

This repository contains scripts (and few documents) used in the study entitled ‘Continuous population-level monitoring of SARS-CoV-2 seroprevalence in a large European metropolitan region’ published in iScience.

Continuous population-level monitoring of SARS-CoV-2 seroprevalence in a large European metropolitan region

Marc Emmenegger^{*,&}, Elena De Cecco, David Lamparter[#], Raphaël P. B. Jacquat[#], Julien Riou, Dominik Menges, Tala Ballouz, Daniel Ebner, Mathias M. Schneider, Itzel Condado Morales, Berre Doğançay, Jingjing Guo, Anne Wiedmer, Julie Domange, Marigona Imeri, Rita Moos, Chryssa Zografou, Leyla Batkitar, Lidia Madrigal, Dezirae Schneider, Chiara Trevisan, Andres Gonzalez-Guerra, Alessandra Carrella, Irina L. Dubach, Catherine K. Xu, Georg Meisl, Vasilis Kosmoliaptsis, Tomas Malinauskas, Nicola Burgess-Brown, Ray Owens, Stephanie Hatch, Juthathip Mongkolsapaya, Gavin R. Screaton, Katharina Schubert, John D. Huck, Feimei Liu, Florence Pojer, Kelvin Lau, David Hacker, Elsbeth Probst-Müller, Carlo Cervia, Jakob Nilsson, Onur Boyman, Lanja Saleh, Katharina Spanaus, Arnold von Eckardstein, Dominik J. Schaer, Nenad Ban, Ching-Ju Tsai, Jacopo Marino, Gebhard F. X. Schertler, Nadine Ebert, Volker Thiel, Jochen Gottschalk, Beat M. Frey, Regina R. Reimann, Simone Hornemann, Aaron M. Ring, Tuomas P. J. Knowles, Milo A. Puhon, Christian L. Althaus, Ioannis Xenarios, David I. Stuart, and Adriano Aguzzi^{*}

[#] Equal contribution

[&] Lead contact

^{*} To whom correspondence should be addressed

The scripts are partitioned into different folders:

1. TRABI + SEROPREVALENCE

1.1. MAIN (lead: David Lamparter)

- *Coded in R*
- *Establishment of TRABI technology (see Fig. 1C-D, Fig. S2, Table S1)*
- *QDA- and LDA-based approaches and computation of posterior probability (see Fig. 2A-B, Fig. S5, Fig. 6A)*
- *Assay reproducibility (see Fig. 2D-E, Fig. S6)*
- *Bootstrapping*
- *Computation of seroprevalence per month (see Table 1)*
- *Association of TRABI data with clinical and demographic information (see Fig. S4)*

1.2. REPRODUCTION IN PYTHON (lead: Raphaël Jacquat)

- *Coded in Python*
- *Establishment of TRABI technology (see Fig. 1C-D, Fig. S2, Table S1)*
- *QDA- and LDA-based approaches and computation of posterior probability (see Fig. 2A-B, Fig. S5, Fig. 6A)*
- *Assay reproducibility (see Fig. 2D-E, Fig. S6)*
- *Bootstrapping*
- *Computation of seroprevalence per month (see Table 1)*
- *Association of TRABI data with clinical and demographic information (see Fig. S4)*

2. REFINED SEROPREVALENCE + DEMOGRAPHY + CUMULATIVE INCIDENCE

2.1. MAIN SCRIPTS (lead: Julien Riou)

- *Coded in R*
- *Refined seroprevalence assessment based on main script shown in 1.1 (see **Fig. 2C**)*
- *Population-based age- and sex-adjustment on cohorts and exclusion of COVID-19 patients (see **Fig. 3A-C, Fig. S7A-B**)*
- *Characterisation of cohorts used in seroprevalence study (see **Fig. S3, Fig. S4C**)*
- *Analysis of antibody waning (see **Fig. 3D**)*
- *Calculation of cumulative incidence using modified SEIR model (see **Fig. 3E**)*

2.2. REPORT (lead: Julien Riou)

- *Coded in R*
- *A main report script refers to data prepared in subscripts, mostly for figure generation*

3. DISEASE MINING

3.1. BAYESIAN LOGISTIC REGRESSION (lead: Julien Riou)

- *Coded in R*
- *Deployment of multiple logistic regression model in Bayesian frame to correlate ICD-10 codes, age, and sex with posterior probability of being SARS-CoV-2 seropositive using STAN (see **Fig. 5A**)*

3.2. DIMENSIONALITY REDUCTION (lead: Marc Emmenegger)

- *Coded in R*
- *Deployment of PCA on ICD-10 codes for seropositive and seronegative individuals (see **Fig. S8**)*
- *Deployment of UMAP on ICD-10 codes for seropositive and seronegative individuals (see **Fig. S8**)*

4. SAMPLE PROVENANCE MAP

4.1. MUNICIPALITY DATA (lead: Raphaël Jacquat)

- *Used to generate the maps*

4.2. SCRIPT (lead: Raphaël Jacquat)

- *Coded in Python*
- *Generation of geographic maps to detail the provenance of hospital patients (see **Fig. 4**)*

5. ONLINE HEALTH SURVEY

5.1. QUESTIONNAIRE CODEBOOK (lead: Dominik Menges, Tala Ballouz, Marc Emmenegger)

- *The codebook containing all questions asked in the follow-up online health survey*

5.2. SCRIPT (lead: Dominik Menges, Tala Ballouz)

- *Coded in R*
- *Characterisation of the cohort (see **Table 2**)*
- *Analysis of the questionnaire (see **Fig. 5D-J, Tables S2-S5**)*

The methods related to the scripts are detailed in the study. The provisioning of code may enable the development of analytical tools based on those we have published. If bits of code shall be used for other studies, please

- get in touch with the lead contact of the study published in iScience.
- cite both the repository on Zenodo as well as the study.