Estimating the size of undetected cases of the SARS-CoV-2 outbreak in Europe: An upper bound estimator

Irene Rocchetti

Statistical Office - Consiglio Superiore della Magistratura

Dankmar Böhning

Southamption Statistical Sciences Research Institute

University of Southampton

Heinz Holling

Department of Methods and Statistics, Faculty of Psychology and Sports

University of Münster

Antonello Maruotti

Dipartimento di Giurisprudenza, Economia, Politica e Lingue Moderne

Libera Università Ss Maria Assunta

Department of Mathematics

University of Bergen

June 17, 2020

Abstract

Background: While the number of detected SARS-CoV-2 infections are widely available, an understanding of the extent of undetected cases is urgently needed for an effective tackling of the pandemic. The aim of this work is to estimate the true number of SARS-CoV-2 (detected and undetected) infections in several European Countries. The question being asked is: How many cases have actually occurred?

Methods: We propose an upper bound estimator under cumulative data distributions, in an open population, based on a day-wise estimator that allows for heterogeneity. The estimator is data-driven and can be easily computed from the distributions of daily cases and deaths. Uncertainty surrounding the estimates is obtained using bootstrap methods.

Results: We focus on the ratio of the total estimated cases to the observed cases at April 17th. Differences arise at the Country level, and we get estimates ranging from the 3.93 times of Norway to the 7.94 times of France. Accurate estimates are obtained, as bootstrap-based intervals are rather narrow.

Conclusions: Many parametric or semi-parametric models have been developed to estimate the population size from aggregated counts leading to an approximation of the missed population and/or to the estimate of the threshold under which the number of missed people cannot fall (i.e. a lower bound). Here, we provide a methodological contribution introducing an upper bound estimator and provide reliable estimates on the *dark number*, i.e. how many undetected cases are going around for several European Countries, where the epidemic spreads differently.

Keywords: Capture-recapture methods; COVID-19; Geometric distribution; Chao's lower bound.

1 Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has become a pandemic within few weeks. The number of detected cases increased day-by-day, at an exponential rate at the beginning, and now follows a logistic distribution [1, 2]. Cases of SARS-CoV-2 might have

been vastly under-reported in official statistics. It is widely acknowledged that the majority of the cases are asymptomatic and, thus, not observed or recorded [3–5]. In other words, the available data just tell us a part of the story: individuals may be already infected but are not aware of it, maybe because of the absence of symptoms, or cases may be under symptomatic suspicion but the disease has not been diagnosed yet (due to the delay in getting swab results). The total number of cases is thus unknown, and general comments on the spread of the epidemic are thus partial as based on a (relatively small) fraction of the total cases. Some studies have used simulation-based approaches to infer reasonable estimates of total number of cases, but often these estimates are surrounded by poor uncertainty measures, leading to too wide confidence intervals [6]. Here, we are proposing a simple and effective method to obtain reasonable point and interval estimates of the total number of SARS-CoV-2 infections in several European countries. In detail, we introduce a novel estimator based on a capture recapture (CR) approach. The capture-recapture method should be considered the gold standard for counting when it is impossible to identify each case and large undercounts will occur [7]. CR methods were originally developed in the ecological setting with the aim of estimating the unknown size of a (possibly elusive) population and then they started to be applied also to epidemiological and health sectors (see [8, 9]). Many CR estimators have been proposed in the literature (see e.g. [10–12]), and some of them can be used to identify lower bounds [13] of the population size. In the analysis SARS-CoV-2 infections, official data are available at the aggregated level, whereas individual data are not available to the general or the academic public. Hence, it is not possible to get the exact distribution of the number of infected individuals observed exactly one day, exactly two days and so on until m days. The population is open, subjected to deaths, and this may further complicate the analysis [14]. A lower bound of the total number of infected cases is computed by [3] modifying the Chao estimator [15] to address issues related to the data at hand. This is a relevant result as it provides reasonable information to the policy makers about the undetected cases and the magnitude this phenomenon may have at least, so that national health systems may be aware of the minimum number of cases that may demand

health care services. At this stage of the spread of the epidemic, governments are willing to relax restrictive measures and several researches address issues related to the epidemic [16–18]. To calibrate the new interventions, an estimate of the lower bound of the number of infections may not be enough, as SARS-CoV-2 has already shown to spread around the population very quickly [19–21]. This contribution aims at providing an approximated upper bound for the total number of SARS-CoV-2 cases, to better appreciate the dimension of the epidemic, under the worse scenario. Such an estimate is obtained from a non parametric CR model, providing an upper bound estimate of the total number of infections regardless of the true data generating process.

This contribution is organized as follows. In Section 2, we introduce the basic notation and how we are going to work with the data at hand. A brief summary of the modified Chao lower bound is also discussed. These notions are then used to compute the upper bound, details of which are provided in Section 3, along with the computation of the uncertainty surrounding the estimates. In Section 4 we show the empirical application of the proposal on data from several European countries. A discussion showing other interesting insights concludes.

2 Methods

2.1 Preliminaries

Let us denote with $N(t)$ the cumulative count of infections at day t where $t = t_0, \cdots, t_m$. Hence $\Delta N(t) = N(t) - N(t-1)$ are the number of new infections at day t where $t = t_0 + 1, \cdots, t_m$. Also, let $D(t)$ denote the cumulative count of deaths at day t where $t = t_0, \cdots, t_m$. t_0 defines the beginning of the observational period and t_m defines the end. We assume the trivial assumption $t_m > t_0$, so that the observational window is not empty. Again, we denote with $\Delta D(t) = D(t) - D(t-1)$ the count of new deaths at day t where $t = t_0 + 1, \cdots, t_m$.

The question arises how this can be linked to a capture-recapture approach. Let X_i denote the number of identifications for each infected individual i typically provided by the days the

individual will surely remain infected. Let denote τ_x the probability of identifying an individual x times where $x = 0, \dots$ A lower bound estimator of the unobserved frequency f_0, say $\hat{f}_0$, can be estimated by using the observed frequency of those identified exactly once, f_1, and of those identified twice, f_2, [13, 15]:

$$\hat{f}_0 = f_1^2/f_2. \tag{1}$$

It is thus crucial to relate f_1 and f_2 with the data at hand. In detail, at each day t, $f_1(t)$ represents the infected people identified just once, i.e. the new infections, whose number is given by $\Delta N(t)$. Similarly, $f_2(t)$ represents the infected people detected at time $(t-1)$ and still infected at time t. This can be computed as $\Delta N(t-1) - \Delta D(t)$. Hence the estimate for the number of hidden infections at day t is

$$\hat{f}_0(t) = \frac{[\Delta N(t)]^2}{\Delta N(t-1) - \Delta D(t)}. \tag{2}$$

By applying the estimator (1) day-wise we get the modified Chao lower bound estimator (see [3]):

$$\hat{f}_0 = \sum_{t=t_0+1}^{t_m} \frac{[\Delta N(t)]^2}{\Delta N(t-1) - \Delta D(t)}. \tag{3}$$

In practice, however, the bias-corrected form of (3) suggested by [22] is used:

$$\hat{f}_0 = \sum_{t=t_0+1}^{t_m} \frac{\Delta N(t)[\Delta N(t)-1]}{1 + \Delta N(t-1) - \Delta D(t)}. \tag{4}$$

We define the understanding that $\Delta N(t-1) - \Delta D(t)$ is set to 0 if it becomes negative, in other words we use $max\{0, \Delta N(t-1) - \Delta D(t)\}$. The final estimate of lower bound (LB) of the total number of infection is then given as what has been observed at the end of the observational window t_m and the estimate of the hidden numbers:

$$\mathrm{N}_{LB} = N(t_m) + \hat{f}_0 \tag{5}$$

2.2 The Upper Bound estimator

The lower bound is helpful as an indication of the minimum number of people having had SARS-CoV-2 and answers to a fundamental open question: "How many undetected cases are at least

going around?". Nevertheless, this information may be treated as a starting point whenever interventions and tools to dampen the spread of the epidemic are rolled out. The proposed upper bound estimator extends the research on the undetected cases and helps policy makers to evaluate the SARS-CoV-2 epidemic situation locally and at the current phase of its development. An estimate of the worse possible scenario is provided.

Following a similar strategy as in Section 2.1, this is achieved by firstly estimating daily-specific upper-bounds and then summing up all the estimates to get the final point-estimate of the maximum number of undetected cases. This daily–wise based upper bound approach provides an approximation of the data generation process.

Let us introduce the cumulative distribution function

$$\pi_{ij} = Pr(X_i \leq j) = Pr(X_i = 0) + Pr(0 < X_i \leq j) = \pi_{i0} + (1 - \pi_{i0})p_{ij}, \tag{6}$$

where homogeneity in the probability of being infected at a certain date t is assumed, i.e. $\pi_{ij} = \pi_j$, with $p_{ij} = p_j$ being the cumulative zero-truncated probability distribution. Equation (6) represents the probability that an individual is infected for at most j days, and it is function of π_0 and p_j; but π_0 is not observed. The quantities $p_j (j = 1, 2, 3)$ in equation (6) at each time t may be approximated as

$$\begin{aligned} p_1(t) &= f_1(t)/n^*_{obs}(t), \\ p_2(t) &= (f_1(t) + f_2(t))/n^*_{obs}(t), \\ p_3(t) &= (f_1(t) + f_2(t) + f_3(t))/n^*_{obs}(t) \end{aligned}$$

where $f_1(t)$ and $f_2(t)$ have been introduced in the previous section and

$$f_3(t) = \Delta N(t-2) - \Delta D(t-1) - \Delta D(t).$$

and $n^*_{obs}(t)$ is the number of current infected individuals observed at each time. We think that it is reasonable, for each day t, to consider the number of individuals affected by SARS-CoV-2 for the day t, for day t and the day before, and, for day t and the two days before, as $m = 3$ is the minimum number of consecutive days of new infections necessary for the upper bound estimator

to be computed. Furthermore, considering more than 3 days for an individual to be observed as affected by SARS-CoV-2 would lead to the risk of not observing the number of people affected by SARS-CoV-2 for exactly four, five and so on times because of the higher risk of overlapping cases.

Since π_0 is unknown, to compute the probabilities in (6), we substitute it with

$$\hat{\pi}_0(t) = \frac{\hat{f}_0(t)}{f_1(t) + f_2(t) + \hat{f}_0(t)}.$$

where $\hat{f}_0(t)$ is the *lower bound* probability of undetected cases derived from the Chao estimator in its bias corrected form, computed at each time t (see Equation 2). This also explains why a lot of detail was devoted to the lower bound estimator in the previous section as it is very much needed here. In other words, based on the Chao lower bound estimator of the undetected cases, we derive the *complete* count distribution and calculate the upper bound for the population size on such a complete distribution. Now, it follows that Equation (6) takes the form

$$\hat{\pi}_j(t) = \hat{\pi}_0(t) + (1 - \hat{\pi}_0(t))p_j(t)$$

when theoretical probabilities are replaced by their now available estimates. In order to provide an upper bound estimator we use the main results of [23]:

$$\pi_j \leq p_j \left[1 - (1 - p_j) \left(\frac{p_{j+1} - p_j}{p_{j+1} - p_j \frac{\hat{\pi}_j}{\hat{\pi}_{j+1}}} \right) \right]^{-1}.$$

For $j = m - 2$, and by some algebra, we get the equivalent condition

$$\pi_0 \leq \frac{p_{m-1} - p_{m-2}}{\left(1 - \frac{\hat{\pi}_{m-2}}{\hat{\pi}_{m-1}}\right) + p_{m-1} - p_{m-2}} = \hat{\pi}_0^{UB};$$

that makes clear why at least $m = 3$ days should be considered. The right-hand side $\hat{\pi}_0^{UB}$ of the above inequality provides an upper bound estimate of the population size based on the Horvitz–Thompson estimator:

$$\hat{f}_0^{UB}(t) = n^*_{obs}(t) \frac{\hat{\pi}_0^{UB}}{1 - \hat{\pi}_0^{UB}}.$$

However we deal with a day-wise upper bound approximation of $\pi_0(t)$ which is given by

$$\hat{\pi}_0^{UB}(t) = \frac{p_2(t) - p_1(t)}{\left(1 - \frac{\hat{\pi}_1(t)}{\hat{\pi}_2(t)}\right) + p_2(t) - p_1(t)}.$$

To get an estimate for the missed SARS-CoV-2 infections $\hat{f}_0(t)$ at each time t we compute the Horvitz–Thompson (HT) estimator at each time t and ultimately we sum it up over all times, reaching thus the final upper bound for the missed SARS-CoV-2 cases n_0 as follows

$$\hat{f}_0^{UB} = \sum_{t=t_0+2}^{t_m} \left(\frac{n^*_{obs}(t)}{1-\pi^*_0(t)} - n^*_{obs}(t) \right). \tag{7}$$

Hence, the approximated upper bound of the total number of infected people, $\hat{N}_{UB}$, in the time window from t_0 to t_m is then given by

$$\hat{N}_{UB} = \hat{f}_0^{UB} + N_{t_m}.$$

2.3 Uncertainty estimation

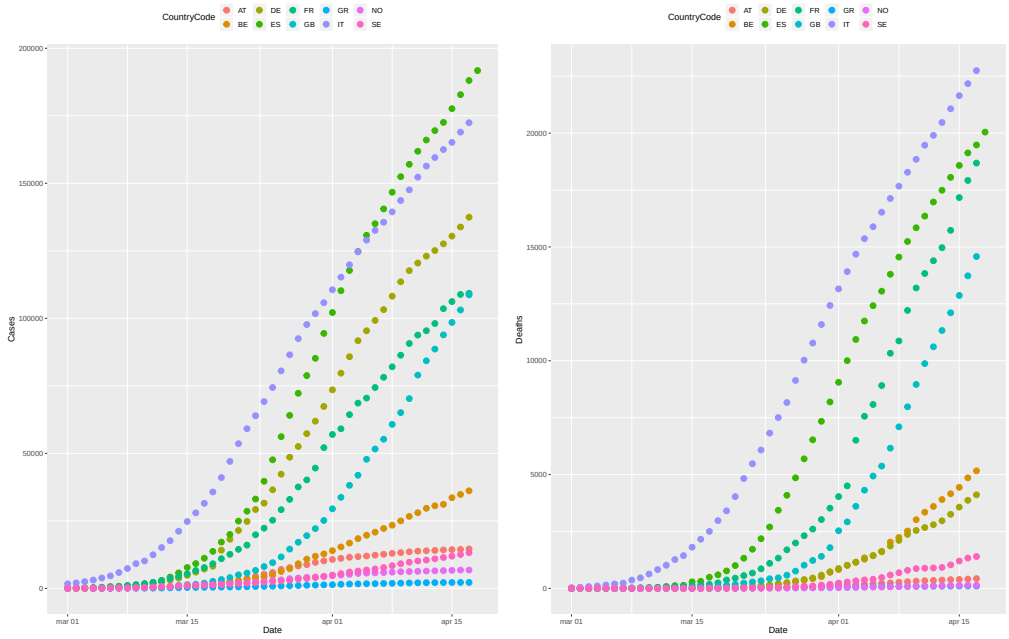
A fundamental issue in general CR analyses is the quantification of uncertainty surrounding the estimates of the unknown population size. An estimation of the population size can be correctly computed, but if the associated estimation of variance is poor, then coverage by the 95% confidence interval may falsely indicate poor estimation by the point estimator, i.e. the point estimator may result in a poor coverage rate. Focusing on the proposed upper-bound estimator, we attempt here to investigate bootstrap methods as a robust and general approach to estimate variances and confidence intervals. Various bootstrap methods have been considered to estimate uncertainty in CR analyses with respect to other estimators [24–27]. In the following, we consider two different bootstrap approaches to approximate the uncertainty surrounding the point estimate: the imputed and the reduced bootstrap approaches.

Under the imputed bootstrap approach, we draw 1000 bootstrapped samples of size N_{UB} generated according to a multinomial model whose probabilities are given by

$$\left\{ \hat{\pi}_0^{UB}(t) = \frac{\hat{f}_0^{UB}(t)}{N_{UB}(t)}, \frac{f_1(t)}{N_{UB}(t)}, \frac{f_2(t)}{N_{UB}(t)}, \frac{f_3(t)}{N_{UB}(t)} \right\},$$

where $N_{UB}(t) = \hat{f}_0^{UB}(t) + f_1(t) + f_2(t) + f_3(t)$.

Differently, under the reduced bootstrap approach, each of the bootstrapped samples contains $n^*_{obs}(t) = f_1(t) + f_2(t) + f_3(t)$ observations generated according to a multinomial model whose



(a) Cases (b) Deaths

Figure 1: Cases and deaths for the analyzed countries

probabilities are given by $\left\{\frac{f_1(t)}{n^*_{obs}(t)}, \frac{f_2(t)}{n^*_{obs}(t)}, \frac{f_3(t)}{n^*_{obs}(t)}\right\}$. For each of the two approaches, the upper bound N_{UB} is computed for each bootstrapped sample, by summing up over the time period. Of course, for the imputed bootstrap the fraction of undetected cases is dropped and considered unknown when computing the population size. We report the 2.5% and 97.5% values of N_{UB} distribution. This allows us to overcome issues often encountered in the construction of the symmetric confidence intervals [28]: the sampling distribution could be skewed, the coverage probabilities may be unsatisfactory, etc.

3 Data Analysis

The example provided here relies on European data. The time series of cumulative cases and deaths up to 17/04/2020 are considered and are taken from `https://github.com/open-covid-19/data`. A graphical representation of the data at hand is shown in Figure 1.

Data from the day which we record the first death are analyzed only. We obtain the estimates

Table 1: Estimated hidden and total cases of Sars-Cov-2 for several European countries, at 17/04/2020

Country	observed cases	upper bound for total number of cases	(2.5% – 97.5%) bootstrap values (IB)	(2.5% – 97.5%) bootstrap values (RB)	total/observed	total/observed (2.5% – 97.5%) IB	total/observed (2.5% – 97.5%) RB
Italy	172434	780704	(777690–784121)	(778080 – 783895)	4.53	(4.51–4.58)	(4.51–4.55)
Austria	14603	62403	(61631–63465)	(61549 – 63474)	4.27	(4.22–4.37)	(4.21–4.35)
Germany	137439	650841	(647138–655236)	(646974 – 655056)	4.74	(4.71 – 4.77)	(4.71–4.77)
Spain	188068	871660	(868136–875570)	(868615 – 874953)	4.63	(4.61–4.66)	(4.62–4.65)
France	109252	867214	(814767–944686)	(811082 – 952137)	7.94	(7.46–8.65)	(7.42–8.72)
UK	108692	504652	(501972–508031)	(501982 – 507713)	4.64	(4.62–4.67)	(4.62–4.67)
Greece	2207	9586	(9262–10311)	(9243 – 10316)	4.34	(4.20–4.67)	(4.19–4.67)
Belgium	36138	186633	(182715–191609)	(182744 – 191383)	5.16	(5.06–5.30)	(5.06–5.30)
Norway	6791	26680	(26199–27456)	(26197 – 28344)	3.93	(3.86–4.04)	3.86–4.03
Sweden	13216	56917	(56120–58001)	(56103 – 58004)	4.31	(4.25–4.39)	(4.25–4.39)

of an upper bound for undetected cases for several European countries (see Table 1). The last column in Table 1 shows the ratio of the total estimated cases to the observed cases. The ratio of the total estimated cases (in the worse scenario) to the observed cases is interesting in itself. A ratio of 4.5 would mean that for every observed patient there are 3.5 infected persons unseen. The reason for this can be manifold as these unseen cases might be without symptoms or show very mild signs of infection.

As expected, the undetected cases represent a relevant portion of the total number of cases. This is in line with a few existing works and discussions on the topic, see e.g. [29–31]. The number of total number of cases are at most approximately 4.5 times the observed cases. Of course, differences arise at the Country level, and heterogeneous estimates ranging from the 3.93 times of Norway to the 7.94 times of France, see Table 1. These differences are due to different heterogeneity structures in the cases and deaths time series at the country level. These results are telling us that SARS-COV-2 outbreak was more prevalent than described by the official data, though a significant number of individuals that are infected actually remain asymptomatic.

Point estimates can be used to synthetically describe the SARS-COV-2 outbreak, but they may be rather uncertain. In Table 1, we also provide uncertainty measures, based on the boot-

strap procedures described in Section 3.1. It is also possible to compare the two employed bootstrap approaches. They perform rather similarly (see also [26]) and the bootstrap intervals are rather narrow, with France only showing a rather wide interval to indicate that its point estimate should taken with caution.

4 Conclusions

Different capture-recapture approaches have been used to estimate the size of a partially observed population; many parametric or semi-parametric models have been developed to estimate the population size from aggregated counts leading to an approximation of the missed population and/or to the estimate of the threshold under which the number of missed people cannot fall (i.e. a lower bound). While several proposals for the latter exist, the estimation of an upper bound in capture recapture methods has been often overlooked, with the exception of the recent work of [23]. We propose an extension of the upper bound estimator under cumulative data distributions, in an open population, such that a day-wise estimator varying over time. The approach results in a time-aggregated approximation for f_0 and thus for N. The proposed upper bound estimator has been applied to registered cases in some European Countries; confidence intervals for N have been provided by employing bootstrap approaches. We consider, for each country, data up to the 17 of April, by assuming, given also the day wise nature of the estimator, that the recoveries are negligible; however when dealing with cases and deaths at a more recent date, given the increased percentage of immune people, recoveries should be taken into account in the computation.Another issue which should be considered is the one concerning the role of the deaths: even when the number of confirmed cases for two different Countries are close to each other the upper bounds can be different according to the deaths size wrt the cases (i.e. France and Spain). The length of the observation window plays an important role in this context and according to the distribution of SARS-CoV-2 cases observed more than once, the distribution can be less or more stable. It appears necessary to analyze this issue more deeply and we propose

to do this in a future work.

References

[1] Petropoulos, F., Makridakis, S. (2020) Forecasting the novel coronavirus COVID-19. PLoS ONE 15(3): e0231236.

[2] Sebastiani, G., Massa, M., Riboli, E. (2020) Covid-19 epidemic in Italy: evolution, projections and impact of government measures. *European Journal of Epidemiology* **35**: 341–345.

[3] Böhning, D., Rocchetti, I., Maruotti, A., Holling, H. (2020). Estimating the undetected infections in the Covid-19 outbreak by harnessing capture-recapture methods. *International Journal of Infectious Diseases*, to appear.

[4] Tuite, A. R., Ng, V., Rees, E., Fisman, D. (2020) Estimation of COVID-19 outbreak size in Italy. *The Lancet Infectious Disease*, 20: 537.

[5] Yue, M., Clapham, H. E., Cook, A. R. (2020) Estimating the Size of a COVID-19 Epidemic from Surveillance Systems, *Epidemiology*, doi: 10.1097/EDE.0000000000001202

[6] Flaxman, S., Mishra, S., Gandy, A., Unwin, H., Coupland, H. et al. (2020) Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries, *http://hdl.handle.net/10044/1/77731*

[7] Lange, J.H., LaPorte, R.E. (2003) Capture-recapture method should be used to count how many cases of SARS really exist. *BMJ*, 326: 1396.

[8] Böhning, D., van der Heijden, P.G.M., Bunge, J. (2019). Capture-Recapture Methods for the Social and Medical Science. CRC Press.

[9] McRea, R.S, Morgan, B.J.T. (2015) Analysis of Capture-Recapture Data. CRC Press.

[10] Tilling, K. (2001) Capture-recapture methods–useful or misleading? *International Journal of Epidemiology*, 30: 12–14.

[11] Wesson, P. D., Mirzazadeh, A., McFarland, W. (2018). A Bayesian approach to synthesize estimates of the size of hidden populations: the Anchored Multiplier. *International Journal of Epidemiology* 47: 1636-1644.

[12] Wesson, P.D., McFarland, W., Qin, C.C., Mirzazadeh, A. (2019). Software Application Profile: The Anchored Multiplier calculator–a Bayesian tool to synthesize population size estimates, *International Journal of Epidemiology*, 48: 1744–1749.

[13] Chao, A., Colwell, R.K. (2017) Thirty years of progeny from Chao's inequality: estimating and comparing richness with incidence data and incomplete sampling. *SORT Stat Oper Res Trans* 41:3–54

[14] McDonald, T.L., Amstrup S.C. (2001). Estimation of Animal Abundance and Related Parameters. *Journal of Agricultural, Biological, and Environmental Statistics*, 6: 206–220.

[15] Niwitpong, S.A, Boehning, D,. van der Heijden, P.G., Holling, H. (2013) Capture–recapture estimation based upon the geometric distribution allowing for heterogeneity. *Metrika* 76:495–519

[16] Gregori, D., Azzolina, D., Lanera, C., Prosepe, I., Destro, N., Lorenzoni, G., Berchialla, P. (2020). A first estimation of the impact of public health actions against COVID-19 in Veneto (Italy). J Epidemiol Community Health.

[17] Khalatbari-Soltani, S., Cumming, R. G., Delpierre, C., Kelly-Irving, M. (2020). Importance of collecting data on socioeconomic determinants from the early stage of the COVID-19 outbreak onwards. J Epidemiol Community Health.

[18] Lai, F. T. T. (2020). Association between time from SARS-CoV-2 onset to case confirmation and time to recovery across socio–demographic strata in Singapore. J Epidemiol Community Health.

[19] Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W., Shaman, J. (2020). Substantial un-

documented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science*, 368: 489–493.

[20] Zhou, T., Liu, Q., Yang, Z., Liao, J., Yang, K., Bai, W. et al. (2020) Preliminary prediction of the basic reproduction number of the Wuhan novel coronavirus 2019-nCoV. *Journal of Evidence Based Medicine* 13: 3–7.

[21] Zhao, S., Lin, Q., Ran, J., Musa, S.S., Yang, G., Wang, W., et al. (2020) Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. *International Journal of Infectious Diseases*, 92: 214–217.

[22] Chao, A. (1989) Estimating population size for sparse data in capture–recapture experiments. *Biometrics* 45:427–438.

[23] Alfó, M., Böhning, D., Rocchetti, I. (2020). Upper bound estimators of the population size based on ordinal models for capture-recapture experiments. *Biometrics*, https://doi.org/10.1111/biom.13265

[24] Zwane, E., van der Heijden, P. (2003). Implementing the parametric bootstrap in capture-recapture studies. *Statistics & Probability Letters* 65: 121–125.

[25] Norris, J.L., Pollock, K.H. (1996). Including model uncertainty in estimating variances in multiple capture studies. *Environmental and Ecological Statistics* 3:235–244.

[26] Anan, O., Böhning, D., Maruotti, A. (2017) Uncertainty estimation in heterogeneous capture–recapture count data. *Journal of Statistical Computation and Simulation*, 87: 2094–2114.

[27] Buckland, S., Garthwaite, P. (1991) Quantifying precision of mark-recapture estimates using the bootstrap and related methods. *Biometrics*, 47: 255–268.

[28] Chao, A. (1987) Estimating the population size for capture–recapture data with unequal catchability. *Biometrics* 43:783–791.

[29] Day M. (2020) Covid-19: identifying and isolating asymptomatic people helped eliminate virus in Italian village. BMJ, 368:m1165.

[30] La Stampa (2020) Castiglione d–Adda – un caso di studio: –Il 70% dei donatori di sangue – positivo–. lastampa.it. 2020.https://www.lastampa.it/topnews/primo-piano/2020/04/02/news/coronavirus-castiglione-d-adda-e-un-caso-di-studio-il-70-dei-donatori-di-sangue-e-positivo-1.38666481

[31] WHO (2020). Q&A: Similarities and differences – COVID-19 and influenza. https://www.who.int/news-room/q-a-detail/q-a-similarities-and-differences-covid-19-and-influenza