

# COVID-19 Vaccine Predictions Methodology

## Overview

We wanted to know how long it will take to develop vaccines that are safe and effective against COVID-19, and to understand whether the portfolio of vaccine candidates was sufficiently diversified to increase our chances of getting a vaccine against COVID-19 in a timely manner. To do this we interviewed a range of experts about COVID-19 and built this model.

There are more than 200 vaccine candidates against COVID-19 at various stages of development. Assessing the odds of a vaccine being approved relies on much information that is not available; and there are many unknowns for each specific vaccine. Our approach helps to address both of these issues helping users answer questions, (e.g. how likely is it that a well-funded candidate in a pre-clinical development succeeds to approval?) and apply them to the vaccine portfolio. The model comes with default inputs, which are generated by a mixture of expert interviews, desk research, and our internal expertise. These approaches are outlined [in our policy paper](#).

We have created two models that fit together into the same user interface: a research and development model, and a manufacturing model. Our research and development model estimates if and when vaccines will be approved for use, and our manufacturing model looks at how long it will then take before we can start manufacturing these vaccines, and how many can be produced every month once scale up is complete.

## Research and development model

### *Inputs*

Both models work with two types of inputs. The first consists of all the vaccines listed on the Vaccines page on our web tool, this data was compiled using data from the London School of Hygiene and Tropical Medicine, US clinical trial.gov data and extensive research. You can find our full list of vaccines [here](#). The most important values are the institutes involved, the countries involved, the platform being used (inactivated vaccines, live attenuated vaccines, protein subunit, RNA and DNA), the estimated funding category (large external funding, some external funding, large pharma, medium pharma, and small- bio-tech/academic), the start and end dates of the trial phases (Pre-Clinical, Phase I, Phase II, Phase III, and Approval) if known, and an arbitrary vaccine number for cross referencing. For users interested in further customization it is possible to upload a customized list of vaccines if they wish to run in this model.

The second type consists of the values of all the parameters such as success rates at each phase, the timescales for each phase, factors that depend on the platform and funding category, manufacturing timelines, and capacity. The user can edit all of these parameters.

The model takes data on existing COVID-19 vaccines in various stages of clinical trials, and expert opinions as to their likely success, and generates predictions on how many vaccines will be approved by a “[stringent regulator](#),” and on what timescales. The model uses Monte Carlo techniques to randomly decide an outcome given the input parameters and should be run a number of times (runs) to smooth out statistical fluctuations, we recommend running the model at least 200 times. The model currently consists of approximately 1,500 lines of Python. The source code is [available here](#).

### *Initialisation*

After reading the input files, the program checks whether all the parameters are within specifications, e.g., that success rates are between 0 and 1. If any parameters are outside the specifications, the run is aborted.

At the start of every run each vaccine is initialised. Each vaccine is allocated a Probability of Success (PoS) for each phase given in,  $P_i$  where  $i = 0 \dots 4$  is the phase. The PoS for Pre-Clinical, Phase I, Phase II and Approval are the same, whereas for Phase III it depends on which platform the vaccine is on. In addition, these PoS are multiplied by a factor that depends on the funding category the vaccine is in. If the factor is  $f$  then  $P_i$  is multiplied by the 5th root of this factor  $5\sqrt{f}$ . Since the overall PoS is given

by  $P = P_0 \times P_1 \times P_2 \times P_3 \times P_4$ , this ensures that the overall PoS is multiplied by  $f$ . It is assumed that vaccines in the funding category Bio-tech/Academic will be bought out by Large Pharma if they succeed in Phase I and so in this case the square root of  $f$  is used since only two phases will be relevant. This feature can be switched off if required. The PoS values are stored for later use. Note that when multiplying a PoS by a factor, the result is capped at 1 should it exceed 1.

Each vaccine is also allocated a start and end month for each phase. Firstly, all start and end dates in the input vaccines file are converted to months relative to the current month ( $= 0$ ). Dates in the future are ignored by default. Each vaccine is allocated a 'best case', a 'most likely' and a 'worst case' date for each phase based on the input parameter values given in for each phase multiplied by factors that depend on the platform and the funding category.. When phase II and III start is based not only on the length of time for previous trials but also on whether trials are overlapped. The overlap categories are as follows (using the default values):

Almost simultaneous - the next phase starts 1 month after the start of the previous phase.

Mostly overlapped - the next phase starts 2 months after the start of the previous phase.

Phases I & II overlapped - Phase II starts 1 month after the start of Phase I, Phase III starts 1 month after the end of Phase II.

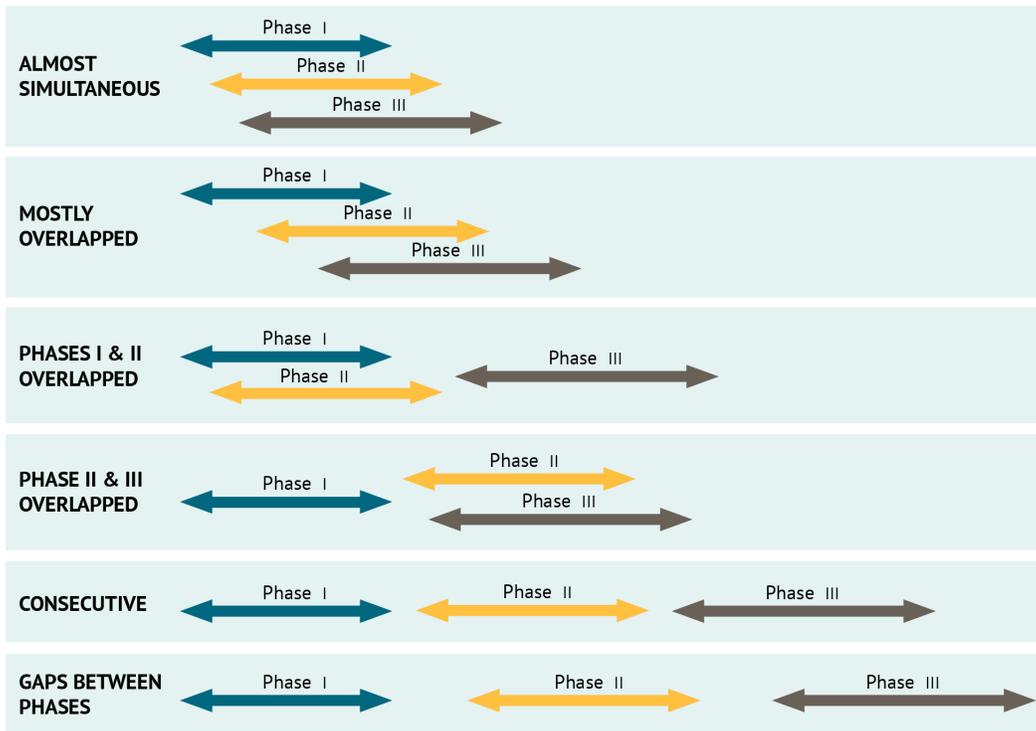
Phases II & III overlapped - Phase II starts 1 month after the end of Phase I, Phase III starts 1 month after the end of Phase II

Consecutive - each phase starts 1 month after the end of the previous phase.

Gaps between phases - each phase starts 6 months after the end of the previous phase.

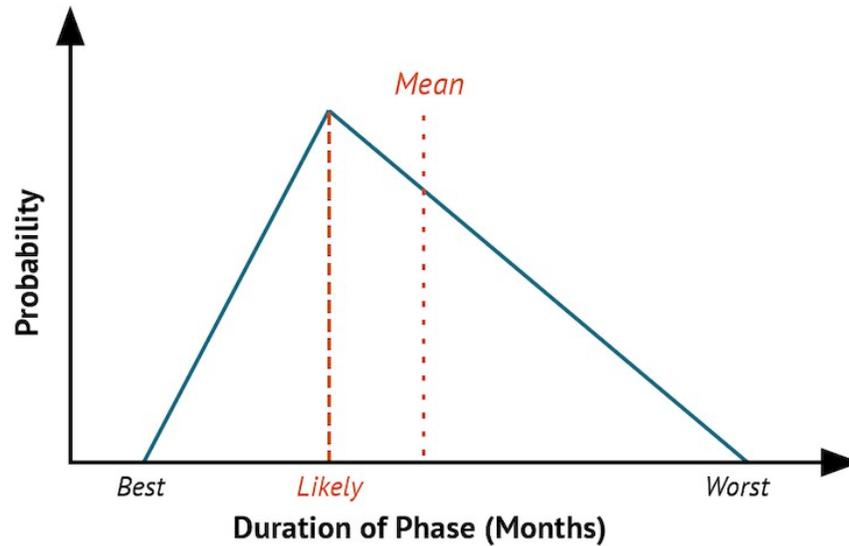
Phase I starts 1 month after the end of Pre-Clinical and Approval starts 1 month after the end of Phase III. These values are shown in of the detailed parameters. The overlap categories are illustrated in Figure1.

**Figure 1. Options for overlapping trials**



Having set the start of each phase, the end of the phase is set by selecting a random phase length from a triangular distribution. Triangular distributions are defined by three points: minimum, most likely, and maximum. Numbers closer to the most likely figure are more likely to be chosen. This is illustrated below in figure 2. In addition, a fraction of vaccines have their Phase III end dates delayed (see discussion below). Any end dates that come out before month 1 are set to month 1 as it is assumed that otherwise these would be in the original vaccines file. The start and end months are stored for later use. Figure 2 is a triangle distribution used to randomly select month lengths. Triangle distributions are used for several of the inputs in the manufacturing models too.

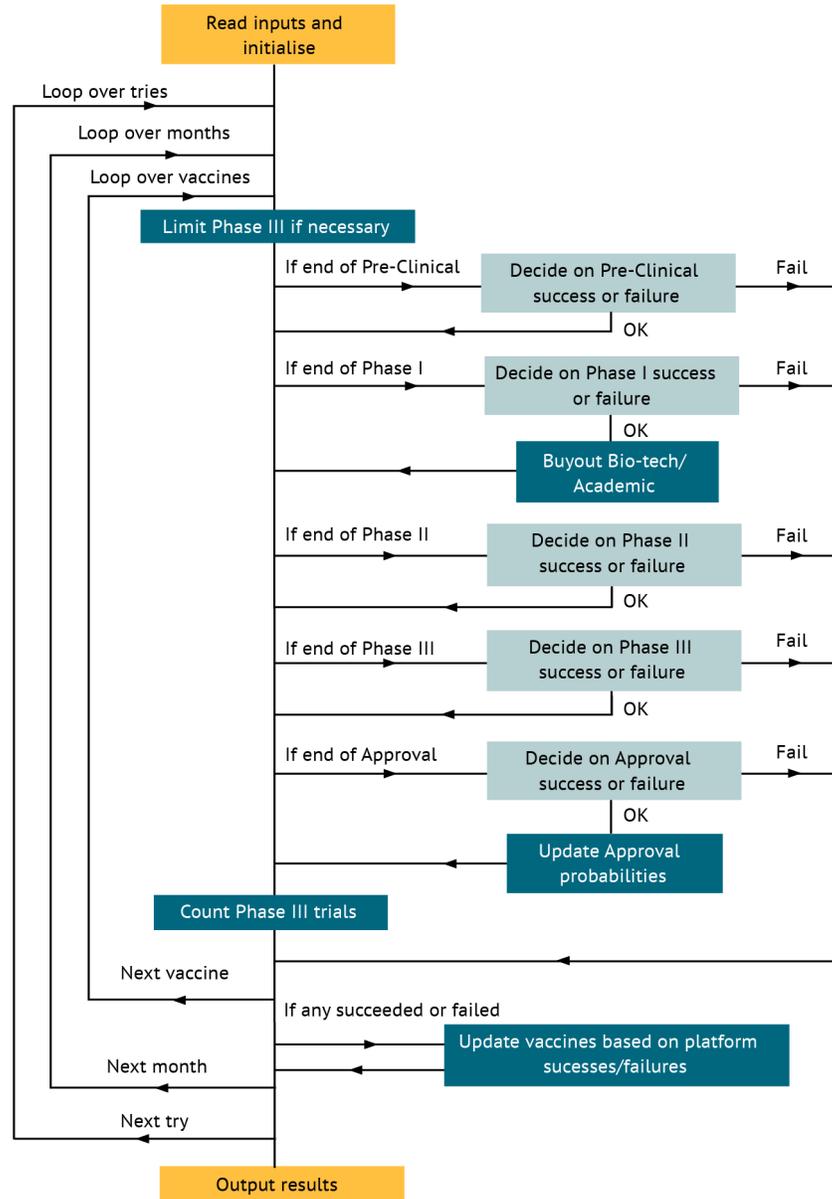
Figure 2. Triangle distribution for randomly selecting month lengths



### *Program Flow*

The program proceeds as shown in figure 3. There is a main loop over tries, then for each try the program loops over each month and for each month loops over each vaccine. To avoid any bias the order of the vaccines is randomised every try. Each month each vaccine is tested to see if a phase ends that month. If so, a random number is used to decide whether the vaccine succeeds or fails dependent on its PoS for that phase. If it succeeds it continues for further months. If a vaccine succeeds the Approval Phase, this information is used to possibly inhibit further approvals as discussed below, and if a Bio- tech/Academic candidate succeeds at Phase I, this vaccine may be bought out as described below. If the vaccine fails, a second random number is used to decide whether the failure is technical or economic. All failures for well-funded vaccines are assumed to be technical whereas all failures for Bio- tech/Academic are assumed to be economic. This information is used to calculate correlations as described below. Figure3 shows a flow chart for the model.

Figure 3. Flow chart for the model



### *Limiting Phase III Trials*

In practice, Phase III trials may be slowed down if infection rates drop across the world. In addition, the overall number of trials might be limited, or they might be slowed down if too many Phase III trials take place simultaneously. At the end of each month the number of vaccines that have successfully passed Phase III is calculated. At the start of the next month, if a vaccine is about to start Phase III and if the number of already-successful vaccines exceeds a limit (by default 6), then by default the end of the Phase III trial is delayed (in addition to the delay described above). There is also the option to prevent the vaccine entering Phase III until the total number drops below the limit. This is done by deferring the start of Phase III till the next month (which may happen several times). Any adjustments to Phase III are also applied to the start of Approval, which will be similarly delayed.

### *Small Bio-tech/Academic Buyout*

Some (by default 100%) small Bio-tech/Academic vaccines that are successful at Phase I are randomly deemed to be bought out by Large Pharma organisations. In this case the funding category is changed, and the vaccine is re-initialised to update its PoS and timelines to reflect the new category. It then proceeds as if it were a Large Pharma candidate.

### *Limiting Approval*

Once we have an adequate number of vaccines, it may become more difficult (or easier) for other candidates to be approved. Above some limit subsequent approvals will be less likely to succeed and the timeline will be longer. Each time a vaccine is successfully approved and the number of approved vaccines is above some limit (by default 3) all other vaccines have their Approval PoS and Approval phase length multiplied by some factor. These factors are cumulative so that it gets harder and harder to approve subsequent vaccines.

### *Correlations between Vaccines*

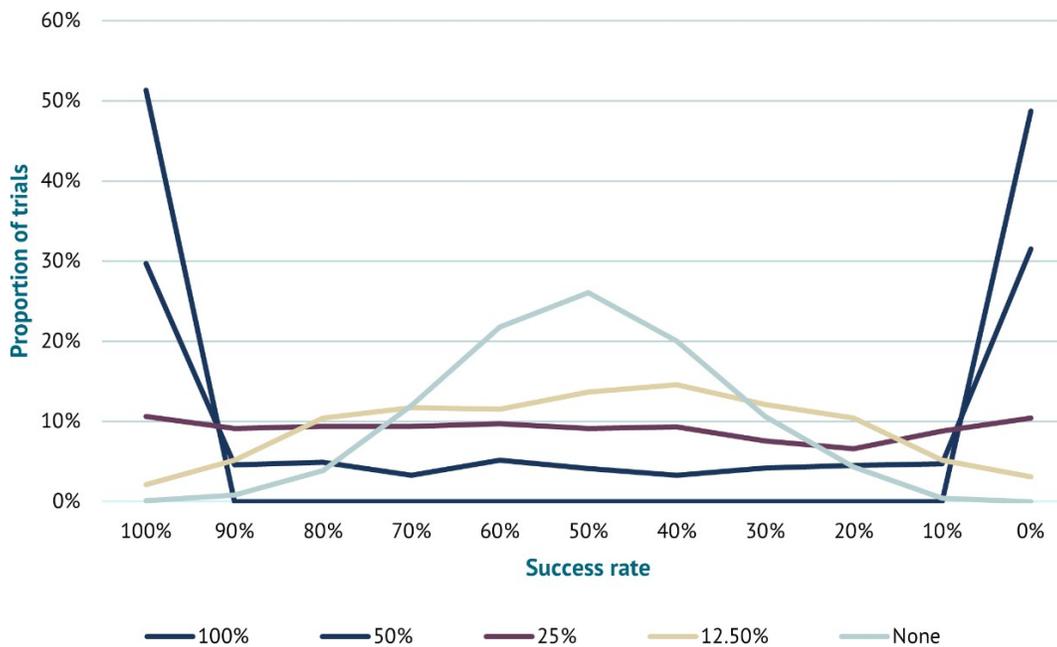
It is likely that success or failure of one vaccine on a platform might impact other vaccines on the same platform, i.e., a failure might indicate that this platform is inherently unsafe. Each time a vaccine succeeds or fails technically at Phase I, Phase II or Phase III the following procedure is enacted. Commercial failures are ignored. The number of successes,  $S$ , and technical failures,  $F$ , at that phase on that platform is used to calculate an aggregate  $A = S \div (S + F)$ . This aggregate is then compared to the  $PoS$ ,  $P_i$ , for that platform and phase as described during initialisation above. A new PoS is then calculated as

$$P'_i = ((A - P_i) \times C) + P_i.$$

This is a Bayesian type approach whereby the initial PoS,  $P_i$  is updated with new information based on  $A$ . The factor  $C$  is a correlation strength that depends on the platform. The ratio  $R = P'_i \div P_i$  between the new and old PoS is then used to multiply the original PoS of all phases of all vaccines on the same platform. There are then adjusted for the funding category as described in the initialisation. In this way if one vaccine on the platform succeeds, the others are more likely to and likewise if one fails, the others are more likely to fail. This introduces a correlation between vaccines on the same platform and drives a divergence of the overall PoS for that platform.

Figure 4 is the output of a 1,000-run simulation that looks at the proportion of the first 10 vaccines in a platform that are successful when different adjustment rates are used. The adjustment is switched off for when the correlation is rated none; variation on this line is the randomisation inherent in Monte Carlo simulations. With 100% correlation you will see that all vaccines in a portfolio either succeed or fail. With no correlations each of the 10 trial outputs in each run are grouped around the mean, and there are very few highly successful or highly unsuccessful platform runs. With a 25% correlation, we found that any combination of successes and failures within a platform were about equally likely, when the input was 50% (other inputs would weight more strongly to either success or failure). 12.5% does have some very successful runs and complete flops, but here the majority of runs are near the mean. It is for this reason that we used 25% correlation as our standard input. We then halved this to 12.5% for lowly correlated vaccine portfolios and increased it to 50% for highly correlated portfolios.

**Figure 4. Changes in probability of success after funding adjustments**



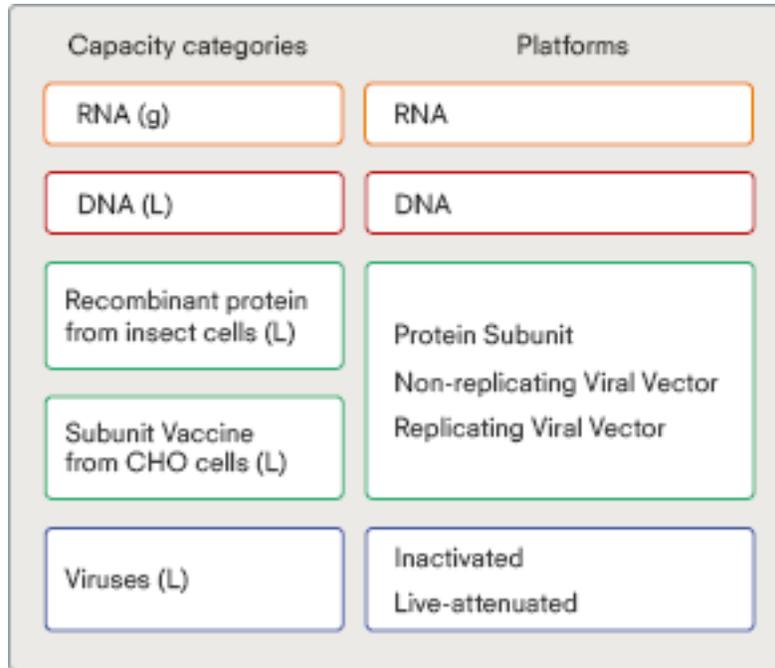
## Manufacturing models

### *Overview and Purpose*

The purpose of the manufacturing model is to identify plausible scenarios for how long it will take to manufacture vaccines to treat the populations in need, and is split into two parts: Manufacturing Preparation, predicting when primary (drug substance) and secondary (drug product) manufacturing could start; and Manufacturing Capacity, predicting when enough doses are made to meet the needs of the target populations. The model uses output from the Research and Development model, expert input and interviews, literature, and manufacturing capacity survey data as inputs. The model uses Monte Carlo techniques to derive the dates at which the vaccine production targets will be met and should therefore be run many times to smooth statistical fluctuations. The source code is [available here](#).

*Preliminary Notes on Vaccine Platforms*

This model uses the same seven platforms as the R&D model. Much of the data used is categorised by vaccine platform. Some production capacity is assumed to be able to be shared between platforms as follows:



The rationale for this sharing of capacity is explained in the [Technical Report](#).

*Inputs and Default Data*

The model uses several sets of default data as a basis for inputs, stored and accessed as JSON files:

- Gantt schedules: data which identify all the key activities and dependencies for manufacturing preparation for each platform, with default pathways and timings derived from expert input and interviews
- Primary/drug substance capacity: based on CEPI survey data of global bioreactor volume (m<sup>3</sup>), separated by country and platform (please see additional information section for mapping of CEPI categories to our platforms)
- Secondary/drug product capacity: based on CEPI survey data of global capacity, separated by country and platform (please see additional information section for mapping of CEPI categories to our platforms)

The default data are used to define single parameters alongside triangular distributions within the model. Triangular distributions are defined by three points: minimum, most likely, and maximum.

These default values are used to derive model inputs based on the user's input in the online tool. For example, if a user provides triangular inputs for the total bioreactor volume for a particular vaccine platform, the calculated total is allocated to each country in the same proportions as for the default values provided by CEPI. Similarly, if a user selects simultaneous scale up and technology transfer for a vaccine platform, the Gantt chart used will alter the critical path to allow this.

The model also uses other variable data such as:

- R&D output: output from the Research and Development model identifying approved vaccines for each run and their month of approval
- Vaccines: the vaccines under consideration and data used to process and classify them including platform and funding category

### *Initialisation*

Model initialisation is performed as follows:

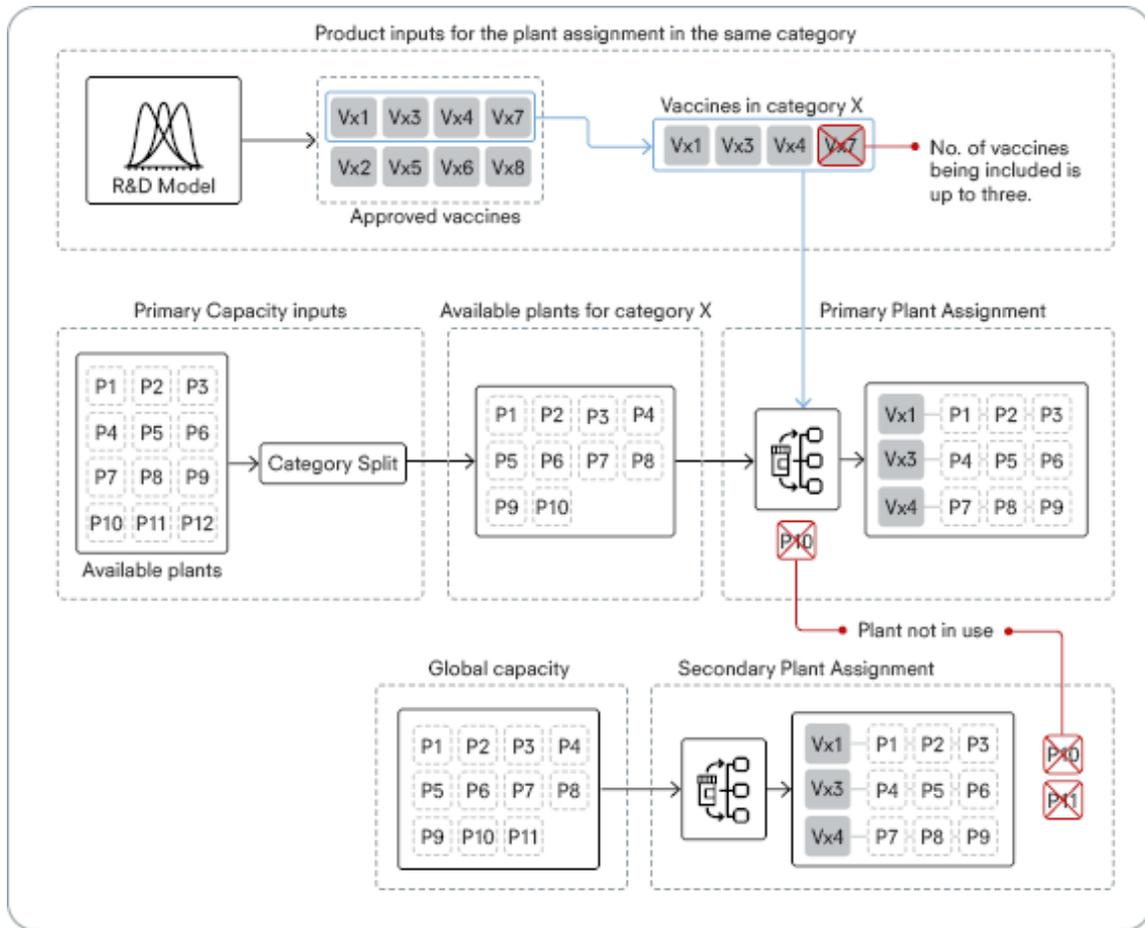
- Load data from JSON files
- Modify default global data based on user-defined parameters and prepare model inputs, including:
  - Vaccine targets – these are set directly from user inputs into the online interface and the total required doses are calculated using waste percentage and doses per person, with default values based on the published WHO allocation framework
  - Primary/drug substance capacity – a global total of doses for each platform is calculated based on the user's distribution inputs of bioreactor capacity, and these totals are allocated to each country proportionally using factors derived from the default CEPI data
  - Secondary/drug product capacity – a global total of doses is calculated based on the user's distribution inputs, and this total is allocated to each country proportionally using factors derived from the default CEPI data
- Keep the first 3 vaccines to be approved within each category, and discard any which are approved later
- Allocate a maximum of 3 countries to each vaccine for both primary/drug substance and secondary/drug product capacity – capacity is allocated in descending order from first approved vaccine to last approved vaccine

The calculation to convert total bioreactor volume to monthly doses of all platforms except RNA is as follows:

Calculation Steps		Calculation
1	Estimate the lowest, most likely and highest volume per month for each platform	$V$
2	Estimate doses per litre of bioreactor volume per batch	$N_{dv}$
3	Estimate batches per bioreactor per year	$B$
4	Calculate doses per $m^3$ of bioreactor volume per month for each platform	$N_v = V \times (N_{dv}/1000) \times (B/12)$

The calculation to convert microgram doses to monthly doses for RNA is as follows:

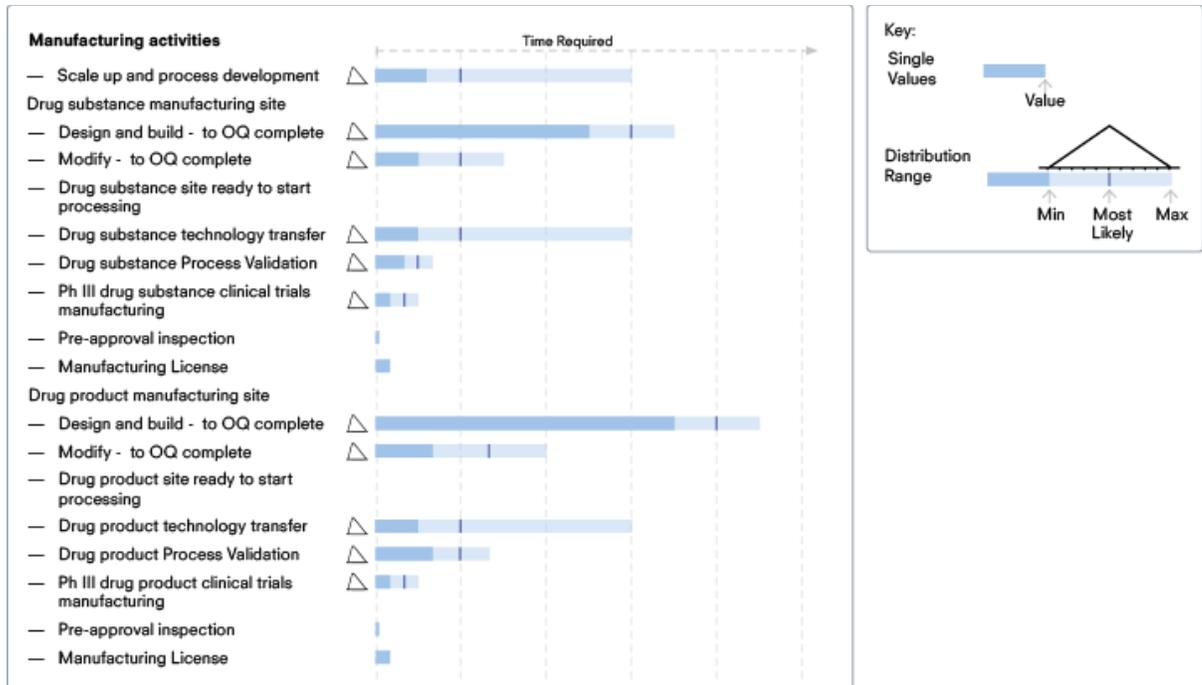
Calculation Steps		Calculation
1	Estimate the lowest, most likely and highest mass per month	$M$
2	Estimate microgram per dose	$N_{dm}$
3	Calculate doses for most likely mass per month	$N_m = M / N_{dm} / 12$



### *Manufacturing Preparation*

Manufacturing preparation time is computed as follows for each trial of the R&D model:

- For each vaccine that is approved:
  - Select the appropriate default Gantt chart based on the vaccine platform and whether the user has selected simultaneous technology transfer and scale up, and whether manufacturing can start before vaccine approval (i.e. different critical path and dependencies between activities)
  - User inputs from the online interface are used to change individual activity timings within the timeline, calculating activity times from triangular distribution inputs and total time per platform
  - User inputs from the online interface are then used to apply a factor to every activity within the timeline, based on the vaccine's funding category
  - Compute the critical path through the activities and determine a start date for primary/drug substance and secondary/drug product manufacture



More detail on the individual activities can be found in the additional information section.

### *Manufacturing Capacity*

Manufacturing capacity is computed as follows for each trial of the R&D model:

- For each vaccine:
  - Use the start time for primary/drug substance and allocated capacity (doses) to compute the monthly production of primary product
  - Use the start time for secondary /drug product and allocated capacity (doses) to compute the monthly production of secondary product
  - Identify the bottleneck process and use this to determine final monthly capacity
- Repeat across all vaccines and sum monthly totals
- For each month, compute a cumulative sum of all doses to that point and compare with vaccine targets
- Record the month at which each target is met

## *Assumptions*

The main assumptions for computing the schedule and capacity are as follows, by category.

### *Product and Process:*

- Drug substance (primary) plant capacities are characterised by bioreactor capacity and downstream processing capacity is assumed to be suitable and to match bioreactor capacity
- All products within a platform have the same drug substance productivity (doses/litre bioreactor)
- For any product the same facility must be used for Phase III trials and for commercial manufacture
- All products are delivered as aseptic liquids
- Each country is treated as one block of capacity for each drug substance platform
- Each country is treated as one block of secondary capacity
- As a vaccine is approved it takes all available drug substance capacity in the correct class up to a maximum of three plants
- No more than three vaccines within any platform are manufactured
- They are selected on a first-come-first served basis as they are approved – the plants with a higher throughput will be assigned first
- Capacity of any type is allocated optimally – there is currently no allowance for national or commercial restrictions

### *Scheduling/Manufacturing Preparation:*

- Manufacturing activities are defined as all those needed to get to approved, effective manufacture excluding research and development and clinical trials. Included in manufacturing activities are:
  - Process development and scale-up
  - Technology transfer
- Primary manufacturing activities always start at risk; they do not wait for positive clinical trial results
- By default, secondary manufacturing starts at risk but dose form commercial manufacture waits for drug approval (user-defined in the online tool).

### *Supply Chain:*

- No upstream supply chain restriction

## *Outputs*

- Timeline graph: Each bar in the graph shows a platform being approved in the R&D model. The R&D approval month, primary start time and secondary start time are averaged based on the number of iterations per family. No primary start time shown in the graph means that, for that product, the primary production starts straight after it has been approved.
- Distribution graphs: The distribution graphs show the time when the global cumulative production can hit each target. The bars in the graph show the number of runs and the line indicates the cumulative percentage of runs to reach the target in the number of iterations recorded.
- Cumulative production graphs: Each line shows the cumulative dose production for a single model run. The model runs are selected from a set of runs from the month in which the cumulative probability of reaching a certain target number of doses reaches the stated percentage.
- Number of doses breakdown by platform: In each pie chart, the cumulative doses for each target is broken down by platform.

## Additional Information:

### *Vaccine Platform Mapping*

The manufacturing processes identified for each platform informed the allocation of CEPI survey categories to model capacity categories and define which product platforms can be made in a category of capacity. We developed two versions of the categorisation. In the first version there are five exclusive categories of capacity and microbial and yeast capacity is not used. In the second version there are two specialised categories of RNA and Viral Vaccine production plus a third, broad capacity category suitable for all other platforms. The five-category version is the more pessimistic and was used for the model.

RNA and DNA vaccine production are considered to be specialised and so each have their own category.

Insect cell and mammalian cell culture-based facilities are considered to be equivalent for the purposes of the model because the upstream and downstream unit operations and equipment are similar; typically, these are single use bioreactors, chromatography and cross-flow filtration systems.

The CEPI classification of Viruses production is interpreted as classical viral vaccine production. Live and attenuated vaccines can only, in the model, use this capacity category as it is assumed that their biosafety categorisation would prevent their location in more general cell-culture facilities.

Viral vector production processes are similar to classical viral vaccine processes. They have been allocated to the mammalian/insect cell culture category, however, as they do not usually have the high biosafety requirements discussed above.

Whilst some individual products may use yeast or microbial expression systems, we considered mammalian and insect cell culture to be the norm and so in the 5-category analysis this capacity is not utilised. The first step in DNA production often uses microbial fermentation but we considered that the downstream processing would be specialised. This is a pessimistic assumption.

### *Manufacturing Preparation Activities*

The transition from R&D to manufacturing is typically carried out by a large multifunctional team and includes process development activities, design and construction activities, and quality assurance/regulatory activities.

We have broken this transition down and simplified it to form a series of steps for any vaccine. For each step, we assigned a probabilistic distribution of durations around a value based on experience of many bio pharmaceutical projects. Steps can overlap.

### *Scale up and process development*

This means all activities required to take a laboratory-based process and make it suitable for mass production. Under normal circumstances this can include the evaluation of alternative manufacturing routes and unit operations, for example changes in downstream processing strategies or cell-line optimisation. Each unit operation is developed in laboratory and pilot facilities to achieve levels of productivity, stability, and reproducibility suitable for regulatory approval and commercial viability. This can be done in-house or using contract development and manufacturing organisations (CDMOs).

Under normal circumstances scale up and process development may be carried out in parallel with clinical development, but the speed and extent are dependent on the drivers and financial capacity of the developing organisation.

For the pandemic emergency circumstances, we have assumed that these activities will be started as early as possible for all developers, with a start date set as 1 March 2020. This is based on public reports and expert input that for major developers this is the case. We have extended this assumption to all vaccine candidates in the manufacturing model. This might be considered optimistic for less well funded candidates. The chances of these reaching the manufacturing stage, however, are already reduced in the R&D model so one could conjecture that a promising “outsider” would garner resources from major manufacturers, which would offset this optimistic assumption.

We have seen this activity tends to be on the critical path and so its sequencing with technology transfer is critical and this is discussed further below.

### *Manufacturing facility preparation*

Each product requires drug substance (primary) and drug product (secondary) manufacturing. These are typically carried out in separate facilities although they may be on the same site. Both drug substance and drug product manufacturing may be carried out in-house or by CDMOs. Multiple sites can be used for the same product and single sites may produce several products.

The steps to bring drug substance and drug product sites into production are described below.

As with to scale-up and process development we have assumed that these activities are carried out as early as possible and do not wait for positive clinical trial results. This implies substantial capital expenditure at risk across the global manufacturing base.

#### *Drug substance manufacturing site*

Design and build to operational qualification (OQ) complete:

Operational qualification (OQ), a mandatory step that means that a facility has been demonstrated to be functioning suitably to manufacture a given product but before any of that product has been made in that facility.

The model was set up with three options regarding the status of a production plant: “design and build”; “modify”; or “existing.” Design-and-build means that a new brownfield or greenfield site is built from scratch. This is obviously the slowest of the three options.

In current model runs we have assumed that this route is, in fact, not used. There is some reporting of new facilities being built; in the UK the VMIC facility was already in progress before COVID-19 struck. A new emergency facility has been reported as being built in record time in China. In the US, a “rapid reaction” concept was developed in the past to counter bio-terrorism threats. As there is sufficient latent capacity available, however, this route does not seem to be widely in use for first generation vaccines. Put another way, first generation COVID-19 vaccines are unlikely to have to wait for new plants to be built.

Modify to OQ complete:

When a new product is introduced into an existing plant, that plant typically requires modification. The extent can vary widely from significant building extensions to additions of minor equipment only. In all cases the modified plant requires procedures and usually software modifications to be prepared to allow it to achieve OQ for the new product.

This is the default option for all products.

Drug substance site ready to start processing:

This milestone flags that the site is ready. Both site and product must be ready for production to start.

Technology transfer:

This describes all activities to take a new product and establish a new seat of manufacture. These comprise scientific, engineering, organisational, regulatory and commercial activities and must be carried out for each site.

Conventionally this will occur after a certain amount of process development and scale up. In some cases, a CDMO will carry out the scale-up and tech transfer for a vaccine developer so that the process is developed with a particular plant and manufacturing approach in-mind.

The extent of overlap between scale-up and process development and technology transfer in the model has a large impact on when a vaccine is available, so we have modelled two options to look at the sensitivity to this factor.

Some respondents have indicated that products are being moved to full scale manufacturing facilities as soon as possible with combined scale-up and tech transfer activities.

#### Drug substance Process Validation:

Following tech transfer, process validation is required to show that the process can produce product at the required quality and that all supporting quality systems are in place.

#### Phase III drug substance clinical trials manufacturing:

We have assumed that the sites used for manufacturing drug substance and drug product will be the same size used for eventual full-scale manufacturing. This is preferred for biologics so that the risk of changes in the product between phase three and routine manufacture is minimised.

#### Pre-approval inspection:

This is a relatively short duration step in which regulatory approval is sought for an individual manufacturing site for a particular product. Successful pre-approval inspection leads to the granting but manufacturing licence.

#### *Drug product manufacturing site*

The same steps are used in the scheduling of the drug product manufacturing site, but durations are different and steps with the same name are not linked unless shown so in the schedule logic.