-19 survival calculator (“Nexoid”). We chose the Nexoid calculator because Nexoid has made anonymous individual-level records publicly available under the “Attribution 4.0 International (CC BY 4.0)” license.

[JS] Each Nexoid record corresponds to one user’s inputs to the risk calculator, alongside their exposure and mortality risk estimates. The dataset includes geographic identifiers, behaviour, demographic and medical information, and excludes identifying information such as e-mail address and date of birth; latitude and longitude are randomly adjusted within 0.1 degrees, and age is reduced to 10-year bands. The validation study population includes all Nexoid users with valid location information located in the United States who entered information and received risk estimates from August 12, 2020 to November 10, 2020 (n=51,799).

[JS] We calculated mortality and exposure risk using 19andMe’s batch API functionality for each Nexoid user record, mapping each record’s IP address latitude and longitude to a zip and FIPS code. Ages were mapped from Nexoid ranges (e.g. 30-40) to the middle values (e.g. 35) for 19andMe. 19andMe defines direct exposure as the number of close contacts within a week, including your household members; we calculate this as the larger of Nexoid’s number of contacts (partners, children, colleagues, customers, patients, etc.) and number of people in the household/apartment. 19andme defines indirect exposure as the number of people your household comes in contact with during a week; as Nexoid lacks a direct analog for this input, we conservatively estimate this value as two contacts per household member. We set 19andMe’s hand hygiene to “True” if the Nexoid user selected “often” or “always” for hand-washing or hand sanitizer use, and personal protective equipment to “True” if the user selected “often” or “always” for mask wearing. Sex, pre-existing conditions, and symptoms were mapped directly from Nexoid to 19andMe.

[JS] We first compare mortality risk estimates generated for each user by 19andMe (death\_risk) and Nexoid (risk\_mortality), establishing similarity via Spearman ranked correlation and characterizing users with discrepancies exceeding 10% in either direction. We classify users into three discrepancy bands: discrepancy is within 10% (“within”), 19andMe is 10% lower (“lower”), 19andMe is 10% higher (“higher”). We focus on the users age 60-plus, as all mortality risk estimates were within 10% for users under age 60. Since 19andMe does not account for race while Nexoid does, we first investigate race as a source of discrepancy, using a chi-squared test to assess whether the distribution of race is the same in all three bands. After identifying that all users in the lower band are Black, we use matching to compare Nexoid mortality scores for similar but non-Black users to the corresponding 19andMe scores. For the higher band users over age 60, we use a Wilcoxon test to compare the mean number of pre-existing conditions to the within band. We test differences in proportions for incidence of diabetes, heart disease, hypertension, immune disease, kidney disease, lung disease, obesity, and smoking, using Bonferroni adjustment for multiple comparisons. We use Chi-squared test for difference in employment types in the higher versus within band, and test for difference in proportions of Covid symptoms.

[JS] We then compare exposure risk estimates for each user by 19andMe and Nexoid. We calculate 19andMe’s per-week exposure risk as

exposure\_risk\_19andMe = [sympt\_covid\_risk + exposure\_risk\*(1 - sympt\_covid\_risk)]

to capture the probability of having symptomatic Covid-19 (sympt\_covid\_risk) plus probability of catching Covid-19 through community transmission (exposure\_risk) if not already infected. Nexoid reports cumulative exposure risk, so we divide by the number of weeks of Nexoid data since March 24, 2020 (nweeks=12.86) to obtain Nexoid’s per-week exposure risk,

exposure\_risk\_nexoid = (risk\_infection / 100) / nweeks.

We again establish similarity using Spearman ranked correlation. Since the range of exposure risks is smaller than mortality risk, we characterize users with discrepancies exceeding 1% in either direction. We classify users into three discrepancy bands: discrepancy is within 1% (“within”), 19andMe is 1% lower (“lower”), 19andMe is 1% higher (“higher”). For users in the lower band, we test for difference of proportions in presence of several conditions (diabetes, as well as heart, kidney, liver, and lung disease), living in a nursing home or assisted care facility, employment in the healthcare sector, public transit use, and workout outside the home. These are factors which Nexoid accounts for, while 19andMe does not. For users in the high band, we use Wilcoxon tests to compare the mean number of primary and secondary contacts in the higher versus within bands, adjusted for multiple comparisons. We also test for differences of proportions in hand-washers and mask-wearers.

**Results**

**Risk calculator**

* Compare exposure risk with prevalence rate [EPR]
* Under-reporting factor has come down over time [EL]
  + Risk scores have not, an objective measure against “pandemic fatigue”
* Web app and API availability [CH]

**Validation**

* ~~Exposure risks in 19andMe and Nexoid [JS]~~
* ~~Mortality risk in 19andMe and Nexoid [JS]~~
* Hospitalization/ICU/Death in 19andMe and COVER [EL]

[JS] Mortality risk estimates were largely consistent between 19andMe and Nexoid (Spearman ranked correlation 0.91). Of the 51,799 users, 0.985 (n=51,024) were within 10%, 0.0003 (n=18) had 19andMe estimates at least 10% lower than Nexoid, and 0.146 (n=757) had 19andMe estimates at least 10% higher than Nexoid (Figure 1A). All of the users in the higher and lower bands were over age 60. For the lower band, all 18 users were Black (Figure 1B). After adjusting the Nexoid mortality risk estimates via matching, the matched estimates were in the within or higher band, in line with other estimates for users over age 60; Nexoid’s adjustment for race explains the discrepancies where 19andMe is at least 10% lower mortality risk. We found evidence that racial distributions differ across bands (p<0.001). For the higher band users over age 60, we found evidence of different mean numbers of pre-existing conditions (p<0.001) (Figure 1B). We also found evidence of difference in proportions for incidence of diabetes, heart disease, hypertension, immune disease, kidney disease, lung disease, obesity, and smoking in the within versus higher band (all multiple comparison-adjusted p-values <0.001). Diabetes, heart disease, immune disease, and lung disease had the most marked differences; these correspond to conditions which have higher mortality adjustment factors in Nexoid [ADD REF]. We did not find evidence of difference in types of employment between the two bands (p=0.080).

[JS] Exposure risk estimates were also correlated between 19andMe and Nexoid (Spearman rank correlation 0.48). Of the 51,799 users, 0.784 (n=40,600) were within 1%, 0.136 (n=7,043) had 19andMe estimates at least 1% lower than Nexoid, and 0.080 (n=4,156) had 19andMe estimates at least 1% higher than Nexoid (Figure 2A). For users in the higher band, we found evidence of different mean numbers of primary and secondary contacts (p<0.001), with the number of primary and secondary contacts larger in the higher band (Figure 2B). We also found evidence that hand-washing and mask-wearing are less prevalent in the higher band (p<0.001 for both hand-washing and mask use). For users in the lower band, we found the following factors contributing to higher Nexoid risk estimates: presence of pre-existing conditions (diabetes, or kidney, liver, or lung disease, p<0.001), employment in the healthcare sector (p<0.001), use of public transit (p<0.001), and working outside the home (p<0.001). These are factors which 19andMe does not account for, which raise exposure risk.

**Discussion**

**Risk calculator**

* Can be used for vaccine allocation
* Living breathing sites that will continue to be updated

**Validation**

* Validation is needed for the public to gain confidence in risk calculators
* This validation study is made possible by digital technology (API) and data sharing by Nexoid

[JS] Validation is key in building public confidence in risk calculators. We find that 19andMe and Nexoid produce largely aligned exposure and mortality risk estimates, which should build confidence and trust in both calculators. Areas of difference are explained by different choices each calculators has made. For mortality risk, Nexoid makes slightly stronger adjustments to risk for ages over 60. 19andMe’s decided not to include race, as it is likely a proxy for exogeneous variables such as access to health care and occupation, as opposed to a mechanistically causal variable. This decision, in combination with the difference in age adjustments, results in lower 19andMe mortality estimates for Black users. Meanwhile, 19andMe is increasing mortality risk estimates for pre-existing conditions, and presence of multiple conditions, to a greater degree than Nexoid.

[JS] For exposure risk, 19andMe and Nexoid are again largely consistent. Areas of discrepancy present opportunities for each application to consider enhancing its functionality. 19andMe’s consideration of both primary and secondary contacts leads to increased exposure risk estimates for more social users, while Nexoid is not decreasing risk as strongly for users who comply with CDC mask and hand-washing guidelines. 19andMe also might consider inclusion of some behavioural and occupational information that Nexoid incorporates, such as working outside the home, public transit use, and healthcare sector employment.

**Limitations**

* Ecological fallacy, unreported confounders for risk of severe health outcomes
* Certain exposure routes are hard to quantify, such as fleeing encounters, or exposure risk as a function of time, risk associated with different activities (going to gym/work/use public transit etc.)

**Conclusion**

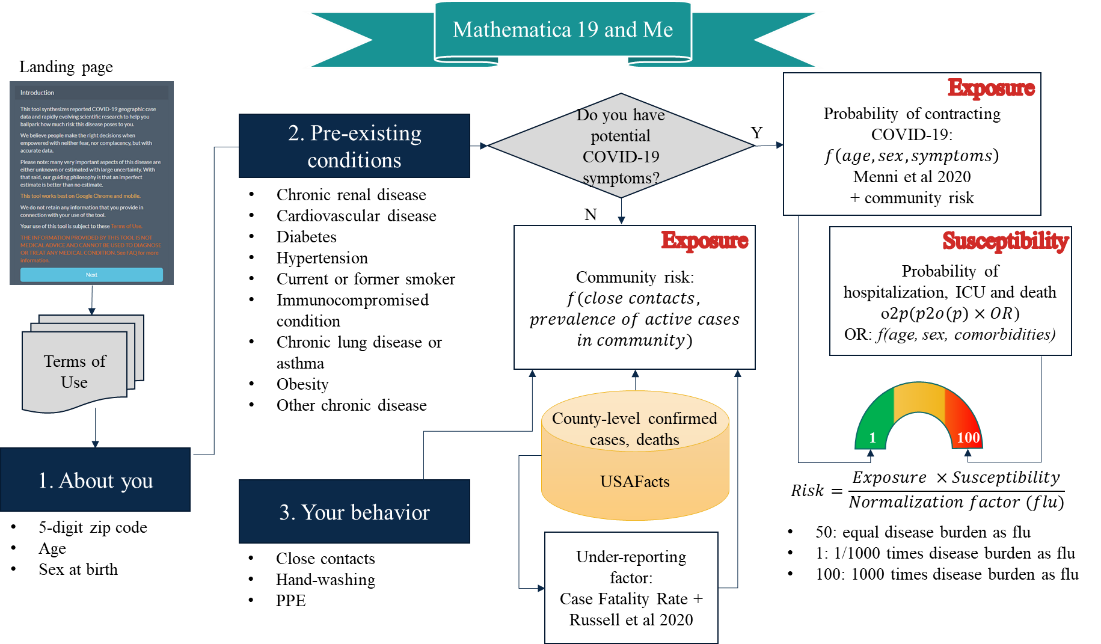
PLACEHOLDER

**Tables and Figures**

**Table 1**

PLACEHOLDER

**Figure 1**



**Data Sharing Statement**

Do we need one? Maybe! Other statements can go here, as necessary.

Yes, and we probably also need Code Sharing Statement somewhere

**References**