https://www.philomaths.org/papers/free_infected_individuals.
Version: TN6-v1-2, Click above url to ensure you are viewing latest version.

A Note on Estimating the Number of Free Infected Individuals Philomaths Technical Note - TN6

C. R. Drane
Philomaths Pty. Limited
chris@philomaths.org

September 21, 2020

1 Introduction

Russell et al. [1] describe a method for estimating the delay-adjusted case fatality in order to give an estimate of the Case Detection Ratio (CDR). This is useful to estimate under reporting. Russell uses the method to estimate the degree of under reporting in many countries around the world. The method has been used in Australia [2] to estimate under reporting in the different Australian states and territories. This note describes how the method can be adapted to estimate the number of Free Infected Individuals (FII), which is how many infected individuals are still in the community. In the process of developing this estimator we replicated the Russell results.

2 Replication of Russell Results

The CDR algorithm developed by Russell was coded in the R programming language, and a public version of the code was made available in a Github respository. Using the original equations described by Russel, we developed the CDR technique using the Ruby programming language and independently downloaded the data from the Johns Hopkins repository. By downloading the Russell CFR repository we were able to feed synthetic case and death

data in the Russell's R language version of the algorithm and our ruby version of the algorithm. The results agreed to four significant figures or better.

Russell applied the algorithm by using case and deaths data from the Johns Hopkins web site [3], and published the results of the algorithm in table form for April 30, 2020. We independently downloaded the relevant data from the John Hopkins data and calculated the CFR for each country using our ruby code. We found the results broadly consistent, taking into account different timezones and preprocessing of the raw data.

Although the above does not represent an in-depth replication it does provide some confidence that the R implementation of the algorithm is correct and that the ruby version is consistent with the R version.

We calculate the CDR on a daily bases for all the countries in the Johns Hopkins website. Without attempting a validation, it can only be said that in most cases the results appear to be plausible. One exception is for Singapore. On June 30th, 2020, our calculation of the lower CI, estimate and upper CI was [1.00, 1.00, 1.00]. Russell reported exactly the same CDR range on that date. At that time there was significant prevalence of the disease in Singapore with roughly 400 cases a day being reported. Common sense dictates that there should have been at least some cases not being detected. This mismatch may be due to the Singapore's extraordinarily low death rate. This Strait Time [4] article explains the death rate in terms of the young demographic of infections, which is mainly amongst the migrant worker community. As well, it is claimed that it also partly due to the superior hospital care in Singapore.

3 Estimation of Free Infected Individuals

In Australia, and no doubt in many other jurisdictions, health officials endeavor to remove any infected individual as soon as they are detected. This can be done by requesting the individual self-isolates or in some cases, placing them in quarantine. This means the daily case counts gives a false impression of the number of Freely Infected Individuals (FII) who are moving about the community. We define FII as the number of individuals who are infected but who have not been detected and are not already in quarantine. Health officials are very concerned with identifying these individuals as they are source of community transmission. One indirect indicator of the number of FII is the size of community transmission, i.e. how many of the currently detected cases have an unknown source.

We propose a method of estimating the number of FII, using the CDR. Suppose we have a time series, $i_0, i_1, ..., i_N$ of daily new infected cases. Over time, the known infected individuals will recover, so the actual number of currently infected individuals on the k^{th} day, p_k is found by convolving the complimentary serial interval cumulative distribution function with the new cases i.e the prevalance on the k^{th} day is

$$p_k = \sum_{j=0}^{j=k} i_{k-j}(1 - s_j) \tag{1}$$

where s cumulative distribution function of the serial interval. For the k^{th} day the number of FII is given by

$$f_k = p_k * (\frac{1}{r_k} - 1) \tag{2}$$

where r_k is the CDR on the k^{th} day.

4 Confidence Interval

If the estimate of the CDR also provides an upper and lower confidence interval, I_U and I_L, then the lower value of the CI for r_k can be approximated by substituting, I_U for r_k in equation 2. The use of the upper CI of the CDR to calculate the lower CI for FII is due to the reciprical relation between CDR and FII. Similarly, the upper value for the CI can be found using substituting I_L in equation 2.

A more consistent estimate of the CI would take into account the uncertainty in calculating the prevalence. However the CI for the CDR calculated using the Russell method is dominated by the uncertainty in the case fatality ratio, so it is not clear that including the uncertainty in the prevalence would have a useful contribution to the CI.

5 Conclusion

This note reports on the replication of the results of the Russell CDR estimator and then applies that to developing an estimator of the number of FII and of Immunity. It also describes an alternative estimator for FII using community transmission data. An example of the usage of the FII and

Immunity estimators can be seen at this website [5] which calculates the FII and Immunity with confidence intervals for all the countries contained in the JHU database [3]. This is done on a daily basis. Compared to the Russell paper, the website uses an Infection Fatality Rate from Meyerowitz and Merone [6], which is being also being used by the Centers for Disease Control and Prevention [7]. This means that the estimate of FII is based on an Infection Detection Ratio rather than a CDR, so will include individuals that are asymptomatic.

A Resources

The resources for this technical note are available for access at https://github.com/philomaths-org/covid-19. The free_infected_individuals folder contains the pdf for this paper and relevant code. You can access resources for earlier versions of this note on Github by clicking on the tag corresponding to the earlier technical note's version number.

References

[1] T. W. Russell, J. Hellewell, S. Abbott, C. Jarvis, K. van Zandvoort, C. nCov working group, S. Flasche, A. Kucharski, *et al.*, "Using a delay-adjusted case fatality ratio to estimate under-reporting." https://cmmid.github.io/topics/covid19/global_cfr_estimates.html, 2020.

[2] N. Golding, F. M. Shearer, R. Moss, P. Dawson, L. Gibbs, E. Alisic, J. McVernon, D. J. Price, and J. M. McCaw, "Estimating temporal variation in transmission of covid-19 and adherence to social distancing measures in australia,"

[3] J. H. C. for Systems Science and Engineering, "Cssegisanddata/covid-19." https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data, 2020.

[4] S. Khalik, "Why singapore has relatively low covid-19 death rate." https://www.straitstimes.com/singapore/why-spore-has-relatively-low-covid-19-death-rate.

[5] "Philomaths ctracker." https://cvd-tracker.herokuapp.com.

[6] G. Meyerowitz-Katz and L. Merone, "A systematic review and meta-analysis of published research data on covid-19 infection-fatality rates." https://doi.org/10.1101/2020.05.03.20089854, 2020.

[7] CDC, "Covid-19 pandemic planning scenarios." https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html. Using meyerowitz2020systematic as basis of overal IFR.