European Medicine Agency
User Manual

February 15, 2022

This manual was commissioned by UHasselt to the European Medicine Agency (EMA) in 2022

Date of last version of the final study report :
Final version expected in February 2022

EU PAS register number :
EUPAS44229

Authors:
Steven Abrams, Jonas Crèvecoeur, Niel Hens, Neilshan Loedy, Geert Molenberghs, Johan Verbeeck, Lander Willem.

Contents

List of Tables

List of Abbreviations

Abbreviation	**Meaning**
ARS	Agenzia Regionale di Sanità della Toscana
BIFAP	Base de Datos para la Investigación Farmacoepidemiológica en Atención Primaria
COVID-19	Coronavirus Disease 2019
ECDC	European Centre for Disease Prevention and Control
EEA	European Economic Area
EMA	European Medicines Agency
EU	European Union
EudraVigilance	European Union Drug Regulating Authorities Pharmacovigilance
GISAID	Global Initiative on Sharing Avian Influenza Data
ICU	Intensive Care Unit
IR	Incidence Rate
mRNA	Messenger Ribonucleic Acid
PM	Probabilistic Model
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus-2
TESSy	the European Surveillance System
VoCs	Variants of Concerns

1 Introduction

1.1 How to use this manual

This manual serves as a reference guide for working with your project website. The manual is split into three parts.

1. **Section I**
   Contains general information from the introduction, such as the purpose of the document, a description of the document, and the methodologies used for the analysis.

2. **Section II**
   Introduces you to the data sources, summary and structure of the data, and the models parameters used for the analysis.

3. **Section III**
   Explains the R Shiny web application and its objects, including the user profile, interface and how the user can interact with the website's interactive objects. This section also elucidate the description, interpretation, and further actions of each feature available in the dashboard.

4. **Section IV**
   Provides user(s) with information on how they can get in touch with the Authors of the user manual.

For easy navigation refer to the content table.

1.2 Overview

The Benefit-Risk Contextualisation of COVID-19 Vaccines in the EU Tool is an R-based tool that employs statistical methods to provide adjusted estimates from vaccine's safety and effectiveness while considering the issues of missing data and uncertainty concerning the various ingredients of the risk-benefit measure. While it does not replace the ability to analyze data and make adjustments, it is designed for routine application in surveillance as no significant programming skills are required.

The Benefit-Risk contextualisation of COVID-19 vaccines in the EU tool, is an interactive web application build with Shiny, which is part of R programming language. This manual provides you with steps and instructions on how to run the project website.

In case you can not find answers to your questions in this manual, feel free to contact us.

1.3 Objectives

The main objective addressed in this user manual is to provide users with a set of relevant pages on:

- **Description**
  provides steps on how to operate the quantification of both the benefits and the risks related to COVID-19 vaccines in the EU, such as the ability to change the study object into the country-level or EU-level, vaccine's type, the emerging variant of concerns, and statistical outputs.
- **Interpretation**
  explaining the significance of the output
- **Further actions**
  what to do if there are any issues with the output. This could entail performing a new analysis or altering the data.

2 Limitations

Hereunder we list some general limitations of the model used for the benefit and risk calculation.

Benefits

- The benefit-side of vaccination provided in the tool is based on a probabilistic model (PM) (see below for more details).
- Alternatively, a more complex and refined stochastic compartmental model (SCM) can be considered in which a more natural description of disease dynamics is included. In the PM, some simplifying assumptions are made. In order to rely on the SCM instead of the PM, one needs access to granular time- and age-specific incidence data on confirmed COVID-19 cases, hospitalizations, ICU admissions and deaths.
- Due to data sparseness and computational complexity it remains challenging to inform the SCM and to achieve converge for all model parameters and all countries.
- Although the PM is based on a direct quantification of the prevented cases, hospitalizations, ICU admissions and deaths, the current application of the PM does not provide an assessment of uncertainty. However, this feature is implicitly available through the repeated use of the Shiny R app enabling a user-defined Monte Carlo simulation including a range of user-specified parameter values.

- The PM does not account for build-up of natural immunity in the absence of vaccine-induced immunity. The inclusion of this feature would imply a recursive assessment of the impact of vaccination over time, given that the prevention of (confirmed) COVID-19 cases at calendar time t in age group a, will have an impact on natural immunity levels at subsequent time points. Although such a recursive assessment would be possible and the inclusion of natural immunity is methodologically straightforward, such country-specific computations are computer-intensive thereby potentially jeopardising the usefulness of the Shiny R app.

- As individual-level data on the time between consecutive vaccination doses is missing, we approximate the protection induced by vaccination against infection, hospitalization, ICU admission and death at the population level. More specifically, a gradual build-up of vaccine protection after first, second and booster vaccination is accounted for, though the aggregation of protection over time is performed without explicitly accounting for time between consecutive doses. This implies that we assume that the time between such doses is random.

- In the absence of time- and age-dependent information concerning hospitalizations, ICU admissions and deaths at the country-level, the quantification of prevented hospitalizations, ICU admissions and deaths in the PM is based on a time-invariant age distribution estimated for each of these endpoints separately over the entire course of the epidemic. However, the prevented burden is therefore potentially overestimated as the temporal protective and beneficial effects of vaccination in age groups with high vaccine uptake are averaged out.

- **This further underlines the importance of the availability of detailed age- and time-specific information to improve estimation of the prevented burden.**

Risks

- A **limitation in the risk assessment** is the, potentially arbitrary, manually chosen weight to assign more value to one source of background incidence rates compared to another source. To exclude the arbitrariness, an alternative could be to use a variation on the weighted least square method, where the inverse variance weight is multiplied with a factor that is based on Cochran's Q heterogeneity measure [1]. A formal sensitivity analysis can be performed by providing several weighting scenarios to the Shiny R application.

- A **limitation in the myocarditis risk assessment** is the exclusion of vaccine coverage data with unknown age or vaccine in the ECDC vaccine coverage data set. The exclusion on a country level is quite substantial for Poland (∼92,000 vaccines excluded) and Croatia (∼16,000) slightly less for Austria (∼4,000), France (∼3,200) (Bulgaria (∼2,000) and The Netherlands (∼1,300), and negligible for Denmark, Estonia, Greece, Ireland, Iceland, Lithuania, Latvia, Portugal and Slovakia (between 2 and 462). The

actual vaccine coverage is thus actually higher than the coverage used in the risk assessment and the Observed-Expected ratios potentially smaller than reported. The impact on the conclusions is, however, deemed minor, given it concerns only 0.07% of all administered vaccines (∼120,000 on ∼178,000,000). To include the administered doses with unknown age or vaccine, a redistribution as applied before can be applied.

2.1 Methodological background

2.2 Benefit: Probabilistic Model (PM)

Quantification of the benefits associated with COVID-19 vaccination can be done directly by comparing the observed number of confirmed COVID-19 cases, hospitalizations and deaths with the expected number of the aforementioned events in the counterfactual case that no COVID-19 vaccination would have been present. In order to do so, we perform a by-country assessment of the observed events and the probability of being protected through vaccination at that moment in time. More specifically, we consider a probabilistic model (PM) that accounts for differential vaccine effectiveness after vaccination as a function of time since vaccination, emergence of new variants of concern and age-specific and temporal differences in disease dynamics.

We use a so-called "leaky" vaccination approach. For example, vaccination with 50% vaccine effectiveness (against infection), implies that for a vaccinated individual the likelihood to acquire infection is 50% less as compared to a non-vaccinated individual of the same age and at the same calendar time. As such, we do include breakthrough infections after vaccination in our analysis.

2.2.1 Basic implementation

The implementation of the PM is based on the following reasoning. Let $n(a,t)$ denote the number of confirmed COVID-19 cases at time t and age group a. The (counterfactual) number of confirmed cases in the absence of vaccination would then be equal to:

$$n^*(a,t) = \frac{n(a,t)}{1-\phi(a,t-d)},$$

where $\phi(a,t-d)$ represents the proportion of protected individuals through vaccination of age (or age group) a at time $t-d$, and with d a fixed delay between time of SARS-CoV-2 infection and COVID-19 confirmation. In the ensuing analysis, we consider the delay d to be equal to zero.

The aforementioned expression can be obtained as follows:

$$n(a,t) = [c_{t-d}(a)n^*(a,t)] + \{[1 - v_{t-d}(a)]\, n^*(a,t)\},$$

with the first term representing the contribution of breakthrough infections after vaccination up to time $t-d$, a proportion denoted by $c_{t-d}(a)$, and the second term being the contribution of unvaccinated individuals (until time $t-d$) becoming infected, i.e., $1 - v_{t-d}(a)$, and

$$c_t(a) = \sum_{k=0}^{t} p_k(a)(1 - \mathrm{ve}_{(t-k)}),$$
$$v_t(a) = \sum_{k=0}^{t} p_k(a),$$

where

- $p_k(a)$ represents the proportion of vaccinated individuals in age group a at time k (with discrete time steps of one day), and
- $\mathrm{ve}_{(t-k)}$ is the vaccine effectiveness as a function of the time since vaccination equal to $t-k$.

Consequently, we have

$$\begin{aligned} n(a,t) &= \{c_{t-d}(a) + [1 - v_{t-d}(a)]\}\, n^*(a,t) \\ &= [1 - \phi(a, t-d)]n^*(a,t). \end{aligned}$$

The proportion of protected individuals in age group a at time t is computed as follows:

$$\phi(a,t) = v_t(a) - c_t(a) = \sum_{k=0}^{t} p_k(a)\mathrm{ve}_{(t-k)}.$$

The proportion of vaccinated individuals in the age group a at time t is derived from available vaccine uptake data. The vaccine effectiveness function is considered to have the following functional form:

$$\mathrm{ve}_{(t-k)} = \mathrm{ve}^* \left\{1 + \exp\left(-k_0^* \left[(t-k) - x_0\right]\right)\right\}^{-1} \exp\left(-\omega \left[(t-k) - 2x_0\right]_+\right).$$

This implies a logistic growth after vaccination to reach a level of protection equal to ve^*, after which an exponential decay takes place with waning rate equal to ω. The operator $[\cdot]_+$ equals zero if the argument is negative and takes the value of the argument

when positive. The parameter x_0 represents the midpoint of the logistic growth curve.

Based on the difference between $n^*(a,t)$ and $n(a,t)$ one can estimate the prevented number of COVID-19 cases in age group a at time t. More specifically, the number of prevented cases can be expressed as $\phi(a,t-d)n^*(a,t)$.

2.2.2 Vaccine types

The European regulatory body has granted emergency use authorization to several vaccines, including those developed by Pfizer-BioNTech, Janssen, Moderna, and Oxford-AstraZeneca. Clinical trials and evaluations of mass vaccination campaigns have shown that two doses given three to four weeks apart, or a single dose for Janssen, can provide high levels of protection against symptomatic and severe disease [2]. In order to take the protection after second dose vaccination into account, we allowed for differential vaccine effectiveness estimates for the mRNA and adeno-based vaccines. Note that in the toolkit, a vaccine brand specific analysis is included, without grouping these vaccine brands into mRNA and adeno-based vaccines, but for the demonstration of the methodology we keep the aforementioned terminology in place.

Consequently, we have

$$\begin{aligned}\phi(a,t) &= \phi^{(1)}(a,t) + \phi^{(2)}(a,t) \\ &= \sum_{k=0}^{t} \left\{ p_k^{(1)}(a)\mathrm{ve}_{(t-k)}^{(1)} + p_k^{(2)}(a)\mathrm{ve}_{(t-k)}^{(2)} \right\}, \end{aligned} \tag{1}$$

with superscripts referring to (1) mRNA- or (2) adeno-based vaccination.

Next to an overall assessment of the impact of vaccination on confirmed cases, hospitalizations, ICU admissions and deaths, a vaccine-type specific approach could be considered as well. Based on the terminology introduced above, we established the relation

$$n(a,t) = \left[1 - \phi(a,t-d)\right] n^*(a,t),$$

where $\phi(a,t-d) = \phi^{(1)}(a,t-d) + \phi^{(2)}(a,t-d)$. Assume that we want to study the impact of vaccination with mRNA vaccines on the number of confirmed COVID-19 cases. In that case, the counterfactual number of confirmed cases, in the absence of mRNA vaccination, is equal to

$$n^{(2)*}(a,t) = \left[1 - \phi^{(2)}(a,t-d)\right] n^*(a,t),$$

and the prevented number of confirmed cases by mRNA vaccines only equals

$$n^*(a,t) - n^{(2)*}(a,t) = \phi^{(2)}(a,t-d)n^*(a,t).$$

Similarly, the prevented number of confirmed cases by adeno-based vaccination equals

$$n^*(a,t) - n^{(1)*}(a,t) = \phi^{(1)}(a,t-d)n^*(a,t).$$

As a result, the total number of prevented confirmed COVID-19 cases, i.e., $\phi(a,t-d)n^*(a,t)$ (see Section 2.2.1), is exactly equal to the prevented cases using mRNA and adeno-based vaccines separately, i.e.,

$$\begin{aligned} n^*(a,t) - n^{(1)*}(a,t) + n^*(a,t) - n^{(2)*}(a,t) &= 2n^*(a,t) - \left[n^{(1)*}(a,t) + n^{(2)*}(a,t)\right] \\ &= 2n^*(a,t) - n^*(a,t)\left[2 - \phi(a,t-d)\right] \\ &= n^*(a,t)\left\{2 - \left[2 - \phi(a,t-d)\right]\right\} \\ &= \phi(a,t-d)n^*(a,t) \end{aligned}$$

A similar strategy can be considered for other endpoints such as hospitalizations, ICU admissions and deaths by decomposing protection against these clinical outcomes by vaccine type.

2.2.3 Variants of Concern

Vaccines can help to reduce disease burden in a variety of ways, including preventing infection, making infected people less infectious, and avoiding severe outcomes in those who are infected, all of which reduce the overall number of infected people and/or the proportion of symptomatic infections [3]. Nonetheless, the emergence of VoCs has resulted in lower vaccine effectiveness, hence, leading to re-emergence of the disease. In the calculation of the prevented confirmed COVID-19 cases, a change in vaccine effectiveness is allowed for in relation to the prevalence of α and δ VoCs in the population under study. The following correction factor was used:

$$\psi(a,t) = \{1 - p_{\alpha,t}(a) - p_{\delta,t}(a)\} + p_{\alpha,t}(a)\frac{\text{ve}^*_\alpha}{\text{ve}^*} + p_{\delta,t}(a)\frac{\text{ve}^*_\delta}{\text{ve}^*}.$$

These prevalences $p_{\alpha,t}(a)$ and $p_{\delta,t}(a)$ are estimated based on available GISAID data [4, 5, 6]. The GISAID EpiCoV dataset includes genomic surveillance data obtained from sequencing COVID-19 positive samples. In order to estimate the prevalences $p_{\alpha,t}(a)$ and $p_{\delta,t}(a)$ we consider a generalized additive model (GAM) for binary outcome data (e.g., α VoC or not) with a logit-link function [7].

Hence, using the correction factor, the proportion of protected individuals in age group a at time t is equal to

$$\phi(a,t) \times \psi(a,t),$$

thereby providing a differential baseline vaccine effectiveness as compared to ve* for different variants of concern.

2.2.4 Partial vaccine-induced protection after first vaccination

After first dose COVID-19 vaccination individuals gain already (partial) protection against infection, hospitalization and death. However, upon initiation of second dose vaccination in the population, the contribution of first and second dose vaccination to the overall protection in the population should be accounted for. The proportion of protected individuals of age a at calendar time t after first and second vaccination are equal to

$$\phi_1(a,t) = \psi_1(a,t) \sum_{k=0}^{t} p_{1,k}(a) \text{ve}_{1,(t-k)},$$
$$\phi_2(a,t) = \psi_2(a,t) \sum_{k=0}^{t} p_{2,k}(a) \text{ve}_{2,(t-k)},$$

respectively. Here, we have for $l = 1, 2$, dose-specific vaccine effectiveness functions $\text{ve}_{l,(t-k)}$ for which

$$\text{ve}_{1,(t-k)} = \text{ve}_{(t-k)}$$
$$\text{ve}_{2,(t-k)} = \frac{\text{ve}_2^*}{\text{ve}^*} \times \text{ve}_{1,(t-k)},$$

implying that the vaccine effectiveness after the second dose is proportional to the effectiveness after first dose with time since vaccination. Furthermore, the functions $\psi_1(a,t)$ and $\psi_2(a,t)$ allow for a different correction factor for each of the doses regarding protection against variants of concern.

Note that an extension to include mRNA and adeno-based vaccination is immediate (see above). For ease of presentation, we confine attention to the one vaccine type formulation first. A more detailed description of the expressions in case of multiple vaccine types is presented below.

In order to weigh the contribution of each of the aforementioned expressions in the overall protection in age group a at calendar time t we rely on the proportion of individuals of

age a being vaccinated only once or twice in the population, hence, implying that

$$\phi(a,t) = w_{1,t}(a)\phi_1(a,t) + \phi_2(a,t),$$

where $w_{1,t}(a)$ represents the proportion of vaccinated individuals in age group a at time t with a first dose only. More specifically, we have

$$w_{1,t}(a) = \frac{\sum_{k=0}^{t} p_{1,k}(a) - \sum_{k=0}^{t} p_{2,k}(a)}{\sum_{k=0}^{t} p_{1,k}(a)} = \frac{v_{1,t}(a) - v_{2,t}(a)}{v_{1,t}(a)},$$

in terms of the proportion of vaccinated individuals with a first and second dose respectively. Note that in the absence of protection after the first dose, or when all individuals received their second dose, the formula reduces to $\phi(a,t) = \phi_2(a,t)$. Alternatively, we can write

$$\phi(a,t) = \phi_1(a,t) + \phi_2(a,t)\left\{\frac{v_{1,t}(a) - v_{2,t}(a)\dfrac{\phi_1(a,t)}{\phi_2(a,t)}}{v_{1,t}(a)}\right\},$$

implying an initial increase in protection induced by an increase in protection after second dose vaccination countered by a gradual decrease in protection after first (and second) dose vaccination with time since vaccination. Needless to say, the alternative expression implies that $\phi_2(a,t) \neq 0$. In the absence of second dose vaccination, i.e., $\phi_2(a,t) = 0$, the formula reduces to $\phi(a,t) = \phi_1(a,t)$ as required.

The above formula can be interpreted as follows: the total protection induced by vaccination is the sum of the dose-specific contributions, weighted according the proportion of individuals with a given vaccination uptake (either one dose or two doses).

Considering two vaccine types, i.e., mRNA and adeno-based vaccines, the aforementioned probabilities should be considered vaccine type-specific, leading to

$$\phi(a,t) = \phi^{(1)}(a,t) + \phi^{(2)}(a,t),$$

with, for $j = 1, 2$ and $l = 1, 2$ denoting mRNA and adeno-based vaccines, and first and

second dose, respectively,

$$\phi^{(j)}(a,t) = w_{1,t}^{(j)}(a)\phi_1^{(j)}(a,t) + \phi_2^{(j)}(a,t),$$
$$\phi_l^{(j)}(a,t) = \psi_l^{(j)}(a,t)\sum_{k=0}^{t} p_{l,k}^{(j)}(a)\mathrm{ve}_{l,(t-k)}^{(j)}.$$

Note that this is a generalization of expression (1) in which vaccine types are assumed not be mixed. The functions $\psi_l^{(j)}(a,t)$ are allowed to differ by vaccine dose and vaccine type:

$$\psi_l^{(j)}(a,t) = \{1 - p_{\alpha,t}(a) - p_{\delta,t}(a)\} + p_{\alpha,t}(a)\frac{\mathrm{ve}_{l,\alpha}^{*(j)}}{\mathrm{ve}^{*(j)}} + p_{\delta,t}(a)\frac{\mathrm{ve}_{l,\delta}^{*(j)}}{\mathrm{ve}^{*(j)}}.$$

2.2.5 Additional booster vaccinations

In the presence of additional booster vaccinations, we can rely on a similar approach as considered in Section 2.2.4. More specifically, booster vaccination (using mRNA-based vaccines), denoted hereunder by $b_1, \ldots, b_B$, implies an increase in immunity according to functions $\phi_{b_1}(a,t), \ldots, \phi_{b_B}(a,t)$, with

$$\phi_{b_j}(a,t) = \psi_{b_j}(a,t)\sum_{k=0}^{t} p_{b_j,k}(a)\mathrm{ve}_{b_j,(t-k)}. \tag{2}$$

Note that vaccine properties captured by $\mathrm{ve}_{b_j,(t-k)}$ can be defined differently for booster vaccination $j = 1, \ldots, B$ and that protection is adapted for VoCs depending on their prevalence using $\psi_{b_j}(a,t)$. Note that booster vaccination is presumed to be only of the mRNA type, hence, there is no need to differentiate between booster vaccine types. Consequently, the total immunity in age group a at calendar time t, after first dose booster vaccination, is given by:

$$\phi(a,t) = w_{1,t}^{(1)}(a)\phi_1^{(1)}(a,t) + w_{2,t}^{(1)}(a)\phi_2^{(1)}(a,t) + w_{1,t}^{(2)}(a)\phi_1^{(2)}(a,t) + w_{2,t}^{(2)}(a)\phi_2^{(2)}(a,t) + \phi_{b_1}(a,t),$$

where the weights $w_{1,t}^{(j)}(a)$ and $w_{2,t}^{(j)}(a)$ define the probabilities of having received only a first dose of vaccine type j, or two doses of vaccine type j among vaccinated individuals in age group a at time t. A generalization towards more doses is straightforward by computing the appropriate probabilities.

Calculating the age- and vaccine-type specific burden in the presence of first, second and booster dose vaccination is straightforward and similar to the approach described in Section 2.2.2. For example, the prevented number of cases (first and second dose) as

a result of adeno-based or mRNA vaccination is given by $\left\{w_{1,t}^{(2)}(a)\phi_1^{(2)}(a,t) + w_{2,t}^{(2)}(a)\phi_2^{(2)}(a,t)\right\} n^*(a,t)$ or $\left\{w_{1,t}^{(1)}(a)\phi_1^{(1)}(a,t) + w_{2,t}^{(1)}(a)\phi_2^{(1)}(a,t)\right\} n^*(a,t)$, respectively. Consequently, the additional effect of the booster vaccination equals

$$\{\phi_{b_1}(a,t)\}\, n^*(a,t).$$

Essentially, this implies that the prevented burden as a result of first and second dose mRNA and adeno-based vaccination, and the additional prevented burden due to booster vaccination are relative shares of the total prevented burden.

2.2.6 Prevented hospitalizations, ICU admissions and deaths

In the description of the PM methodology, we mainly focused on COVID-19 confirmed cases. However, next to averted cases, we are also interested in prevented hospitalizations, ICU admissions and COVID-19 related deaths as a result of COVID-19 vaccination efforts. Although different approaches are possible to calculate, for example, the prevented number of hospitalizations (e.g., including an approach translating directly the prevented number of confirmed cases into prevented hospitalizations), we will focus here on an approach directly accounting for the reported number of hospitalizations (ICU admissions or deaths). More specifically, we use an approach which is similar to the approach considered to compute the counterfactual number of confirmed COVID-19 cases.

Using input parameters for vaccine effectiveness against hospitalization, we compute $\phi^H(a,t)$, the proportion of individuals in age group a that are protected against hospitalization at calendar time t. The counterfactual number of hospitalizations is again computed as

$$n^{H*}(a,t) = \frac{n^H(a,t)}{1-\phi^H(a,t-d^H)}.$$

The delay parameter d^H, denoting the delay between hospital admission and infection, is for simplicity set to 0.

The proportion of individuals protected against hospitalization can be expressed as:

$$\phi^H(a,t) = \sum_{k=0}^{t} p_k(a)\mathrm{ve}_{(t-k)}^H,$$

where $\mathrm{ve}_{(t-k)}^H$ is the vaccine effectiveness against hospitalization for time since vaccination equal to $t-k$. Vaccine effectiveness against hospitalization is considered to evolve according to a function $\mathrm{ve}_{(t-k)}^H$ which can be considered different from the evolution of vaccine

effectiveness against infection with time since vaccination. In the current implementation, a time- and age-invariant re-scaling according to a factor η (i.e., $\text{ve}^H_{(t-k)} = \eta \text{ve}_{(t-k)}$) is considered:

$$\eta = \frac{\text{ve}^{H*}}{\text{ve}^*},$$

with ve^{H*} and ve^* denoting the reported maximal vaccine effectiveness against hospitalization and infection, respectively.

Moreover, the impact of VoCs on the vaccine protection against hospitalization is included in a similar way as for confirmed cases, i.e.,

$$\phi^H(a,t) = \psi^H(a,t) \sum_{k=0}^{t} p_k(a) \text{ve}^H_{(t-k)},$$

with

$$\psi^H(a,t) = \{1 - p_{\alpha,t}(a) - p_{\delta,t}(a)\} + p_{\alpha,t}(a)\frac{\text{ve}^{H*}_\alpha}{\text{ve}^{H*}} + p_{\delta,t}(a)\frac{\text{ve}^{H*}_\delta}{\text{ve}^{H*}},$$

with correcting factors potentially being dependent on vaccine dose and vaccine type (see Section 2.2.4).

This method is readily generalizable to other clinical endpoints, however, it requires parameters for vaccine effectiveness against hospitalization (ICU admission, death) and relies on the availability of the number of hospitalizations (ICU admissions, deaths) per age group over time.

2.2.7 Imputation in case of missing data

In order to calculate the counterfactual number of confirmed cases, hospitalizations, ICU admissions and deaths, we require detailed information regarding vaccine uptake (including information on the number of administered doses for each vaccine brand) as well as age- and time-specific information on confirmed cases and hospitalizations. However, for some countries such information is lacking. In the PM, we opted to impute such missing information based on information for other countries for which the required data is available.

Imputing vaccine uptake for Liechtenstein

We re-scaled the reported age- and vaccine-type-specific uptake from Austria to the population of Liechtenstein.

Imputing vaccine brands

For nine countries, there is no vaccine brand data available. Currently, this information is imputed based on the proportion of COVID-19 vaccines of a given brand over time inferred based on the available vaccine uptake data for countries with such data.

Imputing the number of cases per age group

If no age-specific data is available for a country, use the overall age-distribution for all countries.

Imputing the number of hospitalizations per age group

For some of the countries, age-specific hospitalization, ICU admission or mortality data is lacking. Consequently, we need to rely on (imputation) methods to overcome these issues and to estimate the prevented number of hospitalizations, ICU admissions and deaths by age group.

2.2.8 Methods used for Risk Contextualization of COVID-19 Vaccines in the EU

2.2.8.1 Missing Age and/or Gender Information in Reported EudraVigilance Cases

In a small but relevant proportion of the reported EudraVigilance cases information on age and/or gender is missing. By taking the appropriate age and gender proportions of the completely reported EudraVigilance cases the missing information is imputed.

2.2.8.2 Pooling of the Background Rate Estimates Over All Data Sources

When risk background incidence rate (IR) estimates are available from different data sources, a benefit risk assessment can be performed with each of these estimates or the background rates can be pooled. Both the single source as the pooled background rate analyses are feasible in the toolkit. The toolkit automatically calculates the pooled background rates of any number of provided background incidence rate sources using the methodology detailed in this section.

The IR estimates from distinct data sources are pooled by inverse variance (σ^2) weighting so that estimates from a large data source, which are expected to be more precise, will be allowed to have a larger influence on the weighted mean [8]. Additionally, we allow for a manually chosen weight, x, so that more importance can be assigned, if needed, to a data source that is believed to produce more precise estimates (e.g., a primary care data source versus a primary and secondary care data source).

The variance for the incidence for each age category, gender and data source i combination:

$$\mathrm{IR}_i = \frac{c_i}{t_i} 100,000$$

is obtained from the symmetric Wald confidence limits for count c_i and person-years t_i. The Wald standard deviation $sd_i(c) = \sqrt{c_i}$, such that

$$\sigma_i^2(\mathrm{IR}) = \frac{c_i}{t_i^2} 100,000^2.$$

While the assumption of symmetry may not be entirely verified, for sufficiently large data sets, this is a viable approximation.

A weighted mean $\overline{\mathrm{IR}}$ of the different estimates per data source i can be obtained by:

$$\overline{\mathrm{IR}} = \frac{\sum_{i=1}^{n} \mathrm{IR}_i w_i}{\sum_{i=1}^{n} w_i},$$

with n the number of data sources and

$$w_i = \frac{x_i}{\sigma_i^2}.$$

If $x_i = 1$ for each i then each estimate contributes equally. By setting for instance $x_i = 10$ for one data source, the contribution of this data source to the mean will be more important. Finally, for a single source analysis of data source i, the weight for this data source is set to 1, while all other data sources receive a weight of zero.

The variance of the weighted mean is obtained by:

$$\mathrm{Var}(\overline{\mathrm{IR}}) = \frac{\sum_{i=1}^{n} x_i^2}{\left(\sum_{i=1}^{n} w_i\right)^2}.$$

2.2.8.3 Redistribution of Vaccination Coverage

Vaccination coverage data is available via two sources:

- The weekly updated ECDC vaccine tracker, includes weekly vaccine coverage from 27 European countries and Norway, Iceland and Liechtenstein by vaccine brand, dose and age (Age 0–4; **Age 5–9; Age 10–14; Age 15–17**; Age 18–24; **Age 25–49**; Age 50–59; Age 60–69; Age 70–79; Age 80+). Some countries report the vaccine coverage <18 years as an aggregate, while others report only an aggregate of the coverage ≥18 years. Gender information is absent.

- Solicited request (30 September 2021) to European Member States, includes aggregated vaccine coverage of 19 European countries for Vaxzevria (AstraZeneca), Comirnaty (Pfizer), Spikevax (Moderna) (ECDC countries minus Bulgaria (BG), Cyprus (CY), Denmark (DK), Germany (DE), Hungary (HU), Italy (IT), Liechtenstein (LI), Luxembourg (LU), Malta (MT), Poland (PL), Slovakia (SK)) and 15 countries for Janssen (19 countries minus Finland (FI), France (FR), Norway (NO), Sweden (SE)) by vaccine brand, dose, **gender** and age (**Age 5–11; Age 12–15; Age 16–17**; Age 18–24; **Age 25–29; Age 30–39; Age 40–49**; Age 50–59; Age 60–69; Age 70–79; Age 80+). Greece (GR or EL) uses age categories 12–14y and 15–17y rather than 12–15y and 16–17y.

As the ECDC data source contains the most up to date information on the vaccine coverage per country, but lacks gender information, the coverage per gender is imputed from gender proportions obtained from the Member States data. Additionally, a redistribution over some age categories is required as the age categories in the two data sources do not coincide and may differ from the age categories required for the risk assessment. The coverage per gender and required age category is obtained in a pre-processing step outside the toolkit, via two methodologies:

- **Fixed proportions**: Coverage redistribution per age category of countries that report only aggregates to ECDC by the age proportions of the countries that do report coverage per age categories. Subsequently, coverage redistribution by gender and the required age categories by using the Member States gender and age coverage proportions.
- **Multiple imputation**: Coverage redistribution of the ECDC data via multiple imputation to gender and the required age categories by using the Member States age and gender coverage proportions.

Let $n_{c,a_r,s}$ be the vaccination coverage in country c, required age category a_r and gender s. The vaccination coverage is a three-way table of which for some countries only marginal (or aggregated) counts n_{+c,a_r} are observed or in an overlapping age category.

To redistribute the vaccine coverage in the ECDC vaccine tracker to the gender and required age category, the observed vaccine coverage is multiplied with the age gender proportion $\pi_{a,s}$ taken over all countries:

$$n_{c,a_r,s} = n_{c,a_i,s}\pi_{a_r,s},$$

where a_i can be in the required age category a_r or an aggregated or overlapping age category a_k.
For a given age category a, the vaccine coverage $n_{c,a_r,s}$ follows a multinomial distribution with probability $\pi_{a_r,s}$, which can be estimated from the reported marginal (aggregated) counts:

$$\widehat{\pi}_{a_r,s} = \frac{n_{+a_r,s}}{n_{++a_r}}$$

with variance

$$\text{Var}(\widehat{\pi}_{a_r,s}) = \frac{\widehat{\pi}_{a_r,s}(1-\widehat{\pi}_{a_r,s})}{n_{++a_r}}.$$

In the **fixed proportion** redistribution, it is assumed that the vaccine coverage follows the same multinomial distribution in all countries, where the proportion $\widehat{\pi}_{a_r,s}$ is estimated from the Member States data. For countries, where coverage is reported aggregated over age categories a_k, the proportion is additionally multiplied with $\widehat{\pi}_{a_r} = \frac{n_{+a_r}}{n_{+a_k}}$, the proportion of the multinomial age distribution obtained from the ECDC countries with coverage reported by a_r age categories.

In the **multiple imputation** redistribution, it is assumed that the multinomial age-gender distribution of the vaccine coverage in all countries may or may not be equal to the estimate over the Member State countries with reported age categories. The uncertainty in the age-gender distribution, is allowed by drawing $m = 1, \ldots, M$, times $\pi_{s,a}^{(m)}$ from $N\left(\widehat{\pi_{s,a}}, \frac{\widehat{\pi_{s,a}}(1-\widehat{\pi_{s,a}})}{n_{++a}}\right)$ and by subsequently drawing $n_{c,a}$ copies from Binom $\left(\pi_{s,a}^{(m)}\right)$ for the gender redistribution. Or by drawing from a multivariate normal distribution with a diagonal variance-covariance matrix and subsequently a multinomial distribution for gender and age category redistribution. Both result in M estimates $n_{c,a,s}^{(m)}$ of the vaccine type coverage in that country, age group and gender.

2.2.8.4 Observed-Expected Ratio

In case the **fixed proportions** are used for the redistribution of the vaccine coverage data, the Observed-Expected risk ratio is obtained by:

$$R_{a,s} = \frac{EV_{a,s}100,000}{\text{IR}_{a,s}n_{a,s}t}$$

with $EV_{a,s}$ the reported risk cases to EudraVigilance, $n_{a,s} = \sum^{c} n_{c,a,s}$ and t the time horizon in years. The confidence interval of the Observed-Expected ratio is obtained by applying the Wilson and Hilferty [9] approximation for chi-square percentiles [10]:

$$LL = \left(1 - \frac{1}{9EV_{a,s}} - \frac{z_{1-\alpha/2}}{3\sqrt{EV_{a,s}}}\right)^3 \frac{EV_{a,s}100,000}{\text{IR}_{a,s}n_{a,s}t}$$

$$UL = \left(1 - \frac{1}{9(EV_{a,s}+1)} + \frac{z_{1-\alpha/2}}{3\sqrt{EV_{a,s}+1}}\right)^3 \frac{(EV_{a,s}+1)100,000}{\text{IR}_{a,s}n_{a,s}t}$$

In the case **multiple imputation** is used for the redistribution of the vaccine coverage, we have per multiple imputation m, the coverage $n_{c,a,s}^{(m)}$ per vaccine type, age, gender

and country and its variance $\text{Var}(n_{c,a,s}^{(m)}) = n_{c,a,s}^{(m)}\pi_{s,a}^{(m)}\left(1-\pi_{s,a}^{(m)}\right)$.

For each multiple imputation m, the Observed-Expected ratio $R_{a,s}^{(m)}$ per age and gender is obtained by summing $n_{c,a,s}^{(m)}$ and $\text{Var}(n_{c,a,s}^{(m)})$ over all countries, assuming coverage between countries is independent, and by:

$$R_{a,s}^{(m)} = \frac{EV_{a,s}100,000}{\text{IR}_{a,s}n_{a,s}^{(m)}t}$$

The variance of $R_{a,s}^{(m)}$ consists of the variability of the imputed $n_{a,s}^{(m)}$, with or without the variability of the pooled $IR_{a,s}$ which are considered independent:

$$\text{Var}\left(R_{a,s}^{(m)}\right) = \left(\frac{EV_{a,s}100,000}{t}\right)^2 \frac{\text{Var}(\text{IR}_{a,s})\text{Var}\left(n_{a,s}^{(m)}\right) + \text{Var}(\text{IR}_{a,s})\left(E\left(n_{a,s}^{(m)}\right)\right)^2 + \text{Var}\left(n_{a,s}^{(m)}\right)(E(\text{IR}_{a,s}))^2}{\left(E(\text{IR}_{a,s})E\left(n_{a,s}^{(m)}\right)\right)^4}.$$

If it is decided not to pool the background risk incidence rates from the various data sources, the $\text{Var}\left(IR_{a,s}\right)$ becomes zero in the above formula.

Finally, the $R_{a,s}^{(m)}$ are combined using Rubin's rule to obtain the observed-risk ratio per age category and gender $\overline{R_{a,s}}$, averaged over the multiple imputation estimates:

$$\overline{R_{a,s}} = \frac{\sum_{m=1}^{M} R_{a,s}^{(m)}}{M},$$

with variability:

$$\text{Var}(R_{a,s}) = \frac{\sum_{m=1}^{M}\text{Var}\left(R_{a,s}^{(m)}\right)}{M} + \left(\frac{M+1}{M}\right)\left(\frac{\sum_{m=1}^{M}\left(R_{a,s}^{(m)} - \overline{R_{a,s}}\right)^2}{M-1}\right).$$

2.2.9 Expected prevented or additional burden

Vaccination may cause risk events or prevent risk events associated with both vaccination and COVID-19. The expected prevented or additional burden of vaccination is expressed as the difference between the observed and expected risk events per age category and gender:

$$D_{a,s} = EV_{a,s} - \frac{\text{IR}_{a,s}n_{a,s}t}{100,000}.$$

3 Prerequisites

3.1 Data sets

The tool uses a total of 8 data sets (5 for the benefits and 3 for the risks). For each of these data sets the user can either upload a new data set or use one of the default data sets provided by the application owner.

3.1.1 Default data sets

The following data sets are available as default in the application:

- **Observed risks after vaccination**: The observed risk events after vaccination by vaccine, age and gender, recorded by EudraVigilance for myocarditis on 13 October 2021 and for TTS on 25 July 2021. For myocarditis, 139 of 4635 cases with missing age and or gender information (24 missing gender, 111 missing age and 4 missing both gender and age) have been redistributed as outlined in 2.2.8. For TTS 20 of 503 TTS events, 30 days after Vaxzevria vaccination, with missing age and or gender information (19 missing age and 1 missing gender) have been redistributed.
- **Background rate risks**: The background incidence risk for myocarditis per age group and gender are provided by EMA dd. 10 November 2021, including estimates from three data sources (ARS, BIFAP PCHOSP and BIFAP PCCOVID) in the period before the COVID-19 pandemic and during the COVID-19 pandemic prior to vaccination.
- **Aggregated vaccine data (imputed)**: The aggregated vaccine coverage by vaccine, age group and gender aggregated over time and countries. The default data contains vaccination coverage data from the European Union/European Economic Area (EU/EEA) countries' Vaccine Tracker submissions to ECDC via the European Surveillance System (TESSy), de dato 10 February 2021, but restricted to 3 October 2021 for myocarditis and 11 July for TTS, redistributed according to two methods as described in 2.2.8, where the multiple imputation method uses the vaccination coverage requested to the Member States by EMA, de dato 30 September 2021. An R-script has been made available to EMA to impute the coverage via the multiple imputation method.
- **Country demographics data**: The population per country and 10 year bands age categories from EUROSTAT in December 2021.
- **Granular COVID19 data**: The default data set contains case-based data between January until the end of March 2021 by 10-year age categories and countries for the incidence of confirmed COVID-19 cases, and limited information on hospitalizations, ICU admissions, and COVID-19 related deaths. The data is obtained from both the EMA as well as the ECDC datasets for 20 countries.

- **Aggregated COVID19 data**: The default data set contains the aggregated number per country per day for cases, ICU admissions, hospitalisations, mortality between 26 February 2020 and February 9 2022.
- **COVID19 variants of concern data**: The proportion of each variant of concern per number of sequenced SARS-CoV-2 strain, by day and 30 country, derived from the GISAID EpiCoV database and TESSy since 11 May 2020.
- **Granular vaccine data**: daily new vaccinations per vaccine, dose, date, country and age group.

3.1.2 Adding a default data set (for the application owner)

The application owner can add a default data set if needed. Data stored in the default data directory is available to all users of the application. For the application to automatically recognize these data files they should be named

[name data set]_[name displayed in app].rds,

with the [name data set] and [name displayed in app] combination as detailed in Table 1. Default data sets have to be uploaded in '.rds' format following the specification listed in Section 3.1.4 and 3.1.5.

Table 1: Default data set names

[name data set]	[name displayed in app]
age_specific_covid_data	Granular COVID-19 data
age_specific_vaccine_data	Granular vaccine data
aggregated_covid_data	Aggregated COVID-19 data
background_risks_data	Background rate risks
observed_risks_data	Observed risk after vaccination
population_data	Country demographics data
vaccine_aggregated_data	Aggregated vaccine data
variant_data	COVID-19 variants of concern

3.1.3 Uploading data within the application

As a user you can upload data within the application. Uploaded data files can be in the '.rds' or '.csv' format. Uploaded '.csv' files should use a semicolon (;) as the delimiter. When a new data set is uploaded the application checks the requirements listed in Section 3.1.4 and 3.1.5.

3.1.4 Benefit

Country demographics data File containing the number of inhabitants per country and age group. This file determines for which countries and age groups the benefit analysis is performed. Requirements:

- Required columns: country, age_group, population
- Required format age group $XX-YY$, $-XX$, $\leq XX$, $\geq YY$, $YY+$, with XX and YY numeric. For example: 0-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80+

Table 2: Country demographics data

No	Variable name	Description	Required values
1	country	Country names used in the analysis	No requirements
2	age_group	Age group for specific time period	Allowed formats $XX-YY$, $-XX$, $\leq XX$, $\geq YY$, $YY+$, with XX and YY numeric
3	population	Age-specific population for the country	Numeric

Granular COVID-19 data File containing the age distribution of cases, hospitalizations, ICU admission and mortality per country. Requirements:

- Required columns: country, age_group, cases, hospitalization, icu, mortality
- Data required for all countries specified in the country demographics data
- Data required for all age groups specified in the country demographics data
- Sum of the proportion of cases, hospitalization, icu and mortality over the age groups can not add up to more than one per country

Table 3: Granular COVID-19 data

No	Variable name	Description	Required values
1	country	Country names used in the analysis	values from country demographics
2	age_group	Age group for specific time period	values from country demographics
3	cases	proportion of age group in the total number of cases	Proportion
4	hospitalization	proportion of age group in the total number of hospitalizations	Proportion
5	icu	proportion of age group in the total number of ICU admissions	Proportion
6	mortality	proportion of age group in the total number of deaths	Proportion

* = optional

Aggregated COVID-19 data File containing the total number of cases, hospitalizations, ICU admission and mortality per date per country. Requirements:

- Required columns: date, country, cases, hospitalization, icu, mortality
- Data required for all countries specified in the country demographics data

- The model can only be used for the date range for which all countries report cases, hospitalizations, icu admission and deaths

Table 4: Aggregated COVID-19 data

No	Variable name	Description	Required values
1	date	Date of reporting	yyyy-mm-dd
2	country	Country names used in the analysis	values from country demographics
3	cases	total number of cases over all age groups	Numeric
4	hospitalization	total number of hospitalizations over all age groups	Numeric
5	icu	total number of ICU admissions over all age groups	Numeric
6	mortality	total number of deaths over all age groups	Numeric

COVID-19 variants of concern File containing for each date the distribution of variants of concern per country. Requirements:

- Required columns: date, country, variant, proportion
- Data required for all countries specified in the country demographics data
- Allowed values for variant are Original, Alpha and Delta
- The model can only be used for the date range for which all countries report the distribution of variants of concern

Table 5: COVID-19 variants of concern data

No	Variable name	Description	Required values
1	date	Date of reporting	yyyy-mm-dd
2	country	Country names used in the analysis	values from country demographics
3	variant	COVID-19 variants	Original, Alpha or Delta
4	proportion	$\frac{\text{number of detections variant}}{\text{sequenced numbers}}$	Proportion

Granular vaccine data File containing for each date, country and age group the number of new vaccinations per vaccine and dose. Requirements:

- Required columns: date, country, age_group, vaccine, dose, count
- Data required for all countries specified in the country demographics data
- Data required for all age groups specified in the country demographics data
- Allowed values for dose are 1, 2 and 3
- Vaccine names should match with vaccine names specified in the risk data sets
- For a given dose, the total number of vaccines in a given country and age_group cannot exceed the total population size registered in the country demographics data

- The model can only be used for the date range for which all countries report vaccine uptake

Table 6: Granular vaccine data

No	Variable name	Description	Required values
1	date	Date of reporting	yyyy-mm-dd
2	country	Country names used in the analysis	values from country demographics
2	age_group	Age group for specific time period	values from country demographics
4	vaccine	Name of the vaccine	match values risk data
5	dose	administered dose	1, 2 or 3
6	count	Number of vaccine administered to individuals on this date	Numeric

3.1.5 Risk

Three input data sets for the risk assessment are required. A data set containing the reported risk events, a data set with risk background incidence rate estimates and a data set with the vaccine coverage. The required attributes/variables and their description are presented in Table 7 - 9.

Observed risk after vaccination File containing for each combination of age group, vaccine and sex the total number of reported cases per condition aggregated over all countries included in the benefit analysis and over the date range risk reporting is available. Multiple conditions can be included in this file and the age groups are allowed to differ per condition. Requirements:

- Required columns: age_group, sex, cases, condition, vaccine
- Required format age group $XX - YY$, $-XX$, $\leq XX$, $\geq YY$, $YY+$, with XX and YY numeric. For example: 0-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80+. The age group specified in this file per condition determines the age groups used in the benefit-risk analysis. Age groups can differ per condition.
- Allowed values for sex are either F and M or Female and Male
- Vaccine names should match those in the benefit data

Table 7: Observed risks after vaccination data

No	Variable name	Description	Required values
1	age_group	Age group for specific time period	Allowed formats $XX - YY$, $-XX$, $\leq XX$, $\geq YY$, $YY+$, with XX and YY numeric
2	sex	Gender of the reported cases	M/F or Male/Female
3	cases	Number of reported risk cases	Numeric
4	condition	Type of risk	f.e. myocarditis, TTS,...
5	vaccine	Type of vaccine	match values benefit data

Background rate risks File containing cumulative number of risk cases observed per person years by age category and gender. Multiple data sources estimating the background risk, multiple risks and multiple periods in which the background risk is estimated can be joined in a single input data set. Requirements:

- Required columns: source, condition, age_group, sex, person_years, cases and period
- Condition should match those present in the 'observed risk after vaccination' data. If no background rate is available for a condition, data should be included with cases set to zero
- Per condition age_group should match the age groups specified in 'observed risk after vaccination'

Table 8: Background rate risk data

No	Variable name	Description	Required values
1	source	Source of background risk estimation	f.e. ARS, BIFAP PCCOVID,..
2	condition	Type of risk	f.e. myocarditis pericarditis; TTS
3	age_group	Age category	same as in observed risk data
4	sex	Gender of the estimated incidence rate	same as in observed risk data
5	person_years	Denominator for the incidence rate	Numeric
6	cases	Number of reported cases of the condition Numerator for the incidence rate	Numeric
7	period	period in which background rate is estimated	f.e. pre pandemic, intra-pandemic,..

Aggregated vaccine data File containing the aggregated vaccine coverage by vaccine, age group and gender aggregated over the analysis period and over the countries included in the benefit data. Two methods for redistribution of the vaccine coverage for missing gender or age are included 'multiple imputation' and 'fixed proportion'. Results from both redistribution methods are uploaded in a single data set. Requirements:

- Required columns: vaccine, sex, age_group, imputation_id, cov, var, imputation_method
- Values for vaccine, sex and age_group should match those included in the 'observed risk after vaccination' data.
- Imputation_id has to be set to 1 when imputation_method = fixed proportion

Table 9: Aggregated vaccine data (imputed)

No	Variable name	Description	Required values
1	vaccine	Type of vaccine	match observed risk data
2	sex	Gender of the vaccine coverage	match observed risk data
3	age_group	Age category	match observed risk data
4	imputation_id	=1 for fixed proportion method =1,..,m for m multiple imputations	Numeric
5	cov	Number of vaccines delivered	Numeric
6	var	Variance of the coverage due to multiple imputation =1 for fixed proportion method	Numeric
7	imputation_method	Method of redistribution of coverage	multiple imputation, fixed proportion

3.2 Benefit parameters

In this section, we present the default vaccine effectiveness parameters with regard to clinical outcome (i.e., infection, hospitalization, ICU admission and mortality) used throughout the analysis. In Table 10, these vaccine properties are listed by vaccine brand, endpoint and variant of concern. For endpoints for which detailed information was lacking, we consider parameter values from a less severe endpoint respecting the order infection, hospitalization, ICU admission and finally death. As vaccination appears to protect against serious disease caused by all of the major virus types, most vaccinations are less effective against symptomatic sickness due to the delta form than the alpha and Wuhan variant, the delta variant still has significant vaccine efficacy against both symptomatic and severe disease. However, evidence from other coronaviruses suggests that immunity against SARS-CoV-2 may be short-lived, lasting 12–18 months depending on a variety of conditions [11]. Needless to say, these parameter values can be altered in the toolkit (based on novel insights described in the scientific literature) to explore the impact thereof on the final benefit-risk quantification.

Table 10: Overview of the default parameters used in the analysis and toolkit. These parameter values correspond to the ve^*, ve^*_α and ve^*_δ parameters mentioned above (allowing for differential protection depending on the vaccine dose).

Vaccine type	Vaccine dose	Variant	Clinical outcome	Value	Source
Comirnaty	Dose 1	Wuhan	Infection	0.52	Creech et al. (2021) [12]
		Alpha		0.48	Bernal et al. (2021) [13]
		Delta		0.36	Bernal et al. (2021) [13]
Comirnaty	Dose 1	Wuhan	Hospitalization	0.89	Creech et al. (2021) [12]
		Alpha		0.83	Stowe et al. (2021) [14]
		Delta		0.94	Stowe et al. (2021) [14]
Comirnaty	Dose 1	Wuhan	ICU admission	0.89	–
		Alpha		0.83	–
		Delta		0.94	–
Comirnaty	Dose 1	Wuhan	Death	0.89	–
		Alpha		0.83	–
		Delta		0.94	–
Comirnaty	Dose 2	Wuhan	Infection	0.95	Creech et al. (2021) [12]
		Alpha		0.94	Bernal et al. (2021) [13]
		Delta		0.88	Bernal et al. (2021) [13]
Comirnaty	Dose 2	Wuhan	Hospitalization	0.95	–
		Alpha		0.95	Stowe et al. (2021) [14]
		Delta		0.96	Stowe et al. (2021) [14]
Comirnaty	Dose 2	Wuhan	ICU admission	0.95	–
		Alpha		0.95	–
		Delta		0.96	–
Comirnaty	Dose 2	Wuhan	Death	0.95	–
		Alpha		0.95	–
		Delta		0.96	–
Comirnaty	Dose 3	Wuhan	Infection	0.95	–
		Alpha		0.94	–
		Delta		0.94	Andrews et al. (2022b) [15]
Comirnaty	Dose 3	Wuhan	Hospitalization	0.95	–
		Alpha		0.95	–
		Delta		0.98	Andrews et al. (2022b) [15]
Comirnaty	Dose 3	Wuhan	ICU admission	0.95	–
		Alpha		0.95	–
		Delta		0.98	–
Comirnaty	Dose 3	Wuhan	Death	0.95	–
		Alpha		0.95	–

Table 10: Overview of the default parameters used in the analysis and toolkit. These parameter values correspond to the ve^*, ve^*_α and ve^*_δ parameters mentioned above (allowing for differential protection depending on the vaccine dose).

Vaccine type	Vaccine dose	Variant	Clinical outcome	Value	Source
		Delta		0.98	Andrews et al. (2022b) [15]
Spikevax	Dose 1	Wuhan	Infection	0.92	Creech et al. (2021) [12]
		Alpha		0.90	Bruxvoort et al. (2021) [16]
		Delta		0.77	Bruxvoort et al. (2021) [16]
Spikevax	Dose 1	Wuhan	Hospitalization	0.92	–
		Alpha		0.90	–
		Delta		0.77	–
Spikevax	Dose 1	Wuhan	ICU admission	0.92	–
		Alpha		0.90	–
		Delta		0.77	–
Spikevax	Dose 1	Wuhan	Death	0.92	–
		Alpha		0.90	–
		Delta		0.77	–
Spikevax	Dose 2	Wuhan	Infection	0.94	Creech et al. (2021) [12]
		Alpha		0.98	Bruxvoort et al. (2021) [16]
		Delta		0.87	Bruxvoort et al. (2021) [16]
Spikevax	Dose 2	Wuhan	Hospitalization	0.98	Creech et al. (2021) [12]
		Alpha		0.98	–
		Delta		0.98	Bruxvoort et al. (2021) [16]; Andrews et al. (2022a)
Spikevax	Dose 2	Wuhan	ICU admission	0.98	–
		Alpha		0.98	–
		Delta		0.98	–
Spikevax	Dose 2	Wuhan	Death	0.98	–
		Alpha		0.98	–
		Delta		0.98	–
Spikevax	Dose 3	Wuhan	Infection	0.94	–
		Alpha		0.98	–
		Delta		0.95	Andrews et al. (2022b) [15]
Spikevax	Dose 3	Wuhan	Hospitalization	0.98	–
		Alpha		0.98	–

Table 10: Overview of the default parameters used in the analysis and toolkit. These parameter values correspond to the ve^*, ve^*_α and ve^*_δ parameters mentioned above (allowing for differential protection depending on the vaccine dose).

Vaccine type	Vaccine dose	Variant	Clinical outcome	Value	Source
		Delta		0.98	–
Spikevax	Dose 3	Wuhan	ICU admission	0.98	–
		Alpha		0.98	–
		Delta		0.98	–
Spikevax	Dose 2	Wuhan	Death	0.98	–
		Alpha		0.98	–
		Delta		0.98	–
Vaxzevria	Dose 1	Wuhan	Infection	0.64	Creech et al. (2021) [12]
		Alpha		0.49	Bernal et al. (2021) [13]
		Delta		0.30	Bernal et al. (2021) [13]
Vaxzevria	Dose 1	Wuhan	Hospitalization	0.95	Creech et al. (2021) [12]
		Alpha		0.76	Stowe et al. (2021) [14]
		Delta		0.71	Stowe et al. (2021) [14]
Vaxzevria	Dose 1	Wuhan	ICU admission	0.95	–
		Alpha		0.76	–
		Delta		0.71	–
Vaxzevria	Dose 1	Wuhan	Death	0.95	–
		Alpha		0.80	Sonabend et al. (2021) [17]
		Delta		0.80	Sonabend et al. (2021) [17]
Vaxzevria	Dose 2	Wuhan	Infection	0.70	Creech et al. (2021) [12]
		Alpha		0.75	Bernal et al. (2021) [13]
		Delta		0.67	Bernal et al. (2021) [13]
Vaxzevria	Dose 2	Wuhan	Hospitalization	0.98	Creech et al. (2021) [12]
		Alpha		0.86	Stowe et al. (2021) [14]
		Delta		0.92	Stowe et al. (2021) [14]
Vaxzevria	Dose 2	Wuhan	ICU admission	0.98	–
		Alpha		0.86	–
		Delta		0.92	–
Vaxzevria	Dose 2	Wuhan	Death	0.98	–
		Alpha		0.95	Sonabend et al. (2021) [17]
		Delta		0.95	Sonabend et al. (2021) [17]

3.3 Online tool

The Benefit-Risk Contextualization of COVID-19 Vaccines in the EU tool is available online through Shinyapps at https://dsi-uhasselt.shinyapps.io/covid_vaccine_risks_and_benefits/.

A username and password has been sent in a separate email to EMA. An active Internet connection is required. It is advised to use relatively recent versions of web browsers such as Chrome, Firefox, Internet Explorer, Edge or Safari with support for JavaScript enabled.

4 Getting Started

This section provides detailed, step-by-step system operating instructions.

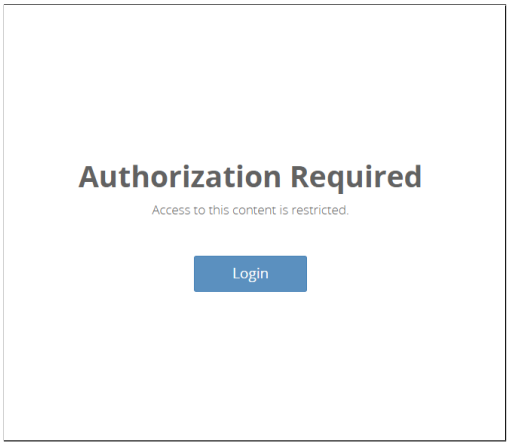
1. **Navigating the Shiny Web Application**
   The risk-benefit contextualization web application is a browser-based interface that you can use to assess the benefit-risk of COVID-19 vaccines. The interface has the following parts:

   (a) Access
   (b) Interface
      i. Banner
      ii. Menu Panel
   (c) Main Panel

2. **Starting the application**
   In this section, we will go over how to use the site address **to enter data as input data**, how to update the data when it needs to be updated, and how to use the static and dynamic reports.

   2.1. **Access**
   Open the application with the following URL address in a web browser (e.g., Google Chrome, Mozilla FireFox, or others):
   https://dsi-uhasselt.shinyapps.io/covid_vaccine_risks_and_benefits/.

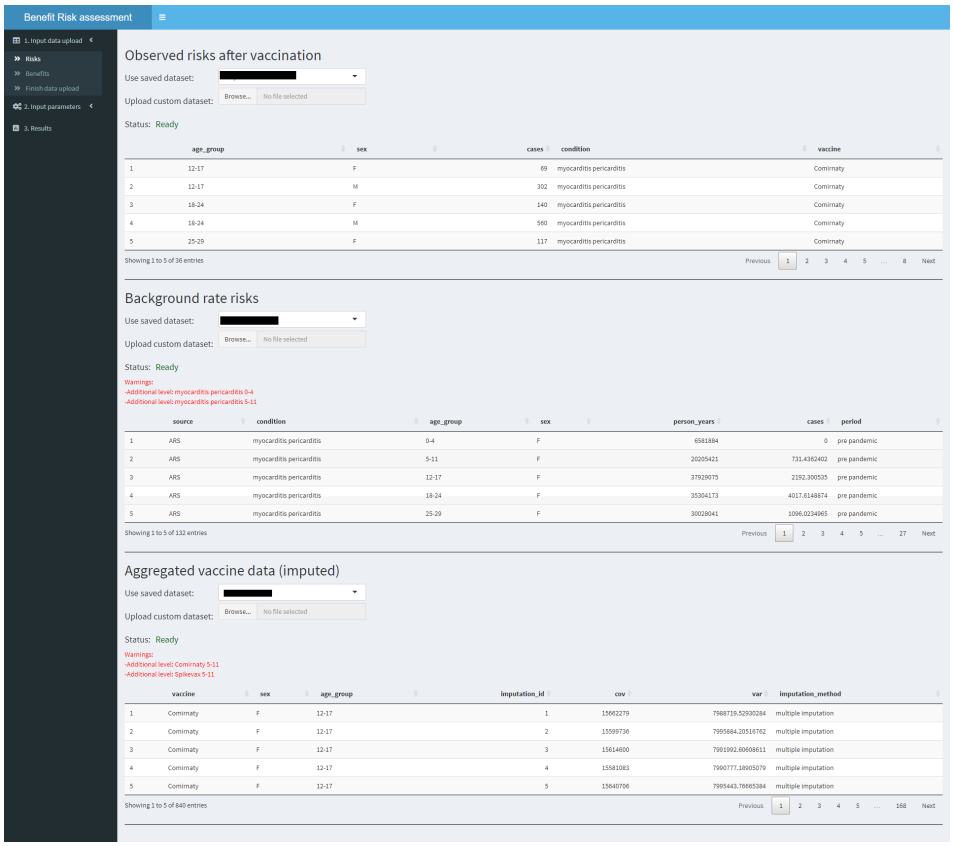
   - An authorization webpage will appear on your screen. If you wish to continue, go to the Login Page by clicking the **"Login"** icon on the webpage



   - EMA users can gain access by signing up for a free RStudio account and contacting our team mentioning access to the Risk-Benefit assessment tool. This e-mail should specify the RStudio account to which access should be granted.

2.2. Interface

As soon as you log in to the system, the dashboard will appear in the main page.

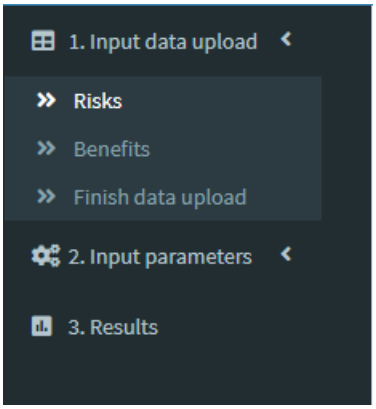


- **Banner**

  The top part of the dashboard is the banner. The banner, shown across the top of the interface and is always displayed.

- **Menu Panel**

  The Menu Panel, which is located on the left side of the dashboard includes three sections, namely, the **"1. Input data upload"**, **"2. Input parameters"**, and **"3. Results"**. The **Arrow Sign** (<) to the right of the designation can be clicked to reveal detailed features.

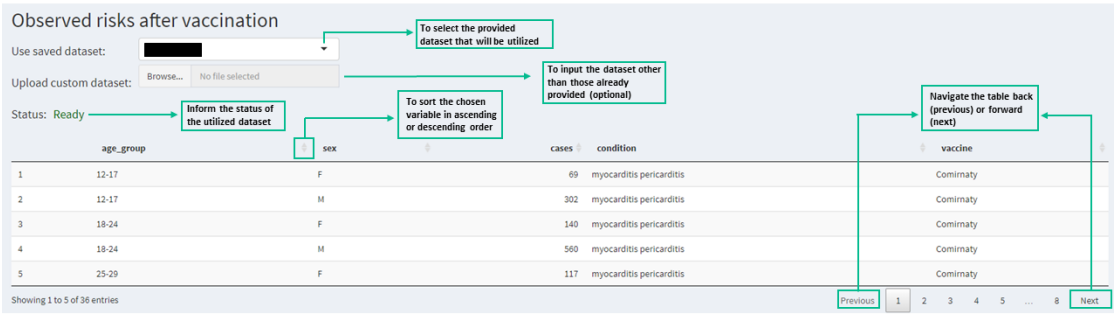


Note: The sections need to be completed in order, and the User cannot return to a previous section once completed. Features within a section can be completed in any order and revisited. Refreshing the page will clear all input entered by the User and return to the initial menu set-up.

1. Input data upload

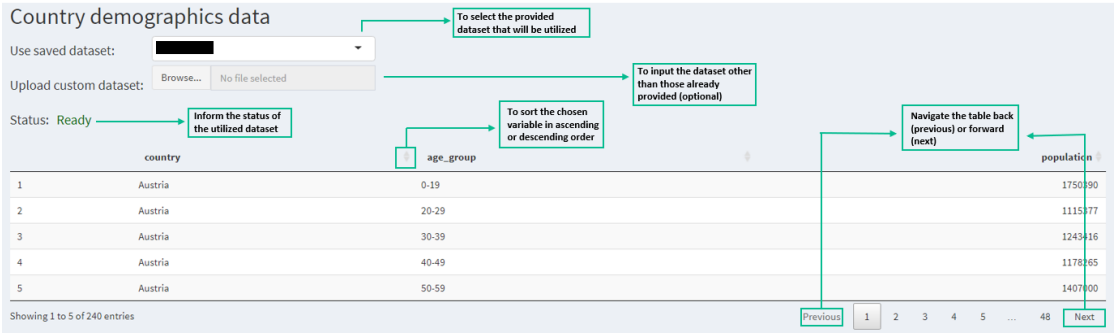
>> Risks

The **Risks** section, requires 3 (three) data sets as input, namely **"Observed risk after vaccination"**, **"Background rate risks"**, and **"Aggregated vaccine data (imputed)"**, which has been described in Sections 3.1.1 and 3.1.5.



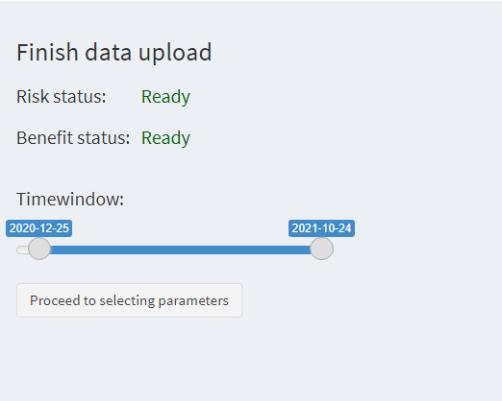
>> Benefits

The **Benefits** section, requires 5 (five) data sets as input, namely **"Country demographics data"**, **"Granular COVID-19 data"**, **"Aggregated COVID-19 data"**, **"COVID-19 variants of concern data"**, and **"Granular vaccine data"**, which have been described in Sections 3.1.1 and 3.1.4.



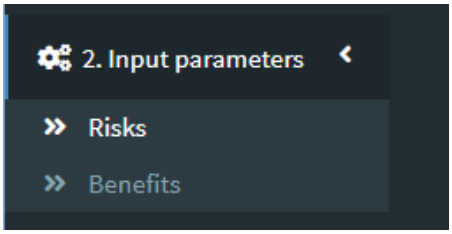
>> **Finish data upload**

When all input data is uploaded or a data set integrated in the tool is selected for both the Risk and Benefit, a **"Ready"** message will be displayed in the **"Finish data upload"**, next to the Risk and Benefit Status. The tool computes the range over which the benefit analysis can be performed. **The user should modify the time window of the benefit analysis to the range over which the benefit data is aggregated.** When all input data sets and the time window for the vaccine benefit estimation are selected, the User may press **"Proceed to selecting parameters"** to go to the next section. Note that the User cannot return the **"Input data upload"** section, once moving to the next section. Note that the time window in this subsection will adjust with new data imputed.



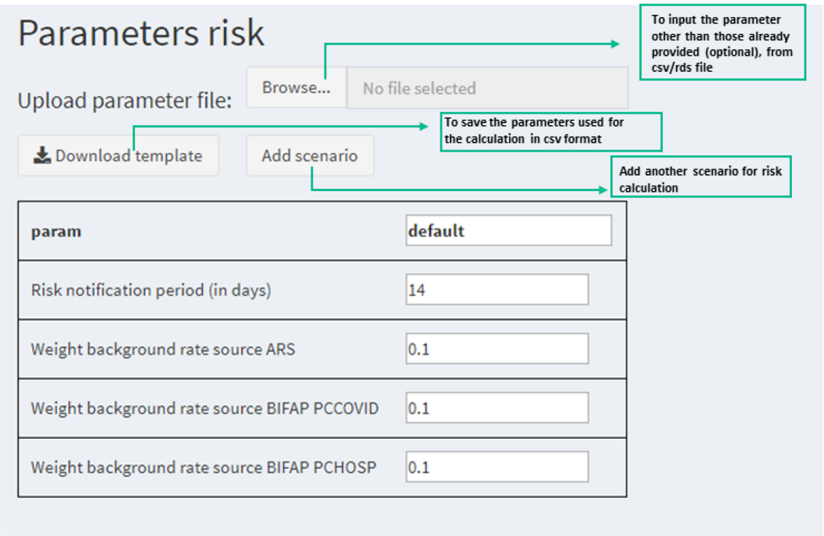
2. **Input parameters**

In this section, Users can specify the parameters used in the risk/benefit analysis. When available the tool will suggest default parameters. These default parameters can be found and edited by the application host in the 'data/' directory. The input parameters are divided into two parts, namely, **"Risks"** and **"Benefits"**.



>> **Risks**

In this part, Users can choose the main parameters for the risk analysis. Via the add scenario button additional scenarios can be added for comparison. All parameters should be completed for the first scenario (baseline). In the additional scenarios parameters can remain blank in which case they will be replaced by the parameter value of the baseline scenario. Users can upload a parameter file in csv or rds format using the **"Browse"** button. Note that uploaded csv files should use a semicolon (;) as the delimiter. A template of the parameter file can be downloaded by clicking on the **"Download template"** button.



>> **Benefits**

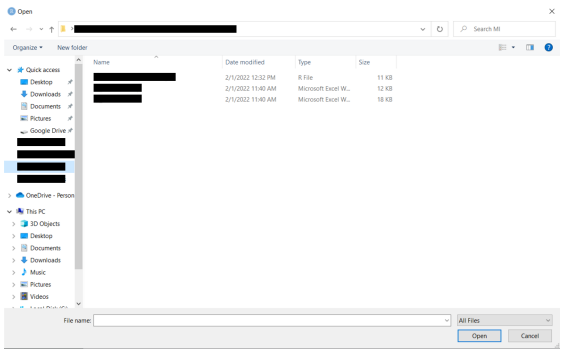
In this part, Users can choose the main parameters for the benefit analysis. Via the add scenario button additional scenarios can be added for comparison. All parameters should be completed for the first scenario (baseline). In the additional scenarios parameters can remain blank in which case they will be replaced by the parameter value of the baseline scenario. Users can upload a parameter file in csv or rds format using the **"Browse"** button. Note that uploaded

csv files should use a semicolon (;) as the delimiter. A template of the parameter file can be downloaded by clicking on the **"Download template"** button.

vaccine	dose	variant	against	parameter	[redacted]
Comirnaty	1	Alpha	cases	efficacy	0.48
Comirnaty	1	Delta	cases	efficacy	0.36
Comirnaty	1	Original	cases	efficacy	0.52
Comirnaty	1	Alpha	hospitalization	efficacy	0.83
Comirnaty	1	Delta	hospitalization	efficacy	0.94
Comirnaty	1	Original	hospitalization	efficacy	0.89
Comirnaty	1	Alpha	icu	efficacy	0.83
Comirnaty	1	Delta	icu	efficacy	0.94
Comirnaty	1	Original	icu	efficacy	0.89
Comirnaty	1	Alpha	mortality	efficacy	0.83
Comirnaty	1	Delta	mortality	efficacy	0.94
Comirnaty	1	Original	mortality	efficacy	0.89
Comirnaty	2	Alpha	cases	efficacy	0.94

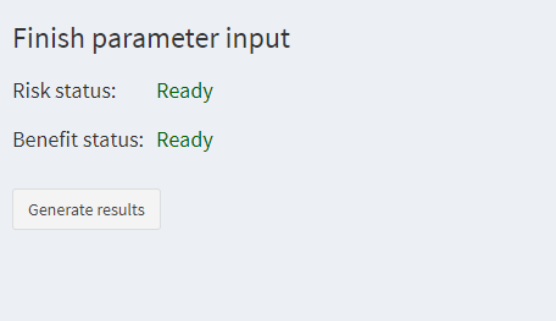
Note:

The **"Browse"** button in each subsection can be used by users who want to add their own data set or parameters to the dashboard. A window will appear, and you can select any file you need for the chosen subsection by clicking it, and then click the **"Open"** button in the right-bottom part of the window.



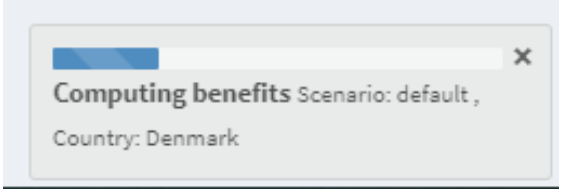
>> **Finish parameter input**

When all parameters are entered or a set of parameters integrated in the tool is selected for both the Risk and Benefit, a **"Ready"** message will be displayed in the **"Finish parameter input"**, next to the Risk and Benefit Status and the User may press **"Generate results"** to go to the Result section. Note that the User cannot return the previous sections, once moving to the Result section.



>> **Results**

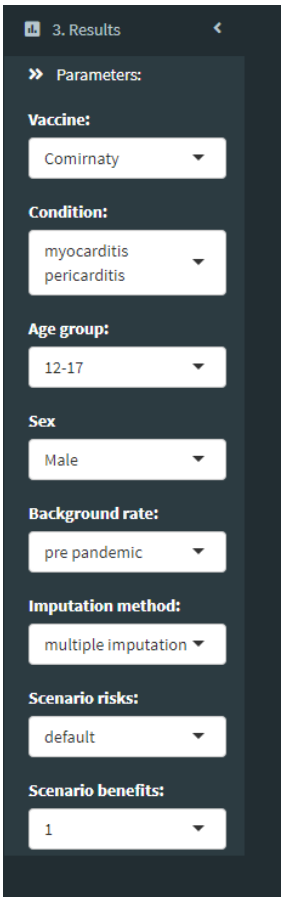
When pressing **"Finish parameter input"** the Results section will open automatically. A little tab will emerge in the bottom-right corner of the screen to indicate the user the loading process of the calculation before the results appear on the screen. The tab indicates the country and scenario that is currently being progressed.



On the left of the screen, drop-down menu's with display options are available for Users to select the presentation of the benefit/risk assessment results by the Users preference. The display options per drop-down menu will appear when pressing the upside down triangle symbol (▼). The display options are divided into several categories, namely:

- ☐ **Vaccine:** All vaccines included in the risk data.
- ☐ **Condition:** All conditions specified in the risk data
- ☐ **Age group:** Age group specified in the risk data set for the selected condition. The available ages in this select box will alter when the condition is changed.

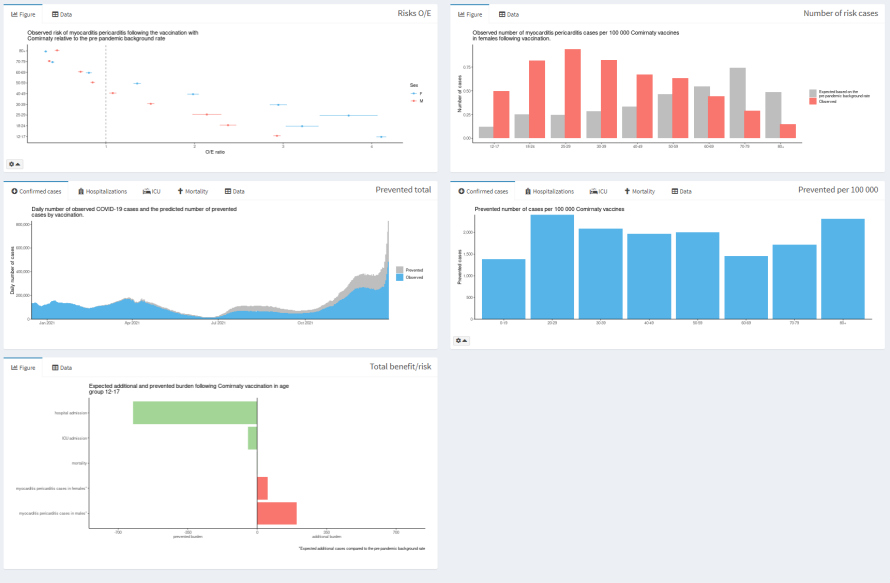
- ☐ **Sex:** Male, Female
- ☐ **Background rate:** Periods entered in the risk input data set, in which background incidences rates are estimated. f.e., Pre-pandemic, intra pandemic
- ☐ **Imputation method:** Two fixed options: Multiple imputation, Fixed proportion. If in the risk input data set only one option is entered, results are available only for this option.
- ☐ **Scenario risks:** Scenarios entered in the risk parameter section. If only one scenario is entered, only this option will be available.
- ☐ **Scenario benefits:** Scenarios entered in the benefit parameter section. If only one scenario is entered, only this option will be available.



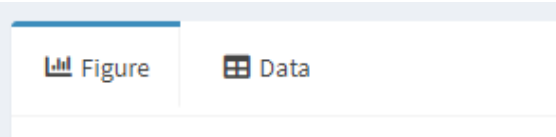
2.3. Main Panel

Description

The main result panel contains the results summarized into table(s) and figure(s) in 5 sub-figure panels.



Per sub-figure panel where this is available, the figure or table output can be chosen by selecting one of the features in the top left (**Figure** for figure and **Data** for table).

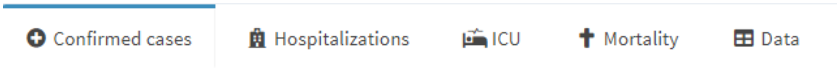


Interpretation

When the User selects different display options in the drop-down menu's on the left, the output is generated automatically.

Further actions

Inspect filtered data. Decide on filtering to be used for adjustments.



Interpretation

When the user selects one of the tabs, the result will alter to the user's pre-

ferred selections. There are 5 (five) different tabs, namely, Confirmed cases (to show the number of COVID-19 cases), Hospitalizations (to show the number of COVID-19 hospitalizations), ICU (to show the number of COVID-19 ICU admissions), Mortality (to show the number of COVID-19 fatalities), and Data (to show the output in table format of the chosen tab).

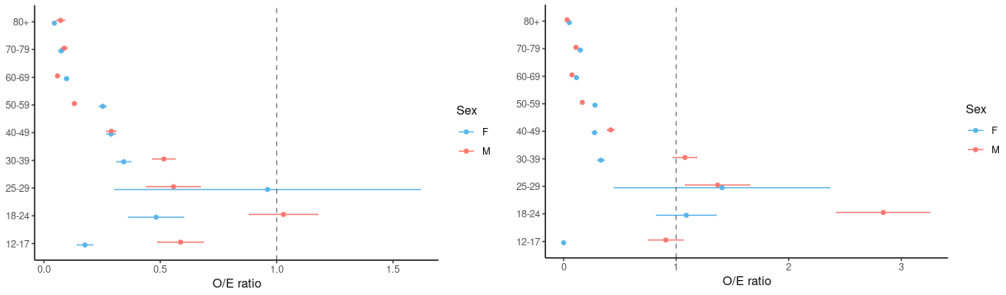
Further actions

Inspect filtered data. Decide on filtering to be used for adjustments.

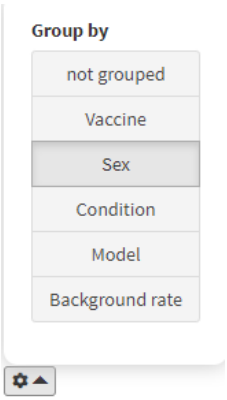
- ## Observed-Expected Ratio

Description

The Observed-Expected ratio, using the pooled background incidence rate during COVID-19 with inverse variance weight and additional (10,1,1) weight and a vertical dashed line, shows an equal observed and expected risk of myocarditis/pericarditis. This analysis can be displayed into several group of output (grouped by sex, vaccine, condition, model. In this user manual, two outputs based on sex were displayed, one when the user selects the Cominarty vaccination (left) and another when the user selects the Spikevax vaccine (right).



Note that there is a **Setting Button** in the left right of the sub-figures, which can adjust the grouping for the observed-expected ratio into not grouped, vaccine, gender, condition, model, and background rate.



Interpretation

The observed-expected ratios at 14 days using the pooled background incidence rate during COVID-19 prior to vaccination, with both inverse variance weighting and additional $(10, 1, 1)$ weight and multiple imputation, show higher observed-expected ratios in the age categories < 40 years for both genders and vaccine types. Only for Spikevax the observed-expected ratio is above 1, indicating a higher observed myocarditis risk then expected. More specifically, the risk is increased for males in age groups 18–24y and less so for 25–39y after Spikevax vaccination. The wide variability for females aged between 25–29 is explained by a zero IR estimate in 2 of the 3 background data sources (see right Figure **??**).

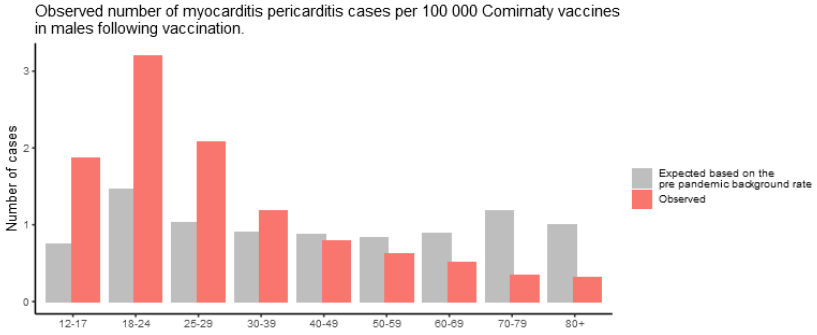
Further actions

Inspect filtered data. Decide on filtering to be used for adjustments.

- ## Observed-Expected Ratio with pre-pandemic background rate

Description

Observed vs Expected based on the pre-pandemic background rate of myocarditis pericarditis or TTS cases per 100,000 vaccines in the chosen gender within two weeks following vaccination. The one shown in this user manual is for the Cominarty vaccine, but users can choose any vaccination they wish to see the output for in the toolkit.



Interpretation

This illustration directly compares the risk of vaccination across different age groups towards myocarditis/pericarditis between the observed cases and the expected cases with pre-pandemic background rate. It can be shown that the highest increase in myocarditis and pericarditis occurs in the 18-24 years age group. In the toolkit, users can select any vaccination for which they want to see the results.

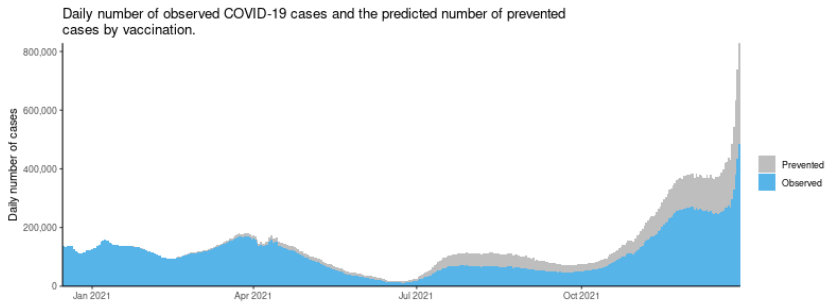
Further actions

Inspect filtered data. Decide on filtering to be used for adjustments.

- **Total prevented cases/ hospitalizations/ ICU admissions/ mortality of COVID-19 in the presence of vaccination**

Description

The benefit of vaccination is illustrated by comparing the number of prevented confirmed COVID-19 cases/ hospitalizations/ ICU admissions/ mortality by the probabilistic model with the number of observed cases. The one shown in this user manual is for the confirmed COVID-19 cases, but users can choose any output they wish to see the output for in the toolkit.



Interpretation

Based on the default model parameters, an estimated total of 13,322,567 confirmed COVID-19 cases have been prevented with mRNA and adeno-based vaccines since the start of the COVID-19 vaccination program in the different European countries.

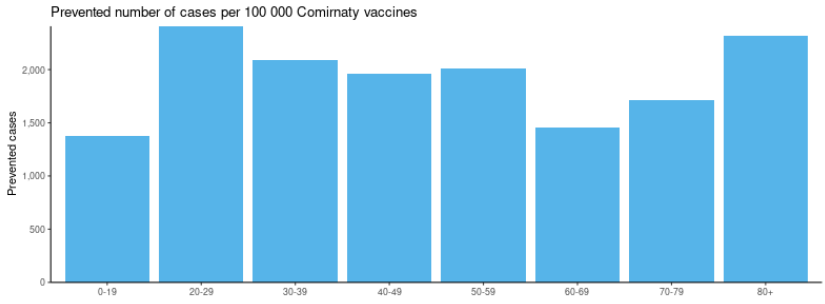
Further actions

Inspect filtered data. Decide on filtering to be used for adjustments.

- ## Prevented cases of COVID-19 per 100,000 individuals

Description

The Figure shows the expected number and confidence intervals of prevented COVID-19 cases between December 13, 2020 and December 31, 2021 when 100,000 individuals would have been vaccinated in the respective age categories at December 31, 2021. The one shown in this user manual is for the confirmed COVID-19 cases with Cominarty vaccine, but users can choose any output and vaccine type they wish to see the output for in the toolkit.



Interpretation

This illustration provides a direct comparison of the impact of vaccination across different age groups given the infection dynamics present within the specific time interval. It can be observed that individuals with the age group of 20-29 and 80+ years old have the most benefit compared to the other age groups.

Further actions

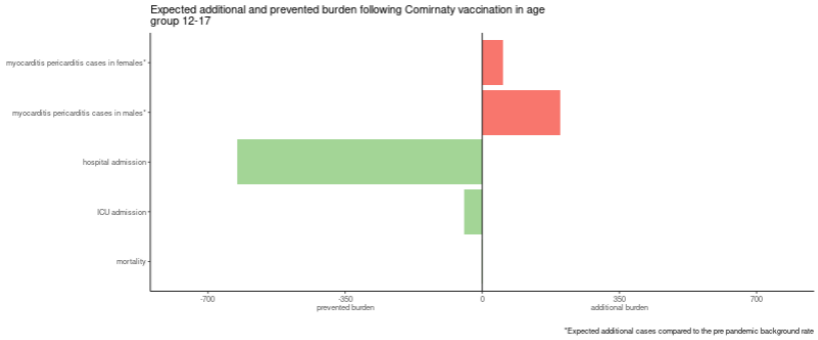
Inspect filtered data. Decide on filtering to be used for adjustments.

- ## Benefit-Risk Comparison

Description

A graphical comparison of the benefits of two dose COVID-19 vaccination (in terms of prevented confirmed COVID-19 cases) and the observed TTS

or myocarditis/pericarditis risks for both male and female associated with vaccination is presented. The one illustrated in this user manual is for Cominarty vaccination within 12-17 age group, however users can select any vaccine type and age group from the toolkit.



Interpretation

The Figure shows the benefits (in terms of prevented confirmed COVID-19 cases) of two dose COVID-19 of Cominarty vaccination and observed TTS or myocarditis/pericarditis risks associated with the vaccine type. This illustration shows that the benefits considerably outweigh the risks association with COVID-19 vaccination.

Further actions

Inspect filtered data. Decide on filtering to be used for adjustments.

5 Further Help

If Users require any additional assistance in using the Benefit-Risk Contextualization of COVID-19 Vaccine in the EU dashboard, please contact our team.

5.1 Contact us

Center for Statistics
Hasselt University
Agoralaan 1, Building D
B-3590 Diepenbeek, Belgium
(https://ibiostat.be)
(www.uhasselt.be/DSI)
(www.uhasselt.be/CenStat)

References

[1] M. Pathaka, S.N. Dwivedib, B. Thakura, and S. Vishnubhatla. Methods of estimating the pooled effect size under meta-analysis: A comparative appraisal. *Clinical Epidemiology and Global Health*, 8:105–112, 2020.

[2] S. M. Moghadas, T. N. Vilches, K. Zhang, S. Nourbakhsh, P. Sah, M. C. Fitzpatrick, and A. P. Galvani. Evaluation of covid-19 vaccination strategies with a delayed second dose. *PLoS biology*, 19(4):e3001211, 2021.

[3] J. M. Caldwell, X. Le, L. McIntosh, M. T. Meehan, S. Ogunlade, R. Ragonnet, G. K. O'Neill, J. M. Trauer, and E. S. McBryde. Vaccines and variants: Modelling insights into emerging issues in covid-19 epidemiology. *Paediatric Respiratory Reviews*, 39:32–39, 2021.

[4] S. Khare, C. Gurry, L. Freitas, M.B. Schultz, G. Bach, A. Diallo, N. Akite, J. Ho, R.T.C. Lee, W. Yeo, et al. Gisaid's role in pandemic response. *China CDC Weekly*, 3(49):1049, 2021.

[5] S. Elbe and G. Buckland-Merrett. Data, disease and diplomacy: Gisaid's innovative contribution to global health. *Global challenges*, 1(1):33–46, 2017.

[6] Y Shu and J McCauley. Eurosurveillance 2017, doi: 10.2807/1560-7917, 2017.

[7] S.N. Wood. *Generalized additive models: an introduction with R*. chapman and hall/CRC, 2006.

[8] J.J. Deeks, J.P.T. Higgins, and Altman D.G. Cochrane handbook for systematic reviews of interventions version 6.2 (updated february 2021), 2021. Available from https:/www.training.cochrane.org/handbook.

[9] E.B. Wilson and M.M. Hilferty. The distribution of Chi-square. *Proc Natl Acad Sci USA*, 17:684–8, 1931.

[10] K.J. Rothman and J.D. Boice. *Epidemiologic Analysis with a Programmable Calculator*. Boston, MA: Epidemiology Resources, Inc, 1982.

[11] P. Krause, T.R. Fleming, I. Longini, A. M. Henao-Restrepo, R. Peto, N.E. Dean, M.E. Halloran, Y. Huang, T.R. Fleming, P.B. Gilbert, et al. Covid-19 vaccine trials should seek worthwhile efficacy. *The Lancet*, 396(10253):741–743, 2020.

[12] C.B. Creech, S.C. Walker, and R.J. Samuels. Sars-cov-2 vaccines. *Jama*, 325(13):1318–1320, 2021.

[13] J.L. Bernal, N. Andrews, C. Gower, E. Gallagher, R. Simmons, S. Thelwall, J. Stowe, E. Tessier, N. Groves, G. Dabrera, et al. Effectiveness of covid-19 vaccines against the b. 1.617. 2 (delta) variant. *New England Journal of Medicine*, 2021.

[14] J. Stowe, N. Andrews, ..., and J. L. Bernal. Effectiveness of COVID-19 vaccines against hospital admission with the delta (B.1.617.2) variant. *preprint*, pages 1–3, 2021.

[15] N. Andrews, J. Stowe, F. Kirsebom, S. Toffa, R. Sachdeva, C. Gower, M. Ramsay, and J. L. Bernal. Effectiveness of covid-19 booster vaccines against covid-19 related symptoms, hospitalisation and death in england. *Nature medicine*, pages 1–1, 2022.

[16] K. J Bruxvoort, L. S. Sy, L. Qian, B. K Ackerson, Y. Luo, G. S. Lee, Y. Tian, A. Florea, M. Aragones, J. E. Tubert, et al. Effectiveness of mrna-1273 against delta, mu, and other emerging variants of sars-cov-2: test negative case-control study. *bmj*, 375, 2021.

[17] R. Sonabend, L. K. Whittles, N. Imai, P. N. Perez-Guzman, E. S. Knock, T. Rawson, K.A.M. Gaythorpe, B. A. Djaafara, W. Hinsley, R.G. FitzJohn, et al. Non-pharmaceutical interventions, vaccination, and the sars-cov-2 delta variant in england: a mathematical modelling study. *The Lancet*, 398(10313):1825–1835, 2021.

References

[1] M. Pathaka, S.N. Dwivedib, B. Thakura, and S. Vishnubhatla. Methods of estimating the pooled effect size under meta-analysis: A comparative appraisal. *Clinical Epidemiology and Global Health*, 8:105–112, 2020.

[2] S. M. Moghadas, T. N. Vilches, K. Zhang, S. Nourbakhsh, P. Sah, M. C. Fitzpatrick, and A. P. Galvani. Evaluation of covid-19 vaccination strategies with a delayed second dose. *PLoS biology*, 19(4):e3001211, 2021.

[3] J. M. Caldwell, X. Le, L. McIntosh, M. T. Meehan, S. Ogunlade, R. Ragonnet, G. K. O'Neill, J. M. Trauer, and E. S. McBryde. Vaccines and variants: Modelling insights into emerging issues in covid-19 epidemiology. *Paediatric Respiratory Reviews*, 39:32–39, 2021.

[4] S. Khare, C. Gurry, L. Freitas, M.B. Schultz, G. Bach, A. Diallo, N. Akite, J. Ho, R.T.C. Lee, W. Yeo, et al. Gisaid's role in pandemic response. *China CDC Weekly*, 3(49):1049, 2021.

[5] S. Elbe and G. Buckland-Merrett. Data, disease and diplomacy: Gisaid's innovative contribution to global health. *Global challenges*, 1(1):33–46, 2017.

[6] Y Shu and J McCauley. Eurosurveillance 2017, doi: 10.2807/1560-7917, 2017.

[7] S.N. Wood. *Generalized additive models: an introduction with R*. chapman and hall/CRC, 2006.

[8] J.J. Deeks, J.P.T. Higgins, and Altman D.G. Cochrane handbook for systematic reviews of interventions version 6.2 (updated february 2021), 2021. Available from https:/www.training.cochrane.org/handbook.

[9] E.B. Wilson and M.M. Hilferty. The distribution of Chi-square. *Proc Natl Acad Sci USA*, 17:684–8, 1931.

[10] K.J. Rothman and J.D. Boice. *Epidemiologic Analysis with a Programmable Calculator*. Boston, MA: Epidemiology Resources, Inc, 1982.

[11] P. Krause, T.R. Fleming, I. Longini, A. M. Henao-Restrepo, R. Peto, N.E. Dean, M.E. Halloran, Y. Huang, T.R. Fleming, P.B. Gilbert, et al. Covid-19 vaccine trials should seek worthwhile efficacy. *The Lancet*, 396(10253):741–743, 2020.

[12] C.B. Creech, S.C. Walker, and R.J. Samuels. Sars-cov-2 vaccines. *Jama*, 325(13):1318–1320, 2021.

[13] J.L. Bernal, N. Andrews, C. Gower, E. Gallagher, R. Simmons, S. Thelwall, J. Stowe, E. Tessier, N. Groves, G. Dabrera, et al. Effectiveness of covid-19 vaccines against the b. 1.617. 2 (delta) variant. *New England Journal of Medicine*, 2021.

[14] J. Stowe, N. Andrews, ..., and J. L. Bernal. Effectiveness of COVID-19 vaccines against hospital admission with the delta (B.1.617.2) variant. *preprint*, pages 1–3, 2021.

[15] N. Andrews, J. Stowe, F. Kirsebom, S. Toffa, R. Sachdeva, C. Gower, M. Ramsay, and J. L. Bernal. Effectiveness of covid-19 booster vaccines against covid-19 related symptoms, hospitalisation and death in england. *Nature medicine*, pages 1–1, 2022.

[16] K. J Bruxvoort, L. S. Sy, L. Qian, B. K Ackerson, Y. Luo, G. S. Lee, Y. Tian, A. Florea, M. Aragones, J. E. Tubert, et al. Effectiveness of mrna-1273 against delta, mu, and other emerging variants of sars-cov-2: test negative case-control study. *bmj*, 375, 2021.

[17] R. Sonabend, L. K. Whittles, N. Imai, P. N. Perez-Guzman, E. S. Knock, T. Rawson, K.A.M. Gaythorpe, B. A. Djaafara, W. Hinsley, R.G. FitzJohn, et al. Non-pharmaceutical interventions, vaccination, and the sars-cov-2 delta variant in england: a mathematical modelling study. *The Lancet*, 398(10313):1825–1835, 2021.