1114609

What is the disease burden of COVID-19 by age-group and occupation category?

Analysis as of 15 Feb 2021

Disease burden in disability-adjusted life-years (DALYs) is already routinely calculated, based on notified cases and deaths in OSIRIS and hospital admission and ICU admission data provided by NICE. In this report we extend these previous estimates [1,2] to account for underascertainment in notifications, and we estimate disease burden as of 31 Dec 2020. We stratify disease burden estimates by age-group and by occupation category, and present both absolute DALYs and DALYs/100,000 persons (a measure of relative burden, that adjusts for denominator population size).

For the per-capita DALY estimates stratified by occupation category, estimates of the denominator – the total number of persons in each category (from CBS), stratified by age-group – are required. As the available information from CBS [3] contains the number of persons in each occupation per 10-year age-group (15-25, ... 65-75) only, assumptions were required to map the 10-year denominator age-groups to 5-year age-groups (see below).

Burden stratified by age-group

For the methodology used for disease burden estimation, see [1,2]. Briefly, the clinical pathway progression is as follows: confirmed SARS-CoV-2 positive cases who develop mild symptomatic COVID-19 can progress to moderate disease (requiring hospital admission), and then to severe disease (requiring ICU admission). Death due to COVID-19 is assumed possible following any of these three disease states (see Table 1). We carry out two sets of analyses: for the period until 24 Sept 2020 (representing the period covered by the PICO3 serosurvey, and coincidentally before the second wave fully took off: 2145 positive cases were notified – and more relevant for disease burden, 7 COVID-19 deaths – on 24 Sept), and for the period until 31 Dec 2020.

Analysis period until 24 September 2020. The cumulative incidence of Mild infections was based on age-group specific seroprevalence from the PICO3 study conducted between 22 Sept and 23 Nov 2020 (the 'index' date of 25 Sept was selected as 90% of participants responded by 9 Oct, with 14 days assumed for development of an IgG response), weighted to adjust for survey representativeness and seroreversion (Fig. 1a) and the estimated age-group specific symptomatic proportion. The latter was derived using PICO2 study data (collected in June/July 2020), where 'symptomatic' is defined according to the ECDC case definition (fever and/or cough and/or shortness of breath and/or loss of smell/taste), and where the observed proportion of seropositive persons reporting symptoms is adjusted for reported symptom occurrence among seronegative persons; for further details see [4]. The age-aggregated symptomatic proportion using this approach and PICO2 data was estimated at 63%).

The cumulative incidence of infection, and of symptomatic infection (SI), with SARS-CoV-2 was estimated at 872,700 and 323,900, respectively (Fig. 4a). This entails that overall ascertainment of estimated cumulative SI incidence by the total number of OSIRIS notifications in this period (n=107,662) was 33%. DALY estimates, as calculated using the approach detailed in [1,2], are shown in Fig. 4b. Very little of the total COVID-19 disease burden (60,900 DALYs; 95% CI: 59,100–62,700) was contributed by morbidity (i.e., YLD accounted for approximately 1.0% of the total DALYs). The highest absolute burden in a given age-group was observed for 75-79 years.

Analysis period until 31 Dec 2020. For this period, as no seroprevalence data an alternative (provisional) approach to estimating cumulative SI incidence for the period 25 Sept through 31 Dec 2020 was required. We pooled nine estimates of the ascertainment of all infected persons by notified cases based on population-level survey data from England (nine occasions when members of a community cohort underwent virological testing, conducted by the ONS between 18-24 Sept and 22-28 Nov 2020). Using these data entailed making two strong assumptions: (i) testing policy, availability of tests, and willingness to be testing in England is broadly similar to the Netherlands over this period, and (ii) ascertainment does not vary age. The pooled age-independent ascertainment estimate is 38.7% (95% CI: 36.1–41.4%). We then estimated cumulative infection incidence for the period 25 Sept through 31 Dec 2020 by synthesising estimates using this approach (while adjusting precision of estimated ascertainment for multiple age-groups) with those from a second approach (for age-groups 30-34 and older only): multiplying age-group specific cumulative hospital admission ratios by the cumulative incidence as of 24 Sept 2020. The second approach is the same general method used in for estimating the prevalent number of infectious persons that is presented on the coronavirus dashboard.

After integration of the estimated symptomatic proportion, we estimated a cumulative SI incidence of 950,600 (95% CI: 897,100–1,009,600) between 27 Feb and 31 Dec (Fig. 1b, Fig. 4c). The cumulative infection incidence over this full analysis period was estimated at 2,571,400 (95% CI: 2,444,900–2,710,700), or 14.8% (95% CI: 14.0-15.6%) of the total population.

The estimated age-aggregated ascertainment of cumulative SI incidence and cumulative infection incidence by the cumulative number of OSIRIS notified cases (*n*=808,791) over this period was 85% (95% CI: 80-90%) and 31% (95% CI: 30-33%), respectively. The SI case ascertainment figure of 85% appears unrealistically high; however there are two factors that should be considered. First, the proportion of all infections that are symptomatic was estimated with respect to the ECDC case definition, which is stricter than the criterion for testing by the GGD (any of a list of symptoms, which includes very mild and non-specific respiratory symptoms). Second, from 1 Dec 2020 testing was expanded to include asymptomatic persons who had travelled abroad or were identifed via contact tracing; an unknown, though likely small, percentage of positive results were recorded among asymptomatic testees.

Total DALYs in this period were estimated at 106,900 (95% CI: 104,600–109,300), of which 1.6% were contributed by YLD (Fig. 4d).

The age-specific burden estimates for the full period (27 Feb through 31 Dec 2020) are quite similar to those for the pre-second wave period (27 Feb through 24 Sept), except for a much higher estimated cumulative SI incidence and a correspondingly higher YLD in the full period. However, the estimated total disease burden in the full period did not increase in proportion to the increase in cumulative incidence of infection (2,571,000 vs 873,000 persons; i.e., 2.9 times the estimated total number of infections but only 1.75 times the total burden), because the YLL contribution to disease burden was lower since the first wave (in part because of the somewhat younger age-distribution of infections and in part because of improvements in patient management and care).

Burden stratified by occupation category

Methods

We first defined occupation categories according to notified case data in OSIRIS (Table 2), and then plotted the distribution over occupation category, also stratified by (fairly broad) age-group (Fig. 2). Estimation of the occupation category denominators required the set of occupation categories in OSIRIS to be mapped to the '4 digit code' categories used by CBS. A perfect match was not possible; in particular for the category 'Other contact professions' (see Table 2 for the adopted mapping).

Two definitions of the period for defining the distribution of notified over occupation category are relevant: (i) the period from 27 Feb 2020 (the date of the first notified case) through 31 Dec 2020 , and (ii) the period with 'open society' and non-priority testing policy (1 Jun to 20 Sep 2020). Note that the 'full period' definition contains periods in which there was restricted testing (i.e., before 1 June priority was given to severe/hospitalised cases) and/or priority testing for certain occupations, such as healthcare workers and the education sector, and so the distribution of occupation categories among notified cases is influenced by access to testing.

The distribution over occupation category during periods of 'open society' and non-priority testing policy will reflect the burden due to potentially higher transmission risks for certain categories, e.g., catering (restaurant/cafe/bar) occupations. The same distribution if calculated from only those cases notified only during those periods of time in which practicing of certain occupations was drastically limited through lockdown measures, such as catering and contact professions, will reflect potentially lower transmission risks for the affected occupations (see Fig. 3b, which suggests a slightly higher proportion of cases for the Catering category when restaurants were generally open compared to the 'full period').

In the main analysis, we apply the occupation category distribution based on notification data (from OSIRIS) during the full period (thus this also reflects the impacts of testing policy, closure of certain parts of the economy, various (sector-specific) preventative measures in place, and the periods in which lockdown was imposed) (Fig. 2a) to estimate the disease burden stratified occupation category. A limitation of this analysis is the assumption that the occupation provided in a notified case's OSIRIS record applied throughout the analysis period (i.e., person was not (temporarily) inactive in their occupation, and did not become unemployed). Because a substantial proportion of notifications had occupation 'Not known'; we apply simple univariate imputation to re-distribute the Not known category among the observed occupation categories.

As an additional analysis, we also apply the occupation category distribution based on notifications (OSIRIS) made during the 'open society' period (Fig. 2b) to estimate the disease burden per occupation category. Note that by applying the occupation distribution derived from the 'open society' period to periods in which strict measures were in place (some occupations could not be practiced; for others, contact patterns and ensuing transmission risk might be quite different), we effectively attempt to estimate the distribution of disease burden over occupation category that *would have been* observed, assuming that the measures were *not* in place. This is an imperfect counterfactual; we recognise that the proportion in category 'education' will not be fully representative of the term-time situation with in-person teaching, due to the (partial) continuation of online teaching after 1st June 2020, and the school vacation period. As well, the proportions in all categories will reflect the effect of ongoing safety measures in place since 1 June 2020 (e.g., Catering: spacing of restaurant tables; Transportation: no contact with bus drivers).

To estimate DALYs stratified by occupation category, we simply apply the occupation category distribution that had been determined on the basis of 10-year age-groups to the narrower, 5-year age-groups used to assign OSIRIS cases to occupation category; e.g. the distribution inferred for 25-34 years is applied to both 25-29 and 30-34 years, and the assumed denominator population for these two 5-year age-groups is the 10-year denominator population weighted according the national population sizes of the 25-29 and 30-34 years age-groups. Importantly, the occupation distribution is calculated separately within each age-group and applied to the DALYs within each age-group.

All burden estimates are restricted to the 'work-eligible' age range (defined as age 20 through 69 years).

Summary of results (analysis period to 31 Dec 2020)

While the absolute burden is greatest for the 'non-working' occupation category (consisting of retired persons, employment seekers and presumably students), largely because of the much higher mortality burden among older aged retirees (Fig 5a, Fig 8a), when the size of the occupation denominator is taken into account (i.e. the DALY/100,000 measure, aggregating over age), the category Healthcare appear to bear a disproportionally high relative burden (Fig. 8b). The higher relative burden for this category holds true also when calculated separately per age-group, as the relative disease burden is notably higher than seen for other occupations starting from age-group 45-49 (Fig. 7). The higher relative burden among healthcare workers is attributable to the relatively high cumulative incidence of symptomatic infection seen across all age-groups for this category (Fig. 6), which presumably reflects a combination of increased workplace exposure and a higher likelihood of being tested (at least during part of the year).

Note that the relative disease burden for a given occupation category, as estimated for the full analysis period, is not necessarily indicative of the recent burden; for instance, widespread availability of PPE and other risk-reducing measures may mean that the proportion of burden experienced by healthcare workers over the last half of the year is now much reduced.

In addition, this approach does not take into account possible variation in the risk of severe disease and/or mortality by occupation, because the occupation distribution (per age-group) is applied to the total burden (for that age-group). For instance, if healthcare workers have better underlying health and therefore better prognosis compared with other occupations, then both the absolute and relative disease burden will have been overestimated for this group. Unfortunately, applying a separate occupation category distribution as observed among fatal cases for the calculation of YLL (which could address this issue) is not viable, due to relatively high level of missingness of occupation information among working age fatal cases (Fig. 3a).

The supplementary analysis, in which the distribution of positive cases over occupation category was determined during the 'open society' period, showed similar patterns of absolute and relative burden (Fig. 9a, Fig. 9b), except that Other contact professions now indicated the second higher relative burden of all occupation categories.

Overall summary

The total disease burden for the period until 31 Dec 2020 presented here is known to underestimate the true burden, mainly because of the underascertaiment of mortality due to COVID-19 in OSIRIS, but also because post-acute health outcomes are not yet included. Nevertheless, we can draw several useful conclusions from this exercise. COVID-19 disease burden is overwhelmingly determined by premature mortality (>98% of DALYs). The absolute disease burden (in DALYs) grew more slowly between the first and second SARS-CoV-2 waves in proportion to the estimate cumulative incidence of infection. This is due to improvements in COVID-19 patient prognosis, but also to changes in the age-distribution of infected persons, with consequence impact on risk of severe or fatal outcomes. Using the relative disease burden measure (DALYs per 100,000 population), we can compare the per-capita burden between different strata of the population. Thus the (age-aggregated) burden experienced by healthcare workers (approximately 660 DALYs per 100,000; Fig. 8b) is an order of magnitude lower than the burden experienced by the oldest segment of the population (e.g., approximately 6000 DALYs per 100,000 for the age-groups 85-89 years and older; Fig. 4e).

References

[1]. Lagerweij G, et al. Ziektelast COVID-19. In *Staat van Infectieziekten in Nederland, 2019*. Bilthoven: RIVM, 2021. To appear.

[2]. Volksgezondheid Toekomst Verkenning (VTV). Verder kijken dan corona, over de toekomst van onze gezondheid. Bilthoven: RIVM, 27 Nov 2020. https://www.volksgezondheidtoekomstverkenning.nl/c-vtv/covid-19/ziekte.

[3]. Central Bureau for Statistics (CBS). Statline. Werkzame beroepsbevolking; beroep. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/82808NED/table. Accessed 16 Nov 2020.

[4]. McDonald SA, Miura F, Vos E, et al. Estimating the asymptomatic proportion of SARS-CoV-2 infection in the general population: Analysis of a nationwide serosurvey in the Netherlands. Manuscript submitted for publication.

1114609

Table 1. Summary of data sources, DALY parameters, and other decisions, for analysis period 27 Feb through 31 Dec 2020 (note that the cumulative incidence of Mild cases in the period 25 Sep through 31 Dec 2020 was estimated differently, text).

Parameter/data	
Analysis period	27 Feb 2020 t/m 31 Dec 2020
Life expectancy	5-year bins, determined based on 1-year CBS values for 2019[a]. The interpolated 1-year LE for the exact midpoint of each age-category is used (eg. LE(82.5) for LE(80-84); LE(97.5) for LE(95+))
Incidence **Mild** cases	Estimated symptomatic infection cases derived from seroprevalence data (from PICO3) and symptomatic proportion (derived using PICO2), with estimated 95% uncertainty interval [Beta distribution].
Underreporting adjustment **Mild**	N/A
Disability duration **Mild**	10 days
Incidence **Moderate** cases	Cumulative NICE non-ICU hospital admissions, per 5-year age-group. Assumed Poisson distributed.
Underreporting adjustment **Moderate**	1.10 (1.06-1.18) [Uniform distribution]
Disability duration **Moderate**	8 days
Incidence **Severe** cases	Cumulative NICE ICU admissions, per 5-year age-group. Assumed Poisson distributed.
Underreporting adjustment **Severe**	1.0
Disability duration **Severe**	19 days (NB. a preceding Moderate phase of 10 days duration is assumed)
Deaths	Cumulative deaths in OSIRIS (per 5-year age-group). Assumed Poisson distributed.
Underreporting adjustment Deaths	1.0
Age-groups	<1, 1-4, 5-9, ... 80-84, 85-89, 90-94, 95+
Disability weights	Mild: 0.051; Moderate: 0.133; Severe: 0.655
Notes	Occupation category distribution for all OSIRIS notified cases (including those admitted to hospital and/or ICU and/or deceased) is determined from the full analysis period, with the distribution derived from the 'open society' period (1 June 2020 t/m 20 Sept 2020) applied in supplementary analysis.

[a] URL: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/37360ned/table?fromstatweb

Table 2. Definition of occupation categories and proposed set of denominator occupations from CBS.

Occupation category	Occupation label(s) in OSIRIS	CBS occupation category(s) for population denominator
Healthcare	(Gezondheids)zorg	1011 Artsen 1012 Gespecialiseerd verpleegkundigen 1033 Verpleegkundigen (mbo) 1034 Medisch praktijkassistenten 1051 Verzorgenden
Education	Onderwijs en kinderopvang	0111 Docenten hoger onderwijs en hoogleraren 0112 Docenten beroepsgerichte vakken 0113 Docenten algemene vakken secundair onderwijs 0114 Leerkrachten basisonderwijs 0115 Onderwijskundigen en overige docenten 0121 Beroepsgroep sportinstructeurs 0131 Leidsters kinderopvang en onderwijsassistenten
Catering	Horecamedewerker	1112 Koks 1113 Kelners en barpersoneel 1122 Keukenhulpen
Transportation	Transport	1211 Dekofficieren en piloten 1212 Chauffeurs auto's, taxi's en bestelwagens 1213 Buschauffeurs en trambestuurders 1214 Vrachtwagenchauffeurs
Other contact professions	Overige contactberoepen Seksindustrie	1013 Fysiotherapeuten 1035 Medisch vakspecialisten 1114 Kappers en schoonheidsspecialisten 1116 Verleners van overige persoonlijke diensten (o.a. rijinstructeurs, prostituees)
Other	Klinisch laboratorium Landbouw Andere sector Werk met dieren of dierlijke producten Groenvoorziening Afvalverwerking Schoonmaakbranche Buitenland	*Denominator calculated as [age-group-specific] 'totale werkzame beroepsbevolking' minus sum of above categories*
Not applicable	N.v.t. (kinderen, gepensioneerden, werkzoekenden)	*Denominator calculated as [age-group-specific] national population size minus sum of all above categories*

Fig. 1a. Estimated cumulative number of patients (as of 24th Sept 2020) with symptomatic and asymptomatic infection per age-group (lower bound indicated). This is based on smoothed seroprevalence – adjusted for survey representativeness and seroreversion – from PICO3, and the estimated proportion of symptomatic infection, derived using PICO2 data. Plot also shows the cumulative notified cases from OSIRIS.

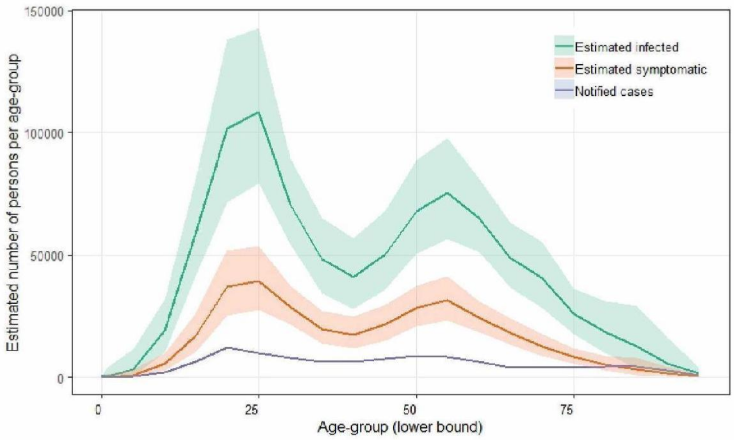

Fig. 1b. Estimated cumulative number of patients (as of 31 Dec 2020) with symptomatic and asymptomatic infection per age-group (lower bound indicated). This is based on smoothed seroprevalence from PICO2 until 24 Sept, and OSIRIS cases adjusted for estimated ascertainment from 25 Sept through 31 Dec. Plot also shows cumulative notified cases from OSIRIS.

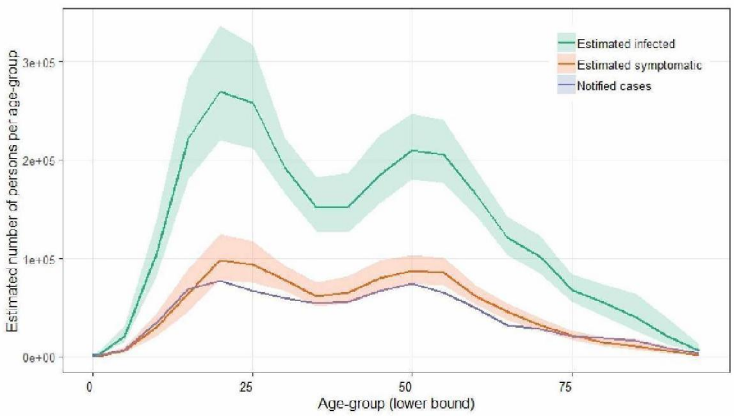

Fig. 2. Distribution over occupation categories (from OSIRIS) stratified by broad age-group, using the 'full' analysis period definition (i.e., 27 Feb 2020 through 31 Dec 2020).

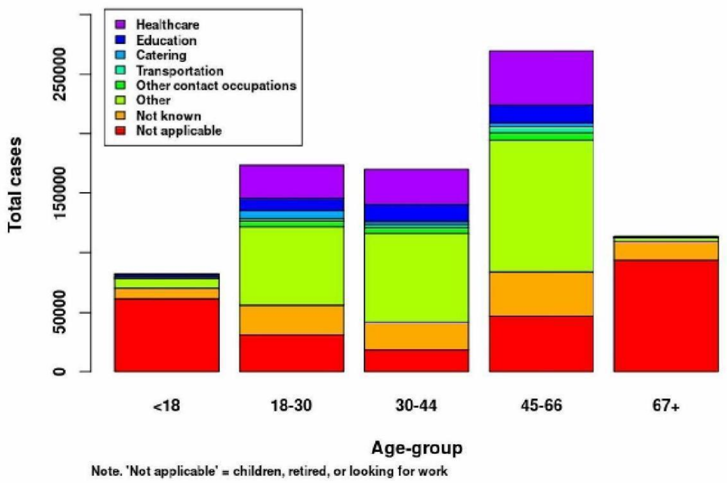

Fig. 3a. Proportion with occupation 'not known' per age-group from OSIRIS notifications, comparing (notified) fatal cases with (notified) cases who are not known to have died, 27 Feb through 31 Dec 2020. Note that for age-groups below 45-49 years, the denominators for the 'fatal' series are very small.

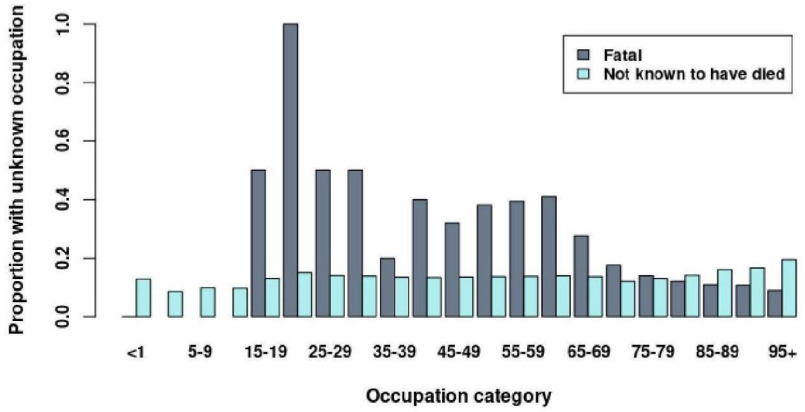

Fig. 3b. Distribution over occupation categories (from OSIRIS notifications, all ages), comparing two analysis period definitions ('full period' = 27 Feb through 31 Dec 2020; 'open society' = 1 June through 20 Sept 2020).

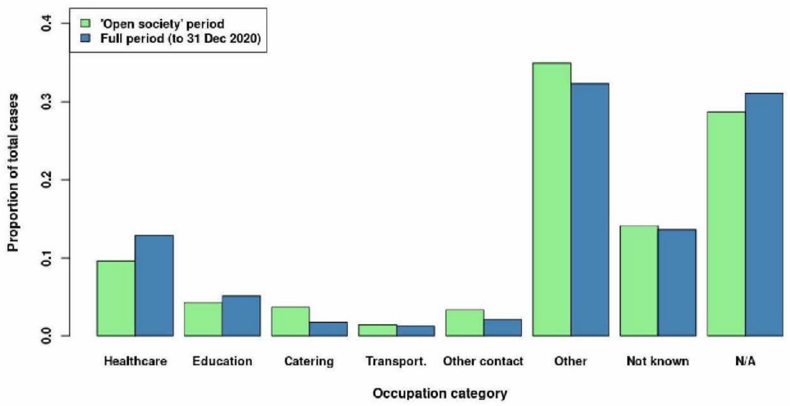

Fig. 4a. Estimated cumulative SI incidence per 5-year age-group with 95% CIs, up to 24 Sept 2020.

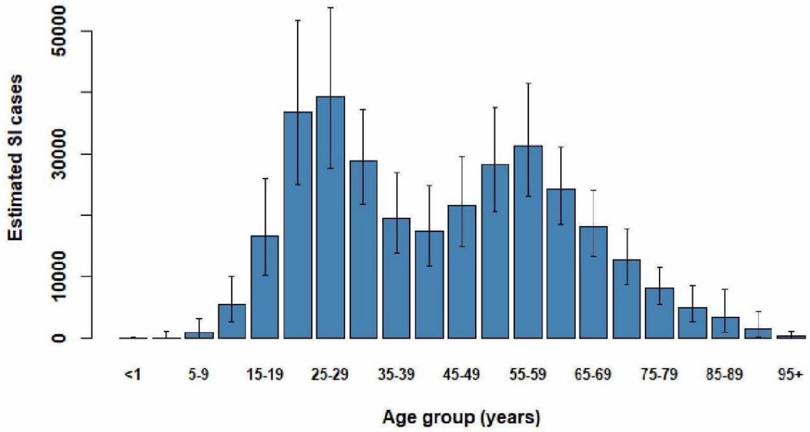

Fig. 4b. Estimated DALY (split into YLD and YLL) per 5-year age-group with 95% CIs, up to 24 Sept 2020.

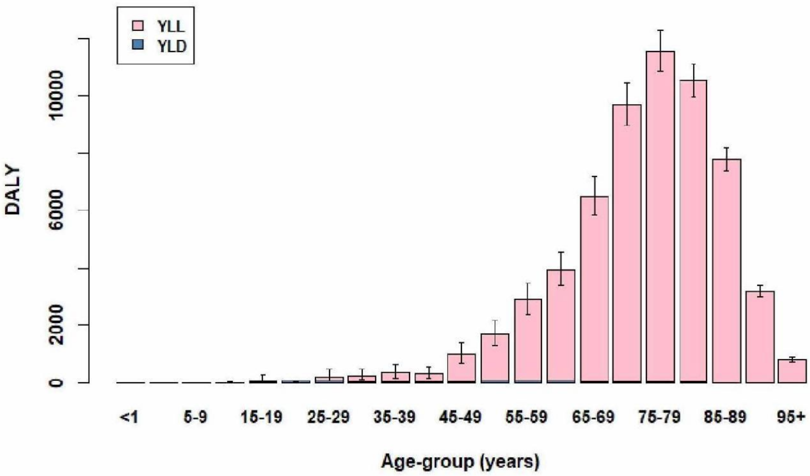

Fig. 4c. Estimated SI cases per 5-year age-group with 95% CIs, up to 31 Dec 2020.

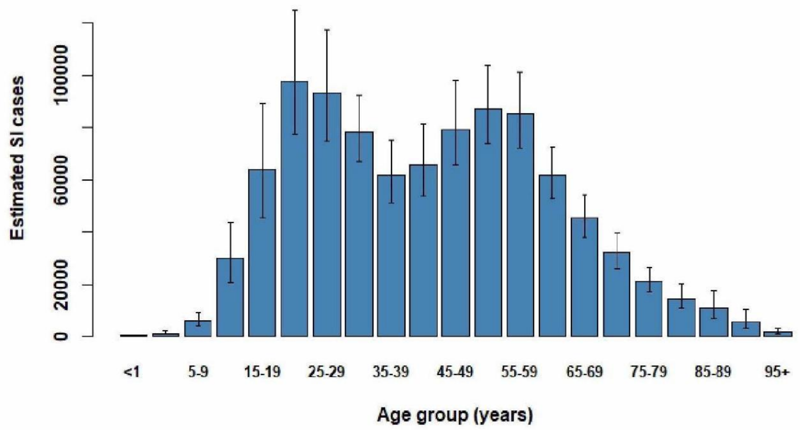

Fig. 4d. Estimated DALY (split into YLD and YLL) per 5-year age-group with 95% CIs, up to 31 Dec 2020.

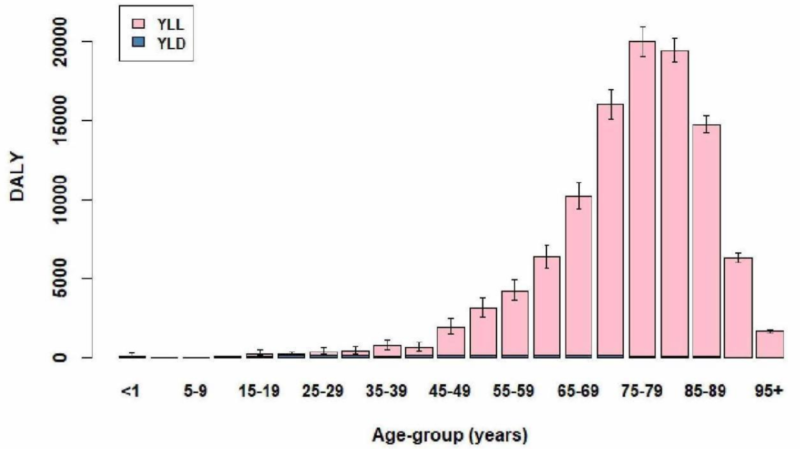

Fig. 4e. Estimated disease burden per 5-year age-group as DALYs per 100,000 persons, up to 31 Dec 2020.

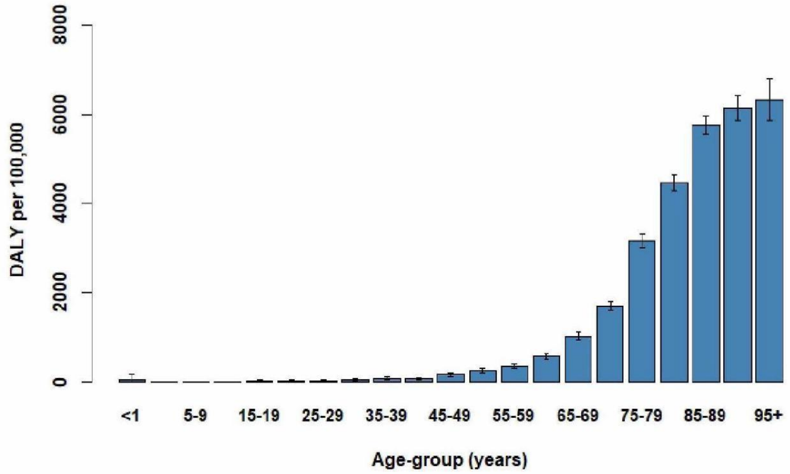

Fig. 5. Estimated absolute disease burden (in DALYs) per occupation category (with occupation 'Not known' imputed) and 5-year age-group, up to 31 Dec 2020.

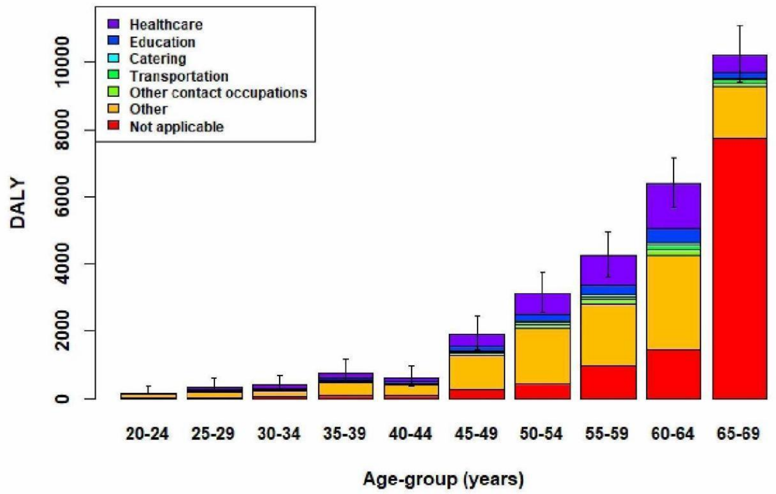

Fig. 6. Estimated cumulative incidence (per 100,000) of symptomatic infection per occupation category and 5-year age-group (as the estimated total number of patients per 100,0000 persons in each category within each age-group), up to 31 Dec 2020 and shown for the age range 20-69 years only. 'Not known' occupation imputed.

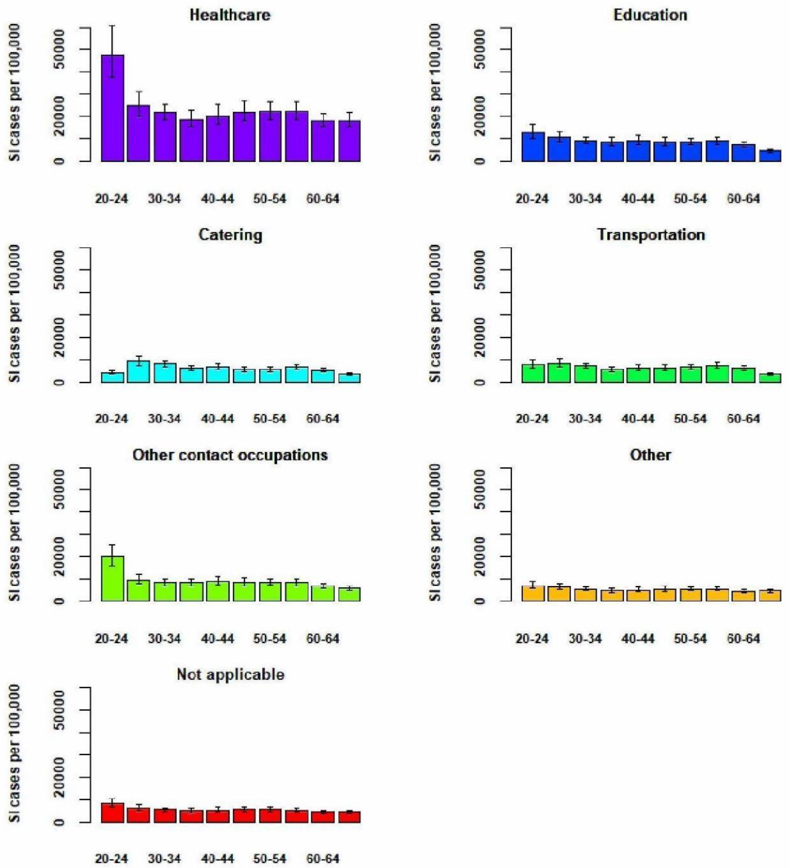

Fig. 7. Estimated disease burden per occupation category and 5-year age-group (as DALYs per 100,0000 persons in each category within each age-group), up to 31 Dec 2020, and shown for the age range 20-69 years only. 'Not known' occupation imputed.

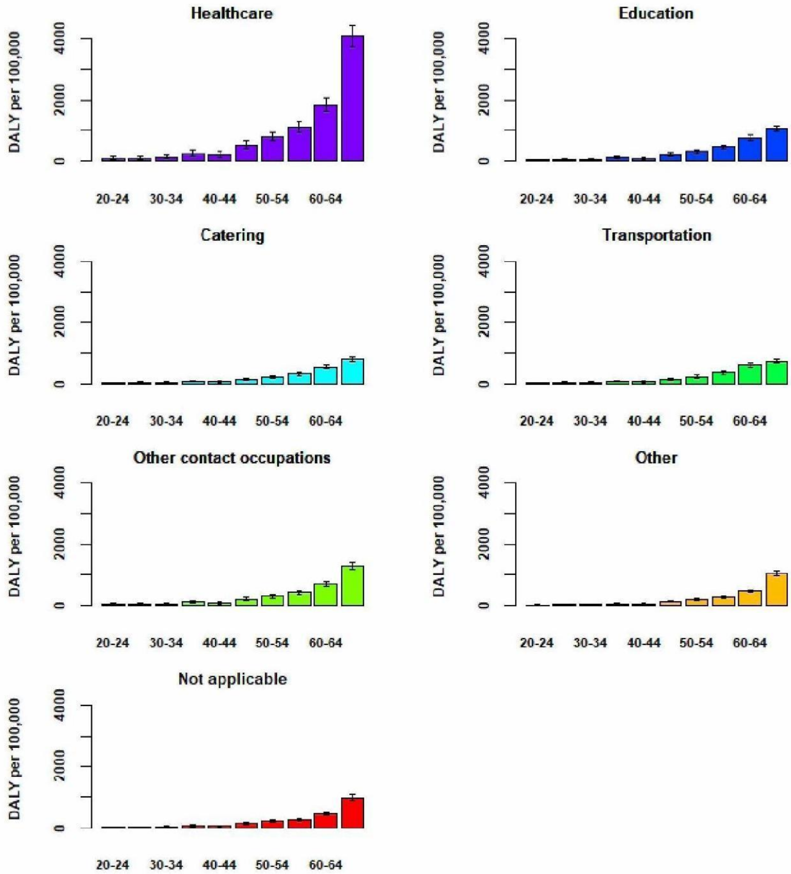

Fig. 8a. Estimated absolute disease burden per occupation category (as DALYs), up to 31 Dec 2020 and within the age range 20-69 years only. 'Not known' occupation imputed.

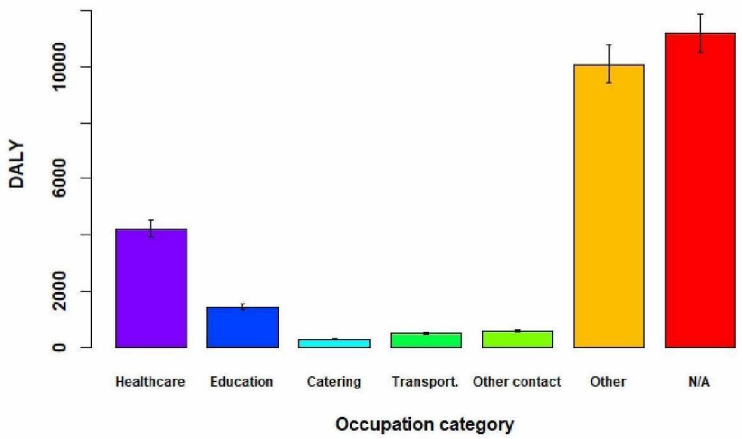

Fig. 8b. Estimated disease burden per occupation category (as DALYs per 100,0000 persons in each category, aggregating over age), up to 31 Dec 2020 and within the age range 20-69 years only. 'Not known' occupation imputed

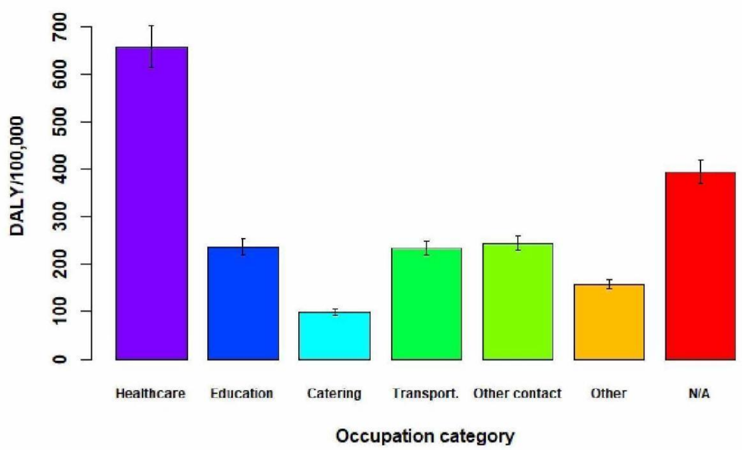

Fig. 9a. Expected absolute disease burden per occupation category (as DALYs), up to 31 Dec 2020 and within the age range 20-69 years only. **'Open society' occupation category distribution used**. 'Not known' occupation imputed.

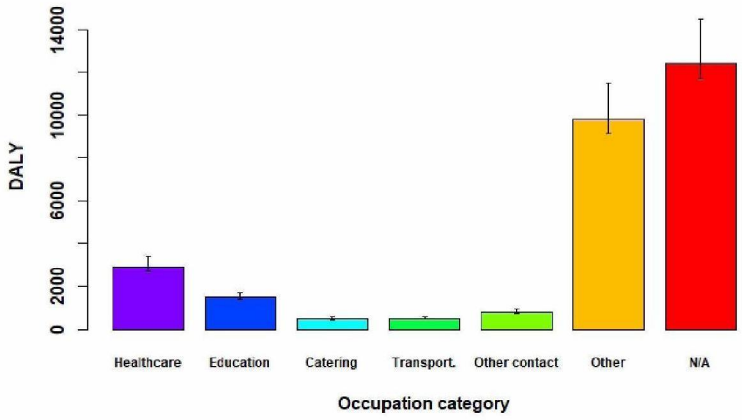

Fig. 9b. Expected relative disease burden per occupation category (as DALYs per 100,0000 persons in each category, aggregating over age), up to 31 Dec 2020 and within the age range 20-69 years only. **'Open society' occupation category distribution used**. 'Not known' occupation imputed.