

## RESEARCH ARTICLE

# Estimating the infection burden of COVID-19 in Malaysia

Vivek Jason Jayaraj<sup>1,2</sup>, Chiu-Wan Ng<sup>1</sup>, Awang Bulgiba<sup>1</sup>, Maheshwara Rao Appannan<sup>2</sup>, Sanjay Rampal<sup>1\*</sup>

**1** Centre for Epidemiology and Evidence-based Practice, Department of Social and Preventive Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia, **2** Ministry of Health Malaysia, Putrajaya, Malaysia

\* [srampal@ummc.edu.my](mailto:srampal@ummc.edu.my)



## OPEN ACCESS

**Citation:** Jayaraj VJ, Ng C-W, Bulgiba A, Appannan MR, Rampal S (2022) Estimating the infection burden of COVID-19 in Malaysia. *PLoS Negl Trop Dis* 16(11): e0010887. <https://doi.org/10.1371/journal.pntd.0010887>

**Editor:** Pedro F. C. Vasconcelos, Universidade do Estado do Para: Universidade do Estado do Para, BRAZIL

**Received:** February 7, 2022

**Accepted:** October 12, 2022

**Published:** November 8, 2022

**Copyright:** © 2022 Jayaraj et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The raw data supporting the conclusions here were extracted from the Ministry of Health Malaysia COVID-19 data repository at <https://github.com/MoH-Malaysia/covid19-public>.

**Funding:** This study is part of the COVID-19 Epidemiological Analysis and Strategies (CEASE) Project with funding from the Ministry of Science, Technology and Innovation (<https://www.mosti.gov.my/en/>). AB and NCW are recipient of the award (UM.0000245/HGA.GV). The funders had no

## Abstract

Malaysia has reported 2.75 million cases and 31,485 deaths as of 30 December 2021. Underestimation remains an issue due to the underdiagnosis of mild and asymptomatic cases. We aimed to estimate the burden of COVID-19 cases in Malaysia based on an adjusted case fatality rate (aCFR). Data on reported cases and mortalities were collated from the Ministry of Health official GitHub between 1 March 2020 and 30 December 2021. We estimated the total and age-stratified monthly incidence rates, mortality rates, and aCFR. Estimated new infections were inferred from the age-stratified aCFR. The total estimated infections between 1 March 2020 and 30 December 2021 was 9,955,000-cases (95% CI: 6,626,000–18,985,000). The proportion of COVID-19 infections in ages 0–11, 12–17, 18–50, 51–65, and above 65 years were 19.9% (n = 1,982,000), 2.4% (n = 236,000), 66.1% (n = 6,577,000), 9.1% (n = 901,000), 2.6% (n = 256,000), respectively. Approximately 32.8% of the total population in Malaysia was estimated to have been infected with COVID-19 by the end of December 2021. These estimations highlight a more accurate infection burden in Malaysia. It provides the first national-level prevalence estimates in Malaysia that adjusted for underdiagnosis. Naturally acquired community immunity has increased, but approximately 68.1% of the population remains susceptible. Population estimates of the infection burden are critical to determine the need for booster doses and calibration of public health measures.

## Author summary

The global scale of COVID-19 is unprecedented. Malaysia has experienced several waves of infection, leading to an overwhelmed health system. Nonetheless, reported cases remain an inaccurate estimate of the true disease burden within the population. We approximated the infection fatality ratio using an estimated adjusted age-specific case fatality ratio. We then estimate the true number of infections that have occurred in Malaysia. A total of 9.95 million cases are estimated to have occurred amongst Malaysians between March 2020 and December 2021. This figure corresponds to 32.8% of the total population of Malaysians. These findings suggest that community immunity remains low,

role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

and a great deal of the population remains susceptible. Booster doses and periodic strong public health measures may be required for the foreseeable future.

## Introduction

The global transmission of COVID-19 is unprecedented and has led to more than 282 million cases and 5.4 million deaths as of 31 December 2021 [1]. Efforts to contain its transmission have focused primarily on public health and social measures (PHSM) that have come at significant economic and social costs [2].

Malaysia has reported 2.75 million cases and 31,485 deaths as of 31 December 2021 [3]. The first large outbreak in Malaysia was managed successfully using movement restrictions between March and April 2020 [4]. However, since September 2020, institutional outbreaks, state elections, and inconsistent implementation of PHSM have led to large periodic outbreaks [5].

Underestimation remains an issue despite the substantial reported burden of disease. Screening strategies and diagnostic test accuracy are two factors that drive this underestimation [6]. Reported cases are biased estimators of true disease burden. The true burden of disease may be estimated using seroprevalence surveys and random sampling [6,7]. Alternative indicators such as hospitalization and emergency room data do not estimate the overall infection rate [7]. A more accurate estimator of the actual COVID-19 infection burden may be COVID-19 mortalities, especially in countries with low excess mortalities [7–9].

Accurately estimating the epidemic size is critical in forming situational awareness in designing and evaluating future public health and social measures. Misclassified estimates of total COVID-19 cases may hamper forecasting and future disease control planning [9,10].

To the best of our knowledge, no studies have yet estimated the true burden of disease at a population level in Malaysia. We aimed to estimate the burden of COVID-19 infections in Malaysia between 1 March 2020 and 31 December 2021.

## Methods

### Ethics statement

This study was registered under the National Medical Research Register with a registration ID (NMRR-20-1208-55087) and obtained ethical approval from the Medical Research and Ethics Committee (MREC), Ministry of Health Malaysia.

### Data source and study population

A retrospective cohort study design was utilized. Data on cases and mortalities were extracted from line lists maintained by the Ministry of Health, Malaysia, between 1 March 2020 and 30 November 2021 [11]. Healthcare facilities adjudicate all deaths as either death due to COVID-19 or death with COVID-19 based on a set of criteria (S1 Appendix) - with only the former contributing to COVID-19 mortality statistics. Age-stratified population projection estimates for 2021 were extracted from the Department of Statistics, Malaysia (DOSM) [12]. Additionally, we extracted line lists of age-stratified all-cause mortalities from the DOSM between 1 January 2020 and 30 September 2021. We excluded non-Malaysians from this study due to possibly differential health-seeking behavior and uncertainty in the enumeration of non-Malaysians in Malaysia due to undocumented immigrants [13].

## Statistical analysis

Data were explored for missingness using descriptive statistics, visualizations, and a univariate and multivariate logistic regression model. A multiple imputation model using expectation-maximization with bootstrapping was utilized to impute age. Age was then categorized as 0–11, 12–17, 18–50, 51–65, and >65 years old.

Cases, deaths, brought-in dead, incidence rates, mortality rates, and adjusted case fatality rates (aCFR) estimates were tabulated cumulatively and stratified by age and time. Death location was classified as either in-hospital death or brought-in-dead. A 95% confidence interval around these parameters was estimated using Wald's bootstrapping approach [14]. The daily incidence and mortality rates were visualized to explore longitudinal trends within Malaysia. The mid-year population was assumed to be the population at risk for the risk set on each day. The incidence density (date of reporting) and mortality rate (date of actual death) are given by:

$$\begin{aligned} &\text{Incidence density (per 100,000 population)} \\ &= \frac{\text{No. of reported COVID-19 cases (in time period)}}{\text{Mid-year population}} \times 100,000 \end{aligned}$$

$$\begin{aligned} &\text{Mortality rate (per 100,000 population)} \\ &= \frac{\text{No. of reported COVID-19 mortalities (in time period)}}{\text{Mid-year population}} \times 100,000 \end{aligned}$$

The reported case fatality rate (CFR) is estimated as the percentage of COVID-19 mortalities on a specific date over the reported number of COVID-19 cases on the date of death. A limitation of this crude reporting is the misspecification of the population at risk resulting in a more prevalent measure rather than an incident measure of risk [15–18]. We estimated an adjusted CFR (aCFR) by first attributing deaths to the date they were reported positive and then calculating the percentage of COVID-19 mortalities over the reported number of COVID-19 cases on the date the death was reported as a case. The aCFR was calculated as the number of deaths (by date of death tested positive) divided by the number of reported cases on the date the deceased cases tested positive. The infection fatality rate (IFR) is the percentage of COVID-19 mortalities over the true number of COVID-19 cases. The CFR, aCFR and IFR are given by:

$$\text{CFR (\%)} = \frac{\text{No. of reported COVID-19 deaths (date of reporting)}}{\text{No. of reported COVID-19 cases (date of reporting)}} \times 100$$

$$\text{Adjusted CFR (\%)} = \frac{\text{No. of reported COVID-19 deaths (date of testing positive)}}{\text{No. of reported COVID-19 cases (date of reporting)}} \times 100$$

$$\text{IFR} = \frac{\text{All COVID-19 deaths}}{\text{All COVID-19 cases}} \times 100$$

We approximated the age-stratified IFR using the lowest-non zero aCFR between 1 October 2020 and 31 October 2021. We utilized this period as few deaths ( $n = 12$ ) occurred before 1 October 2020, and the risk of death is likely different after 31 October 2021 due to the National COVID-19 Immunization Program. The age-stratified aCFR was estimated over a moving 3-month period to stabilize the approximated IFR as there was a low number of deaths reported in some age strata. The lowest age-stratified non-zero aCFR was compared to reported pooled estimates [19, 20].

The ratio of CFR to IFR will equal 1 when all cases are ascertained ( $\frac{CFR}{IFR} = 1$ ). As the number of underestimated cases increases, the CFR also increases, and as such  $\frac{CFR}{IFR} > 1$ . We corrected the observed number of cases to reflect the true number of cases by calibrating the observed cases against an adjustment factor obtained from the ratio of  $\frac{CFR}{IFR}$  which is given by:

$$E_{tp} = O_{tp} * \left( \frac{CFR_{tp}}{IFR_p} \right), \text{ where}$$

$O_{tp}$  = observed cases at time  $t$  in age group  $p$

$IFR_p$  = Lowest non-zero CFR in age group  $p$ ,

$E_t$  = expected number of cases in age group  $p$

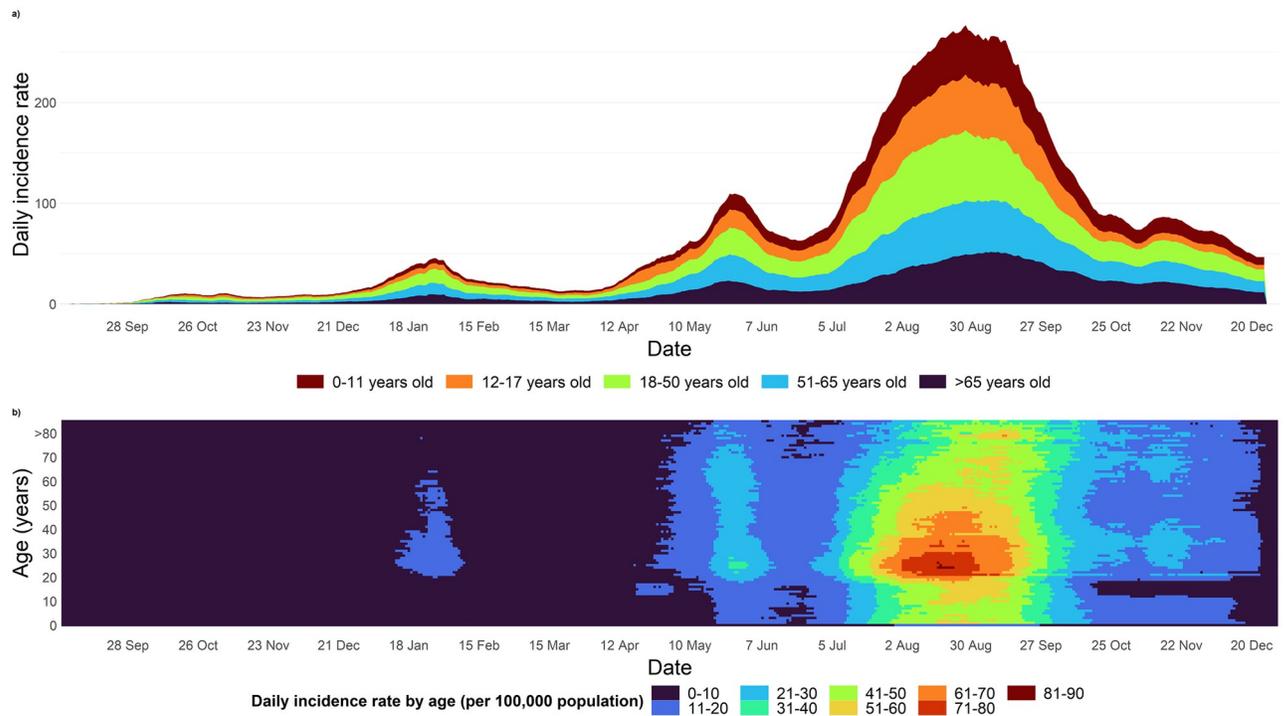
The expected distribution of COVID-19 cases from COVID-19 mortalities assumes: i) the estimated lowest non-zero CFR approximates the unknown IFR, ii) All COVID-19 deaths are reported, iv) Exclusion of brought-in-dead mortalities will account for any excess mortality secondary to poor healthcare accessibility, and v) IFR is constant through time. Hospital protocols during this period likely resulted in high ascertainment of COVID-19 related deaths among all deaths. In addition, all reported deaths outside the hospital were tested for SARS-CoV2 using RT-PCR tests. All adjustment factors below one were replaced with one as it is impossible to estimate fewer cases than reported and were likely due to a small number of observed cases. The expected number of infections was rounded to the nearest thousand to ease interpretation. Reported cases and the estimated infections were tabulated by months. A sensitivity analysis was carried out to quantify the effect of hospital death underreporting on aCFR and case estimation. The Farrington algorithm was utilised to model daily age-specific excess mortality counts between January 2020 and September 2021 in Malaysia using age-specific all-cause mortality data from January 2016 to December 2019 [21–27]. A back-projection model following a Poisson process was carried out to estimate the unobserved age-specific death curve on the day a case was reported positive using an empirically estimated time-lagged age-specific delay distribution from reporting to death [28–30]. Excess mortality counts were then utilised in estimating the aCFR and subsequent degree of underestimation. All analysis was carried out using the “tidyverse”, “zoo”, “epitools”, “prevalence”, “boot”, “amelia” and “surveillance” packages in R 4.3.1.

## Results

A total of 2.31 million cases were reported in Malaysians between 1 March 2020 and 30 December 2021. Incidence trends were quadrimodal with peaks on 26 March 2020 ( $n = 235$ ), 30 January 2021 ( $n = 5,728$ ), 29 May 2021 ( $n = 9,020$ ) and 12 August 2021 ( $n = 21,668$ ). Cases aged 18 and 50 had the highest daily reported case density between August and September 2021 (Fig 1).

A total of 27,711 deaths were reported between 1 March 2020 and 30 December 2021 (Table 1). Mortality trends were quadrimodal with peaks reported on 29 March 2020 ( $n = 7$ ), 18 February 2021 ( $n = 25$ ), 2 June 2021 ( $n = 126$ ), and 9 August 2021 ( $n = 360$ ). The age-specific mortality rate is highest among individuals aged above 60 between August and September 2021 (Fig 2). The distribution of cases by age remained similar across time.

A CFR of 1.0 (95% CI: 0.9–1.0) was reported for Malaysia cumulatively across the study period. Cumulative CFR across the study period is 0.01, 0.02, 0.34, 2.47, and 7.79 in individuals aged between 0–11, 12–17, 18–50, 51–65, and more than 65 years, respectively. The lowest non-zero age specific aCFR utilised for the assumed IFR were 0.002, 0.02, 0.07, 0.85, and 4.3 79 in individuals aged between 0–11, 12–17, 18–50, 51–65, and more than 65 years, respectively.



**Fig 1. Cases in Malaysia between 1 October 2020–30 December 2021 in Malaysia.** a) Daily incidence rate by age categories across time, b) Heatmap of daily incidence rate by age categories across time (Data before 1 September 2020 were removed as deaths were relatively lower to increase visualization quality)

<https://doi.org/10.1371/journal.pntd.0010887.g001>

This assumed IFR approximated the pooled age-specific IFR from the literature. The aCFR trends are unimodal in individuals aged 0–11 (peak in October 2020) and 12–17 years old (peak in July 2021). The aCFR trends are bimodal in individuals aged 51–65 and above 65, respectively, with peaks in July 2020 and 2021. The aCFR trends are trimodal in individuals aged 18–50, with peaks in January 2020, July 2020, and July 2021. The aCFR is consistently highest in individuals above 65 years old (Fig 3).

A total of 9,955,000-cases (95% CI: 6,626,000–18,985,000) cases were estimated to have occurred in Malaysia, corresponding to a prevalence of 32.8%. The relative percentage of cases in 0–11, 12–17, 18–50, 51–65, and more than 65 years age groups were 19.9% (n = 1,982,000, 95% CI: 687,000–3,393,200), 2.4% (n = 236,000, 95% CI: 177,000–3,130,000), 66.1% (n = 6,577,000, 95% CI: 4,766,000–10,525,000), 9.1% (n = 901,000, 95% CI: 783,000–1,063,000), and 2.6% (n = 256,000, 95% CI: 210,000–332,000), respectively. Prevalence of infections was highest in those aged 18–50 years (43.9%), followed by those aged 0–11 years (31.9%), 51–65 years (21.9%), above 65 years (12.6%), and 12–17 years (7.9%) (Table 2).

Overall time trends of the age-specific monthly incidence rate are similar to reported infections with four major peaks. Underestimation is highest between June and September 2021. Underestimation of infections was highest in those aged 0–11 years (Adjustment Factor, AF = 5.5; 95%CI: 1.9–10.9), followed by those aged 18–50 years (AF = 5.0; 95% CI: 3.6–7.9), 51–65 years (AF = 2.9; 95% CI: 2.5–3.5), above 65 years (AF = 1.8; 95% CI: 1.9–10.9), and 12–17 years (AF = 1.3, 95% CI:1.0–17.6) The largest number of estimated cases in individuals aged in 0–11, 12–17, 18–50, 51–65, and more than 65 years is observed in August 2021 (n = 450,000, 95% CI: 150,000–900,000), August 2021 (n = 60,000, 95% CI: 45,000–900,000), August 2021 (n = 2,166,000, 95% CI: 1,557,000–3,476,000), July 2021 (n = 258,000, 95% CI: 223,000–305,000) and August 2021 (n = 52,000, 95% CI: 42,871–68,366), respectively (Fig 4)

Table 1. Monthly COVID-19 reported deaths between March 2020 and December 2021 in Malaysia

	Total	Age groups (%)				
		0–11 years old	12–17 years old	18–50 years old	51–65 years old	>65 years old
<b>Total</b>	27,711	55	42	5,652	8,932	13,030
<b>Mar-20</b>	47	0 (0)	0 (0)	14 (29.8)	10 (21.3)	23 (48.9)
<b>Apr-20</b>	52	0 (0)	0 (0)	3 (5.8)	20 (38.5)	29 (55.8)
<b>May-20</b>	12	0 (0)	0 (0)	3 (25)	5 (41.7)	4 (33.3)
<b>Jun-20</b>	5	0 (0)	0 (0)	0 (0)	2 (40)	3 (60)
<b>Jul-20</b>	3	0 (0)	0 (0)	0 (0)	1 (33.3)	2 (66.7)
<b>Aug-20</b>	3	0 (0)	0 (0)	1 (33.3)	1 (33.3)	1 (33.3)
<b>Sep-20</b>	8	0 (0)	0 (0)	1 (12.5)	1 (12.5)	6 (75)
<b>Oct-20</b>	102	1 (1)	0 (0)	13 (12.7)	36 (35.3)	52 (51)
<b>Nov-20</b>	83	2 (2.4)	1 (1.2)	9 (10.8)	34 (41)	37 (44.6)
<b>Dec-20</b>	121	0 (0)	0 (0)	17 (14)	47 (38.8)	57 (47.1)
<b>Jan-21</b>	357	0 (0)	0 (0)	52 (14.6)	119 (33.3)	186 (52.1)
<b>Feb-21</b>	288	2 (0.7)	0 (0)	29 (10.1)	84 (29.2)	173 (60.1)
<b>Mar-21</b>	124	0 (0)	0 (0)	12 (9.7)	33 (26.6)	79 (63.7)
<b>Apr-21</b>	243	1 (0.4)	0 (0)	39 (16)	61 (25.1)	142 (58.4)
<b>May-21</b>	1,574	1 (0.1)	3 (0.2)	158 (10)	490 (31.1)	922 (58.6)
<b>Jun-21</b>	2,313	1 (0)	1 (0)	335 (14.5)	790 (34.2)	1,186 (51.3)
<b>Jul-21</b>	5,697	6 (0.1)	6 (0.1)	1,321 (23.2)	2,235 (39.2)	2,129 (37.4)
<b>Aug-21</b>	7,368	12 (0.2)	13 (0.2)	2,074 (28.1)	2,468 (33.5)	2,801 (38)
<b>Sep-21</b>	4,860	10 (0.2)	10 (0.2)	1,054 (21.7)	1,322 (27.2)	2,464 (50.7)
<b>Oct-21</b>	2,217	9 (0.4)	7 (0.3)	271 (12.2)	574 (25.9)	1,356 (61.2)
<b>Nov-21</b>	1,347	4 (0.3)	1 (0.1)	144 (10.7)	366 (27.2)	832 (61.8)
<b>Dec-21</b>	887	6 (0.7)	0 (0)	102 (11.5)	233 (26.3)	546 (61.6)

<https://doi.org/10.1371/journal.pntd.0010887.t001>

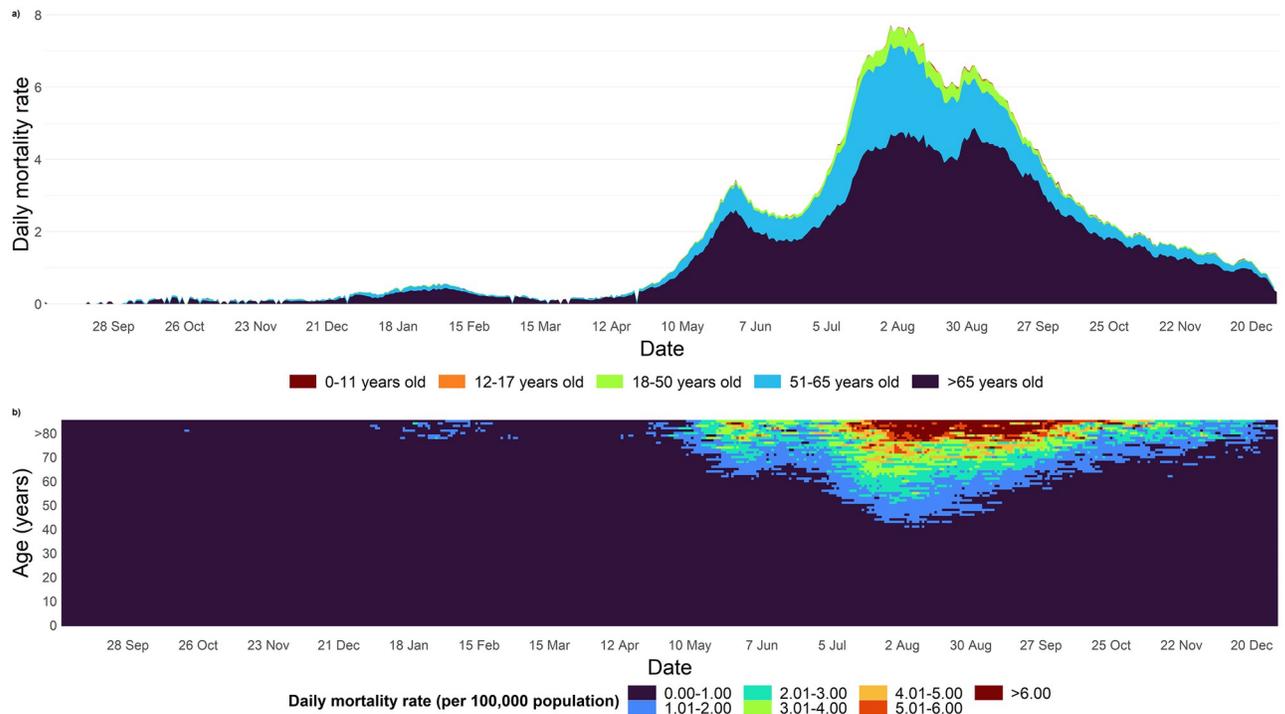
Population prevalence between 1 January 2020 and 30 September 2021 is estimated to be 25.8% (95% CI: 19.6–54.4) corresponding to 7,835,000-cases (95% CI: 6,665,000–10,994,000), when excess-counts based IFRs are utilised. This compared to a population prevalence of 28.8% (95% CI: 22.0–36.3) corresponding to 8,741,000-cases (95% CI: 5,952,000–16,504,000) when reported-deaths based IFRs were utilised over the same period (Table 3).

## Discussion

An estimated 32.8% (9.95 million) of the population are likely to have been infected, and 23.2% of COVID-19 infections were reported between March 2020–December 2021. The adjustment factor for the burden of illness varied by time and age group. These results suggest that community immunity is higher than expected.

The number of infections is estimated to be, on average, 4.3 times (Range = 1–8.8) the number of reported cases in Malaysians, with variations by period and age group. The overall underestimation in Malaysia is comparable to estimations in the United States (US) across time and age strata [31, 32]. Underestimation was estimated to be nine times (90% CI: 4–14) higher than reported cases in another IFR-based adjustment in the US [9]. Another global modelling study reported that the true number of infections was 1.4 to 18 times higher than reported cases with heterogeneity between countries [15]. The degree of underestimation is comparable in many of these settings to the findings observed in our study [6,7,33].

Over the study period, an estimated 32.8% of the population is approximated to have been infected. Estimates of seroprevalence over smaller geographical localities and periods are



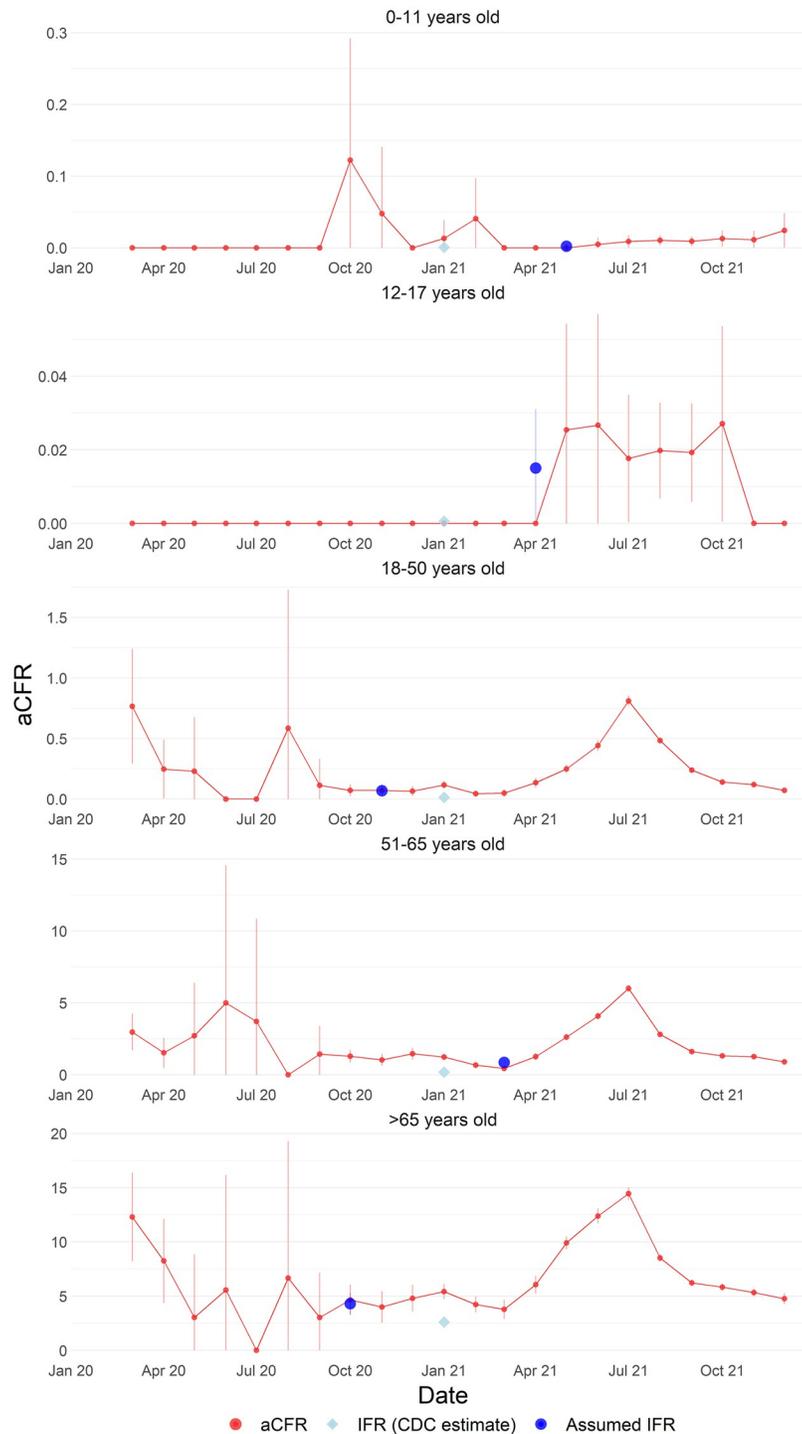
**Fig 2. Deaths in Malaysia between 1<sup>st</sup> October 2020-30<sup>th</sup> December 2021 in Malaysia.** a) Daily deaths by age categories across time, b) Heatmap of daily mortality rate by age categories across time (Data before 1 September 2020 were removed as deaths were relatively lower to increase visualization quality).

<https://doi.org/10.1371/journal.pntd.0010887.g002>

consistent with estimates here [15,34–36]. However, comparisons of the national-level period prevalence estimates were not carried out due to the sparse availability of published literature.

Age-specific prevalences between Malaysia and the United States (US) are comparable except for those above 65 years [31,32]. This may be due to lower aCFR estimations when reported deaths are utilised, as an increase in the number of estimated cases in individuals aged above 65 is observed when excess mortality counts were utilised. Nonetheless, per capita differences remain even when excess mortality counts were utilised. This residual difference of higher prevalence of infection among those above 65 years in the US may be due to the higher proportion of elderly under nursing or institutional care. Malaysia also reports high prevalences in the youngest age groups, with a significant degree of these infections being underestimated [37]. Explanations for this phenomenon include the possibility of preferential testing of older age groups due to their higher risk of severe disease [38]. Changes in this age dynamic have also been reported in countries that have changed testing strategies [39]. Lower susceptibility in younger age groups and a higher propensity to be asymptomatic have also been proposed as an explanation for this dynamic [40]. One final possibility is the use of long-term school closures within many settings, particularly in lower-income countries, which has been suggested to be very effective in reducing transmission [41]. The lack of human-human interaction between younger individuals over a prolonged period, such as has been observed in Malaysia, could potentially amplify this age dynamic.

Published estimates of the true burden of disease have utilized diverse methodologies, and these published findings have primarily been within the global north. These include: i) the use of random population sampling [6,42], ii) seroprevalence studies [43,44], iii) ILI surveillance-based models to estimate prevalence [9,45], iv) crowdsourced data [46], v) testing data based



**Fig 3. Comparison of reported aCFR of COVID-19 in Malaysia by age categories from 1 March 2020–30 December 2021 in Malaysia compared to assumed IFR and literature-derived IFR**

<https://doi.org/10.1371/journal.pntd.0010887.g003>

adjustments using probabilistic bias analysis [33] or post-sampling stratification and reweighting [47], and vi) mortality-data based methods such as mechanistic disease models [15], vii) statistical curve fitting [48], viii) mortality mapping using Bayesian frameworks [9], and ix) combination methods; combining mechanistic models, random sampling, and other data

Table 2. Monthly reported COVID-19 cases and estimated COVID-19 cases between March 2020 and December 2021 in Malaysia

	Total		Age categories									
	Reported cases	Estimated cases (95% CI)	0–11 years old		12–17 years old		18–50 years old		51–65 years old		>65 years old	
	Reported cases	Estimated cases (95% CI)	Reported cases	Estimated cases (95% CI)	Reported cases	Estimated cases (95% CI)	Reported cases	Estimated cases (95% CI)	Reported cases	Estimated cases (95% CI)	Reported cases	Estimated cases (95% CI)
<b>Total</b>	2,308.1	9,955 (6,626, 18,985)	360.5	1,982 (687, 3,932)	177.8	237 (178, 3,130)	1319.6	6,577 (4,767, 10,526)	308.6	901 (784, 1,064)	141.6	257 (211, 333)
<b>March-September 2020</b>	7.9	31 (24, 47)	0.5	0 (0, 0)	0.5	1 (1, 1)	4.7	25 (18, 40)	1.5	4 (4, 5)	0.6	1 (1, 2)
<b>October-December 2020</b>	58.3	210 (107, 385)	6.2	152 (52, 302)	3.8	4 (4, 4)	37.5	38 (38, 60)	8	12 (11, 14)	2.7	3 (3, 4)
<b>January-March 2021</b>	152.8	327 (203, 533)	16.6	154 (54, 304)	9	9 (9, 9)	96.8	128 (107, 179)	21.6	26 (24, 29)	8.8	10 (9, 12)
<b>April-June 2021</b>	351.8	1267 (932, 2,474)	46.8	75 (47, 125)	29.5	46 (30, 606)	200	904 (650, 1,451)	52.5	186 (161, 219)	23	55 (45, 71)
<b>July-September 2021</b>	1,270.1	6906 (4,686, 13,065)	208.2	1,000 (333, 2,000)	109.7	140 (110, 2,100)	725.5	5,048 (3,628, 8,100)	157.5	578 (500, 683)	69.3	141 (114, 182)
<b>October-December 2021</b>	467.3	1213 (675, 2,480)	82.2	600 (200, 1,200)	25.3	37 (25, 410)	255.1	433 (327, 695)	67.5	96 (84, 113)	37.2	47 (39, 61)

Notes: All figures are in thousands (,000's)

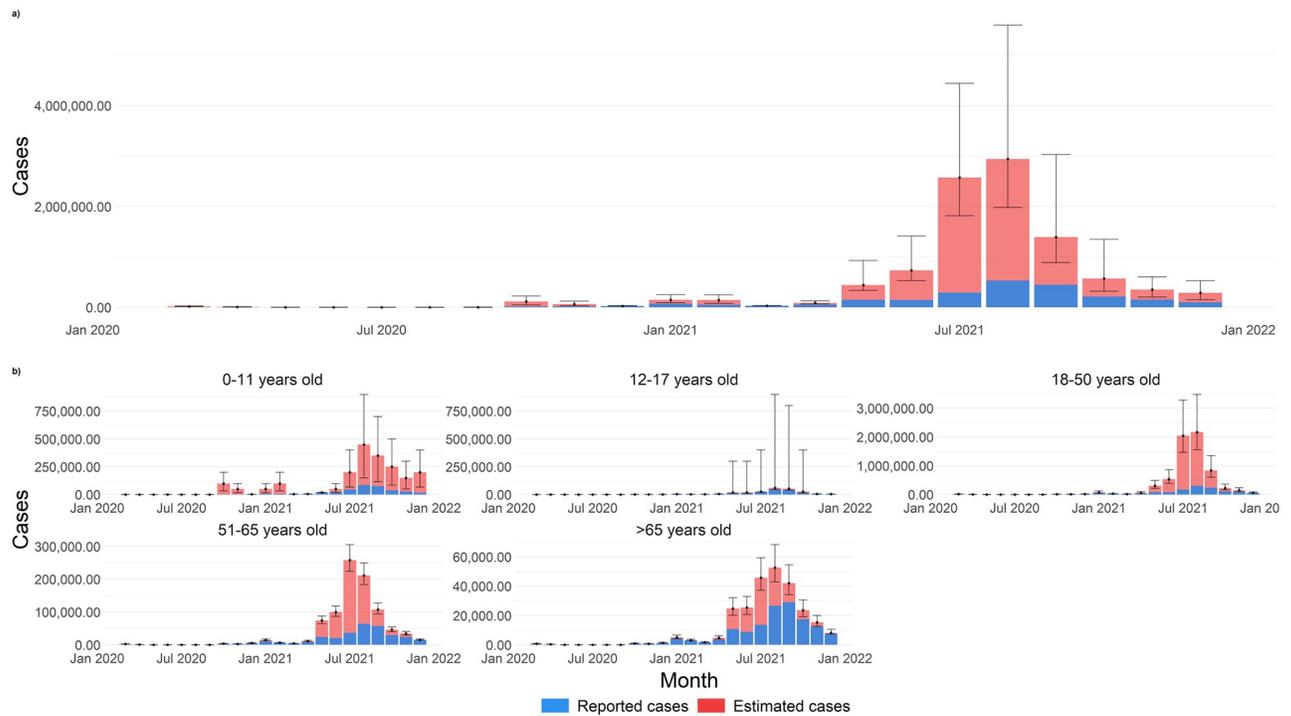
<https://doi.org/10.1371/journal.pntd.0010887.t002>

sources with the IFR [7]. The methodology proposed here is advantageous to the global South, where estimates of the true burden of disease and resources remain scarce.

There are several limitations to this analysis. Firstly, the CFR may deviate away from the unknown IFR by changes in the testing strategy, unreliable vital statistics data, saturation of surveillance systems, the fluidity of disease transmission, virulence, virus genotype, inability to provide adequate care, availability of resources including human capital, heterogeneities in the distribution of medical comorbidities, and changes in immunity due to vaccination [49]. We utilized an approximated IFR estimated from the lowest non-zero, age-specific aCFR as the CFR has been shown to approximate the IFR when the CFR is smaller [50]. These approximated IFR estimates were similar to published pooled IFR estimates [51–54].

Second, we did not quantify the individual effect of various factors that drive the underestimation of COVID-19 incidence in Malaysia. Selection biases, including issues of access, may lead to varying levels of testing over time and space. Testing strategies may also modify the detection of disease within a population. Asymptomatic and pre-symptomatic transmission, variations in transmissibility, and long-tailed incubation periods further complicate the ascertainment of disease [51–54]. Misclassification biases are driven by the accuracy of tests, with studies suggesting nucleic acid amplification (NAA) test sensitivity ranging from 63% to 89% and NAA test specificity of almost 99% [6,9]. These biases limit a surveillance system's ability to ascertain the true burden of COVID-19 infections [45].

Third, the aCFR was estimated using reported deaths instead of the number of excess deaths. The estimated infection may be underestimated using the aCFR compared to excess deaths in countries with poor COVID-19 specific mortality reporting. However, the Malaysian National Death register has a high ascertainment coverage [55,56], and sensitivity analysis



**Fig 4. Comparison of reported COVID-19 cases to estimated COVID-19 infections.** a) cumulatively and b) by age categories from 1 March 2020–30 December 2021 in Malaysia.

<https://doi.org/10.1371/journal.pntd.0010887.g004>

revealed small differences in the prevalence estimate using excess counts (25.8%) and reported deaths (28.8%). Fourth, the incidence density estimates utilised the mid-year population as the population at risk. However, this assumes that infections and vaccinations do not provide complete immunity to the SARS-CoV2 virus. Violation of this assumption may lead to under-estimation of the Incidence density. Finally, we also used a multiplier model approach instead of a mechanistic approach, limiting its utility in forecasting future disease dynamics.

### Conclusion

The characterization of the true burden of disease is essential in developing and implementing policy measures and allocating resources. These estimations highlight a more accurate

**Table 3. Comparison of reported COVID-19 cases with estimated COVID-19 cases based on reported mortality and excess counts between March 2020 and October 2021 in Malaysia**

Age categories, years	Observed Prevalence, %	Estimated Prevalence Model 1, % (95% CI)	Estimated Prevalence Model 2, % (95% CI)
Overall	7.6	28.8 (19.6–54.4)	25.8 (22.–36.3)
0–11	5.8	22.2 (7.8–43.9)	21.7 (15.8–36.9)
12–17	6.0	6.7 (5.2–91.3)	42.6 (27.3–102.3)
18–50	8.8	41. (29.6–65.6)	25.9 (24.5–27.6)
51–65	7.5	19.6 (17.–23.1)	18.8 (17.–21.2)
>65	6.9	10.3 (8.4–13.3)	27.7 (24.6–31.6)

Notes: Model 1 utilises reported-deaths based approximations of the IFR Model 2 utilises excess-mortality counts based approximations of the IFR  
All figures are in thousands (,000's)

<https://doi.org/10.1371/journal.pntd.0010887.t003>

infection burden in Malaysia. It provides the first national-level estimates of prevalence in Malaysia that are adjusted for underdiagnosis. The higher underestimation of infections during April-September 2021 coincided with sustained higher community transmission and higher healthcare utilization.

Naturally acquired community immunity is still low but is likely to increase secondary to an Omicron-fuelled infection surge. Booster doses may further hasten an equilibrium between the SARS-CoV-2 virus and its human host. Such an equilibrium should mark the start of an endemic state. Future variants may upend this equilibrium and necessitate periodic mitigation of disease transmission. Population estimates of the infection burden are critical to determine the need for booster doses and public health measures.

## Supporting information

**S1 Appendix. Adjudication of deaths in Malaysia.**  
(DOCX)

## Acknowledgments

The authors would like to thank the Director-General of Health Malaysia for his permission to publish this article.

## Author Contributions

**Conceptualization:** Vivek Jason Jayaraj, Chiu-Wan Ng, Sanjay Rampal.

**Data curation:** Maheshwara Rao Appannan.

**Formal analysis:** Vivek Jason Jayaraj.

**Funding acquisition:** Awang Bulgiba, Sanjay Rampal.

**Investigation:** Sanjay Rampal.

**Methodology:** Vivek Jason Jayaraj, Chiu-Wan Ng, Sanjay Rampal.

**Resources:** Awang Bulgiba, Maheshwara Rao Appannan.

**Software:** Vivek Jason Jayaraj.

**Supervision:** Chiu-Wan Ng, Awang Bulgiba, Sanjay Rampal.

**Validation:** Vivek Jason Jayaraj, Maheshwara Rao Appannan.

**Visualization:** Vivek Jason Jayaraj.

**Writing – original draft:** Vivek Jason Jayaraj, Sanjay Rampal.

**Writing – review & editing:** Vivek Jason Jayaraj, Chiu-Wan Ng, Awang Bulgiba, Maheshwara Rao Appannan, Sanjay Rampal.

## References

1. World Health Organisation. WHO Coronavirus (COVID-19) Dashboard. 13 July 2021 [cited 15 July 2021]. [https://covid19.who.int/?gclid=Cj0KCQjwub-HBhCyARIsAPctr7wt4KGje6JUmyWQljSMQWGhr8jvXTZSdsSZR5ay9OAp2p1uOWMFJb0aAmAXEALw\\_wcB](https://covid19.who.int/?gclid=Cj0KCQjwub-HBhCyARIsAPctr7wt4KGje6JUmyWQljSMQWGhr8jvXTZSdsSZR5ay9OAp2p1uOWMFJb0aAmAXEALw_wcB)
2. Nicola M, Alsafi Z, Sohrabi C, Kerwan A, Al-Jabir A, Iosifidis C, et al. The Socio-Economic Implications of the Coronavirus and COVID-19 Pandemic: A Review. *Int J Surg*. 2020 [cited 27 April 2020]. <https://doi.org/10.1016/j.ijsu.2020.04.018> PMID: 32305533

3. Department of Social and Preventive Medicine. COVID-19 Epidemiology for Malaysia. In: COVID-19 Epidemiology for Malaysia [Internet]. 2020 [cited 19 Apr 2021]. <https://spm.um.edu.my/knowledge-centre/covid19-epid-live/>
4. Elengoe A. COVID-19 Outbreak in Malaysia. *Osong Public Health Res Perspect*. 2020; 11: 93–100.
5. Rampal L, Boon Seng L. Malaysia's third COVID-19 wave—a paradigm shift required. *Medical Journal of Malaysia*. 2020; 76: 1–4.
6. Sahu I, Whittaker AB. Obtaining Prevalence Estimates of Coronavirus Disease 2019: A Model to Inform Decision-Making. *Am J Epidemiol*. 2021; 190: 1681–1688. <https://doi.org/10.1093/aje/kwab079> PMID: 33831172
7. Irons NJ, Raftery AE. Estimating SARS-CoV-2 infections from deaths, confirmed cases, tests, and random surveys. *Proc Natl Acad Sci U S A*. 2021; 118: e2103272118. Available: <https://www.pnas.org/content/118/31/e2103272118> <https://doi.org/10.1073/pnas.2103272118> PMID: 34312227
8. Yarmol-Matusiak EA, Cipriano LE, Stranges S. A comparison of COVID-19 epidemiological indicators in Sweden, Norway, Denmark, and Finland: <https://doi.org/10.1177/1403494820980264>. 2021; 49: 69–78. <https://doi.org/10.1177/1403494820980264> PMID: 33413051
9. Lu FS, Nguyen AT, Link NB, Davis JT, Chinazzi M, Xiong X, et al. Estimating the Cumulative Incidence of COVID-19 in the United States Using Four Complementary Approaches. *medRxiv*. 2020. <https://doi.org/10.1101/2020.04.18.20070821> PMID: 32587997
10. Wang Y, Xu C, Yao S, Zhao Y, Li Y, Wang L, et al. Estimating the Prevalence and Mortality of Coronavirus Disease 2019 (COVID-19) in the USA, the UK, Russia, and India. *Infect Drug Resist*. 2020; 13: 3335–3350. <https://doi.org/10.2147/IDR.S265292> PMID: 33061481
11. Ministry of Health Malaysia. Official data on the COVID-19 epidemic in Malaysia. In: COVID-19 Public MOH Malaysia Github [Internet]. 27 Dec 2021 [cited 27 Dec 2021]. <https://github.com/MoH-Malaysia/covid19-public>
12. Department of Statistics Malaysia. Population Distribution and Basic Demographic Characteristics. 2015 May. [https://www.dosm.gov.my/v1/index.php?r=column/cthem&menu\\_id=L0pheU43NWJwRWVVSZklWdzQ4TlhUUT09&bul\\_id=MDMxdHJjWTK1SjFzTzNkRXZzcVZjdz09](https://www.dosm.gov.my/v1/index.php?r=column/cthem&menu_id=L0pheU43NWJwRWVVSZklWdzQ4TlhUUT09&bul_id=MDMxdHJjWTK1SjFzTzNkRXZzcVZjdz09)
13. Lee H-A, Leng KY. Counting Migrant Workers in Malaysia: A Needlessly Persisting Conundrum. *Kuala Lumpur*; 2018 Apr. [https://www.iseas.edu.sg/images/pdf/ISEAS\\_Perspective\\_2018\\_25@50.pdf](https://www.iseas.edu.sg/images/pdf/ISEAS_Perspective_2018_25@50.pdf)
14. Devleeschauwer Brecht, Torgerson Paul, Charlier Johannes, Levecke Bruno, Praet Nicolas, Roelandt Sophie, et al. Package “prevalence”: Tools for Prevalence Assessment Studies. CRAN; 2015. pp. 1–35. Available: <http://prevalence.cbra.be/>
15. Russell TW, Golding N, Hellewell J, Abbott S, Wright L, Pearson CAB, et al. Reconstructing the early global dynamics of under-ascertained COVID-19 cases and infections. *BMC Medicine* 2020 18:1. 2020; 18: 1–9. <https://doi.org/10.1186/s12916-020-01790-9> PMID: 33087179
16. Baud David, Qi Xiaolong, Karin Nielsen-Saines Didier Musso, Pomar Léo, Favre Guillaume. Real estimates of mortality following COVID-19 infection. *Lancet Infect Dis*. 2020; 323: 1–1. [https://doi.org/10.1016/S1473-3099\(20\)30195-X](https://doi.org/10.1016/S1473-3099(20)30195-X) PMID: 32171390
17. Shim E. Delay-Adjusted Age-Specific COVID-19 Case Fatality Rates in a High Testing Setting: South Korea, February 2020 to February 2021. *International Journal of Environmental Research and Public Health* 2021, Vol 18, Page 5053. 2021; 18: 5053. <https://doi.org/10.3390/ijerph18105053> PMID: 34064703
18. Rothman JE, Eidelberg D, Rothman SL, Holford TR, Rothman DL. Analysis of the time course of COVID-19 cases and deaths from countries with extensive testing allows accurate early estimates of the age specific symptomatic CFR values. *PLoS One*. 2021; 16: e0253843. <https://doi.org/10.1371/journal.pone.0253843> PMID: 34407073
19. Levin AT, Cochran K, Walsh S, Meyerowitz-Katz G. Assessing the Age Specificity of Infection Fatality Rates for Covid-19: Systematic Review, Meta-Analysis, & Public Policy Implications. 2020. <https://papers.ssrn.com/abstract=3661089>
20. Center for Disease Control. COVID-19 Pandemic Planning Scenarios. 3 Autumn 2021 [cited 5 Dec 2021]. <https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html>
21. Farrington C. P., Andrews N. J., Beale A. D., Catchpole M. A. A Statistical Algorithm for the Early Detection of Outbreaks of Infectious Disease. *Royal Statistical Society*. 1996; 159: 547–563.
22. Höhle M, Mazick A. Aberration Detection in R Illustrated by Danish Mortality Monitoring. *Biosurveillance*. 2010.
23. Noufaily A, Enki DG, Farrington P, Garthwaite P, Andrews N, Charlett A. An improved algorithm for outbreak detection in multiple surveillance systems. *Stat Med*. 2013; 32: 1206–1222. <https://doi.org/10.1002/sim.5595> PMID: 22941770

24. al Wahaibi A, Al-Maani A, Alyaquobi F, al Harthy K, Al-Jardani A, al Rawahi B, et al. Effects of COVID-19 on mortality: A 5-year population-based study in Oman. 2021 [cited 19 Apr 2021]. 10.1016/j.ijid.2020.12.054
25. Statistics Canada. Estimation of excess mortality. 2020. [https://www.statcan.gc.ca/eng/statistical-programs/document/3233\\_D5\\_V1](https://www.statcan.gc.ca/eng/statistical-programs/document/3233_D5_V1)
26. Liu J, Zhang L, Yan Y, Zhou Y, Yin P, Qi J, et al. Excess mortality in Wuhan city and other parts of China during the three months of the covid-19 outbreak: Findings from nationwide mortality registries. *The BMJ*. 2021;372. <https://doi.org/10.1136/bmj.n415> PMID: 33627311
27. Kawashima T, Nomura S, Tanoue Y, Yoneoka D, Eguchi A, Ng CFS, et al. Excess all-cause deaths during coronavirus disease pandemic, Japan, January–May 2020. *Emerging Infectious Diseases*. Centers for Disease Control and Prevention (CDC); 2021. pp. 789–795. <https://doi.org/10.3201/eid2703.203925> PMID: 33622468
28. Marschner IC. Back-projection of COVID-19 diagnosis counts to assess infection incidence and control measures: analysis of Australian data. *Epidemiol Infect*. 2020;148. <https://doi.org/10.1017/S0950268820001065> PMID: 32418559
29. Yip PSF, Lam KF, Xu Y, Chau PH, Xu J, Chang W, et al. Reconstruction of the Infection Curve for SARS Epidemic in Beijing, China Using a Back-Projection Method. *Communications in Statistics—Simulation and Computation*. 2008; 37: 425–433. <https://doi.org/10.1080/03610910701792562>
30. Becker NG, Watson LF, Carlin JB. A method of non-parametric back-projection and its application to AIDS data. *Stat Med*. 1991; 10: 1527–1542. <https://doi.org/10.1002/sim.4780101005> PMID: 1947509
31. Center for Disease Control. Estimated COVID-19 Burden. 16 Nov 2021 [cited 10 Jan 2022]. <https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/burden.html>
32. Reese H, Iuliano AD, Patel NN, Garg S, Kim L, Silk BJ, et al. Estimated Incidence of Coronavirus Disease 2019 (COVID-19) Illness and Hospitalization—United States, February–September 2020. *Clinical Infectious Diseases*. 2021; 72: e1010–e1017. <https://doi.org/10.1093/cid/ciaa1780> PMID: 33237993
33. Wu SL, Mertens AN, Crider YS, Nguyen A, Pokpongkiat NN, Djajadi S, et al. Substantial underestimation of SARS-CoV-2 infection in the United States. *Nature Communications* 2020 11:1. 2020; 11: 1–10. <https://doi.org/10.1038/s41467-020-18272-4> PMID: 32908126
34. Sam IC, Chong YM, Tan CW, Chan YF. Low postpandemic wave SARS-CoV-2 seroprevalence in Kuala Lumpur and Selangor, Malaysia. *J Med Virol*. 2021; 93: 647–648. <https://doi.org/10.1002/jmv.26426> PMID: 32790206
35. Sam IC, Chong YM, Abdullah A, Fu JYL, Hasan MS, Jamaluddin FH, et al. Changing predominant SARS-CoV-2 lineages drives successive COVID-19 waves in Malaysia, February 2020 to March 2021. *J Med Virol*. 2021 [cited 13 Jan 2022]. <https://doi.org/10.1002/jmv.27441> PMID: 34757638
36. Sam I, Chong YM, Tan CW, Chan YF. Low postpandemic wave SARS-CoV-2 seroprevalence in Kuala Lumpur and Selangor, Malaysia. *J Med Virol*. 2020; jmv.26426. <https://doi.org/10.1002/jmv.26426> PMID: 32790206
37. Bauer P, Brugger J, König F, Posch M. An international comparison of age and sex dependency of COVID-19 deaths in 2020: a descriptive analysis. *Scientific Reports* 2021 11:1. 2021; 11: 1–11. <https://doi.org/10.1038/s41598-021-97711-8> PMID: 34580322
38. Sudharsanan N, Didzun O, Bärnighausen T, Geldsetzer P. The Contribution of the Age Distribution of Cases to COVID-19 Case Fatality Across Countries: A 9-Country Demographic Study. *Ann Intern Med*. 2020; 173: 714–720. <https://doi.org/10.7326/M20-2973> PMID: 32698605
39. Malmgren J, Guo B, Kaplan HG. Continued proportional age shift of confirmed positive COVID-19 incidence over time to children and young adults: Washington State March–August 2020. *PLoS One*. 2021; 16: e0243042. <https://doi.org/10.1371/journal.pone.0243042> PMID: 33760808
40. Davies NG, Klepac P, Liu Y, Prem K, Jit M, Eggo RM. Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nature Medicine* 2020 26:8. 2020; 26: 1205–1211. <https://doi.org/10.1038/s41591-020-0962-9> PMID: 32546824
41. Liu Y, Morgenstern C, Kelly J, Lowe R, Jit M. The impact of non-pharmaceutical interventions on SARS-CoV-2 transmission across 130 countries and territories. *BMC Medicine* 2021 19:1. 2021; 19: 1–12. <https://doi.org/10.1186/S12916-020-01872-8> PMID: 33541353
42. Gudbjartsson DF, Helgason A, Jonsson H, Magnusson OT, Melsted P, Norddahl GL, et al. Spread of SARS-CoV-2 in the Icelandic Population. 2020; 382: 2302–2315. <https://doi.org/10.1056/NEJM0A2006100> PMID: 32289214
43. Tanne JH. Covid-19: US cases are greatly underestimated, seroprevalence studies suggest. *BMJ*. 2020; 370: m2988. <https://doi.org/10.1136/bmj.m2988> PMID: 32709608
44. Mohanan M, Malani A, Krishnan K, Acharya A. Prevalence of SARS-CoV-2 in Karnataka, India. *JAMA*. 2021; 325: 1001–1003. <https://doi.org/10.1001/jama.2021.0332> PMID: 33538774

45. Yue M, Clapham HE, Cook AR. Estimating the Size of a COVID-19 Epidemic from Surveillance Systems. *Epidemiology*. 2020; 31: 567–569. <https://doi.org/10.1097/EDE.0000000000001202> PMID: 32324625
46. Garcia-Agundez A, Ojo O, Hernández-Roig HA, Baquero C, Frey D, Georgiou C, et al. Estimating the COVID-19 Prevalence in Spain With Indirect Reporting via Open Surveys. *Front Public Health*. 2021; 0: 306. <https://doi.org/10.3389/fpubh.2021.658544> PMID: 33898383
47. Bassi F, Arbia G, Falorsi PD. Observed and estimated prevalence of Covid-19 in Italy: How to estimate the total cases from medical swabs data. *Science of The Total Environment*. 2021; 764: 142799. <https://doi.org/10.1016/j.scitotenv.2020.142799> PMID: 33066965
48. Flaxman S, Mishra S, Gandy A, Unwin JT, Coupland H, Mellan TA, et al. Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries. Imperial College London. 2020 [cited 4 May 2020]. <https://doi.org/10.25561/77731>
49. Shen C, VanGennep D, Siegenfeld AF, Bar-Yam Y. Unraveling the flaws of estimates of the infection fatality rate for COVID-19. *J Travel Med*. 2021; 28: 1–3. <https://doi.org/10.1093/jtm/taaa239> PMID: 33398337
50. Staerk C, Wistuba T, Mayr A. Estimating effective infection fatality rates during the course of the COVID-19 pandemic in Germany. *BMC Public Health*. 2021; 21: 1–9.
51. Levin AT, Hanage WP, Owusu-Boaitey N, Cochran KB, Walsh SP, Meyerowitz-Katz G. Assessing the age specificity of infection fatality rates for COVID-19: systematic review, meta-analysis, and public policy implications. *Eur J Epidemiol*. 2020; 35: 1123–1138. <https://doi.org/10.1007/s10654-020-00698-1> PMID: 33289900
52. Perez-Saez J, Lauer SA, Kaiser L, Regard S, Delaporte E, Guessous I, et al. Serology-informed estimates of SARS-CoV-2 infection fatality risk in Geneva, Switzerland. *Lancet Infect Dis*. 2021; 21: e69–e70. [https://doi.org/10.1016/S1473-3099\(20\)30584-3](https://doi.org/10.1016/S1473-3099(20)30584-3) PMID: 32679085
53. Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect Dis*. 2020; 20: 669–677. [https://doi.org/10.1016/S1473-3099\(20\)30243-7](https://doi.org/10.1016/S1473-3099(20)30243-7) PMID: 32240634
54. Lu FS, Nguyen AT, Link NB, Davis JT, Chinazzi M, Xiong X, et al. Estimating the Cumulative Incidence of COVID-19 in the United States Using Four Complementary Approaches. *medRxiv*. 2020 [cited 30 Aug 2021]. <https://doi.org/10.1101/2020.04.18.20070821> PMID: 32587997
55. Rao C, Omar MA, Ganapathy SS, Tamin NSI. Strengthening Mortality Statistics for Health Programs in Malaysia: Lessons from the Field. *Dr Sulaiman Al Habib Medical Journal*. 2019; 1: 52.
56. Lai SL, Tey NP. Deficiency in civil registration and vital statistics reporting in remote areas: the case of Sabah, Malaysia. *Genus*. 2021; 77: 1–19.