**Analyzing COVID-19 Case Report Forms Data from the All of Us Program**

Appendix

**Materials and Methods**

The All of US (AoU) program has three levels of access, the *public* tier, which requires no login and includes summary statistics, the *registered* tier, which requires login credentials and has some redaction and the *controlled* tier which requires login credentials and further regulatory approval and has minimal redaction. For our research we use the registered tier of access. The data is accessible via the cloud-based researcher workbench and allows or querying the data base through SQL and analysis using R language.

We defined a data element as numerical if it had a value in the value\_as\_number column for more than 50% of the data rows with responses.

**Results and Discussion**

The COPE survey has been answered multiple times by some participants, as seen by a ratio of responses to participants of greater than one, for questions that can only have one response each time the survey is taken. On average each participant has completed the survey 1.71 times, as seen by the ratio of responses to participants for the most questions(40 of 104, 38.5%). When forming our cohort from the question ‘Do you think you have had COVID?’ we saw the question was asked at different times to the same participants, as 742 (20.1%) participants who answered ‘Yes’ at some point, had answered ‘No’ when asked at a different time and 613 (16.6%) had answered ‘Maybe’.

For useful analytic resources pertaining to the r4aou package and a complete analysis of all data elements included in the COPE survey and overall in AoU, visit our general project repository at <https://github.com/lhncbc/CRI/tree/master/AoU/>