Irene Rocchetti [2]; Dankmar Bohning[1]; Heinz Holling[3]; Antonello Maruotti[4]

[1] Statistical Office Superior Council of Judiciar, Rome (IT)
[2] University of Southampton (Uk)
[3] University of Munster (WWU)
[4] Libera Università Maria SS. Assunta, Rome (IT)

Abstract:
The available data just tell us part of the story: individuals may be already infected but are not aware of it, maybe because of the absence of symptoms, or there is symptomatic suspicion but the disease has not been diagnosed yet (due to the delay in 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). Many studies have also claimed that the number of undiagnosed cases is much higher than the official number (Li et al. 2020; Pollan et al. 2020). Here, we are proposing a simple and effective method to obtain reasonable point and interval estimates of the total number of COVID-19 infections in several European countries. We introduce a novel estimator based on a capture recapture (CR) method which is considered as the gold standard for counting when it is impossible to identify each case and large undercount will occur (Lange and LaPorte 2003).
We propose an upper bound estimator under cumulative data distributions, in an open population, based on a day-wise approach allowing 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.
We focus on the ratio of the total estimated cases to the observed cases at April 17th. Differences arise at the country level; we get accurate estimates ranging from the 3.93 times of Norway to the 7.94 times of France. Bootstrap-based intervals are rather narrow.

Keywords:
Covid19 unknown cases, Capture-recapture; Upper bound estimator; population size; bootstrap based intervals.