

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

4

Irene Rocchetti

Statistical Office - Consiglio Superiore della Magistratura, Italy

Dankmar Böhning

Southampton Statistical Sciences Research Institute

University of Southampton

Antonello Maruotti

Dipartimento di Giurisprudenza, Economia, Politica e Lingue Moderne

Libera Università Ss Maria Assunta

Department of Mathematics

University of Bergen

Heinz Holling

Department of Methods and Statistics, Faculty of Psychology and Sports

University of Münster

5 June 6, 2020

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 (Petropoulos and Makridakis, 2020; Sebastiani et al., 2020). 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 (Yue et al., 2020; Tuite et al., 2020; Böhning et al., 2020). 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 (Flaxman et al., 2020). 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 (Lange and LaPorte, 2003). 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 Böhning et al., 2019; McCrea and Morgan, 2015). Many CR estimators have been proposed in the literature (see e.g. Tilling, 2001; Wesson et al., 2018, 2019), and some of them can be used to identify lower bounds (Chao and Colwell, 2017) 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 (McDonald and Amstrup, 2001). A lower bound of the total number of infected cases is computed by Böhning et al. (2020) modifying the Chao estimator (Niwitpong et al., 2013) to address issues related to the data at hand. This is an important 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. 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 (Li et al., 2020b; Zhao et al., 2020; Zhou et al., 2020). 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 $t = t_0, \dots, t_m$. Hence $\Delta N(t) = N(t) - N(t-1)$ are the number of new infections at day t where $t = t_0 + 1, \dots, t_m$. Also, let $D(t)$ denote the cumulative count of deaths at day t where $t = t_0, \dots, t_m$. 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 $t = t_0 + 1, \dots, t_m$.

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 , (Chao and Colwell, 2017; Niwitpong et al., 2013):

$$\hat{f}_0 = f_1^2 / f_2. \quad (1)$$

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

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

By applying the estimator (1) day-wise we get the modified Chao lower bound estimator (see Böhning et al., 2020):

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

76 In practice, however, the bias-corrected form of (3) suggested by Chao (1989) is used:

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

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

$$N_{LB} = N(t_m) + \hat{f}_0 \quad (5)$$

81 2.2 The Upper Bound estimator

82 The lower bound is helpful as an indication of the minimum number of people having had SARS-
 83 CoV-2 and answers to a fundamental open question: “How many undetected cases are at least
 84 going around?”. Nevertheless, this information may be treated as a starting point whenever
 85 interventions and tools to dampen the spread of the epidemic are rolled out. The proposed
 86 upper bound estimator extends the research on the undetected cases and helps policy makers to
 87 evaluate the SARS-CoV-2 epidemic situation locally and at the current phase of its development.
 88 An estimate of the worse possible scenario is provided.

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

93 Let us introduce the cumulative distribution function

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

94 where homogeneity in the probability of being infected at a certain date t is assumed, i.e.
 95 $\pi_{ij} = \pi_j$, with $p_{ij} = p_j$ being the cumulative zero-truncated probability distribution. Equation
 96 (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 Covid 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 Covid would lead to the risk of not observing the number of people affected by Covid 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 Alfó et al. (2020):

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

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

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

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:

$$\hat{n}_{0_{UB}}(t) = n_{obs}^*(t) \frac{\hat{\pi}_0^{UB}}{1 - \hat{\pi}_0^{UB}}.$$

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

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

114 To get an estimate for the missed Covid infections $n_0(t)$ at each time t we compute the Horvitz–
 115 Thompson (HT) estimator at each time t and ultimately we sum it up over all times, reaching
 116 thus the final upper bound for the missed Covid cases n_0 as follows

$$\hat{n}_{0_{UB}} = \sum_{t=t_0+2}^{t_m} \left(\frac{n_{obs}^*(t)}{1 - \pi_0^*(t)} - n_{obs}^*(t) \right). \quad (7)$$

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

$$\hat{N}_{UB} = \hat{n}_{0_{UB}} + N_{t_m}.$$

119 2.3 Uncertainty estimation

120 A fundamental issue in general CR analyses is the quantification of uncertainty surrounding
 121 the estimates of the unknown population size. An estimation of the population size can be
 122 correctly computed, but if the associated estimation of variance is poor, then coverage by the
 123 95% confidence interval may falsely indicate poor estimation by the point estimator, i.e. the point
 124 estimator may result in a poor coverage rate. Focusing on the proposed upper-bound estimator,

Figure 1: Cases and deaths for the analyzed countries

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 (Buckland and Garthwaite, 1991; Norris and Pollock, 1996; Zwane and van der Heijden, 2003; Anan et al., 2017). 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 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 (Chao, 1987): 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.

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
Italy	172434	780704	(777690–784121)	(778080 – 783895)	4.53
Austria	14603	62403	(61631–63465)	(61549 – 63474)	4.27
Germany	137439	650841	(647138–655236)	(646974 – 655056)	4.74
Spain	188068	871660	(868136–875570)	(868615 – 874953)	4.63
France	109252	867214	(814767–944686)	(811082 – 952137)	7.94
UK	108692	504652	(501972–508031)	(501982 – 507713)	4.64
Greece	2207	9586	(9262–10311)	(9243 – 10316)	4.34
Belgium	36138	186633	(182715–191609)	(182744 – 191383)	5.16
Norway	6791	26680	(26199–27456)	(26197 – 28344)	3.93
Sweden	13216	56917	(56120–58001)	(56103 – 58004)	4.31

145 Data from the day which we record the first death are analyzed only. We obtain the estimates
146 of an upper bound for undetected cases for several European countries (see Table 1). The last
147 column in Table 1 shows the ratio of the total estimated cases to the observed cases. The ratio
148 of the total estimated cases (in the worse scenario) to the observed cases is interesting in itself.
149 A ratio of 4.5 would mean that for every observed patient there are 3.5 infected persons unseen.
150 The reason for this can be manifold as these unseen cases might be without symptoms or show
151 very mild signs of infection.

152 As expected, the undetected cases represent an important portion of the total number of
153 cases. This is in line with a few existing works and discussions on the topic, see e.g. Day (2020);
154 WHO (2020); La Stampa (2020). The number of total number of cases are at most approximately

155 4.5 times the observed cases. Of course, differences arise at the Country level, and heterogeneous
156 estimates ranging from the 3.93 times of Norway to the 7.94 times of France, see Table 1. These
157 differences are due to different heterogeneity structures in the cases and deaths time series at
158 the country level. These results are telling us that SARS-COV-2 outbreak was more important
159 *prevalent?* than described by the official data, though a significant number of individuals that
160 are infected actually remain asymptomatic.

161 Point estimates can be used to synthetically describe the SARS-COV-2 outbreak, but they
162 may be rather uncertain. In Table 1, we also provide uncertainty measures, based on the boot-
163 strap procedures described in Section 3.1. It is also possible to compare the two employed
164 bootstrap approaches. They perform rather similarly (see also Anan et al., 2017) and the boot-
165 strap intervals are rather narrow, with France only showing a rather wide interval to indicate
166 that its point estimate should taken with caution.

167 4 Conclusions

168 Different capture-recapture approaches have been used to estimate the size of a partially observed
169 population; many parametric or semi-parametric models have been developed to estimate the
170 population size from aggregated counts leading to an approximation of the missed population
171 and/or to the estimate of the threshold under which the number of missed people cannot fall (i.e.
172 a lower bound). While several proposals for the latter exist, the estimation of an upper bound
173 in capture recapture methods has been often overlooked, with the exception of the recent work
174 of Alfó et al. (2020). We propose an extension of the upper bound estimator under cumulative
175 data distributions, in an open population, such that a day-wise estimator varying over time. The
176 approach results in a time-aggregated approximation for f_0 and thus for N . The proposed upper
177 bound estimator has been applied to registered cases in some European Countries; confidence
178 intervals for N have been provided by employing bootstrap approaches. We consider, for each
179 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 Covid cases observed more than once, the distribution can be less or more stable. It appears necessary to analyse this issue more deeply and we propose to do this in a future work.

References

- 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>
- Anan, O., Böhning, D., Maruotti, A. (2017) Uncertainty estimation in heterogeneous capture-recapture count data. *Journal of Statistical Computation and Simulation*, 87: 2094–2114.
- Böhning, D., Rocchetti, I., Maruotti, A. and Holling, H. (2020). Estimating the undetected infections in the Covid-19 outbreak by harnessing capture-recapture methods. *International Journal of Infectious Diseases*, to appear.
- Böhning, D., van der Heijden, P.G.M., Bunge, J. (2019). Capture-Recapture Methods for the Social and Medical Science. CRC Press.
- Buckland, S., Garthwaite, P. (1991) Quantifying precision of mark-recapture estimates using the bootstrap and related methods. *Biometrics*, 47: 255–268.
- Chao, A. (1987) Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* 43:783–791.

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

206 Chao, A., Colwell, R.K. (2017) Thirty years of progeny from Chao’s inequality: estimating and
207 comparing richness with incidence data and incomplete sampling. *SORT Stat Oper Res Trans*
208 41:3–54

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

211 Flaxman, S., et al. (2020) Estimating the number of infections and the impact of non-
212 pharmaceutical interventions on COVID-19 in 11 European countries: technical description
213 update. *arXiv:2004.11342 [stat.AP]*

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

216 Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W. and Shaman, J. (2020). Substantial
217 undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2).
218 *Science*, 368: 489–493.

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

220 McDonald, T.L., Amstrup S.C. (2001). Estimation of Animal Abundance and Related Parame-
221 ters. *Journal of Agricultural, Biological, and Environmental Statistics*, 6: 206–220.

222 Niwitpong, S.A, Boehning, D., van der Heijden, P.G., Holling, H. (2013) Capturerecapture
223 estimation based upon the geometric distribution allowing for heterogeneity. *Metrika* 76:495–
224 519

225 Norris, J.L., Pollock, K.H. (1996). Including model uncertainty in estimating variances in mul-
226 tiple capture studies. *Environmental and Ecological Statistics* 3:235–244.

227 Petropoulos, F., Makridakis, S. (2020) Forecasting the novel coronavirus COVID-19. PLoS ONE
 228 15(3): e0231236.

229 Sebastiani, G., Massa, M. and Riboli, E. Covid-19 epidemic in Italy: evolution, projections and
 230 impact of government measures. *European Journal of Epidemiology* **35**: 341345.

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

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

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

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

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

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

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

250 Zhao, S., Lin, Q., Ran, J., Musa, S.S., Yang, G., Wang, W., et al. (2020) Preliminary estimation
 251 of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to

- 252 2020: A data-driven analysis in the early phase of the outbreak. *International Journal of*
253 *Infectious Diseases*, 92: 214–217.
- 254 Zhou, T., Liu, Q., Yang, Z., Liao, J., Yang, K., Bai, W. et al. (2020) Preliminary prediction of the
255 basic reproduction number of the Wuhan novel coronavirus 2019-nCoV. *Journal of Evidence*
256 *Based Medicine* 13: 3–7.
- 257 Zwane, E., van der Heijden, P. (2003). Implementing the parametric bootstrap in capture-
258 recapture studies. *Statistics & Probability Letters* 65: 121–125.