

## Data dictionary for Bloodbank.csv

- **donor\_id:** Unique id for each donor, ranging from 1 to the number of donors.
- **blood\_center:** Blood center where sample was collected. The blood centers are: HEMOAM (Manaus), FPS (São Paulo), HEMOCE (Fortaleza), HEMOMINAS (Belo Horizonte), HEMOPE (Recife), HEMEPAR (Curitiba), HEMOBA (Salvador), HEMORIO (Rio de Janeiro).
- **sex:** Sex of the donor.
- **idade:** Age group of the donor (-1 if unknown). To preserve anonymity, age was binned into the following groups: 16-24, 25-34, 35-44, 45-54, 55-70, and donor's age was substituted by the smallest age in the corresponding age group (16, 25, 35, 45 or 55).
- **week:** Week of sample collection. Week 1 corresponds to January 1<sup>st</sup> 2020 to January 7<sup>th</sup> 2020, and week  $k$  corresponds to the time interval between  $7(k - 1)$  days after January 1<sup>st</sup> 2020 to  $7(k - 1)$  days after January 7<sup>th</sup> 2020.
- **month:** Number of months passed since January 1st 2020. For instance, samples collected in March 2020 and March 2021 have respectively month = 3 and month = 15.
- **result:** Signal-to-cutoff obtained with the anti-N assay. A test is considered positive if result > threshold. In the paper, we use threshold = 0.49.

## Data dictionary for repeat\_blood\_donors.csv

- **donorid:** Unique id for each donor, ranging from 1 to the number of donors.
- **sex:** Sex of the donor.
- **age:** Age group of the donor (-1 if unknown). To preserve anonymity, age was binned into the following groups: 16-24, 25-34, 35-44, 45-54, 55-70, and donor's age was substituted by the smallest age in the corresponding age group (16, 25, 35, 45 or 55).
- **week:** Week of sample collection. Week 1 corresponds to January 1<sup>st</sup> 2020 to January 7<sup>th</sup> 2020, and week  $k$  corresponds to the time interval between  $7(k - 1)$  days after January 1<sup>st</sup> 2020 to  $7(k - 1)$  days after January 7<sup>th</sup> 2020.
- **assay:** The assay used in the test. If the field is empty or is equal to "COV-2 IgG", the sample was tested with the Abbott anti-N assay. If the value is "COV-2IgGII", the sample was tested with the Abbott anti-S assay.
- **covid\_result:** The signal-to-cutoff obtained from the test, concatenated with the string "Index" for the anti-N assay and "AU/mL" for the anti-S assay.

### Data dictionary for convalescent\_plasma\_longitudinal\_roche.csv

- **donor\_id**: Unique id for each donor, ranging from 1 to the number of donors.
- **sex**: Sex of the donor.
- **age**: Age group of the donor (-1 if unknown). To preserve anonymity, age was binned into the following groups: 16-24, 25-34, 35-44, 45-54, 55-70, and donor's age was substituted by the smallest age in the corresponding age group (16, 25, 35, 45 or 55).
- **hospitalized**: Boolean variable indicating whether the donor was hospitalized (1 for hospitalized, 0 for non-hospitalized).
- **week**: Week of sample collection. Week 1 corresponds to January 1<sup>st</sup> 2020 to January 7<sup>th</sup> 2020, and week  $k$  corresponds to the time interval between  $7(k - 1)$  days after January 1<sup>st</sup> 2020 to  $7k$  days after January 1<sup>st</sup> 2020.
- **abbott\_sc**: The signal-to-cutoff obtained with the Abbott anti-N assay.
- **seropositive\_at\_baseline**: A redundant variable indicating if the first test result after infection (baseline) was positive.
- **followup\_visit**: A redundant variable indicating if the row corresponds to the first (baseline), second (first follow-up) or third (second follow-up) sample collected for that plasma donor.
- **days\_post\_symptoms**: The difference in days between date of symptoms onset and date of sample collected (NA if asymptomatic).
- **roche\_sc**: The signal-to-cutoff obtained with the Roche assay (not used in the paper).

### Data dictionary for prepandemic\_cohort.csv

- **donorid**: Unique id for each donor, ranging from 1 to the number of donors.
- **sex**: Sex of the donor.
- **idade**: Age group of the donor (-1 if unknown). To preserve anonymity, age was binned into the following groups: 16-24, 25-34, 35-44, 45-54, 55-70, and donor's age was substituted by the smallest age in the corresponding age group (16, 25, 35, 45 or 55).
- **covid\_valor**: The signal-to-cutoff obtained with the Abbott Anti-N assay.