THE LANCET
Infectious Diseases

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Perez-Saez J, Lauer S A, Kaiser L, et al. Serology-informed estimates of SARS-CoV-2 infection fatality risk in Geneva, Switzerland. *Lancet Infect Dis* 2020; published online July 14. https://doi.org/10.1016/S1473-3099(20)30584-3.

Supplementary Material
Serology-informed estimates of SARS-COV-2 infection fatality risk in Geneva, Switzerland

Javier Perez-Saez[1,2], Stephen A. Lauer[1], Laurent Kaiser[3,4,5], Simon Regard[6,7], Elisabeth Delaporte[7], Idris Guessous[5,8], Silvia Stringhini[5,8,*], and Andrew S. Azman[1,2,*], for the Serocov-POP Study Group[9]

[1] *Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA*
[2] *Institute of Global Health, Faculty of Medicine, University of Geneva, Switzerland*
[3] *Geneva Center for Emerging Viral Diseases and Laboratory of Virology, Geneva University Hospitals, Switzerland*
[4] *Division of Laboratory Medicine, Geneva University Hospitals, Switzerland*
[5] *Faculty of Medicine, University of Geneva, Switzerland*
[6] *Geneva University Hospitals, Switzerland*
[7] *Cantonal Health Service, General Directorate for Health, Geneva, Switzerland*
[8] *Unit of Population Epidemiology, Division of Primary Care medicine, Geneva University Hospitals, Switzerland*
[9] *List of authors and affiliations given on page 9.*
[*] *co-senior authorship*

1 Deriving the infection fatality risk

We derive the infection fatality risk (IFR) from available epidemiological data. Let $D(t)$ be the observed cumulative number of deaths up to time t, $I(t)$ the cumulative number of infections, $C(t)$ that of reported cases and $I^{sero}(t)$ the number of seroconverted people in the population at time t. We note that $I^{sero}(t)$ is estimated from serosurvey data and incorporates uncertainty in test specificity and sensitivity [12]. We have that the cumulative number of deaths up to time t is:

$$D(t) = IFR \int_0^\infty I(t-\tau) f_D(\tau) d\tau,$$

where $f_D(t)$ is the probability density function (PDF) of the time from infection to death. Similarly we have:

$$I^{sero}(t) = \int_0^\infty I(t-\tau) f_{sero}(\tau) d\tau,$$

where $I^{sero}(t) = \theta_t \cdot P$ is the seroconverted population given by the seroprevalence at time t, θ_t, and the population P, and $f_{sero}(t)$ is the PDF of the time from infection to seroconversion.

Taking the ratio between these two equations we have that:

$$IFR = \frac{D(t)}{I^{sero}(t)} \frac{\int_0^\infty I(t-\tau) f_{sero}(\tau) d\tau}{\int_0^\infty I(t-\tau) f_D(\tau) d\tau}. \tag{1}$$

$I(t)$ is unobservable, however we can reconstruct it using $C(t)$:

$$C(t) = \alpha \int_0^\infty I(t-\tau) f_C(\tau) d\tau,$$

where α is the probability of infection reporting (proportion of infections that lead to symptomatic and detected COVID-19 cases), and f_C is the PDF of time from infection to reporting, which accounts both for the incubation period and the delay between symptom onset to reporting. An estimate of $I(t)$ up to a constant of proportionality can be obtained by inverting the convolution:

$$\begin{aligned}
\mathcal{F}\{C\} &= \mathcal{F}\{\alpha I\}\mathcal{F}\{f_I\} \\
\mathcal{F}\{\alpha I\} &= \mathcal{F}\{C\}/\mathcal{F}\{f_I\} \\
I &= \frac{1}{\alpha}\mathcal{F}^{-1}\left\{\mathcal{F}\{C\}/\mathcal{F}\{f_I\}\right\} \\
I &= \frac{1}{\alpha} I^*,
\end{aligned}$$

were $\mathcal{F}$ and $\mathcal{F}^{-1}$ are respectively the Fourier transform and its inverse. In the presence of noise the deconvolution can be numerically unstable due to noise amplification caused by large values of $1/\mathcal{F}\{f_I\}$ at high frequencies. We therefore regularize $\mathcal{F}\{f_I\}$ by applying a threshold under which the values are set to the threshold while preserving their phase, also called water level regularization [1, Chapter 8.3]. Here we use the threshold value of 0.05.

The infection fatality ratio can therefore be expressed as:

$$IFR = \frac{D(t)}{I^{sero}(t)} \frac{\int_0^\infty I^*(t-\tau) f_{sero}(\tau) d\tau}{\int_0^\infty I^*(t-\tau) f_D(\tau) d\tau}, \tag{2}$$

were α cancel out. In practice we do not have continuous values of epidemiological variables, but counts by discrete time periods. In this study data were available at a daily time step (Fig. 1). We therefore replace integrals in eq. 2 with sums over discrete time delays as:

$$IFR = \frac{D(i)}{I^{sero}(i)} \frac{\sum_{j=0}^{T} I^*(i-j) p_{sero}(j)}{\sum_{j=0}^{T} I^*(i-j) p_D(j)}, \tag{3}$$

where i is the day on which deaths and seroprevalence are measured, T is the total number of days since the start of the epidemic, and $p_{sero}(j)$ and $p_D(j)$ are respectively the probabilities of seroconversion and death during day j after infection computed using the distribution functions of seroconversion and death F_{sero} and F_D: $p_{sero}(j) = F_{sero}(j+1) - F_{sero}(j)$ and $p_D(j) = F_D(j+1) - F_D(j)$. The values of I^* are also computed using discrete time steps.

To infer the IFR we considered a binomial likelihood for the number of deaths $D(i)$ occurring among the fraction of the infected population at risk of dying on day i, $I^{sero}(i)\phi(i)$, where $\phi(i) = \frac{\sum_{j=0}^{T} I^*(i-j)p_D(j)}{\sum_{j=0}^{T} I^*(i-j)p_{sero}(j)}$ as:

$$\mathcal{L}(IFR|\theta_{1...T}) = \prod_{i=1}^{T} \binom{I^{sero}(i)\phi(i)}{D(i)} IFR^{D(i)}(1-IFR)^{I^{sero}(i)\phi(i)-D(i)}, \tag{4}$$

were $\mathcal{L}$ is the likelihood of IFR given the data and the seroprevalences at each sampling time i, $\theta_{1...T}$, recalling that $I^{sero}(t) = \theta_t \cdot P$ with P the population. To incorporate uncertainty in the seroprevalence estimates one can integrate over the seroprevalence posterior at time t, $f_t(\theta)$:

$$\mathcal{L}(IFR) = \prod_{t=1}^{T} \int \mathcal{L}(IFR|\theta_t) f_t(\theta) d\theta$$

We approximate the integral by Monte Carlo integration using M posterior draws θ_t^m from our seroprevalence analysis [12]:

$$\mathcal{L}(IFR) \approx \prod_{t=1}^{T} \frac{1}{M} \sum_{m=1}^{J} \mathcal{L}(IFR|\theta_t^m).$$

Finally, the log-likelihood, ll is:

$$ll \approx \sum_{t=1}^{T} -\log(M) + \log\left\{ \sum_{m=1}^{J} \mathcal{L}(IFR|\theta_t^m) \right\}.$$

2 Inference

2.1 Bayesian framework

We aim at inferring the IFR by age class. We use the age-classes in our previous analysis: 5-9, 10-19, 20-50, 50-65, and 65+ [12]. Inference is drawn using a Bayesian framework, where we assume that the IFR for age class a, IFR_a, has a Beta prior distribution with age-specific parameters α_a and β_a:

$$IFR_a \sim Beta(\alpha_a, \beta_a).$$

We reparametrize the prior following [6, Chapter 5]:

$$\gamma_a = \frac{\alpha_a}{\alpha_a + \beta_a}$$
$$\lambda_a = \alpha_a + \beta_a,$$

with hyper-priors:

$$\gamma_a \sim Beta(1, 6.5)$$
$$\lambda_a \sim Pareto(0.1, 1.5).$$

We use for the mean of the IFR prior, γ, a beta distribution which has a median of ≈ 0.1 to account for the fact that current estimates situate around 1%, with more vulnerable age classes around 10%.

Posterior draws were obtained using a Hamiltonian Monte Carlo sampler as implemented in the Stan programming language [3], through the package `rstan` [11] in R. Chain convergence was assessed using the Gelman-Rubin $\hat{R}$ statistic [5]. The code used in the analysis is available at `https://github.com/HopkinsIDD/sarscov2-ifr-gva`.

2.2 Data

Epidemiological data for each age class was provided by the canton of Geneva's public health authority, the Direction Générale de la Santé (DGS) (Fig. 1). Population data [9], and the number of people living in assisted care centers [8] were obtained from the statistics office of the canton of Geneva. The parameter values for the delay distributions are given in Table 1.

2.3 Population-level post-stratification

The population-level IFR was estimated by post-stratification using the estimates by age class a, IFR_a:

$$IFR_{pop} = \frac{1}{I^{sero}} \sum_a I_a^{sero} \cdot IFR_a, \tag{5}$$

where I^{sero} is the estimated number of seropositives in the canton of Geneva on the last available serosurvey week (May 6th), and I_a^{sero} is the estimated number of seropositives in age class a from our previous analysis [12]. We note that this estimate accounts for differences in attack rates across age classes.

Table 1: Paremetrization of delay distributions. The distribution of delays were parameterized using log-normal distributions either as reported in the cited references, or computed to match reported mean and standard deviation (denoted with †). Delay combinations (like infection to symptom onset and symptom onset to reporting), were computed by convolution (denoted by $*$). All distributions are shown in Fig. 2.

	Parameters			
Distribution	$\log \mu$ (mean)	$\log \sigma$ (sd)	Description	Source
f_{inc}	1.57 (5.94)	0.65 (4.31)	Incubation period	[2]
f_{report}	1.50 (5.60)	0.45 (4.20)	Symptom onset to reporting	[10]†
$f_{sympsero}$	2.34 (11.2)	0.38 (4.40)	Symptom onset to seroconversion	[12]
$f_{reportdeath}$	2.1 (11.9)	0.87 (12.7)	Reporting to death	DGS‡
f_C	$f_{inc} * f_{report}$		Infection to reporting	-
f_{sero}	$f_{inc} * f_{sympsero}$		Infection to seroconversion	-
f_D	$f_{inc} * f_{report} * f_{reportdeath}$		Infection to death	-

‡Data from Geneva's public health authority (Direction Générale de la Santé)

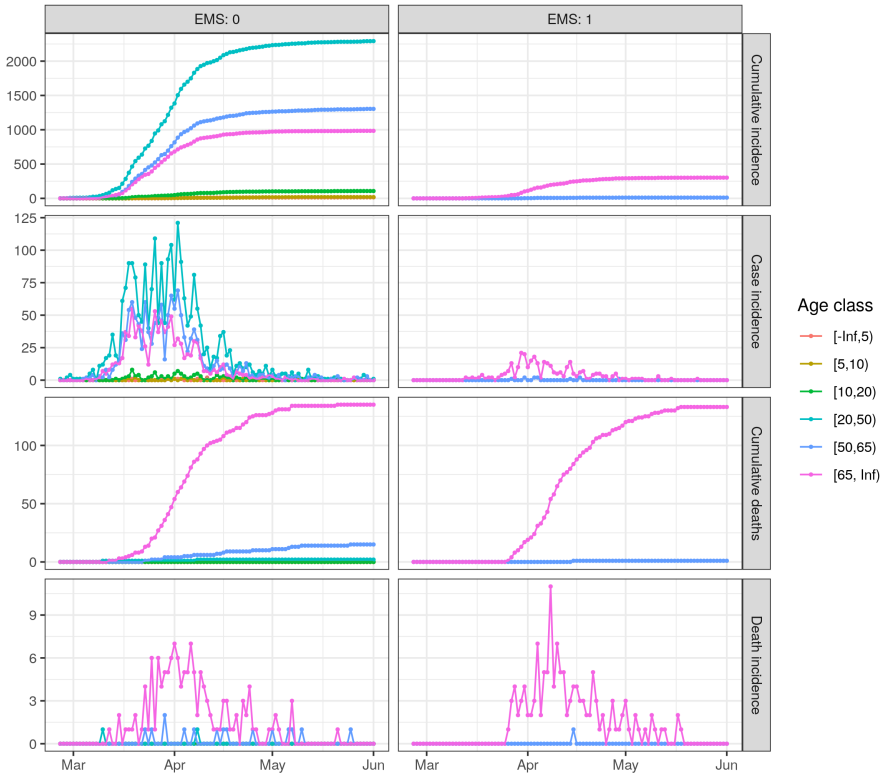

Figure 1: Age-stratified COVID-19 daily epidemiological data in the canton of Geneva, Switzerland in the general population (left column) and within assisted care facilities (EMS) (right column).

3 Results

Primary results. Results discussed in the main text are presented in Table 1 of the main text, with IFR posterior draws given in Fig. 3.

Accounting for assisted care facilities/nursing homes. The true seroprevalence in assisted care facilities remains unknown. If we consider only deaths and infections that occurred in the general population older than 65, the age-specific IFR for this group decreases to 2.7% (95% CrI 1.6-4.6). Excluding this population (both infections and deaths) leads to an overall IFR estimate in Geneva of 0.32% (95% CrI 0.17-0.56), half of what we estimate in our primary analyses (Table 2).

Sensitivity analyses. We fit the model using a uniform instead of a Beta-distributed IFR prior and results were very similar, except for a wider 95% CrI for the 5-9 age class. Finally we fit the model using only the last serosurvey week which yielded IFR posteriors with the same means but wider 95% CrIs. Uncertainty in the delay distributions was accounted for in the main analysis by sampling over estimates of the log-normal distribution parameters (Fig .2).

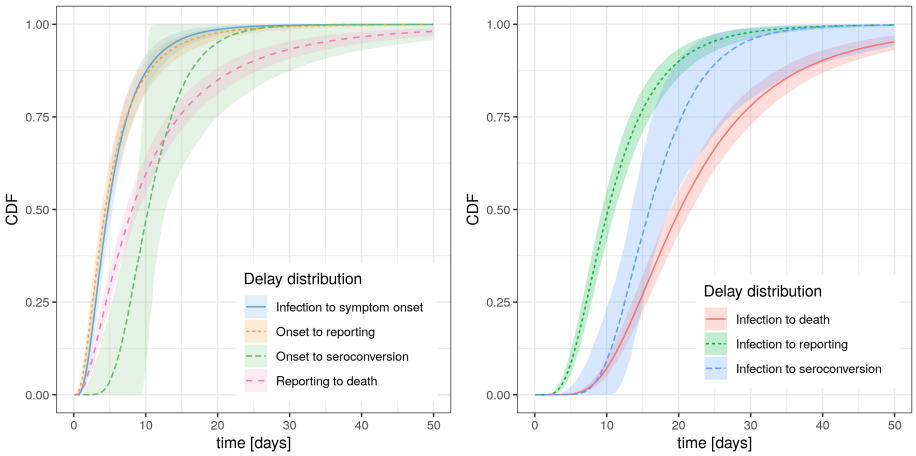

Figure 2: Cumulative probability distributions of delays to key events. *Left:* Delay distributions for which estimates were available from data. *Right:* Un-observable delays estimated using convolutions (Table 1). Distributions are shown in terms of the MLE parameter estimates (lines) and the 95% confidence intervals (shadings).

Reported COVID-19 deaths and excess mortality Analysis of excess mortality trends have been suggested to provide a more accurate estimate of COVID-19 mortality due to confirmed death under-reporting [7]. The cantonal department of public health (DGS) reported the number of excess mortality during the weeks of the epidemic, estimating to 261 excess deaths during weeks 12-18 among the 65 and older age group, and no statistically significant excess mortality in the younger population [4]. In the same period there were 255 reported COVID-19 deaths in the 65+ age class. This suggests that COVID-19-related deaths were not significantly under-reported.

Table 2: Age-stratified estimates of the IFR without accounting for the deaths in assisted care facilities. Results for age classes younger than 65 are the same as in Table 1 of the main text, and estimates for the 65+ age class and the overall population were recomputed by not considering the population living in assisted care facilities.

Age class	Population	Seroconverted population as of May 6th (95% CrI)	Deaths as of June 1st	IFR [%] (95% CrI)
65+	79'509	5'400 (3'000- 8'400)	135	2.7 (1.6-4.6)
all	502'700	54'800 (41'300-70'700)	152	0.32 (0.17-0.56)

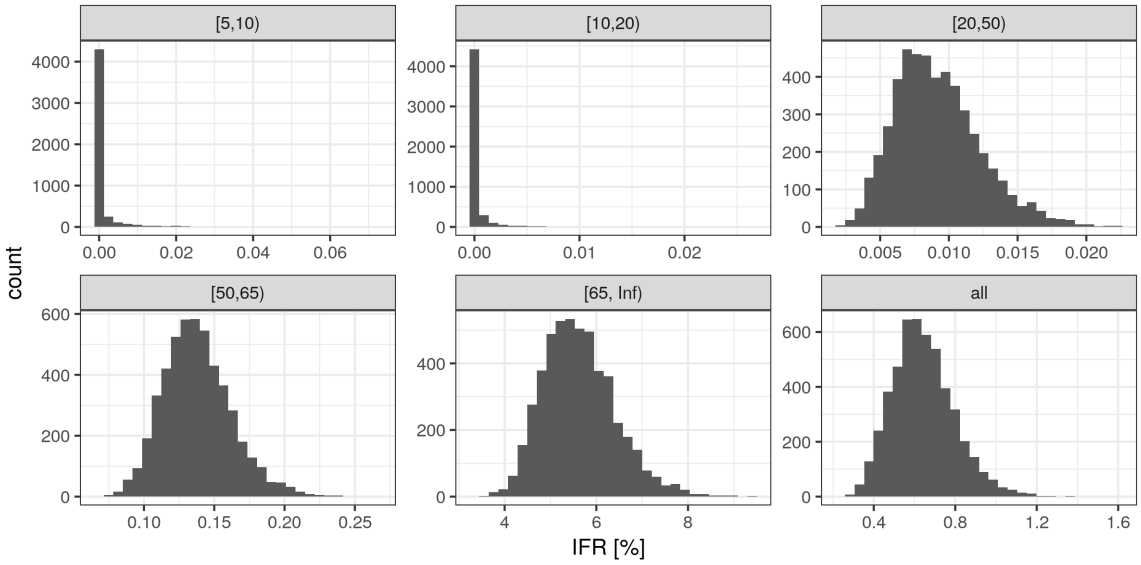

Figure 3: IFR posterior distributions by age class.

References

[1] Richard C Aster, Brian Borchers, and Clifford H Thurber. *Parameter estimation and inverse problems*. Elsevier, 2018.

[2] Qifang Bi et al. "Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study". In: *The Lancet Infectious Diseases* (2020).

[3] Bob Carpenter et al. "Stan: A probabilistic programming language". In: *Journal of statistical software* 76.1 (2017).

[4] Direction générale de la santé. *COVID-19 - Point épidémiologique hebdomadaire Canton de Genève - Situation semaine 21*. Direction générale de la santé, Genève, CH. June 2020. URL: `https://www.ge.ch/document/19696/telecharger` (visited on 2020-06-10).

[5] Andrew Gelman, Donald B Rubin, et al. "Inference from iterative simulation using multiple sequences". In: *Statistical science* 7.4 (1992), pp. 457–472.

[6] Andrew Gelman et al. *Bayesian data analysis*. CRC press, 2013.

[7] David A Leon et al. "COVID-19: a need for real-time monitoring of weekly excess deaths". In: *The Lancet* 395.10234 (2020), e81.

[8] Office cantonal de la statistique. *Informations Statistiques - Les établissements de santé à Genève: Résultats 2017*. Office cantonal de la statistique, Genève, CH. Jan. 2019. URL: `https://www.ge.ch/statistique/tel/publications/2019/informations_statistiques/autres_themes/is_etablissements_sante_01_2019.pdf` (visited on 2020-06-03).

[9] Office cantonal de la statistique. *Population résidante selon l'origine, le sexe, le groupe d'âges et l'état matrimonial, en 2019*. Office cantonal de la statistique, Genève. Dec. 2019. URL: `http://www.ge.ch/statistique/tel/domaines/01/01_01/T_01_01_8_01.xls` (visited on 2020-06-03).

[10] Jérémie Sciré et al. "Reproductive number of the COVID-19 epidemic in Switzerland with a focus on the Cantons of Basel-Stadt and Basel-Landschaft". In: *Swiss Medical Weekly* 150.19-20 (2020), w20271.

[11] Stan Development Team. *RStan: the R interface to Stan*. R package version 2.19.3. 2020. URL: `http://mc-stan.org/`.

[12] Silvia Stringhini et al. "Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland (SEROCoV-POP): a population-based study". In: *The Lancet* (2020).

SEROCoV-POP STUDY Group: Davidovic Alioucha[1], Isabelle Arm-Vernez[2], Andrew S Azman[3], Sultan Bahta[1], Jonathan Barbolini[1], Hélène Baysson[1], Rebecca Butzberger[1], Sophie Cattani[1], François Chappuis[1,4], Alison Chiovini[1], Prune Collombet[1,4], Delphine Courvoisier[4], David De Ridder[1], Eugénie De Weck[5], Paola D'ippolito[4], Antoine Daeniker[1], Odile Desvachez[6], Yaron Dibner[4], Céline Dubas[1], Joséphine Duc[1], Isabella Eckerle[1,2,7], Céline Eelbode[1], Nacira El Merjani[4], Benjamin Emery[8], Benoit Favre[1], Antoine Flahault[1,4,9], Natalie Francioli[4], Laurent Gétaz[1,4], Alice Gilson[1], Acem Gonul[1], Julie Guérin[1], Idris Guessous[1,4], Lina Hassar[1], Aurélia Hepner[1], Francesca Hovagemyan[1], Samia Hurst[10], Laurent Kaiser[1,2,7,11], Olivia Keiser[9], Melis Kir[1], Gaëlle Lamour[12], Pierre Lescuyer[7], Fanny Lombard[1], Amélie Mach[1], Yasmina Malim[4], Eva Marchetti[1], Kailing Marcus[4], Soraya Maret[1], Chantal Martinez[4], Kourosh Massiha[1], Virginie Mathey-Doret[1], Loan Mattera[6], Philippe Matute[4], Jean-Michel Maugey[12], Benjamin Meyer[1,13], Tom Membrez[1], Natacha Michel[1], Aleksandra Mitrovic[4], Emmanuelle Marie Mohbat[1], Mayssam Nehme[4], Natacha Noël[4], Hugo-Ken Oulevey[1], Febronio Pardo[12], Francesco Pennacchio[4], Dusan Petrovic[4,8], Attilio Picazio[4], Giovanni Piumatti[4,10], Didier Pittet[14], Jane Portier[4], Géraldine Poulain[7], Klara Posfay-Barbe[1,15], Jean-François Pradeau[12], Caroline Pugin[4], Rakotomiaramanana Barinjaka Rakotomiaramanana[4], Aude Richard[4,9], Christiane Rocchia Fine[4], Irine Sakvarelidze[1], Lilas Salzmann-Bellard[4], Magdalena Schellongova[1], Stephanie Schrempft[4], Mélanie Seixas Miranda[16], Milena Stimec[1], Silvia Stringhini[1,4,8], Michel Tacchino[12], Sophie Theurillat[17], Mélissa Tomasini[4], Kor-Gael Toruslu[1], Nawel Tounsi[1], Didier Trono[18], Natacha Vincent[1], Guillemette Violot[19], Nicolas Vuilleumier[1,7], Zoé Waldmann[1], Sylvie Welker[1], Manon Will[1], Ania Wisniak[9], Sabine Yerly[2,7], Maria-Eugenia Zaballa[4], Alenka Zeballos Valle[1]

1. Faculty of Medicine, University of Geneva, Geneva, Switzerland
2. Geneva Center for Emerging Viral Diseases and Laboratory of Virology, Geneva University Hospitals, Geneva, Switzerland
3. Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, USA
4. Division of Primary Care, Geneva University Hospitals, Geneva, Switzerland
5. Education Structure, University of Geneva, Geneva, Switzerland
6. Campus Biotech, Geneva, Switzerland
7. Division of Laboratory Medicine, Geneva University Hospitals, Geneva, Switzerland
8. University Centre for General Medicine and Public Health, University of Lausanne, Lausanne, Switzerland
9. Institute of Global Health, Faculty of Medicine, University of Geneva, Geneva, Switzerland
10. Faculty of BioMedicine, Università della Svizzera italiana, Lugano, Switzerland
11. Division of Infectious Diseases, Geneva University Hospitals, Geneva, Switzerland
12. Information Systems Division, Geneva University Hospitals, Geneva, Switzerland
13. Centre for Vaccinology, Department of Pathology and Immunology, University of Geneva, Geneva, Switzerland
14. Infection Prevention and Control program and World Health Organization (WHO) Collaborating Centre on Patient Safety, Geneva University Hospitals, Geneva, Switzerland
15. Division of General Pediatrics, Geneva University Hospitals, Geneva, Switzerland
16. Division of Diagnostics, Geneva University Hospitals, Geneva, Switzerland
17. Human Ressources Departement, Geneva University Hospitals, Geneva, Switzerland
18. School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
19. Division of Communication, Geneva University Hospitals, Geneva, Switzerland