**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]
  + 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.

[JS] Nexoid makes data publicly available under the “Attribution 4.0 International (CC BY 4.0)” license. We use records from their master\_dataset.csv file, where each record represents one individual’s inputs to the risk calculator alongside their exposure and mortality risk estimates. This 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.

[JS] The validation study population includes all Nexoid calculator users located in the United States who entered information and received risk estimates from August 12, 2020 to November 10, 2020 (n=51,799). 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] Mortality validation

[JS] Exposure validation

**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]

**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

**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**

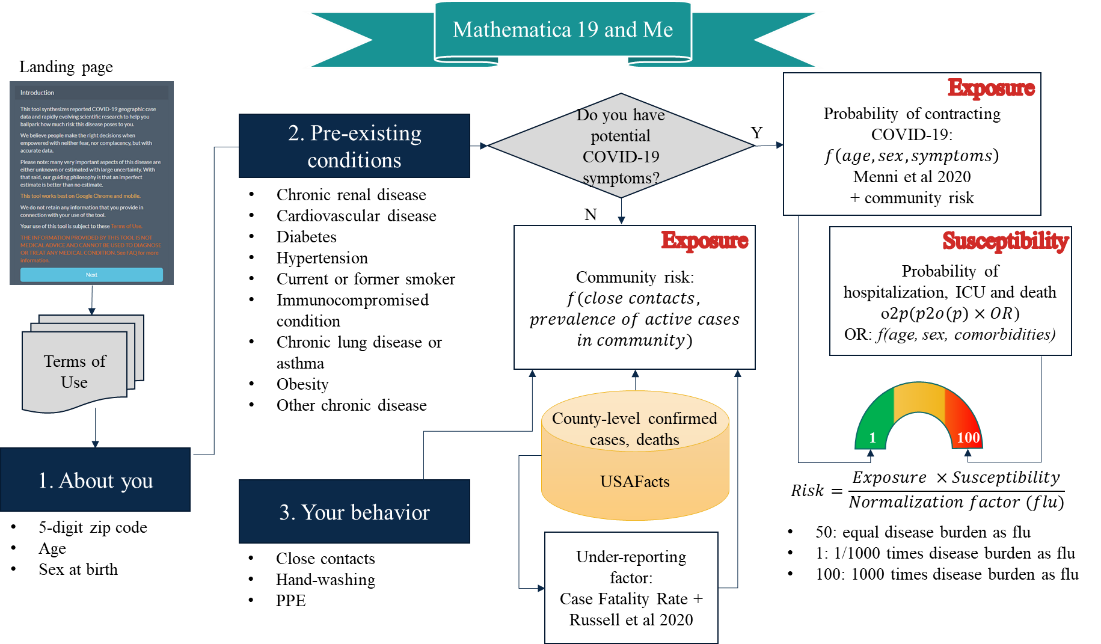
PLACEHOLDER

**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**