



Burden of disease of COVID-19

PROTOCOL FOR COUNTRY STUDIES

Contents

INTRODUCTION	2
Purpose of this protocol	2
How to use this protocol	2
BURDEN OF DISEASE STUDIES.....	3
The disability-adjusted life year (DALY).....	3
Building a disease model	3
IMPLEMENTING A BURDEN OF DISEASE STUDY OF COVID-19	5
Planning a national study	5
Setting.....	5
Team	5
Data access and permissions	5
Disease model.....	5
Data requirements	6
Demographic data	6
Date requirements for years lived with disability	7
Data requirements for years of life lost.....	9
CALCULATIONS	10
Uncertainty.....	11
NOTES ON KNOWLEDGE TRANSLATION	12
Use of COVID-19 disease burden estimates.....	12
Presentation of results (estimates of disease burden)	12
Presentation of other findings	12
Communicating uncertainties	14
REFERENCES	14

Introduction

Coronavirus disease 2019 (COVID-19), caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is an infectious disease for which the spread of disease was categorized as a pandemic by the World Health Organization in March 2020. The outbreak was first identified in December 2019 in Wuhan, China, and has since then spread globally. Estimating the burden of the disease at national level is useful for comparing its impact with other diseases in the population and across populations.

Burden of disease studies quantify the health impact of diseases and risk factors in comparable metrics. The disability-adjusted life year (DALY) is one of the most used summary measure of population health, and the key metric in the Global Burden of Disease study¹.

The [European Burden of Disease Network](#) (*burden-eu*)² was established in 2019 to act as a technical platform for integrating and strengthening capacity in burden of disease assessment across Europe and beyond. It is structured in technical and disease-focused working groups focus. This protocol was developed by the core group of *burden-eu*.

Purpose of this protocol

This protocol provides guidance for anyone planning to assess the burden of disease of COVID-19 at national level using DALYs. It focuses on the direct health impact of the disease, and describes the data requirements, the methods, and considerations for communicating results. It also aims to foster harmonization of methodologies for estimating the burden of disease of COVID-19 across countries.

How to use this protocol

This protocol is divided into two parts. In the first, we provide a generic description of the metric used disability adjusted life year (DALY), and of the main steps towards its estimation. The second part focuses on practical guidance for implementing a national burden of disease study of COVID-19.

¹ The Global Burden of Disease study (GBD) is a worldwide observational epidemiological study led by the Institute for Health Metrics and Evaluation (IHME) that provides a tool to quantify health loss from hundreds of diseases, injuries, and risk factors, and to track health progress within and between countries. Read more at <http://www.healthdata.org/gbd/about>.

² <https://www.burden-eu.net/>

Burden of disease studies

The disability-adjusted life year (DALY)

The DALY is a health gap metric, measuring the healthy life years lost due to diseases, injuries or risk factors (Murray, 1994). DALYs are calculated by adding the number of years of life lost due to premature mortality (YLL) and the number of years lost due to disability (YLD):

$$DALY = YLL + YLD$$

YLL is the product of the number of deaths (M) and the average remaining life expectancy (RLE) at the time of death:

$$YLL = M \times RLE$$

YLD, calculated from an incidence perspective, is defined as the product of the number of incident cases (N), the average duration until recovery or death (D), and the disability weight (DW), which reflects the reduction in health- on a scale from 0 (no impact on full health) to 1 (death):

$$YLD_{inc} = N \times D \times DW$$

Diseases caused by an infectious agent may consist of one or more health outcomes, which can be acute or chronic, and with varied durations varying from a day or less to life-long. Thus, an incidence-based approach is considered the most suitable approach for estimating the burden of infectious diseases. In this approach, all health outcomes and the resulting burden, including those in future years, are assigned to the initial event, i.e. the infection with the agent.

Building a disease model

DALYs allow combining the impact of the different acute symptoms and sequelae associated with an infectious disease, to obtain a comprehensive estimate of its health impact. A key step in a DALY calculation process is therefore the definition of the disease model, i.e., a schematic overview of the outcomes of the infection and their frequency or probability of occurrence (Figure 1). This implies establishing the definition of a case (e.g., a person who tested positive for COVID-19), and defining the health outcomes associated with the infection. The latter may be identified through published epidemiological studies, or through analysis of national clinical data.

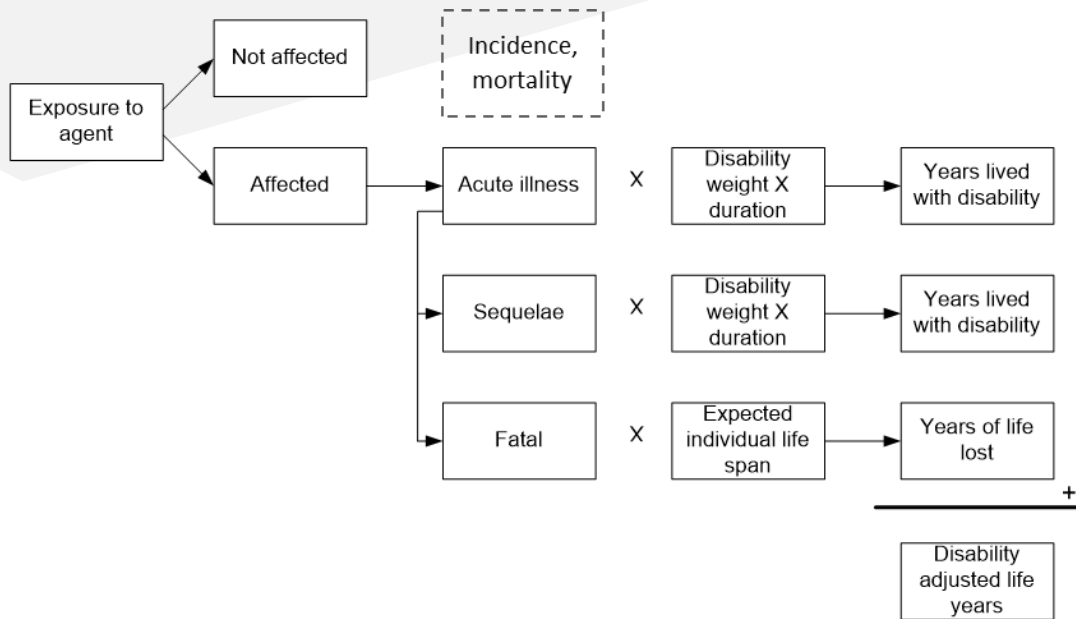


Figure 1. Generic template for a disease model for the calculation of the burden of infectious diseases. “Not affected” may be equivalent to “asymptomatic”.

The disease model helps us understand which “health states” are associated with the infectious disease under study. As shown in Fig. 1, each health state contributes YLDs or YLLs; in other words, we need to derive incidence, duration, and DW, or mortality for each health outcome, and remaining life expectancy at the time of death for each case. The DALYs for the infectious disease are then given by the sum of YLDs and YLLs across the n associated health states:

$$DALY = YLD_{total} + YLL_{total} = \sum_{i=1}^n YLD_i + \sum_{i=1}^n YLL_i$$

Implementing a burden of disease study of COVID-19

Planning a national study

Setting

In an initial step, the setting of the burden of disease study needs to be defined. This implies defining the population of the study, possible subpopulations, and the time period of the study (e.g., one specific year, or a range of years).

Team

Participants in the study team need a range of skills, including in epidemiology, demography, data management, statistics, and modelling. If the required skills and knowledge are not present in the

core team, a network of experts who can provide ad hoc input can be useful. The *burden-eu's* group is available to provide technical guidance.

The *burden-eu's* group is available to provide technical guidance to anyone or any team estimating the national burden of COVID-19. Contact us at info@burden-eu.net, or directly to Sara M. Pires (smpi@food.dtu.dk) or Brecht Devleesschauwer (Brecht.Devleesschauwer@sciensano.be).

Data access and permissions

Access to national health registries' data may require official permission from the data holder authority(ies) in the country, and in some cases require setting up data privacy plans. The process of applying and being granted permission may be lengthy in some countries, and all steps to be followed should be planned to avoid delays.

Disease model

A disease model is a conceptual representation of the clinical progression pathway, from infection until recovery (or death). Because COVID-19 is a recently emerging disease, epidemiological studies identifying all possible health outcomes and the proportion of cases developing such outcomes have not yet been conducted, and an exhaustive disease model cannot yet be specified. The frequency of the health outcomes comprising COVID-19 in the population may be collected from national health data (i.e. surveillance and electronic medical records). If country-specific clinical data are not available, data from other countries may be used to compensate for country-specific data gaps. The completeness of the disease model will depend on the available evidence.

Burden-eu has developed a *consensus* disease model for COVID-19 (Figure 2). This model can be modified to include country-specific data, and/or more data once they become available.

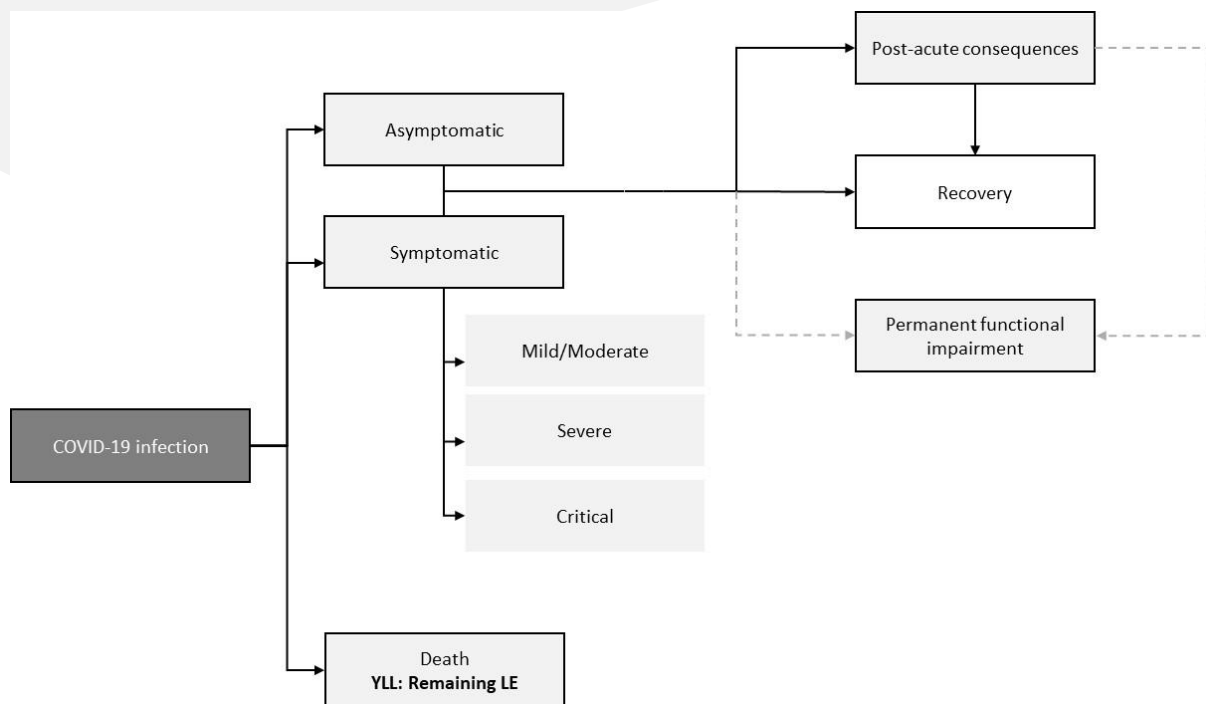


Figure 2. Consensus disease model for COVID-19.

Our model should be adapted to suit individual circumstances to reflect the availability of country-specific data. For example, availability of hospital or intensive care may have been restricted. This appreciates the fact that the distribution of occurrence across health states is likely to vary by location (Wyper et al., 2020). As new data inputs that provide further certainty around estimates become available, the model should be updated accordingly.

Data requirements

Demographic data

General demographic data needed to the estimation of burden of disease of COVID-19 include national population by sex and age, and if relevant and available further distinguished for subpopulations. These data may be collected through national statistics or global health datasets³. The population data should match the prespecified reference year(s), in order for the burden estimates to refer to the reference period.

³E.g. <https://www.who.int/healthinfo/survey/en/>; <http://www.indepth-network.org/data-stats/>;

Date requirements for years lived with disability

For each health state, as categorized in defined health outcomes (see Table 1), data on the frequency, duration and severity will be required. Outcome may be recovery or death. To the extent possible, all of these data should be stratified by age and sex and representative for the reference population and reference period.

Incidence

Incidence is the number of new infections (cases) or the subsequent cases for each health outcome (as shown in the disease model) in the population, stratified by age and sex.

The disease model specifies that incidence data are required for each of the included health states. There are however different pathways to obtain such data, which may differ across health states, in function of the available data. In general, we can distinguish three main pathways (ref Devleeschauwer 2014, DALY calculation in practice):

1. **Direct approach:** the incidence of the health state may be directly available, e.g., direct estimates of the number of hospitalized COVID-19 patients.
2. **Transition approach:** starting from the overall incidence of COVID-19 (e.g. number of newly reported cases in the surveillance system), the incidence of a specific health state may be obtained by multiplying the overall infection incidence with a transition probability, i.e., the proportion of patients that will develop the specific health state. This may be appropriate when population level data on the health state are lacking, but information on health state occurrence is available from a sample. Transition probabilities may be defined as probability distributions, and may differ for age and sex category.
3. **Attribution approach:** starting from the overall incidence of the health outcome, the incidence of a specific health state may be obtained by multiplying the overall outcome incidence with an attributable proportion, i.e., the proportion of patients with the outcome that were affected by COVID-19.

Infection incidence data only contribute to YLD if cases are symptomatic. For this reason, it is reasonable to use data from cases that have sought healthcare because they had symptoms, and have been diagnosed positive. These will vary from between countries testing regimes and healthcare capacity.

Burden of disease studies aim to quantify the true health impact of diseases on the population. It is therefore crucial that the incidence estimates have complete coverage for the reference population, and are corrected for underestimation when needed. For instance, if sentinel systems are used to

monitor infections or hospitalizations, the resulting estimates should be corrected for the incomplete coverage of these systems. Underreporting of cases is a consequence of any failure in the process of diagnosing and notifying a case of infection: not all infected patients will access healthcare, be tested, correctly diagnosed and registered in the surveillance system. The degree of underreporting will also vary between countries. Disease transmission models might be used to estimate the incidence of community cases, including those cases that did not seek healthcare.

Disability weights

We recommend adopting disability weights for each health outcome from the GBD when available. The GBD DWs are the most comprehensive set of internally comparable DWs, and is widely used in global and national BOD studies, supporting comparability of burden estimates. Table 1 shows DWs for identified health states of COVID-19.

Table 1. COVID-19 health states and disability weights.

Type	Name	Description	Data input proxy	Disability weight (95% uncertainty interval)
Acute, infectious disease	Asymptomatic	Has infection but experiences no symptoms	Estimates of suspected asymptomatic community cases	Nil
	Moderate	Has a fever and aches, and feels weak, which causes some difficulty with daily activities.	Positive (and/or suspected) cases community cases	0.051 (0.032-0.074)
	Severe	Has a high fever and pain, and feels very weak, which causes great difficulty with daily activities.	Positive (and/or suspected) requiring a non-intensive care hospitalisation	0.133 (0.088-0.190)
	Intensive care	Intensive care unit admission	Positive (and/or suspected) requiring intensive care hospitalisation	0.655 (0.579-0.727)

Chronic, infectious disease	Post-acute consequences (fatigue, emotional lability, insomnia)	Is always tired and easily upset. The person feels pain all over the body and is depressed.	Not estimated	0.219 (0.148-0.308)
-----------------------------	---	---	---------------	------------------------

These health state names, descriptions and DWs were based on those from the Global Burden of Disease (GBD) 2017 study, with the exception of those requiring intensive care, which were defined by the European Disability Weight study (Haagsma et al., 2015)

Data requirements for years of life lost

Mortality

Mortality is the number of deaths for each of the fatal health states. Depending on data availability, there can be one overall COVID-19 mortality health state, or different mortality health states in function linked to specific sequelae. The number of deaths registered shall also be segregated by age and sex.

Some countries may have timely mortality data available by mutually exclusive causes of death, in which case the underlying cause of death should be used. If ill-defined deaths cannot also be assigned to COVID-19, then using a sensitivity of an upper limit of COVID-19 related deaths would help gauge the degree of potential underestimation of just using cause-specific deaths. For countries where this information is not available, there are other alternatives. In countries where COVID-19 has become a notifiable disease or dedicated COVID-19 mortality surveillance systems were set up, published data can be used to establish mortality counts. Furthermore, the applied case definitions may result in an underestimation of the number of COVID-19 deaths, e.g., if only in-hospital deaths are counted, or only lab-confirmed cases. In this situation, a correction for underestimation would be crucial to obtain valid and comparable population level burden estimates. If COVID-19 mortality counts are not available, estimates of the number of excess deaths during the established time period could be used as a proxy for COVID-19 deaths, providing that mortality rates are stable. A major uncertainty is the extent to which these will also include a proportion due to the indirect consequences of COVID-19, for example the closure of vital services during national lockdowns, thus leading to possible overestimation of COVID-19 deaths.

Life expectancy

Life expectancy represents how long people can expect to live, and can be derived from life tables, which show the mortality and survival patterns in a population. Typically, life tables are constructed for men and women separately, because their mortality rates differ e.g. due to varying life style related risks, or by biological factors.

To estimate YLL, however, deaths are weighted in function of an ideal life expectancy, based on the lowest possible mortality observed globally. The use of a standard life table is therefore preferred. Different standard life tables have been developed for the WHO global health estimates (GHE)⁴ (WHO, n.d.) and the GBD studies of the Institute for Health Metrics and Evaluation (IHME) (GBD, n.d.), and have been updated over time as new data became available.

Calculations

As you engage in a burden of disease study, it will soon become clear that the collection of epidemiological data is the most difficult and time-consuming step. Once all data are collected, calculating DALYs is a relatively simple step.

For all calculations, it will be important to stratify data by age and sex. Describing the process step-by-step:

- Calculate YLD are calculated for each health outcome separately
- Sum all calculated YLD
- Calculate YLL
- Calculate DALY (YLD+YLL)

A variety of software packages are available for performing calculations and modelling, including burden of disease models. These may be general software tools, such as Microsoft Excel or R (R Core Team, 2018), or tools that have been specifically designed for

Technical support by the *burden-eu*'s group can be provided in various forms. You can contact us directly (info@burden-eu.net; Sara M. Pires - smpi@food.dtu.dk; Brecht Devleesschauwer - Brecht.Devleesschauwer@sciensano.be). You can also apply for a **short-term scientific mission**, funded by our network, and become a member of *burden-eu* and of one or more working groups: <https://www.burden-eu.net/join>. See all updates on COVID-19 relevant studies and events at: <https://www.burden-eu.net/outputs/covid-19>.

⁴ https://www.who.int/healthinfo/global_burden_disease/en/

burden of disease calculations, such as DisMod, DisModII (Barendregt et al., 2003) and the DALY calculator.⁵

Uncertainty

It is important to identify and address uncertainties in DALY calculations in order to demonstrate the strength of the evidence generated and allow valid comparisons between studies. The uncertainties in burden of disease estimates can be linked to the quality and representativeness of the data, and to the specifications and assumptions of the disease model used. Uncertainties in estimations of the burden of COVID-19 are particularly relevant, due to the limited knowledge that we still have about the virus and its impacts, especially about the potential the long term effects of the disease.

Where possible, sources of uncertainty in DALY calculations should be identified, quantified, analyzed and reported to aid with the understanding of estimates and to support effective knowledge transfer.

The most useful method for quantifying and analyzing uncertainty is probabilistic sensitivity analysis, also called uncertainty analysis or uncertainty propagation. Probabilistic analysis represents the uncertain parameters by uncertainty distributions. It uses Monte Carlo simulations to sample random values from the specified uncertainty distributions. In each iteration, the sampled values are used to calculate a DALY estimate. The combination of iterations results in an empirical distribution of DALY estimates, reflecting the joint uncertainty in the input parameters, which can be summarized by its mean and a 95% uncertainty interval.

Probabilistic assessment can be complemented with a variable importance analysis or sensitivity analysis, where the contribution of each of the uncertain input variables to the output variable (i.e., DALY estimate) is assessed. A second useful method for quantifying uncertainty are scenario analyses, where the calculations are repeated under multiple, alternative scenarios. This is particularly useful for defining the uncertainty linked to normative assumptions, alternative data sources, or alternative disease models.

⁵ <https://cran.r-project.org/web/packages/DALY/index.html>

Knowledge translation

Use of COVID-19 disease burden estimates

Disease burden estimates only become meaningful when used in a comparative way – e.g., to compare impact across diseases, across countries or subnational entities, age groups, sexes, socio-economic strata, etc. Quantifying the disease burden of COVID-19 would therefore allow putting its impact in perspective.

Disease burden estimates reflect the magnitude of observed health problems, but do not provide insights on how to avert these problems, or how large the impact would have been without measures. A low disease burden therefore does not necessarily mean that the concerned disease should be ignored. It can also reflect successful control measures, without which the disease burden would have been much higher.

Presentation of results (estimates of disease burden)

Disease burden estimates should ideally be presented using a combination of tabular and graphical means, depending on the intended audience. In general, it is useful to present a few key outputs in the main text of the report, possibly supported by infographics, while a detailed set of outputs could be made available in an online appendix.

We suggest that at minimum stratification by sex and age group should be presented. Stratification by health outcome and/or by severity level is often useful. For meaningful comparisons among different population, presenting DALYs per 100,000 persons is essential. If age-standardized rates are presented, the reference population should be explicitly mentioned. For comparison of disease burden due to COVID-19 with other infectious diseases, it is sometimes insightful to calculate DALYs per case in addition to the absolute burden of disease.

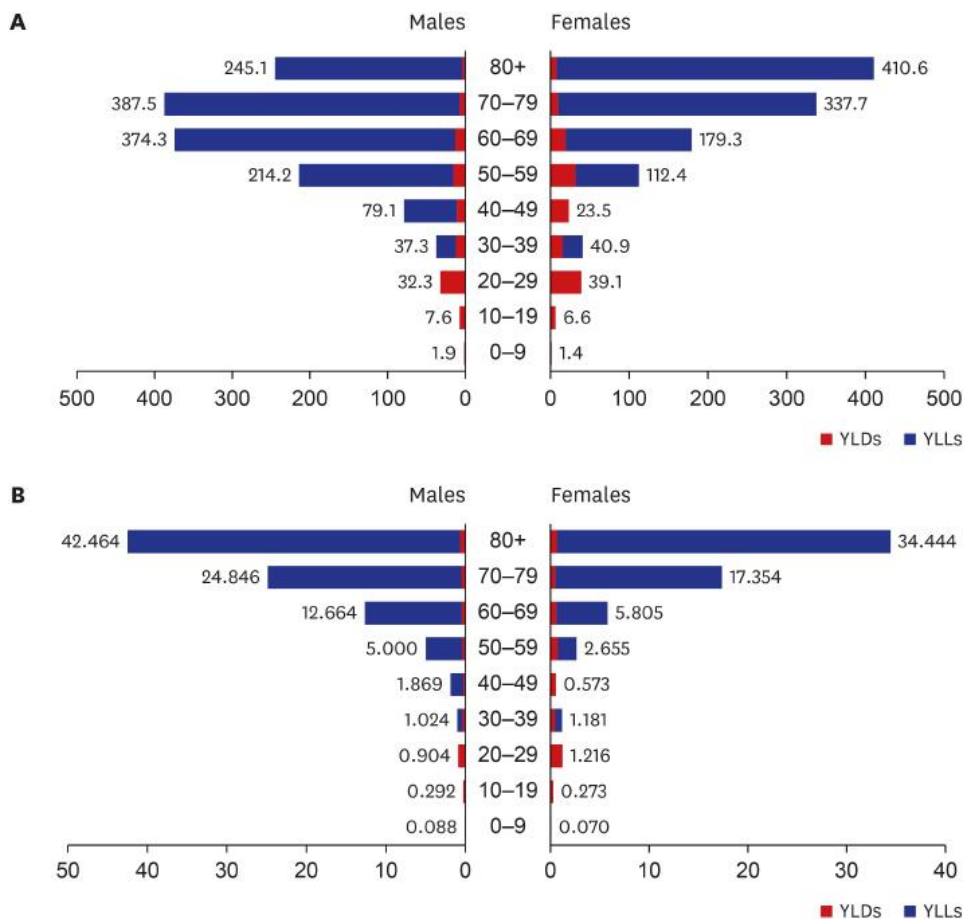
Presentation of other findings

Particularly useful for emerging infectious diseases is the reporting of data gaps, and a description and evaluation of the approaches that were taken to adjust for these gaps. Such an overview will guide the collection of needed data when it becomes available, and support the interpretation of the results in the light of current uncertainties and assumptions.

Burden of Disease due to COVID-19 in Korea

Used data on confirmed cases and deaths due to COVID-19 between January 20 and April 24, 2020, [Jo et al. \(2020\)](#) calculated the burden of disease due to COVID-19 in Korea. Disability-adjusted life years (DALYs) were determined by sex and age. Morbidity was estimated directly among the confirmed, cured, and fatal cases. Disability weights were adopted from previous similar causes on the severity of COVID-19, and years of life lost were calculated using the standard life expectancy from the 2018 life tables for each sex and age.

Results showed that YLD were higher in females (155) than in males (105), but YLL were higher in males (1,274) than in females (996). The total disease burden attributable to COVID-19 in Korea during the study period was estimated to be 2,531 DALYs, or 4.9 DALYs per 100,000 population. The YLLs constituted 90% of the total DALY.



DALY for COVID-19 disease 2019 by sex and age group. (A) DALY; (B) DALY per 100,000 population. DALY = disability-adjusted life years.

Communicating uncertainties

Communication of uncertainties associated with estimates of the burden of disease of COVID-19 will be essential to support the evidence-based movement of public health priorities, as well as the development of a pathway to improve accessibility to data.

To support such communication, the following can be included:

- A description and streamlining of registry databases used (hospital statistics, birth and death records, other databases).
- An overview of methodologies assumptions made and implications in obtained results. These can also be illustrated using scenario analysis.
- Presentation of all knowledge and data gaps, and eventual recommendations of strategies to overcome these.
- Quantitative description of associated uncertainties.
- Graphical representation of the degree to which the result is sensitive to the specified variables, using e.g. a tornado plot.
- Interpretation of results in the light of currently available knowledge.

In general, thorough documentation and transparency will be essential for the correct interpretation and best usage of the study's results.

Final remarks

Once you have an overview of the data requirements and the methodology to estimate the burden of disease of COVID-19, have collected the data, and made sure all resources needed are available, you can start with the actual calculations. We hope this guide will be useful in the process. We also encourage you to contact us at any point, and to share your experiences with the burden-eu network. With each national study launched and each set of results available, we will be able to compare the health impact of COVID-19 across populations and across diseases.

Contact us at info@burden-eu.net; Sara M. Pires - smpi@food.dtu.dk; Brecht Devleesschauwer - Brecht.Devleesschauwer@sciensano.be. To become a member of *burden-eu* and of one or more working groups, fill in the information at: <https://www.burden-eu.net/join>. See all updates on COVID-19 relevant studies and events at: <https://www.burden-eu.net/outputs/covid-19>.

References

Barendregt JJ, van Oortmarsen GJ, Vos T, Murray CJL. A generic model for the assessment of disease epidemiology: The computational basis of DisMod II. *Popul Health Metr* 2003; 1:4.

GBD. Global Burden of Disease (GBD) | Institute for Health Metrics and Evaluation n.d. <http://www.healthdata.org/gbd> (accessed June 22, 2020).

Haagsma JA, de Noordhout CM, Polinder S, Vos T, Havelaar AH, Cassini A, et al. Assessing disability weights based on the responses of 30,660 people from four European countries. *Popul Health Metr* 2015; 13:10.

Jo M-W, Go D-S, Kim R, Lee SW, Ock M, Kim Y-E, et al. The Burden of Disease due to COVID-19 in Korea Using Disability-Adjusted Life Years. *J Korean Med Sci* 2020; 35.

Murray CJL. Quantifying the burden of disease: The technical basis for disability-adjusted life years. *Bull World Health Organ* 1994; 72:429–45.

R Core Team. R: a language and environment for statistical computing 2018. <https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing> (accessed October 15, 2018).

WHO. WHO | Global Health Estimates n.d. https://www.who.int/healthinfo/global_burden_disease/en/ (accessed June 22, 2020).

Wyper GMA, Assunção R, Cuschieri S, Devleeschauwer B, Fletcher E, Haagsma JA, et al. Population vulnerability to COVID-19 in Europe: A burden of disease analysis. *Arch Public Heal* 2020; 78.

Barendregt JJ, van Oortmarsen GJ, Vos T, Murray CJL. A generic model for the assessment of disease epidemiology: The computational basis of DisMod II. *Popul Health Metr* 2003; 1:4.

GBD. Global Burden of Disease (GBD) | Institute for Health Metrics and Evaluation n.d. <http://www.healthdata.org/gbd> (accessed June 22, 2020).

Haagsma JA, de Noordhout CM, Polinder S, Vos T, Havelaar AH, Cassini A, et al. Assessing disability weights based on the responses of 30,660 people from four European countries. *Popul Health Metr* 2015; 13:10.

Jo M-W, Go D-S, Kim R, Lee SW, Ock M, Kim Y-E, et al. The Burden of Disease due to COVID-19 in Korea Using Disability-Adjusted Life Years. *J Korean Med Sci* 2020; 35.

Murray CJL. Quantifying the burden of disease: The technical basis for disability-adjusted life years. *Bull World Health Organ* 1994; 72:429–45.

R Core Team. R: a language and environment for statistical computing 2018. <https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing> (accessed October 15, 2018).

WHO. WHO | Global Health Estimates n.d.
https://www.who.int/healthinfo/global_burden_disease/en/ (accessed June 22, 2020).

Wyper GMA, Assunção R, Cuschieri S, Devleeschauwer B, Fletcher E, Haagsma JA, et al. Population vulnerability to COVID-19 in Europe: A burden of disease analysis. *Arch Public Heal* 2020; 78.