First back-calculation and infection fatality multiplier estimate of the hidden prevalence of COVID-19 in Ireland

Catherine M. Comiskey ⓘ , Anne Snel, Prakashini S. Banka

School of Nursing and Midwifery, Trinity College Dublin, Dublin, Ireland

Correspondence: Catherine Comiskey, School of Nursing and Midwifery, Trinity College Dublin, 24 D'Olier Street, Dublin 2, Ireland, Tel: +353 1896 2776, Fax: +353 1 8963001, e-mail: catherine.comiskey@tcd.ie

Background: To date computer models with multiple assumptions have focussed on predicting the incidence of symptomatic cases of COVID-19. Given emerging vaccines, the aim of this study was to provide simple methods for estimating the hidden prevalence of asymptomatic cases and levels of herd immunity to aid future immunization policy and planning. We applied the method in Ireland. **Methods:** For large scale epidemics, indirect models for estimating prevalence have been developed. One such method is the benchmark multiplier method. A further method is back-calculation, which has been used successfully to produce estimates of the scale of a HIV infected population. The methods were applied from March to October 2020 and are applicable globally. **Results:** Results demonstrated that the number of infected individuals was at least twice and possibly six times the number identified through testing. Our estimates ranged from ∼100 000 to 375 000 cases giving a ratio of 1–6 hidden cases for every known case within the study time frame. While both methods are subject to assumptions and limitations, it was interesting to observe that estimates corroborated government statements noting that 80% of people testing positive were asymptomatic. **Conclusions:** As Europe has now endured several epidemic waves with the emergence globally of new variants, it essential that both policy makers and the public are aware of the scale of the hidden epidemic that may surround them. The need for social distancing is as important as ever as we await global immunization rollout.

Introduction

Classical mathematical epidemiology has highlighted the need to identify a critical population size for an epidemic to drive across a community. This threshold depends not only on the nature of the epidemic but also on the scale of the available susceptible population.[1–4]

In the early stages of a new epidemic, where no vaccine is available, all persons are susceptible. As the epidemic progresses and the number of infectious individuals increases the number of susceptible individuals will decrease. The time between exposure to a virus (becoming infected) and symptom onset, is known as the incubation period. During this period, also known as the 'pre- symptomatic' period, some infected persons can be contagious. Therefore, transmission from a pre-symptomatic case can occur before symptom onset.[5] When an epidemic can produce pre-symptomatic, asymptomatic and symptomatic cases the identification of the numbers infected becomes more challenging. Yet, it is the estimates of this very number that is required to enable the planning of a vaccination strategy, decisions on when a community has reached its critical threshold points and when policy makers and planners can advise on school openings, safety for nursing homes and protection of multiple vulnerable communities.

Internationally, organizations from the United Nations to the Centre for Disease Control to the World Health Organization are coming together to fight the now global and continually expanding pandemic of COVID-19. According to the European Centre for Disease Control clinical presentations of COVID-19 can range from no symptoms (asymptomatic) to severe pneumonia and death. There are also notifications of cases remaining asymptomatic throughout the full duration of laboratory and clinical monitoring. Furthermore, no significant difference in viral load in asymptomatic and symptomatic patients has been reported, indicating the equally significant potential of virus transmission from asymptomatic patients. Asymptomatic cases in infants and children have also been reported.[5]

Without the resources to test and enumerate asymptomatic cases mathematical modelling can be used to provide estimates of these cases. For example, across varying regions of France and within Ireland, statistical modelling approaches using what is known as the Susceptible-Exposed-Infected-Removed (SEIR) models incorporating susceptible, infected, infectious and recovered groups of individuals have been used to estimate the impact of mitigation measures, such as local and wider travel restrictions and the closure of non-essential retail.[6] Indirect estimation methods including multiplier methods and mathematical and statistical models of back-calculation have been used successfully both internationally and in Ireland to produce estimates of the scale of a hidden infected population within HIV/AIDS, heroin use, and more recently bioterrorism, where the comparatively short incubation periods are particularly applicable to COVID-19.[7–14] Working with observed symptomatic cases and the known incubation period, these back-calculation models predict through the incubation period distribution the total numbers of infected and asymptomatic cases the observed cases arose from.

The aim of this study was to provide the first indirect estimates of the hidden prevalence of the population of asymptomatic COVID-19 cases in Ireland using both the back-calculation and the infection fatality multiplier methods. Methods developed while applicable globally were applied nationally from March to October 2020 and prior to the implementation of a national immunization programme.

Methods

This research was approved by the Faculty of Health Sciences Research Ethics Committee Trinity College Dublin, the University of Dublin, Ireland.

Infection fatality multiplier

As case finding methods are not feasible for large scale epidemics indirect mathematical and statistical models for estimating prevalence have been developed. One such method is the benchmark multiplier method, which is recommended by the European Monitoring Centre for Drugs and Drug Addiction for the estimation of the hidden numbers of people who use drugs.[15] In the context of problem drug use the total population of people who use drugs, given by T is unknown (partly hidden population). Given a sample of size B of the population in question (benchmark) and the probability c for someone of this unknown population to be member of the sample, the total population T can be estimated from

$$T = B/c,$$

where B is the number of identified people who use drugs (sample or benchmark) and c is a parameter giving the probability of a drug user (unknown target population) to be a member of the identified sample B. In our example, T is the total number or prevalence of COVID-19 positive both diagnosed and undiagnosed cases, B is the benchmark of the known number of COVID-19 deaths in the time period and the multiplier, $1/c$ is 1/the infection fatality rate, where the infection fatality rate is defined as the proportion of deaths among all infected individuals not just known cases.[16] This method assumes a constant linear relation between prevalence of all cases and the number of deaths. Cases estimated using this method may be defined as all infected individuals in the defined time period.

Back-calculation

The method of back-calculation is also an indirect method for estimating hidden prevalence and has been used successfully both internationally and in Ireland to produce estimates of the scale of an infected population within HIV/AIDS and heroin use.[7,9–12,14] Working with observed symptomatic cases and the known incubation period, these models predict through the incubation period distribution the total numbers of infected and possibly asymptomatic or hidden cases these observed cases arose from. Cases estimated using this method can be broadly defined as cases that are similar to the known diagnosed cases in the defined time period.

In its simplest form, the back-calculation model is given by,

$$C_T(t) = \int_0^t C_U(t-s)f(s)ds. \tag{1}$$

Where $C_T(t)$ is the known COVID-19 cases, $f(s)$ is the incubation period distribution and $C_U(t)$ is the unknown COVID-19 cases we wish to solve for. Given varying forms in the growth of the treated cases $C_T(t)$ and the incubation period $f(s)$ the back-calculation model can be transformed into Volterra integral equations and solved analytically as in Comiskey,[6,7] Comiskey and Hay[12] and Dempsey and Comiskey[9,10] or numerically as in Comiskey and Ruskin.[12]

The incubation period $f(t)$ may be described by the Gamma distribution given by $\Gamma(\alpha, \lambda)$, where

$$f(x : \alpha, \lambda) = \frac{\lambda(\lambda x)^{\alpha-1}e^{-\lambda x}}{\Gamma(\alpha)}, \tag{2}$$

where f is the probability of an incubation period of duration x given a mean incubation period of mean, $\mu = \alpha/\lambda$. The details of the parameters of the incubation period distribution for COVID-19 are provided by Banka and Comiskey.[17] The solution of the back-calculation model when $\alpha = 2$ and $\alpha = 3$ have been provided by Comiskey C.M. (1991)[3] and Dempsey and Comiskey.[9,10]

The solution formula for the unknown number of cases $C_U(t)$ when $\alpha = 3$ is based on a combination of the first, second and third derivative of the number of known cases $C_T(t)$ and is as follows:

$$C_U(t) = \frac{1}{\lambda^3}\frac{d^3(C_T(t))}{dt^3} + \frac{3}{\lambda^2}\frac{d^2(aC_T(t))}{dt^2} + \frac{3}{\lambda}\frac{d(C_T(t))}{dt} + C_T(t). \tag{3}$$

The solution formula for the unknown cases when $\alpha = 2$ is similar and is given by:

$$C_U(t) = \frac{1}{\lambda^2}\frac{d^2(aC_T(t))}{dt^2} + \frac{2}{\lambda}\frac{d(C_T(t))}{dt} + C_T(t). \tag{4}$$

The prevalence of the unknown asymptomatic cases is then given by the integral of the solution over the defined time period.

We have

$$\int_0^t C_U(t)dt. \tag{5}$$

Results

Estimates of hidden prevalence using the infection fatality multiplier

According to WHO[16] the infection fatality ratio for COVID-19 is estimated to be in the region of 0.5–1.0%. The total number of deaths unadjusted for reporting delays in Ireland from 1 March to the 26 of October was 1882 and the total number of known cases was 57 128.[18] Given this benchmark and the infection fatality multiplier, we have the estimates of prevalence in table 1 below.

Table 1 Estimates of the hidden prevalence of COVID-19 in Ireland from 1 March to 26 October 2020 using the infection fatality multiplier

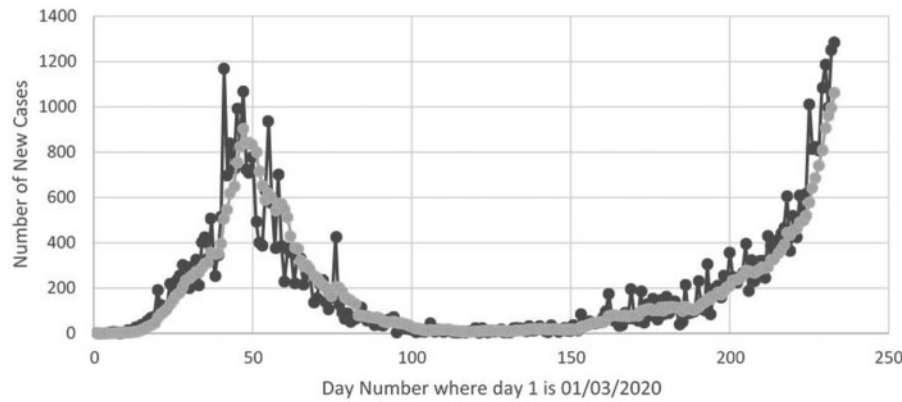| Infection fatality ratio (%)	Known deaths	Estimated total prevalence	Known prevalence	Estimated hidden prevalence	Ratio of unknown cases to each known case
0.5	1882	376 400	57 128	319 272	5.6–1
0.75	1882	250 933	57 128	193 805	3.4–1
1	1882	188 200	57 128	131 072	2.3–1

Figure 1 Known COVID-19 cases from 1st March (Day 1) up to 19th October (Day 233) in Ireland. Dark grey represents new cases and light grey represents new cases smoothed

Table 2 Best fitting curve $C_T(t)$ to the known number of COVID-19 cases across the three phases of the epidemic from March to October 2020 in Ireland

Epidemic wave	Curve	R squared	F, P
Wave 1, increasing phase	$C_T(t) = 0 \cdot 565t^2 - 11 \cdot 194t + 48 \cdot 328$	0.968	656.686, <0.001
Wave 1, decreasing phase	$C_T(t) = 0 \cdot 419t^2 - 76 \cdot 745t + 3581 \cdot 727$	0.985	1403.117, <0.001
Wave 2, increasing phase	$C_T(t) = 0 \cdot 136e^{0 \cdot 037t}$	0.946	2266.953, <0.001

Estimates of hidden prevalence using the back-calculation method

The growth in the known diagnosed cases of COVID-19 in Ireland from March to mid- October 2020 can be seen in figure 1.[19]

We can see from this figure that the epidemic to date has occurred in two waves or three phases. Epidemic Wave 1, Phase 1 was the rapid increase in cases to mid-April followed by epidemic Wave 1, Phase 2, where new cases decreased as a result of mitigation measures, which included working from home, all but essential retail closures and all school closures. These two phases of the first wave lasted until 31st of May when mitigation measures were reduced, schools and childcare facilities were reopened, and people's movement and the economy was reactivated. The second epidemic wave started after the reduction of the mitigation measures. Using simple regression techniques, we fitted three separate curves $C_T(t)$ to the known cases of these three phases. Details of the best fitting curves amongst all curves fitted are provided in table 2 below.

Using the fitted curves for the known cases, the back-calculation Equation (1) was solved three times for the unknown cases $C_U(t)$ given the known incubation period distribution of Banka and Comiskey,[17] who have estimated $\Gamma(\alpha, \lambda) = \Gamma(2 \cdot 810, 0 \cdot 419)$ and the exact solutions in Equations (3) and (4) for $C_U(t)$ with $\alpha = 3$ and with $\alpha = 2$. These solutions were used for the estimates below, where $\Gamma(\alpha, \lambda) = \Gamma(3 \cdot 0, 0 \cdot 45)$ and $\Gamma(\alpha, \lambda) = \Gamma(2 \cdot 0, 0 \cdot 30)$ both of which corresponded to the required mean incubation period of 6.7 days. Table 3 provides a summary of these estimates.

Discussion

Within this article, we have provided explicit formulae for the computation of estimates of the hidden prevalence of COVID-19 asymptomatic cases using initially mortality data and the benchmark multiplier method and latterly the back-calculation method with regression models of known cases and known information on the incubation period. Applying this to the Irish cases from March to October 2020,[18,19] we found that with the back-calculation method it was estimated that for every case identified through testing there was ~1 other asymptomatic, unknown case. This result was similar despite varying the parameters of the incubation period. Given this estimate was derived directly from the known numbers of people who have had a positive COVID-19 test it can be considered as a minimum estimate of the size of the asymptomatic prevalence. From the benchmark multiplier method, we found, using the infection fatality rate that for every case identified there were between 2.3 and 5.6 asymptomatic cases depending on the value of the infection fatality rate. Given that in Ireland all people who have died and who have had a positive COVID-19 test are classified as a COVID-19 related death regardless of the exact cause of death, this estimate may be considered as a maximum estimate of the asymptomatic prevalence of hidden cases. SEIR models of new waves of the epidemic in Ireland estimated that the number of asymptomatic infectious was of the same order of magnitude as the number of symptomatic infectious but with a larger uncertainty, this reflecting the finding from the back-calculation methods here.[19] Authors demonstrated that the role of new variants with Ireland and globally can have a significant impact on transmission while a population remains unvaccinated and the impact of vaccination on new or as yet to emerge variants is unknown.[20]

Our estimates ranged from ~100 000–375 000 cases giving a ratio of 1–6 hidden cases for every known case within the study time frame. Given an Irish population of 4.9 million these estimates equate to ~2.0–7.7% of the total population. An early serological study carried out between 26 June and 20 July 2020 among 12–69 year olds in Ireland estimated the seroprevalence rate at 1.7% (95% CI: 1.1–2.4%), which overlaps with our findings.[21] For the first wave, estimated serological prevalence in the UK found that 6.0% (95% CI: 5.8–6.1%) of individuals tested positive, of these one-third [32.2% (95% CI: 31.0–33.4%)] reported no symptoms and were asymptomatic.[22] There are two key strengths to this study. Firstly, the study utilizes well established prevalence estimation methods that have been endorsed

Table 3 Estimates of the hidden prevalence of COVID-19 in Ireland from 1 March to 19 October 2020 using the back-calculation method with a Gamma incubation period distribution $\Gamma(3 \cdot 0, 0 \cdot 45)$ and $\Gamma(2 \cdot 0, 0 \cdot 30)$[AQ10]

Epidemic wave	Hidden prevalence, $C_U(t)$ given alpha =3	Known diagnosed prevalence, C_T	Estimated total prevalence	Ratio of unknown cases to each known case
Wave 1, increasing phase 1 March–16 April	15 197	12 547	27 744	1.2–1
Wave 1, decreasing phase 17 April–31 May	9748	12 382	22 130	0.8–1
Wave 2, increasing phase 1 June–19 October	24 312	24 747	49 059	1.0–1
1 March–19 October	**49 257**	**49 676**	**98 933**	**1.0–1**
Epidemic wave	Hidden prevalence, $C_U(t)$ given alpha =2	Known diagnosed prevalence, C_T	Estimated total prevalence	Ratio of unknown cases to each known case
Wave 1, increasing phase 1 March–16 April	15 120	12 547	27 667	1.2–1
Wave 1, decreasing phase 17 April–31 May	9610	12 382	21 992	0.8–1
Wave 2, increasing phase 1 June–19 October	24 203	24 747	48 950	1.1–1
1 March–19 October	**48 933**	**49 676**	**98 609**	**1.0–1**

Note: Significance of bold values indicates total prevalence of all epidemic waves.

and recommended by the European Monitoring Centre for Drugs and Drugs Addiction for the EU wide prevalence estimation of hidden phenomena from substance use to HIV/AIDS.[15] Secondly, the study provides explicit formulae for the computation of the hidden prevalence of asymptomatic cases within a country or region given either the known numbers of fatalities or given a regression equation for the known numbers of symptomatic cases. The study does however have its limitations. Cases of COVID-19 were not adjusted for reporting delays and as a result are likely to be an underestimate of the true number of cases within the stated timeframe. As knowledge of the epidemic improved case notification rates may also improve. The two different methods are also subject to different assumptions. The benchmark multiplier method assumes that there is a linear relationship between fatalities and cases, and it is likely that this is an oversimplification as age or underlying health conditions will undoubtedly impact upon fatality rates.[16] Similarly, as reporting systems and healthcare provision and expertise and equipment improves fatality rates and case notification rates will change. The method is however considered useful for planning purposes where little is known about the scale of the hidden prevalence.[15] Similarly, the method of back-calculation assumes that symptomatic cases are similar to asymptomatic cases in their behaviours and can be back projected through the incubation period without taking into account other factors, such as social mixing, population density rates and viral load. The method has however been used successfully for estimating prevalence as it is not intended to measure the impact of behaviour and this can be seen in the works on estimating the prevalence of HIV/AIDS within populations with varying behaviours from men who have sex with men to people who inject drugs.[7–11]

The study findings when applied to the Irish setting from March to October 2020 illustrate that the level of so-called 'silent and hidden infection' in Ireland remains substantial and undocumented. Our findings are also corroborated by government statements stating that 80% of those testing positive were asymptomatic, indicating that for each symptomatic positive case there were four asymptomatic positive cases.[23] As more detailed data becomes available the methods presented within this study could be applied to more refined groupings within particular settings or age categories to provide specific estimates of prevalence within vulnerable or essential worker groups from healthcare professionals working in acute care to older people living within shared community settings. As Europe endures possible future epidemic waves despite eminent vaccines, it essential that the public are aware of the scale of the hidden epidemic that surrounds them. The need for social distancing is still important as we await national and global immunization rollout within the context of the emerging variants of this disease.

Funding

Key points

- We implemented established methods and used available data to provide an accessible methodology to estimate the hidden and asymptomatic prevalence of COVID-19 across Europe.
- Using this methodology, we found that within Ireland for every known case identified there was a minimum of one other unknown case and a maximum of possibly six unknown cases.
- These findings contribute to the identification of levels of herd immunity within our community prior to vaccination and highlight the level of so-called silent infection surrounding us.
- As we plan for imminent vaccination policies, these estimates can be refined to provide estimates of hidden prevalence among varying vulnerable groups, from healthcare workers in acute settings to older people living in communities.
- These estimation methods can assist in the identification of priorities and levels of vaccination coverage that may be required to ensure future immunity.

References

1   Murray J. *Mathematical Biology*. Berlin, Germany: Springer-Verlag, 2003.

2   Bailey N. *The Mathematical Theory of Infectious Diseases and Its Applications*. London: Griffin, 1975.

3 Medley G, Anderson R, Cox D, Billard L. Incubation period of AIDS in patients infected via blood transfusion. *Nature* 1987;328:719–21.

4 Anderson R, May R. *Infectious Diseases of Humans*. Oxford: Oxford University Press, 2010.

5 European Centre for Disease Prevention Control. Coronavirus Disease 2019 (COVID-19) Pandemic: Increased Transmission in the EU/EEA and the UK–Seventh Update. 2020. Available at: https://www.ecdc.europa.eu/sites/default/files/documents/RRA-seventh-update-Outbreak-of-coronavirus-disease-COVID-19.pdf (12 May 2021, date last accessed).

6 Cazelles B, Comiskey C, Nguyen-Van-Yen B, et al. Parallel trends in the transmission of SARS-CoV-2 and retail/recreation and public transport mobility during non-lockdown periods. *Int J Infect Dis* 2021;104:693–5.

7 Comiskey C. Improvements in integral equation models for estimates of the level of HIV infection in Ireland. *J R Stat Soc Series D (Stat)* 1992;41:329.

8 Comiskey C. Integral equations and estimating the incidence of HIV infection. *Zeitschrift Für Angewandte Mathematik Und Mechanik* 1996;76:505–6.

9 Dempsey O, Comiskey CM. Estimating the prevalence of illegal drug use. *Math Popul Stud* 2014;21:65–77.

10 Dempsey O, Comiskey CM. Estimating the incidence of hidden, untreated opiate use. *Math Popul Stud* 2011;18:172–88.

11 Egan JR, Hall IM. A review of back-calculation techniques and their potential to inform mitigation strategies with application to non-transmissible acute infectious diseases. *J R Soc Interface* 2015;12:20150096.

12 Comiskey C, Ruskin H. AIDS in Ireland: the reporting delay distribution and the implementation of integral equation models. *Comput Appl Biosci* 1992;8:579–81.

13 Comiskey C, Hay G. The method of back calculation as a means of estimating incidence. In: Sharp F, Neaman R, editors. *Modelling Drug Use: Methods to Quantify and Understand Hidden Processes*, United Kingdom: European Monitoring Centre for Drugs and Drug Addiction, 2001: 105.

14 Comiskey CM. Methods for estimating prevalence of opiate use as an aid to policy and planning. *Subst Use Misuse* 2001;36:131–50.

15 European Monitoring Centre for Drugs and Drug Addiction. EMCDDA Recommended Draft Technical Tools and Guidelines. Key Epidemiological Indicator: Prevalence of Problem Ddrug Use Lisbon: EMCDDA. 2004. Available at: http://www.emcdda.europa.eu/html.cfm/index65519EN.html (12 May 2021, date last accessed).

16 World Health Organization. Estimating Mortality from COVID-19. 2020. Available at: https://www.who.int/publications/i/item/WHO-2019-nCoV-Sci-Brief-Mortality-2020.1 (12 May 2021, date last accessed).

17 Banka P, Comiskey CM. The incubation period of COVID-19: a scoping review and meta-analysis to aid modelling and planning. *medRxiv*. 2020. https://doi.org/10.1101/2020.10.20.20216143.

18 Worldometer. Irish COVID-19 Cases. 2020. Available at: https://www.worldometers.info/coronavirus/country/ireland/ (12 May 2021, date last accessed).

19 Our World in Data. Coronavirus Pandemic (COVID-19) – The Data: Our World in Data. 2020. Available at: https://ourworldindata.org/coronavirus-data (12 May 2021, date last accessed).

20 Cazelles B, Nguyen , Van Yen B, Champagne C, Comiskey C.Dynamics of the COVID-19 epidemic in Ireland under mitigation. *BMC Infect Dis 2021*. https://www.researchsquare.com/article/rs-143697/v1.

21 HSE. Preliminary report of the results of the Study to Investigate COVID-19 Infection in People Living in Ireland (SCOPI): A National Sero-Prevalence Study, June-July. 2020. Available at: https://www.hpsc.ie/a-z/respiratory/coronavirus/novelcoronavirus/scopi/SCOPI%20report%20preliminary%20results%20final%20version.pdf (12 May 2021, date last accessed).

22 Ward H, Atchison C, Whitaker M, et al. Antibody Prevalence for SARS-CoV-2 Following the Peak of the Pandemic in England: REACT2 Study in 100 000 Adults. 2020. Available at: https://www.imperial.ac.uk/media/imperial-college/institute-of-global-health-innovation/Ward-et-al-120820.pdf (12 May 2021, date last accessed).

23 McGreevy R. Cork, Letterkenny and Cavan Hospitals Exceed Normal ICU Capacity Over COVID-19. The Irish Times. 2020. Available at: https://www.irishtimes.com/news/ireland/irish-news/cork-letterkenny-and-cavan-hospitals-exceed-normal-icu-capacity-over-covid-19-1.4382204 (12 May 2021, date last accessed).