On the Bias Arising from Relative Time Lag in COVID-19 Case Fatality Rate Estimation

Anastasios Nikolas Angelopoulos[†], Reese Pathak[†], Rohit Varma[◇], and Michael I. Jordan[†, ‡]
[◇] Southern California Eye Institute, CHA Hollywood Presbyterian Medical Center, Los Angeles
[†] Department of Electrical Engineering and Computer Science, UC Berkeley
[‡] Department of Statistics, UC Berkeley
{angelopoulos, pathakr,jordan}@cs.berkeley.edu, rvarma@sceyes.org

Abstract

The novel Coronavirus of 2019 caused an ongoing pandemic with over 400,000 confirmed cases and large variability in its reported case fatality rates (CFRs), which are proportions of fatal cases among demographic groups. The relative CFRs between groups and countries are key ratios that guide policy decisions regarding scarce medical resource allocation. In the middle of an active outbreak, estimating this measure involves correcting for time- and severity- dependent reporting of cases as well as time-lags in observed patient outcomes. In this work, we argue that we must make up for lost information about time when estimating the relative CFR: without inferring the time-dependent balance between reporting rates of fatal and non-fatal cases, CFR estimators can perform badly. To make this argument rigorous, we carry out a theoretical analysis of some current estimators of CFR. We then adapt a previously developed method—based on the well known expectation-maximization (EM) technique—for COVID-19 reporting. Our analysis is supplemented by numerical results and an open-source implementation. This should enable epidemiologists and other analysts to fit likelihood-based models similar to the ones we propose as remedies for the biased nature of naive CFR estimates, permitting more accurate planning of medical resource distribution.

1 Introduction

As of March 23, 2020, the 2019 novel Coronavirus (SARS-COV-2) outbreak claimed at least 18,915 lives out of 422,915 confirmed cases worldwide, of which 108,573 recovered [3]. Because the basic reproduction number R_0 of the virus is high (est. 2 to 3) [10], public health organizations and local, state, and national governments must allocate scarce resources to populations especially susceptible to death during this pandemic. Therefore it is critical to have good estimates of the absolute case fatality ratio (CFR) of COVID-19 and the relative CFRs between different populations. It is widely believed that the naïve estimate, obtained from a simple ratio of deaths to reported cases (and which has a value of 4.4% when applied to the data of March 23, 2020), is biased [6, 2]. Despite this understanding, naïve estimates continue to be used, reported, and cited in major publications [16, 18]. In the current manuscript we review some of the statistical reasons for the bias and we discuss model-based corrections for these forms of bias. In particular, we review a bias-corrected estimate based on approximate maximum likelihood estimation, first introduced in Reich et al. [14], and apply it to the COVID-19 dataset.

To our knowledge, this method has not yet been applied to COVID-19 data that has been made publicly available [3]. This method permits us to directly estimate the relative CFR based on covariates such as geography and age. With some additional assumptions, we estimate bias-corrected absolute CFR over time.

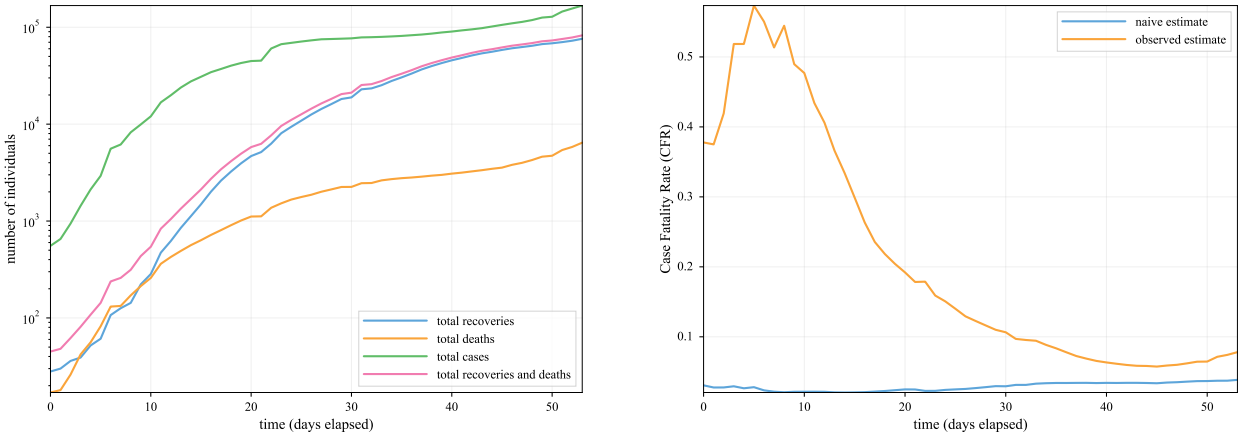

Figure 1: *Left.* Plot of time series of recoveries, R_t, deaths D_t, and cases C_t, plotted against time index t, on a logarithmic scale. We additionally plot the number of total diagnoses, $R_t + D_t$. *Right.* A plot of E_{naive} and E_{obs} as time series.

Our study is motivated by the urgent need to develop data-informed protocols for allocating scarce medical resources geographically. We view the relative CFR as a useful target for such protocols. Indeed, the absolute CFR itself has little intrinsic meaning as it averages out all effects of medical care, age, geography, genetics, and more. Making the conservative assumption that COVID-19 is sufficiently deadly to cause scarcity of medical resources (which is already apparent), the absolute CFR is not useful for planning. Furthermore, even in a given geographic region, the CFR varies in time with the strain on the medical system. One can imagine conditioning the quantity on the covariates above, and more, down to the atomic level, eventually targeting the causal increase in probability that a particular individual will die because of COVID-19. Although this deeper inquiry may be of intrinsic interest, practical decisions will ultimately be made based on averages of these covariates; for example, a region-specific relative CFR may be needed for resource distribution.

Thus, for planning purposes, we target the relative number of deaths among total cases between groups of people as the critical measure of relative risk which informs decisions affecting human lives. Ventilators, test kits, vaccines, and medical personnel should be allocated based on a combination of disease prevalence and relative risk of death (i.e., relative CFR). Fortunately, although the relative CFR varies based on the covariates of the kind mentioned above, it is actually estimable under weaker assumptions than the absolute CFR. Our estimates, explained below, correct the poor performance of the naïve estimates of CFR in a meaningful way which could affect the distribution of medical resources.

2 Notation

We access publically available data courtesy of Johns Hopkins University, consisting of time-series data of recoveries, deaths, and confirmed cases stratified across several dozen groups (in this case, geographic locations) [3]. We denote cohorts or groups of cases by indices g, belonging to a set G. For time points $t = 1, 2, \ldots, T = 41$, we collect daily data that we describe presently: for each

group $g \in G$ we collect R_t^g, D_t^g, and C_t^g, which correspond to the number of recoveries, deaths, and cases reported on day t within group g. We drop the group superscript g for population quantities:

$$R_t := \sum_{g \in G} R_t^g, \quad D_t := \sum_{g \in G} D_t^g, \quad C_t := \sum_{g \in G} C_t^g.$$

3 The naïve estimator

In early March 2020, the WHO estimate of the CFR, 3.4% was widely reported [15, 8]. This estimate is obtained from a naïve estimator;[1] specificially, the raw proportion of deaths among confirmed cases. Formally, as of March 6, 2020,

$$E_{\text{naive}} = \frac{\sum_t D_t}{\sum_t C_t} \approx 3.4\%,$$

As of March 23, 2020, E_{naive} is 4.4%. Unfortunately, as we establish in Appendix A, the naïve CFR is biased for the true CFR of the disease unless the covariance between being diagnosed and dying is zero, and reporting is perfect. Even asymptotically, the performance of this estimator becomes unboundedly bad as reporting goes to zero or the CFR goes to zero.

The bias comes from at least two flawed assumptions. First, at the time of estimation, E_{naive} assumes we have observed all deaths that will happen among diagnosed cases. Second, it assumes the reporting rates of cases is the same at all times regardless of severity. We formalize this argument mathematically in Appendix A. The naïve estimator ignores time-dependent reporting: the denominator of the fraction is growing much faster than the numerator during an active outbreak, as more people become infected (or aware that they are).

It is therefore a challenge to disentangle which recoveries should "count" at which time point t. The efficiency of the estimator is also questionable, as it never takes into account the quantity R_t (the time series of recovered cases).

The naïve estimator requires no complex modeling or tuning parameters and is easy to interpret. As we will see, there is no uniformly best method of measuring the CFR, and the naïve estimator should be viewed as one in a constellation of estimators which give a heuristic idea of the causal CFR. Nonetheless, the naïve estimator can be improved at little cost, and indeed, in this work, we suggest applying a simple correction for time-dependent reporting rates which alleviates the two basic problems with the naïve estimator.

4 Estimation using only observed outcomes

One can view the problems listed in Section 3 as a consequence of "censoring" the data. Methods for handling censored data have been studied for almost 40 years in the statistical literature; in particular, in the context of the bootstrap [4]. Several works have already applied the bootstrap to COVID-19 data to find confidence intervals for other epidemiological parameters such as R_0 [13, 1]. This should also be done for the CFR for COVID-19, as Jewell et al. did for SARS [9]. There is

[1] To be clear, it is not made explicit in the WHO report what the exact form of their case fatality rate estimates are.

also a very simple estimator which only uses observed data, namely

$$E_{\text{obs}} = \frac{\sum_{t=1}^{T} D_t}{\sum_{t=1}^{T} D_t + \sum_{t=1}^{T} R_t} \approx 13.0\%.$$

This estimator accounts for the inflation of the denominator in the naïve estimator, but does not take advantage of any information in $C_{\{1,\ldots,T\}}$. Furthermore, this estimator makes the further assumption that we observe the same fraction of recovered cases and fatal cases at the time of estimation. We formalize this assertion in Appendix B. Note that in all cases, $E_{\text{obs}} \geq E_{\text{naive}}$. Generally it cannot be said that either estimator is an upper or lower bound on the true CFR due to time delay.

The closest work to ours attempts to correct for the time-dependent reporting by picking an single day in the past, and using the denominator of E_{naive} from that day[12]. However, such methods are misleading because they do not compensate for time- and severity-dependent reporting. For example, in the simple case that people do not report the disease until it is in an advanced stage, such a method would break. Instead, we must parameterize a likelihood model which can account for these factors explicitly.

5 Likelihood models

In this section, we explain how, under milder assumptions, the parametric model in Reich et al. [14] can estimate the time-dependent CFR of COVID-19. Accordingly, we show how to outperform E_{naive} and E_{obs} by fitting a parametric model to time-series data. Maximum likelihood methods have been used in the past to fit parametric models of time-dependent epidemiological quantities using data from influenza pandemics such as H1N1 and the Spanish flu [5, 14, 7]. A parametric model allows us to "bake in" assumptions about COVID-19 to improve the estimator. For example, using our model, the (unknown) probability that a person who died from COVID-19 was diagnosed is denoted ψ. ψ will vary depending on the day of symptom onset t_{on} and group g (e.g., due to awareness or availability of testing), so we index it as $\psi_{t_{\text{on}},g}$. Similarly, the time- and group-varying probability that a person who recovered from COVID-19 was diagnosed is denoted $\phi_{t_{\text{on}},g}$. Finally, we denote the true CFR as $p_{t_{\text{on}},g}$. Then, for each person A in group g and onset t_{on}, there are three possibilities:

- Scenario 1: A recovered and was diagnosed, with probability $\phi_{t_{\text{on}},g}(1 - p_{t_{\text{on}},g})$;
- Scenario 2: A died and was diagnosed, with probability $\psi_{t_{\text{on}},g} p_{t_{\text{on}},g}$;
- Scenario 3: A undiagnosed, with probability $(1 - \phi_{t_{\text{on}},g})(1 - p_{t_{\text{on}},g}) + (1 - \phi_{t_{\text{on}},g}) p_{t_{\text{on}},g}$.

Define $N^*_{t_{\text{on}},g}$ as the (unknown) number of total cases infected at time t_{on} and in group g. Then the three scenarios define a multinomial model with three categories. If we had t_{on} and the time of death/recovery for each COVID-19 patient, the model could be fit analytically. But we do not, so we need an extra assumption: that we know, conditioned on the death of a patient, the probability that they die on any day $t_d \geq t_{\text{on}}$. This quantity is defined as:

$$\eta_j := \mathbb{P}[\text{death occurred at time } t_{\text{on}} + j \mid \text{death occurred}].$$

Two assumptions are needed to fit this model. First, p must be small. Second, either: 1) the proportion of people reporting fatal and non-fatal cases is constant; or 2) reporting rates do not

	Iran	Italy	South Korea
E_{naive}	1.94	2.45	0.34
E_{obs}	4.16	10.51	0.76
Ours	**3.29**	**6.97**	**0.14**

Table 1: The relative CFR for different nations using the three different estimators.

vary among covariates; or 3) 100% of fatalities are reported. An extensive mathematical treatment and evaluation of this model is in Reich et al. [14]. They show that as long as p_g stays below 0.05, the estimated CFR is off by less than 10% of its true value. Similarly, it is not very sensitive to misspecification of the distribution of deaths, η. In order to get absolute CFR information, we have to either compare a country to a simulated reference outbreak with a known CFR, or time-lag the country's data against itself. We did the latter. We will now apply this model to the COVID-19 data [3].

6 Results

Using the `coarseDataTools` package in `RStudio`, we applied the above model on the open-sourced COVID-19 data. The code is included for reproducibility.

6.1 Relative CFRs

We report results for the estimation of relative CFRs in three countries in Table 1. The CFRs are calculated relative to China's CFR. Independently, we estimated this absolute CFR to be 2.1%. More relative CFRs can be calculated using our codebase.

6.2 Choosing η

Fourteen vectors of length L, η_1^{14}, were chosen to illustrate the estimator's performance as our assumption about the true distribution of deaths changes. We stopped at 14 because that is the current best estimate for the mean time to death[17]. The $\eta^{(i)}$ were chosen as L evenly spaced points along the domain of a Gaussian probability distribution function with mean i and unit standard deviation:

$$\eta^{(i)}[l] = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}(l-i)^2).$$

For all relative CFR calculations, we used η_{14}.

6.3 Bias-Corrected CFR

We found that the bias-corrected CFR was between 1.7% and 3.6% depending on the chosen η. Table 2 summarizes our results for all values of η. Our results indicate a downward correction of the naïve estimate of 4.4% to 2.4%. More importantly, as more data becomes available, our time-series based methodology will increase in accuracy. It will also be possible to model longer-tailed death distributions as more data points become available.

μ	1	2	3	4	5	6	7	8	9	10	11	12	13	14
CFR (%)	3.6	2.9	2.0	1.7	1.8	2.1	2.3	2.3	2.3	2.6	2.6	2.5	2.8	**2.4**

Table 2: Using our likelihood model, the global, absolute CFR decreases to 2.4 as the mean time to death approaches the true mean of 14. Above, if $\mu = i$, it implies that the experiment was carried out sampling $X \sim N(i, 1)$, as described in section 6.2.

7 Discussion

Our result indicates that after correction for time-dependent reporting rates, the WHO estimate of 3.4% (now 4.4%) is likely an overestimate of the true CFR. This fact motivates the methodology that we use. We note that the motivation will only strengthen as the reporting-rate adjusted estimators for relative CFR begin to converge with more evenly distributed testing in Italy, South Korea, Iran, and the United States. We also show that the epidemiological community should be highly skeptical of both E_{naive} and E_{obs}; they are biased even in generous infinite-data conditions. Our code, released on GitHub with this submission, includes a script which will run the estimator on new COVID-19 data worldwide as it is published.

The estimates in Table 1 numerically confirm the intuitive observation that in Italy and Iran, CFR is much higher than in both China and South Korea, likely due to the differences in availability of medical care. Finer-grained estimates, such as those based on city-level surveillance data, may help allocate personal protective equipment and medical personnel in these dire circumstances.

Our estimate seeks to address only a subset of biases. We explicitly account for the time-dependence of reporting rates, which may differ among covariates. We have separate time-dependent reporting rates for cases which will eventually be fatal or non-fatal, correcting for the fact that reporting is higher among severe cases. Deaths are known to be related to some combination of healthcare quality and age, which can therefore be accounted for by our relative CFR estimates. Reporting delay may be country dependent, which we address, as long as our constant-proportion assumption holds (which it may well not).

Many key biases (in addition to our modeling assumptions) remain unaddressed by the community. Tests are often reserved for severe cases of COVID-19, in order to prescribe treatments to diseased patients. This likely biases any CFR estimate upwards, even with our correction, by causing unpredictable variations in reporting jointly by group and time. Details in the definitions of terms across countries and times also result in severe bias in time series data; for example, on February 12th, the Chinese government changed the definition of "confirmed case" to include symptom-based diagnoses, resulting in a 600% increase in cases that day [11]. Still, China explicitly says they will not count asymptomatic cases towards the confirmed case count (which may actually help, since if the proportion of asymptomatic cases is an intrinsic property of the disease, this could help us estimate the total number of cases). It is questionable whether all data is fully and honestly reported from all countries. Finally, the sensitivity and specificity of COVID-19 assays vary significantly by country. Because many of these biases place higher weight on fatal cases than non-fatal cases, countries with more biased data collection strategies will likely have artificially inflated CFRs. Accounting for these biases may be possible with great effort by many data analysts. However, it may be best for the statistical community to instead channel this effort into a unison clarion call to governments: randomized testing of a closed population, with at-home test kits or at grocery stores, is by far the best way to debias these estimates.

8 Conflicts

The authors declare no conflicts.

9 Acknowledgements

A.N.A. is funded by the National Science Foundation Graduate Research Fellowship Program. R.P. is funded by a UC Berkeley ARCS Fellowship. No funding sources played an involvement in the work. The corresponding author has had full access to all the data in the study and final responsibility for the decision to submit for publication.

A Proof that the naïve estimator of CFR is biased

The naive estimator has two major limitations: 1) It does not compensate for the fact that people who become confirmed cases are biased to become more severely ill. 2) It does not compensate for censoring. The first proof deals with 1), and the second deals with 2).

This proof will have different notation from the rest of the paper, since it deals with individual random variables for each COVID-19 infected person instead of time-series data.

Index the infected population with the integers $\{1, \cdots, N\}$. Let $T_i \sim \text{Ber}(p)$ be a Bernoulli random variable representing whether or not i died, and let $W_i \sim \text{Ber}(q)$ be a Bernoulli random variable representing whether or not i was diagnosed with the virus. Also define $\gamma_1 = \mathbb{P}[W_i = 1 \mid T_i = 1] = (\text{Cov}(T_i, W_i) + pq)/p$. We want to estimate p. But we only have $\sum_{i=1}^N V_i = \sum_{i=1}^N T_i W_i$. Applying the tower property of conditional expectation and using the exchangeability of the (T_i, W_i) pairs, we have:

$$\mathbb{E}\left[\frac{\sum_{i=1}^N V_i}{\sum_{j=1}^N W_i}\right] = \sum_{i=1}^N \mathbb{E}\left[\frac{V_1}{\sum_{j=1}^N W_j}\right] = N\mathbb{E}\left[T_1 \mathbb{E}\left[W_1 \frac{1}{W_1 + \sum_{l=1}^N W_l} \;\middle|\; T_1\right]\right].$$

Since W_1 is independent of $W_{2,\ldots,N}$, we can express the sum in the denominator as a binomial random variable, $B \sim \text{Bin}(N-1, q)$. Note the fact that $\mathbb{E}\left[\frac{1}{1+B}\right] = ((1-(1-q)^N))/Nq$. Then, evaluating the innermost expectation first:

$$N\mathbb{E}\left[T_1 \mathbb{E}\left[W_1 \frac{1}{W_1 + B} \;\middle|\; T_1\right]\right] = N\mathbb{E}\left[T_1 \gamma_1 \mathbb{E}\left[\frac{1}{1+B}\right] \;\middle|\; T_1\right] = p\frac{\gamma_1}{q}(1-(1-q)^N).$$

Finally, substituting for γ_1, we obtain the final form:

$$\mathbb{E}[E_{naive}] = \frac{\text{Cov}(W_i, T_i) + pq}{q}(1-(1-q)^N).$$

Recall that q is the probability of reporting given an infection, and p is the probability of death given an infection. Therefore E_{naive} is biased for p unless both $q = 1$ and $\text{Cov}(T_i, W_i) = 0$, which is clearly false for any real disease. Interestingly this empirical CFR is not constrained to be an underestimate, and can overestimate p if $p \geq \frac{\text{Cov}(W_1, T_1)(1-(1-q)^N)}{q(1-q)^N}$.

Also, if there exists some small probability ϵ of dying independent of the disease and the covariance between T and D is otherwise nonnegative, the (asymptotic) overestimate can get as bad as:

$$\lim_{q \to 0} \frac{\text{Cov}(T_i, W_i) + pq}{q} \geq \lim_{q \to 0} \frac{\epsilon + pq}{q} = \infty.$$

In other words, as the rate of reporting (q) decreases or the covariance between death and reporting increases, the CFR gets worse, ultimately becoming infinitely bad. Simiarly, the ratio $\frac{E_{\text{naive}}}{p}$ can become infinitely bad as the product pq decreases.

A proof of Part 2), that the naïve estimator of the CFR is biased for the true CFR, can be found in Appendix A of Reich et al. [14].

B Proof that the observation-only estimator of CFR is biased

We borrow notation and proof technique from Reich et al. [14]. Specifically, define the time-dependent reporting rate of fatal cases in group g as at time t as $\psi_{t,g} \in [0, 1]$. Similarly, the

time-dependent reporting rate of non-fatal cases is $\phi_{t,g} \in [0,1]$. $D_{t,g}$, $R_{t,g}$, and $C_{t,g}$ are the time series of deaths, recoveries, and confirmed cases, as before. Also define $N^*_{t,g}$ as the total number of cases, reported and unreported, at time t for group g.

In addition, define: $d_{t,g} := \mathbb{E}[D_{t,g} \mid N^*_{t,g}] = N^*_{t,g} p_g \psi_t$ and $r_{t,g} := \mathbb{E}[R_{t,g} \mid N^*_{t,g}] = N^*_{t,g}(1-p_g)\phi_t$. Also, introduce two functions, $F_d(t) \to [0,1]$ and $F_r(t) \to [0,1]$, where F_d represents the fraction of confirmed, fatal cases who have died by time t. Similarly F_r represents the fraction of confirmed, non-fatal cases who have recovered by time t. During an active outbreak, we have $F_d < 1$ and $F_r < 1$. Finally, define T as the current time; all sums over time below have an upper limit of T unless otherwise specified. We seek the asymptotic convergence target of:

$$E_{obs} = \frac{F_d(T)\Sigma_{t_2} D_{t_2,g}}{(F_d(T)\Sigma_{t_1} D_{t_1,g}) + (F_r(T)\Sigma_{t_1} R_{t_1,g})}.$$

By the weak law of large numbers we have:

$$\frac{d_{t,g}}{N^*_{t,g}} \xrightarrow{p} p_g \psi_t,$$

and similarly,

$$\frac{r_{t,g}}{N^*_{t,g}} \xrightarrow{p} (1-p_g)\phi_t.$$

Now we focus on the denominator. We have to introduce a "smoothness" assumption: the number of infected people at each timestep $N^*_{t_1,g}$ has a constant ratio with the number of infected people at each other timestep $N^*_{t_2,g}$. This is a conservative assumption in any real-world scenario. In particular, as $N^*_{t_1,g} \to \infty$ and $N^*_{t_2,g} \to \infty$,

$$\frac{N^*_{t_1,g}}{N^*_{t_2,g}} \to \lambda_{t_1,t_2,g}.$$

Therefore, we also have by Slutsky's theorem that:

$$\frac{d_{t_1,g} + r_{t_1,g}}{N^*_{t_2,g}} \xrightarrow{p} \lambda_{t_1,t_2,g}(p_g\psi_t + (1-p_g)\phi_t).$$

Now, applying the weak law of large numbers and our assumption:

$$\frac{d_{t_2,g} + r_{t_2,g}}{N^*_{t_1,g}} = \frac{N^*_{t_1,g}}{N^*_{t_2,g}}\left(\frac{d_{t_2,g}}{N^*_{t_2,g}} + \frac{r_{t_2,g}}{N^*_{t_2,g}}\right) \xrightarrow{p} \lambda_{t_2,t_1,g}(p_g\psi_t + (1-p_g)\phi_t).$$

Then, by Slutsky's theorem,

$$\frac{d_{t_2,g}}{\Sigma_{t_1} d_{t_1,g} + r_{t_1,g}} = \frac{d_{t_2,g}/N^*_{t_2,g}}{\Sigma_{t_1}(d_{t_1,g} + r_{t_1,g})/N^*_{t_2,g}} \xrightarrow{p} \frac{p_g\psi_t}{\Sigma_{t_1}(\lambda_{t_1,t_2,g}(p_g\psi_t + (1-p_g)\phi_t))}.$$

Finally, applying these results to our estimator (and with one final application of Slutsky's theorem to F_d and F_r),

$$E_{\text{obs}} \xrightarrow{p} \Sigma_{t_2} \frac{F_d(T) p_g \psi_t}{F_d(T)(\Sigma_{t_1}\lambda_{t_1,t_2,g} p_g \psi_t) + F_r(T)(\Sigma_{t_1}\lambda_{t_1,t_2,g}(1-p_g)\phi_t)}$$

This is clearly a biased estimator of p_g.

References

[1] S. J. Akhmetzhanov, B. Yuan, R. Kinoshita, and H. Nishiura. Epidemiological characteristics of novel coronavirus infection: A statistical analysis of publicly available case data. *medRxiv*, 2020.

[2] M. Battegay, R. Kuehl, S. Tschudin-Sutter, H. H. Hirsch, A. F. Widmer, and R. A. Neher. 2019-novel Coronavirus (2019-nCoV): estimating the case fatality rate–a word of caution. *Swiss Medical Weekly*, 150(0506), 2020.

[3] E. Dong, H. Du, and L. Gardner. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*, Online First, 2020.

[4] B. Efron. Censored data and the bootstrap. *J. Amer. Stat. Assc.*, 76(374):312–319, 1981.

[5] K. Ejima, R. Omori, B. J. Cowling, K. Aihara, and H. Nishiura. The time required to estimate the case fatality ratio of influenza using only the tip of an iceberg: joint estimation of the virulence and the transmission potential. *Computational and Mathematical Methods in Medicine*, 2012, 2012.

[6] A. S. Fauci, H. C. Lane, and R. R. Redfield. Covid 19: Navigating the uncharted, 2020.

[7] E. L. Frome and H. Checkoway. Use of poisson regression models in estimating incidence rates and ratios. *American Journal of Epidemiology*, 121(2):309–323, 1985.

[8] T. A. Ghebreyesus. Who Director-General's opening remarks at the media briefing on COVID-19 — 3 march 2020, 2020.

[9] N. P. Jewell, X. Lei, A. C. Ghani, C. A. Donnelly, G. M. Leung, L. Ho, B. J. Cowling, and A. J. Hedley. Non-parametric estimation of the case fatality ratio with competing risks data: an application to Severe Acute Respiratory Syndrome (sars). *Statistics in Medicine*, 26(9):1982–1998, 2007.

[10] Y. Liu, A. A. Gayle, A. Wilder-Smith, and J. Rocklöv. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *Journal of Travel Medicine*, 2020.

[11] National Health Commision of China. How to interpret the 15,152 surge in COVID-19 new cases of february 12. *Accessed on-line*, 2020.

[12] R. Qi, C. Ye, X. Qin, and X. Yu. Case fatality rate of novel Coronavirus disease 2019 in china. *medRxiv*, 2020.

[13] J. M. Read, J. RE Bridgen, D. AT Cummings, A. Ho, and C. P. Jewell. Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. *medRxiv*, 2020.

[14] N. G. Reich, J. Lessler, D. AT Cummings, and R. Brookmeyer. Estimating absolute and relative case fatality ratios from infectious disease surveillance data. *Biometrics*, 68(2):598–606, 2012.

[15] B. Stelter. Trump makes spurious claims about coronavirus in phone call with Sean Hannity, 3 2020.

[16] The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. The epidemiological characteristics of an outbreak of 2019 novel Coronavirus diseases (COVID-19) in China. *China CDC Weekly*, 41(2):145, 2020.

[17] W. Wang, J. Tang, and F. Wei. Updated understanding of the outbreak of 2019 novel coronavirus (2019-nCoV) in wuhan, china. *Journal of Medical Virology*, 92:441–447, 2020.

[18] Z. Wu and J. M. McGoogan. Characteristics of and important lessons from the Coronavirus disease 2019 (COVID-19) outbreak in China: Summary of a report of 72, 314 cases from the Chinese Center for Disease Control and Prevention. *J. Amer. Med. Assc.*, 2 2020.