COVID-19
Multi-Model
Comparison
Collaboration
(CMCC)

TECHNICAL GROUP REPORT: MODEL FITNESS-FOR-PURPOSE ASSESSMENT REPORT

24 AUGUST 2020

Photo by *Anshu A* on *Unsplash*

PREPARED BY
TECHNICAL GROUP, COVID-19 MULTI-MODEL COMPARISON COLLABORATION (CMCC)

PARTNERS

WORLD BANK GROUP

World Health Organization

BILL & MELINDA GATES foundation

MHESI
กระทรวงการอุดมศึกษา วิทยาศาสตร์ วิจัยและนวัตกรรม
Ministry of Higher Education, Science, Research and Innovation

Department for International Development

Norad

USAID
UNITED STATES AGENCY INTERNATIONAL DEVELOPMENT

CDC
CENTERS FOR DISEASE CONTROL AND PREVENTION

LEAVE NO ONE BEHIND

HITAP
Health Intervention and Technology Assessment Program

iDSI
Better decisions. Better health.

Center for Global Development

TABLE OF CONTENTS

TABLES

FIGURES

ABBREVIATIONS

ABM	Agent-Based Model
CMCC	COVID-19 Multi-Model Comparison Collaboration
CoMo	COVID-19 International Modelling Consortium
DALY	Disability-Adjusted Life Year
DHS	Demographic and Health Survey
HITAP	Health Intervention and Technology Assessment Program
ICL	Imperial College London
IDM	Institute for Disease Modeling
ICU	Intensive Care Unit
IFR	Infection Fatality Ratio
IHME	Institute for Health Metrics and Evaluation
LSHTM	London School of Hygiene & Tropical Medicine
QALY	Quality-Adjusted Life Year
SEIR	Susceptible-Exposed-Infectious-Recovered
WHO	World Health Organization

ACKNOWLEDGEMENTS

This report has been prepared by the Technical Group of the COVID-19 Multi-Model Comparison Collaboration (the CMCC). The Technical Group was chaired by Prof. Marc Brisson (Laval University, Canada) and comprised Mr. Nigel Gay (Independent consultant, UK), Dr. Jane Labadin (Universiti Malaysia Sarawak, Malaysia), Prof. Jodie McVernon (The Peter Doherty Institute for Infection and Immunity, Australia), Dr. Paula Mendes Luz (Oswaldo Cruz Foundation, Brasil), Prof. Wilfred Ndifon (African Institute of Mathematical Sciences, Rwanda), Dr. Brooke Nichols (Boston University/Wits Health Consortium, South Africa), Dr. Shankar Prinja (Postgraduate Institute of Medical Education and Research, India), Prof. Joshua Salomon (Stanford University, USA), and Dr. Akhona Tshangela (Africa CDC, Ethiopia). The Secretariat of the Technical Group comprised Dr. Hannah Clapham (National University of Singapore, Singapore), Dr. Mohamed Gad (Imperial College London, UK), Dr. Adrian Gheorghe (Imperial College London, UK), Dr. Raymond Hutubessy (World Health Organization, Geneva), Dr. Itamar Megiddo (University of Strathclyde, UK), Mr. Christopher Painter (HITAP, Thailand), and Mr. Francis Ruiz (Imperial College London, UK).

We are grateful to the COVID-19 modellers who participated in this exercise by sharing information and contributing generously their time to engage with the Technical Group: Dr. Ricardo Aguas (CoMo), Dr. Rosalind Eggo (LSHTM), Prof. Azra Ghani (Imperial College London), Prof. Mark Jit (LSHTM), Dr. Daniel Klein (IDM), Prof. Richard Neher (Basel University), Dr. Bobby Reiner (IHME), Dr. Prashanth Selvaraj (IDM), Prof. Theo Vos (IHME).

We would also like to express our gratitude to other groups of the CMCC – the Policy Group, the COVID-19 Modelling Group and the Management Group – for their support and guidance throughout conducting this work:

- The Policy Group, chaired by Dr. Suwit Wibulpolprasert, Vice Chair of the Health Intervention and Technology Assessment Program Program (HITAP), Thailand, comprised Dr. Mark Blecher (National Treasury, Republic of South Africa), Ms. Sara Cavadid (Mayor's Office City of Bogota, Colombia), Mr. Fanny Collante (Mayor's Office City of Bogota, Colombia), Dr. Juan Cordovez (Mayor's Office City of Bogota, Colombia), Dr. Zulma Cucunba (Mayor's Office City of Bogota, Colombia), Ms. Lina Galeano (Mayor's Office City of Bogota, Colombia), Ms. Luz Genera (Mayor's Office City of Bogota, Colombia), Ms. Maria Gonzales (Mayor's Office City of Bogota, Colombia), Dr. Mahlet Habtemariam (Africa Centre for Disease Control, Ethiopia), Mr. Mohsen Asadi-Lari (Ministry of Health and Medical Education, Iran), Prof. Gabriel Leung (Li Ka Shing Faculty of Medicine, Hong Kong University), Dr. Claudia Lopez (Mayor of the city of Bogota, Colombia), Dr. Sania Nishtar (the Office of the Prime Minister of Pakistan, Pakistan), Dr. Swarup Sarkar (Indian Council for Medical Research, Government of India), Dr. Hamid Sharifi (Kerman University of

Medical Sciences, Iran). The Secretariat of the Policy Group comprised Dr. Y-Ling Chi (Center for Global Development Europe, UK), Ms. Saudamini Dabak (HITAP, Thailand), Mr. Sarin KC (HITAP, Thailand), Mr. Joseph Kazibwe (Imperial College London, UK), and Dr. Yot Teerawattananon (HITAP, Thailand).

- The Modelling Group, chaired by Dr. Marelize Gorgens (the World Bank, USA), comprised Prof. Deborah Ashby (Imperial College London, UK), Dr. Emmanuel Bakare (Federal University Oye Ekiti Nigeria), Dr. Renato Mendes Coutinho (Universidade Federal do ABC, Brazil), Dr. Brittany Hagedorn (Institute of Disease Modelling, USA), Prof. Mark Jit (London School of Hygiene and Tropical Medicine, UK), Dr. Ali Mirzazadeh (University of California San Francisco, USA), Dr. Richard Neher (University of Basel, Switzerland), Dr. Sheetal Silal (University of Cape Town, South Africa), Prof. Anna Vassall (London School of Hygiene and Tropical Medicine, UK), and Prof. Lisa White (University of Oxford, UK). The Secretariat of the Modelling Group comprised Dr. Thomas David Wilkinson (the World Bank, USA) and Ms. Nejma Cheikh (the World Bank, USA).
- The Management Group comprised Dr. Kalipso Chalkidou (Centre for Global Development Europe, UK), Dr. Marelize Gorgens, Dr. Raymond Hutubessy, Dr. Yot Teerawattananon, and Dr. David Wilson (Bill and Melinda Gates Foundation, USA).

The views expressed in this document are solely those of the authors, and do not necessarily reflect the views of the organisations with which they are affiliated or their membership. More information on this project is available here: *https://decidehealth.world/en/cmcc*

All queries should be addressed to Raymond Hutubessy: *hutubessyr@who.int*

Disclosure

CMCC is an independent and non-partisan collaboration of partners. No conditions or limitations on CMCC's independence in research including data collection, analysis, reporting and resultant conclusions, recommendations and publications are attached to any funding received.

PREFACE

As COVID-19 spreads worldwide, national (and sub-national) governments and development partners are making use of a rapidly growing body of evidence to develop policies mitigating against this devastating pandemic. Mathematical models and computational simulation models play a unique role to inform resource planning and policy development (among other uses) through scenario analysis and short-term forecasting. Already in the first six months of this outbreak, we have seen many models at the sub-national, national, regional and global level being developed at an impressive speed.

For many Low- and Middle-Income Countries (LMICs), global and regional models can play an important role in country response planning in the absence of locally developed models. While at this stage of the epidemic, the availability of a diversity of modelling approaches is positive (given the varied needs of decision-makers and uncertainty surrounding this novel threat), without guidance, the modelling landscape has also become very hard to navigate for decision-makers. Moreover, it is crucial to bring in the decision-makers' views early on into model development to ensure that the assumptions, data, outcomes and policy scenarios correspond as much as possible to local characteristics and local policy needs.

A group of national governments, funders and development partners supporting COVID-19 responses in LMICs have come together under the COVID-19 Multi-Model Comparison Collaboration (CMCC). The objective of the CMCC is to enhance the use of mathematical and computational simulation models during the COVID-19 outbreak by ensuring their policy relevance, robustness, and usefulness. The CMCC was convened with funding of the Bill and Melinda Gates Foundation and Ministry of Higher Education, Science, Research and Innovation (MHESI), Royal Thai Government, and brought together the international Decision Support Initiative, the World Bank, World Health Organization and the Thai Ministry of Public Health to provide management and technical expertise to deliver global knowledge products. Other partners of the CMCC include Data 4 Sustainable Development Goals Partnership (Data4SDG), UK Department for International Development (DFID), Norwegian Agency for Development Cooperation (NORAD), The US Centers for Disease Control and Prevention (US CDC), and the United States Agency for International Development (USAID).

This report has been produced through an engagement between a team of independent infectious disease modelling experts and seven modelling groups working on COVID-19. It is a valuable resource that will equip country governments and other model users with an overview and understanding of the participating models' objectives, methods, assumptions, data sources and with guidance for assessing their fitness-for-purpose in context, framed as a dialogue between policymakers, analysts and modellers themselves. We hope that this report will enhance end users' 'command' of models that will lead to better interpretation and use of the models' results and, overall, to an evidence informed use of models in developing positively impactful policies.

EXECUTIVE SUMMARY

Background

The COVID-19 Multi-Model Comparison Collaboration (CMCC) was established to provide country governments, particularly low- and middle-income countries (LMICs), and other model users with an overview of the aims, capabilities and limits of the main COVID-19 models in use. CMCC aims to increase the understanding of model choice for different policy questions, as well as to support the understanding of key assumptions and parameters in the different models and how that could be related to differences in projections.

This report is one of the main deliverables in CMCC's Phase 1, which aims to conduct an initial fitness-for-purpose assessment of current COVID-19 model structures. The report includes two outputs:

1. A flow chart of key questions to assess the fitness-for-purpose of an initial set of COVID-19 epidemiological models, with a special focus on their use in LMICs; and
2. A description and comparison of the aims, methods and reporting standards of the participating COVID-19 models, as well as the level of engagement and expertise required to use or adapt the models to LMIC settings.

Methods

The CMCC's Management Group and Technical Group sought to prioritise the most widely used COVID-19 models for decision-making based on the following criteria:

- The model is applicable or can be adapted to multiple countries and/or regions;
- The model is dynamic and seeks to inform the impact of different COVID-19 policy interventions;
- Results have been published in the public domain; and
- Developers are willing to participate in the comparison study.

Seven models were selected to be a part of this analysis (model name, where applicable):

- Basel University (*Covid-19 Scenarios*);
- COVID-19 International Modelling Consortium (*CoMo*);
- Imperial College London (*Squire*);
- *Institute for Health Metrics and Evaluation*;
- *London School of Hygiene & Tropical Medicine*; and
- Institute for Disease Modeling (*Covasim* and *EMOD*).

A structured questionnaire was developed to facilitate extracting key information regarding each of the models included in the comparison. The questionnaire was initially completed by the Secretariat and members of the Technical Group, then discussed during an interview with each modelling group. These interviews took place in late May and early June 2020. A draft version of the findings of this report were presented to the modelling groups in mid-July to provide them with a further opportunity to verify the information that has been collected on each model and offer feedback on how the findings have been communicated.

Findings

The current versions of the seven models included in the comparison, as of 31 July 2020, share a number of commonalities and differences. Most models are mechanistic, incorporate age-distribution in transmission and mortality parameters, and base many parameter values (e.g. infection fatality ratios, IFRs) on data collected early in the pandemic. At the same time, most models do not account for particular sub-populations or comorbidities, do not include indirect COVID-19 effects on other diseases and do not link these outcomes or the included interventions to economic outcomes.

The models differ to varying degrees in their purposes and aims, in how COVID-19 transmission and contact patterns were implemented, in the breadth of interventions considered, and in how these interventions (particularly contact tracing) were constructed. It is important to note that the models continue to evolve according to the policy questions being asked (which is a function of the phase of the epidemic and setting-specific issues), as the understanding of COVID-19 improves and more epidemiological data become available from different settings. Details of the model description questionnaire, which can be used by researchers and analysts as a basis for other similar exercises, is available in an online appendix.

Based on these findings, the CMCC's Technical Group developed a proposed tool for judging the fitness-for-purpose of alternative models for addressing different policy questions, which may be useful for decision-makers and analysts at both global and country levels. It is hoped that the application of this tool can foster and guide an ongoing dialogue between the two key audiences – policy and technical – through identifying a series of fundamental questions that expose the trade-offs that have to be made when selecting and adapting a model for a given (set of) policy question(s), context and decision constraints at a point in time.

Conclusions

A number of lessons relevant to multi-model comparison exercises also emerged from this process. Keeping up with the rapidly changing policy and behavioural context (e.g. the growing importance of face masks), together with the corresponding evolution of the models themselves, has required maintaining an intense level of engagement with the participating modelling groups throughout this process while trying to minimise demands on their time. Working alongside the Policy Group of the CMCC from the outset has provided the Technical Group with invaluable insights into what matters most for policymakers in LMICs during this pandemic. These insights are reflected in the design of key deliverables, most notably in the model description questionnaire and the fitness-for-purpose flow chart. Finally, discussing intermediary outputs of the comparison with participating COVID-19 modellers has offered opportunities for course correction and has undoubtedly strengthened the exercise.

In the table below, we also propose a number of recommendations for optimising modelling approaches with a view to improving the engagement between policymakers and modellers, ultimately leading to better policy decisions.

Summary of recommendations

For policymakers	For modellers
Question whether modelling tools can address the policy question and if so, whether to use/adapt existing models or to develop completely new ones (if necessary and feasible). Engage local analysts and researchers early on in assessing the individual, organizational and institutional capacity for engaging with models and in the fitness-for-purpose assessment. Make clear to local analysts and modellers the types of decision constraints being faced e.g. time, stakeholder coordination, infrastructure, budget and how the results will be used and disseminated. Conduct the fitness-for-purpose assessment iteratively and not as a one-off exercise because i) models evolve quickly; and ii) depending on context and the specific policy question, the fitness-for-purpose of a given model may also change.	Identify concrete approaches to involve policymakers and analysts in LMICs in developing or adapting the model and user interfaces to maximise relevance for such settings; consider following the collaborative modelling approach in the Policy Group report (Figure 8). For models that are already developed: Clarify in all available model documentation the types of policy questions that the current model version can be used for, and its main limitations; consider using the policy question typology in the Policy Group report (Table 3).
Consider and plan early on for rapid data collection which may minimise uncertainty in model results. Set up a consultative process involving local analysts and researchers for defining, reviewing and validating model assumptions where data are not available. Prefer models that have been calibrated and validated for your setting using appropriate methods given the policy question.	Identify clearly a minimum set of model parameters that should be ideally informed by local data in order for the model to be applied credibly in a given context; refine it continuously based on context-specific uncertainty and sensitivity analyses; support local analysts and policymakers in identifying appropriate data sources and in collecting additional data. Wherever possible engage with local partners and experts to validate key assumptions and the quality of the sources of the setting specific parameters. Ideally, train local analysts on how to use the model themselves.
Seek commitment from modellers to adhere to the recommended reporting trajectories	Commit to the recommended reporting trajectories proposed in the Policy Group report (Figure 6)

INTRODUCTION

Epidemiological mathematical and statistical models have been and continue to be used by high-, low- and middle-income countries' governments to inform their policy responses to and public communication regarding the COVID-19 pandemic. There is an extensive range of model types and approaches designed to address differing policy and analytical questions. The evidence base related to the epidemiological profile of the pandemic in each country and to the effectiveness of interventions is fluid and complex. Considerable uncertainty remains in regard to key parameters of disease transmission. All these factors lead to uncertainty for decision-makers, who currently have limited guidance on how to use findings from such models.

To address this challenge, the COVID-19 Multi-Model Comparison Collaboration (CMCC) was established to provide country governments, particularly low- and middle-income countries (LMICs) which have limited capacity to develop their own models, and other model users with an overview of the aims, capabilities and limits of the main COVID-19 models in use. Moreover, to increase the understanding of model choice for different questions, and of the key assumptions and parameters in the different models, including how that could be related to differences in projections. Through this work, the CMCC aims to help model users to better understand the uses of each model (i.e. mapping out their purpose for decision-making), and better interpret the outputs from these tools for planning and strategic decisions.

Crucially, the aim of this work is not to rank, appraise or approve the models, nor is it to state whether any model is "right or wrong". Instead, it is to :

- Compare the models under a number of dimensions (details below) ;
- Improve the validity and transparency of existing COVID-19 models by facilitating the exchange of information and encouraging discussions between the different groups and stakeholders participating in the exercise ; and
- Guide decision-makers through assessing the fitness-for-purpose of models for policy decisions and recommendations in LMICs.

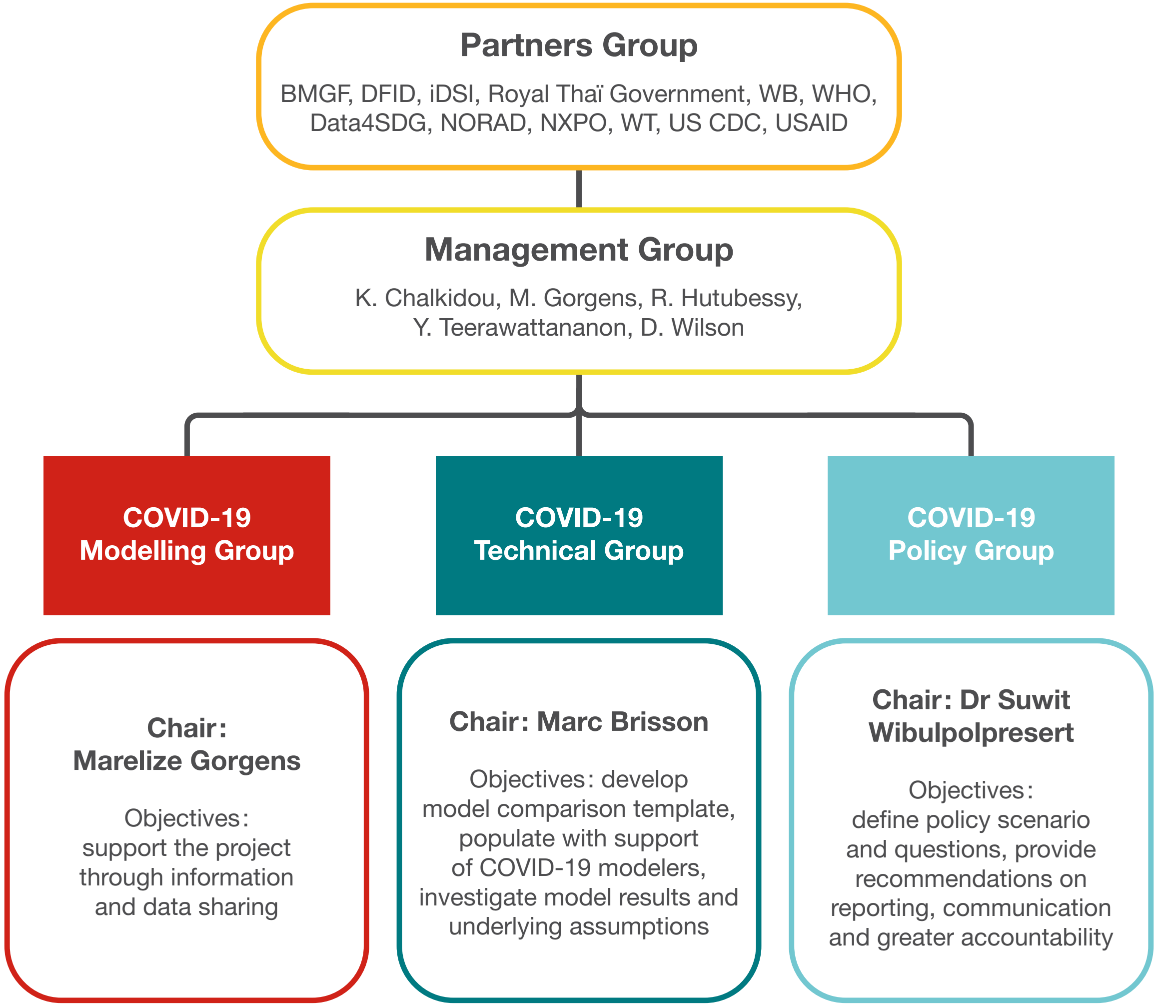
The CMCC's structure is shown below (Figure 1). The composition of CMCC's groups is detailed in the Acknowledgements section. The Technical Group, who produced this report together with the Technical Group Secretariat, comprised ten disease modelling experts independent of the COVID-19 models included in the comparison.

The CMCC's work was structured in two phases: (Phase 1) an initial fitness-for-purpose assessment of current COVID-19 model structures; and (Phase 2) an invitation for modelling groups to participate in the prospective modelling of a few selected policy questions. This report is one of the two main deliverables of Phase 1. It contains the following information about the participating "global" COVID-19 models :

- A fitness-for-purpose tool of COVID-19 epidemiological models, which highlights the linkages to intended policy questions in LMICs ;

- A descriptive comparison of the models that have been used in multiple LMICs in terms of: aims, countries modelled, structure, key parameters, data requirements, outputs, transparency, and level of engagement and expertise required to use or adapt the models to different settings;
- A description of the applicability of the models to LMIC contexts, identifying particular challenges and opportunities for representing the COVID-19 pandemic in resource-constrained health systems and societies;
- Recommendations on how the modelling approaches can be optimised to improve policy engagement and decision-making.

The model descriptions are detailed in language intended for a broad audience of policymakers, public health practitioners and scientists/researchers without having necessarily expertise in epidemiological modelling. The scientific knowledge of COVID-19, understanding of the impact of non-pharmaceutical and pharmaceutical interventions and related policy questions are evolving rapidly. The models assessed are therefore evolving at a rapid pace. Time stamps are thus included to indicate which version of the model was assessed i.e. with model features implemented as of 31 July 2020.



Partners Group

BMGF, DFID, iDSI, Royal Thaï Government, WB, WHO, Data4SDG, NORAD, NXPO, WT, US CDC, USAID

Management Group

K. Chalkidou, M. Gorgens, R. Hutubessy, Y. Teerawattananon, D. Wilson

COVID-19 Modelling Group

Chair: Marelize Gorgens

Objectives: support the project through information and data sharing

COVID-19 Technical Group

Chair: Marc Brisson

Objectives: develop model comparison template, populate with support of COVID-19 modelers, investigate model results and underlying assumptions

COVID-19 Policy Group

Chair: Dr Suwit Wibulpolpresert

Objectives: define policy scenario and questions, provide recommendations on reporting, communication and greater accountability

Figure 1 - Structure of the CMCC

METHODS

Once the membership of the Technical Group was finalised, its members and the Secretariat met 11 times (between 24 April 2020 and 31 July 2020) to discuss and deliberate until there was consensus over how to proceed with the model comparison using the WHO multi-model comparison guide as a reference (Denboon et al. 2019), including deciding upon the inclusion criteria for this analysis.

Inclusion Criteria

The CMCC's Management Group and Technical Group sought to prioritise the most widely used and recognised COVID-19 models for decision-making based on satisfying the following inclusion criteria:

- The model is applicable or adapted to multiple countries and/or regions;
- The model is dynamic and seeks to inform the impact of different COVID-19 policy interventions;
- Results have been published in the public domain; and
- Developers are willing to participate in the comparison study.

Taking into account the resource and time constraints of the CMCC, and to ensure that the findings from this analysis could be disseminated in a timely manner to maximise their impact, seven models were selected to be a part of this analysis.

Selection and Invitation

Firstly, the COVID-19 modellers were invited to participate in the modelling comparison exercise and asked to select one model and version to submit for the comparison. Model versions as of 31 July 2020 were reviewed by the CMCC Technical Group. The modelling groups and models included in the first phase of this analysis are listed below (Table 1).

Table 1 - List of models and modelling groups included in Phase 1 of the CMCC review

Modelling Group	Model name(s), if applicable
University of Basel	*Covid-19 scenarios*
COVID-19 International Modelling Consortium (*CoMo*)	
Imperial College London (ICL)	*Squire*
Institute for Health Metrics and Evaluation (*IHME*)	
London School of Hygiene & Tropical Medicine (*LSHTM*)	
Institute for Disease Modeling (IDM)	*Covasim* and *EMOD*

Data Elicitation and Extraction

A structured questionnaire was developed by the Technical Group to facilitate the extraction of key information regarding each of the models included in the comparison. The primary function of the questionnaire was to understand the objectives and characteristics of the model, the included interventions, methods and any validation[1] or calibration procedures. The questionnaire was then initially completed by the Secretariat and members of the Technical Group using publicly available information regarding each participating model, before being shared with the modelling groups themselves to verify that this information was correct.

The completed and verified questionnaires were then discussed further during an interview with each modelling group, involving at least two members of the Technical Group and representatives of the CMCC Management Group. To ensure consistency, the chair of the Technical Group was present for all interviews. Prior to each interview, the interviewers reviewed the completed questionnaire and identified areas for discussion and further questions. The interview, of approximately 1 hour in duration, aimed to elicit further clarifications about the characteristics, methodological nuances and interpretation of results of the model. In addition, the interview provided an opportunity to better understand how the models are being and have been used, how the models were fitted, calibrated and validated as well as how they were developed, including the extent to which other stakeholders were engaged in this process.

The model developers were also given a chance to describe the limitations of their models and outline the circumstances in which the model is most appropriate for use. These interviews took place in late May and early June 2020. If needed, further clarification on details of the model was sought through email or follow up calls as needed. Draft findings of this report were presented to the modelling groups in mid-July so they could verify the accuracy of the findings and provide feedback on how findings are communicated.

The timeline of the Technical Group's work is presented in Figure 2.

---

[1] Terms in bold are defined in the Glossary at the end of the report.

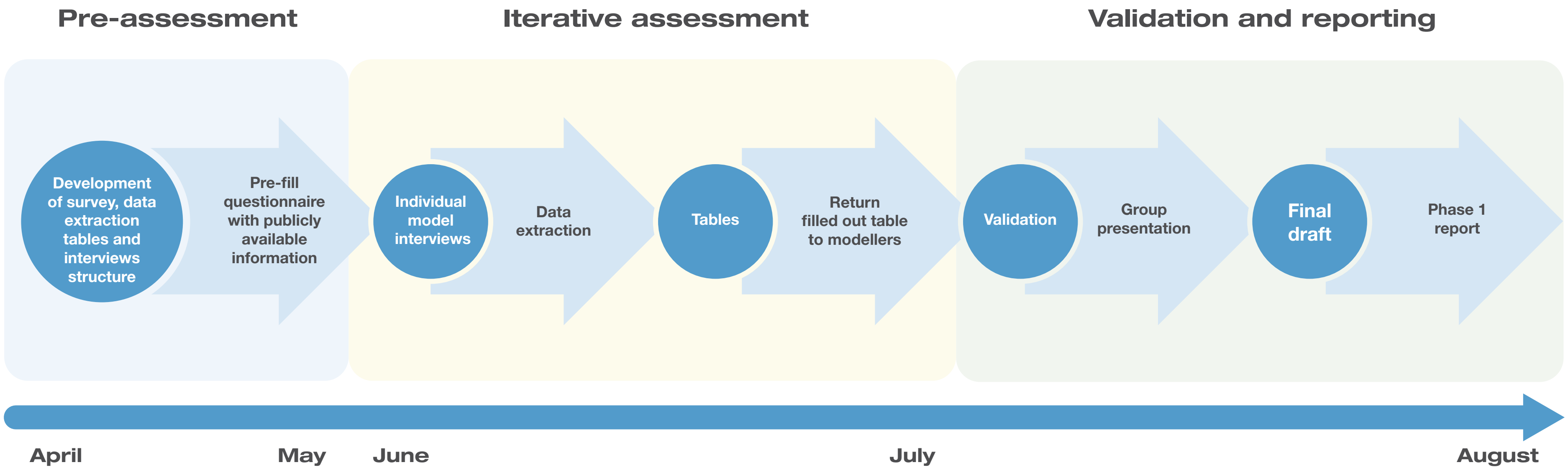


Figure 2 - CMCC's assessment, validation and reporting flow chart and timeline

FINDINGS

As outlined in the Introduction, one of the main purposes of this model comparison is to guide decision-makers through assessing the fitness-for-purpose of models for policy decisions and recommendations. In this section, we first describe how the "fitness-for-purpose" concept was operationalised in this exercise as a series of questions spanning four domains: Policy Aims, Modelling Feasibility, Model Implementation and Model Reporting Commitment. Then, we describe the characteristics of the seven included models following the structure of the fitness-for-purpose questions.

Defining Fitness-for-Purpose: a framework to support model selection

Model users and developers need to account for many considerations when deciding if a model is fit-for-purpose. The question is more «Is the model fit for the current question I would like to answer, given the setting and data availability?". The answer to this will not be general but will depend on a number of factors that we outline here and set out in a decision flow chart (Figure 3). Not all factors can always be fully met, therefore model users and developers will need to balance some of the factors, make trade-offs, and decide where to compromise.

The first thing to consider is what is the purpose of the modelling exercise, or specifically, what one expects the model to be able to do and be used for. In this report, we have covered models that span a range of purposes; for example, predicting case numbers or whether there is sufficient bed capacity. The majority of the models we cover here are able to simulate such outcomes given different interventions.

Next to consider are which interventions to model and whether it is possible to meaningfully incorporate them in the analysis given that such choices will be determined by the type and structure of the model.

Another important issue relates to data needs: what information is required to ensure that a model is meaningful in a given setting, and then, whether such data, suitably contextualised where possible, are actually available. This could be input data such as contact patterns by age or the proportion of cases that are severe or die; it could also be outcome data like the number of cases or deaths that can be used to adjust the model in order to ensure that parameter values are valid for the setting. In the absence of context-specific and/or up-to-date information, assumptions will need to be made, as has been the case with using modelled contact matrices or contact matrices from a "similar country", which shifts the burden onto model users to agree what constitutes an acceptable level of "similarity".

Model users are also interested in the validity of model predictions. The validation expected for a model will depend on the policy question. For instance, if the aim of the modelling analysis is to predict the course of the epidemic in a particular setting, policymakers may ask the extent to which the model performs well at predicting what will happen – in other words, whether it has been validated by comparing past predictions, both in general and in relation to their setting. While a model should not be necessarily criticized for failing to predict the future, a good model should be susceptible to recalibration or respecification incorporating new evidence.

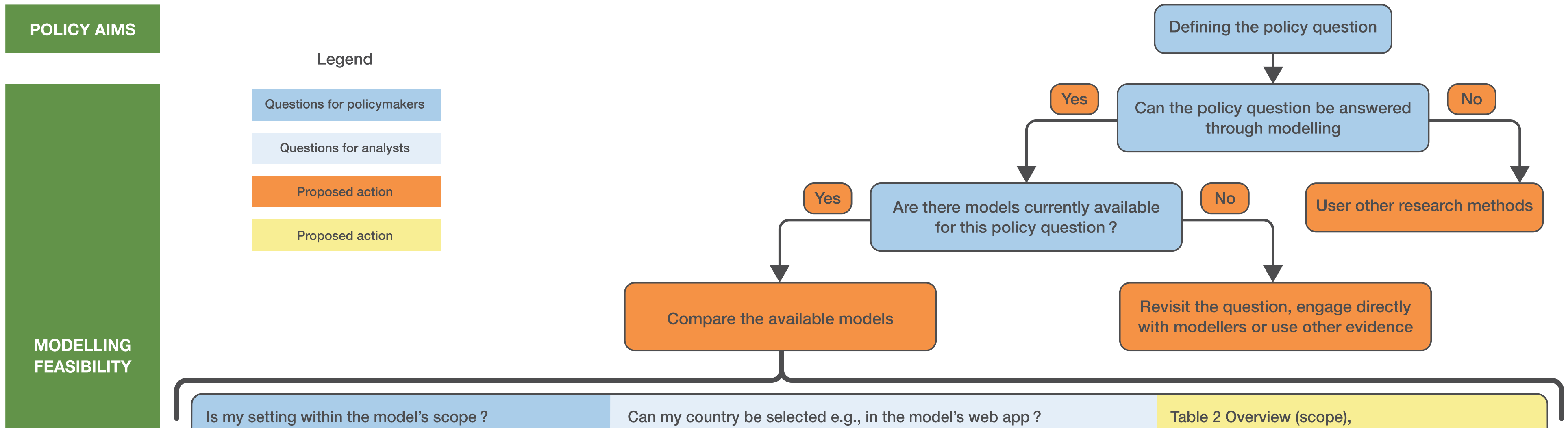
Predictive validity can be difficult because policy action is swift in a crisis situation such as the COVID-19 pandemic and does not wait for model scenarios to play out; therefore, the time window for validation and the counterfactual need careful consideration. Public health and social interventions approximated in model frameworks will likely play out very differently in their 'real world' implementation due to context-varying factors such as imperfect compliance and capacity constraints. Moreover, once a particular policy decision has been made within a given epidemic phase, alternative futures that might have been predicted by a model no longer have opportunity to be observed and compared.

At a minimum, internal model validation (or model verification) is required. That is, models should be carefully tested to ensure their calculations are accurate (e.g. test for programming errors), that the model outputs are consistent with the setting data that are available (calibration), and that local experts and policymakers are consulted to verify whether the results make sense and can be explained at an intuitive level (face validity).

Finally, model users are often faced by decision-time and logistical constraints. How quickly do they have to make a decision? How much technical resources do they have available to interact with the model and its developers in order to implement context-specific adaptations? Some models can be used "off the shelf", usually at the cost of making some simplifying (although not necessarily unreasonable) assumptions, while other models can only be used in direct collaboration with the model developers – in this case, results may be more sensitive to the context and more likely to be correctly interpreted and applied, but such collaboration also takes time and expertise which policymakers in LMICs may or may not have to the required extent.

The flow chart (Figure 3) is primarily addressed to policymakers who are considering the use of infectious disease models to answer specific policy questions related to COVID-19; it aims to guide them through a sequence of questions and decisions which start with assessing the suitability of models to answer the respective policy question(s). If models are appropriate and available, the questions continue in order to compare models in terms of feasibility and implementation. Next, once a model has been selected for use, the questions in the flow chart end with considerations around the recommended model reporting standards (developed by the CMCC's Policy Group). The questions for policymakers are accompanied by corresponding in-depth technical questions for local analysts to answer upon examining the models and in conversation with model developers. The extent to which a model is fit-for-purpose will be ultimately a matter of judgement resulting from an ongoing dialogue between policymakers, local analysts and COVID-19 modellers, informed by the answers to these questions.

Figure 3 - Fitness-for-purpose flow chart



POLICY AIMS

Legend

- Questions for policymakers
- Questions for analysts
- Proposed action
- Proposed action

Defining the policy question

Can the policy question be answered through modelling

Yes: Are there models currently available for this policy question ?

No: User other research methods

Are there models currently available for this policy question ? Yes: Compare the available models

Are there models currently available for this policy question ? No: Revisit the question, engage directly with modellers or use other evidence

MODELLING FEASIBILITY	Is my setting within the model's scope ?	Can my country be selected e.g., in the model's web app ? Can input parameters be specified by the user ?	Table 2 Overview (scope), pages 25-26
	Does the model address the specific policy question ?	Is the model's design suitable for the stated aim(s) ? Are the model's outputs and outcomes suitable for the stated aim(s) ? How are interventions being modelled? E.g. contact matrices, spacial distribution, other adjustment factors Does the model's time horizon cover the horizon of the policy question ?	Table 2 Overview (aims), pages 25-26 Table 3 Outputs, page 27 Table 5 Interventions, page 35-36 Table 2 Overview (Time horizon), pages 25-26
	Are there sufficient time and in-house expertise available to engage with the model ?	How can users engage with the model ? Does it have a user-interface ? What is the programming language ?	Table 6 Transparency, page 39 Table 2 Overview (Policy engagement), pages 25-26
	If YES to the policymakers' questions above, continue below.		
	If NO to any of the questions above, modify existing model or build a new model.		
MODEL IMPLEMENTATION	Are in-country data needed and, if so, are they available or can reasonable approximations be made ?	What data are available for which modelcomponents? E.g. for transmission, for calibration, for stratification, comorbidities.	Table 4, Parameters, pages 30-32
	Has the model been validated ?	How have the results been validated ? What are the validation processes in place ? Has the model been calibrated ? If so, for what parameters ? What are the calibration procedures in place ?	Table 4, Parameters, pages 30-32
	If YES to the policymakers' questions above, select or adapt the model.		
	If NO to any of the questions above, modify existing model or build a new model.		
MODEL REPORTING COMMITMENT	Have the modellers committed to following the reporting standards trajectory ?	How have results been validated ? What are the validation processes in place ? Has the model been calibrated ? If so, for what parameters ? What are the calibration precedures in place ?	Table 2, overview (scope), pages 25-26

Model aims, types and policy uses

In this section we describe the characteristics of the included COVID-19 models. The participating modelling groups validated the summary tables below, which include model features implemented as of 31 July 2020, and also had the opportunity to review table data in relation to other models.

Overview of models

Table 2 provides an overview of the participating COVID-19 models. We will discuss these results as they relate to the questions outlined in Figure 3.

Is my setting within the model's scope ?

The included models have global scope, though they do not quite cover all LMICs in the World Bank classification (July 2019) of such countries (Appendix 1). Most models have been used to simulate a subset of the countries within their scope, and they have the capacity to model additional settings provided contextualising data, calibration, and/or engagement with the modelling group.

Does the model answer my specific policy question ?

Most models aim to predict the epidemiological impact of interventions. Some models provide explicit results (as long-term scenarios and/or short-term forecasts) based on best available parameter estimates and selected scenarios, while others (e.g. Basel, CoMo) may be better understood as tools that allow users to build their own scenarios. The IHME model attempts to also project the short-term epidemiological burden, which requires projecting population and government behaviour. In this model, population behaviour is captured through covariates such as mobility, and government behaviour by providing a few alternative scenarios (similarly to the predict the epidemiological impact of interventions models). Other models can be used to project the short-term burden given the assumption that no behaviour changes occur over the period. However, the model projections beyond a few weeks into the future may not match observed incidence and burden if behaviour changes or additional interventions are put in place.

At the time of writing, the seven models focus mainly on COVID-19 specific epidemiology. They do not include indirect effects on other diseases, or the economic and financial impact of COVID-19 or of the interventions set in place to mitigate the spread of COVID-19. Some of the participating modelling groups have indicated that they explored these aspects separately; it can also be argued that this represents good modelling practice as a single model can hardly answer all potential policy questions of interest. However, this can also cause difficulties for policymakers trying to understand how the scenarios in the epidemiological models link with the economic impacts that are modelled separately. The LSHTM model does incorporate some economic outputs.

The models can be used to run counterfactual scenarios and explore potential future outcomes given combinations of interventions (or no intervention); the possible interventions are shown in Table 5. Many of these models can also explore historical counterfactual scenarios to answer questions about what if scenarios such as, what would have happened if social distancing did not taken place (see time horizon row of Table 2).

Are there sufficient in time and in-house expertise available to engage with the model?

The models' and their developers' differing approach to engagement with users and regional and country experts have implications for each model's fitness-for-purpose. Models that focus on one-way communication, providing users with information, require little computational capacity and technical expertise from users beyond a broad understanding of the model and its validity. However, these models are inflexible: the user's setting/country (or one they deem sufficiently similar) needs to have been modelled ex ante.

Models with user-friendly apps allow users to conduct exploratory analyses and provide varying levels of flexibility depending on the inputs that can be adjusted. These models require higher technical expertise, which can be quite minimal, but depends on the parameter inputs that users can adjust. Users can typically adjust inputs to better represent their setting or a similar one.

Models that do not provide tools for two-way communication or exploratory analysis in user-friendly apps require significantly more technical expertise but are more flexible. Adapting models with user-friendly apps beyond their original scope (e.g. implementing interventions that are not included) similarly requires more technical expertise from users. Model users need to either adapt the model code on their own, provided it is open-source and that they have the computational and technical capacity, or work closely with the modelling group, provided the group has time and capacity for such co-development. The CoMo model, for example, has the mode of working in which groups around the world can join the consortium, with access to the CoMo set up, code and interface and then interact with the CoMo development team for support in answering relevant country questions with data available to the country modellers. Other modelling groups have engaged with local partners through their networks and/or with potential users on online platforms such as GitHub.

As such, some level of technical expertise, time and resources will always be required to engage with modellers. Depending on how a given model is structured and made available, this engagement effort may focus more on checking that the model's data and assumptions are appropriate for a given setting or on adapting the model and populating it with relevant data.

Model types

All models describe disease transmission **mechanistically** to explicitly capture the effect of interventions on transmission and predict the impact of interventions. Out of the seven models, five are mechanistic compartmental models (susceptible-exposed-infected-recovered, SEIR) and two are agent-based models (ABMs). ABMs model the population at the individual level, providing more granularity compared to compartmental SEIR models. Compartmental SEIR models aggregate the population into a few subgroups that describe health states and in some cases characteristics that are deemed important for modelling transmission. Within these subgroups the population is assumed to be homogeneous and well-mixed—infection events are equally likely between any individuals.

Since evidence suggests that contact rates differ between age-groups, most of the models assume an age-structured population, including the ABMs that can model age at the individual level; the exceptions are the IHME and Basel COVID-19 Scenarios models, which assume transmission is equally likely across all ages. The IHME model also differs from other SEIR models by using a statistical model to estimate a time-context-scenario-varying infection rate which has been applied before but not tested to the extent that more conventional compartmental models have been. ABMs offer greater flexibility (e.g. to explicitly model space) than compartmental models and can model complex interactions and specific interventions more explicitly (see Table 5 for more detail on interventions). They can also model more explicitly heterogeneous events such as so-called ‘super spreaders’ in more detail. However, ABMs may require more data when built to more explicitly model heterogeneity, making model adaptation to multiple scenarios more challenging, and implementing or adapting them generally requires greater technical programming and computational capacity.

Several of the models are stochastic, capturing inherent randomness in describing real world events. In contrast, deterministic models will always produce the same outputs for a given set of parameters and initial conditions. The difference between these types of models impacts how uncertainty is reported, which is discussed below (Table 6).

Model outputs

Model outputs, described in Table 3, include a variety of health-related and resource use results and states derived from the initial projected number of infections, and aim to provide information on how, for example, symptomatic cases and/or health care utilization may vary over time. Most models assume a sequential nature to these health states, which in turn require assumptions about the proportion of a cohort in health state that evolves to a subsequent stage; for example, the proportion of symptomatic cases that evolve to hospitalization and, of these, the proportion requiring intensive care unit (ICU) beds, etc. As such, most models provide information on symptomatic cases, the need for hospital beds, the need for ICU beds, and deaths, though the denomination of these health states might differ in each model. For most models, these outputs can be further broken down by age-groups or perhaps even further for ABMs. To date, only the LSHTM model reports using generic health outcome measures such as QALYs, although only in UK applications so far; and health system and household medical costs for a limited number of LMIC settings.

All the models included here only model the direct impact of infection on COVID-19 deaths. They do not model the impact of COVID-19 case counts on hospitals and the collateral impact on other diseases if other health services are reduced. The models also do not include either the health or economic impacts of the control measures. This should be understood rather as a “limitation of purpose” than a failure of “fitness-for-purpose”.

Table 2 - Model overview

	LSHTM	**Basel *COVID-19 Scenarios***	**Imperial *Squire***	**Oxford CoMo**	**IHME**	**IDM *Covasim***	**IDM *EMOD***
SETTINGS modelled							
Countries modelled to date (as of July 31st)	Over 130	Over 120	Over 130	About 20 (3 districts)	All with more than 3 deaths	US (Oregon state, King County, WA), UK, Australia, and Nigeria (Lagos)	Ethiopia, Ghana, South Africa
Countries can be modelled (provided contextualizing data, calibration, engagement)	All	All	All	All	All with more than 3 deaths	All	All
AIM of the MODEL							
Project short-term epidemic course and/or burden of interventions	Yes	No	No	No	Yes	Yes	Yes
Scenario analysis of interventions epidemiological impact	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Scenario analysis of the interventions' economic impact	Yes◆	No	No	No	No	No	No
Time horizon*	1 year*	User defined	1 year*	User defined	4months in the web app, longer durations can be defined	User defined; online tool suggests 90 days	User defined

* *Can be modified*

◆ *The features for exploring these aims are not included in the web-app presently.*

	LSHTM	**Basel *COVID-19 Scenarios***	**Imperial *Squire***	**Oxford CoMo**	**IHME**	**IDM *Covasim***	**IDM *EMOD***
EXPERTISE and ENGAGEMENT REQUIRED							
User-facing app▲	Yes (additional features external to app)	Yes	Yes (additional features external to app)	Yes (additional features external to app)	Yes (one-way communication tool)	Yes (additional features external to app)	No
Level of engagement required with regional experts	High level policy engagement	Ad hoc communication with groups using the model	High level policy engagement	Collaborative modelling with modellers in country	Engagement with WHO regional offices and country teams	Collaborate with partners to setup simulations in LMICs	Collaborate with partners to setup simulations in LMICs
TYPE of MODEL							
Dynamic/Static	Dynamic	Dynamic	Dynamic	Dynamic	Dynamic	Dynamic	Dynamic
Compartmental / Agent-based	Compartmental	Compartmental	Compartmental	Compartmental	Compartmental	Agent-Based	Agent-Based
Natural history●	SEIR	SEIR	SEIR	SEIR	SEIR	SEIR	SEIR
Age-structured	Yes	Disease progression but not transmission	Yes	Yes	No	Yes	Yes
Deterministic / Stochastic	Stochastic + Deterministic	Deterministic	Stochastic	Deterministic	Deterministic + Statistical	Stochastic	Stochastic

▲ *Unless otherwise specified as a one-way communication tool, the apps allow users to adjust inputs; which inputs can be adjusted varies by model. Several of the models with apps also have software packages or code with features (e.g. outputs or interventions) that are not included in the apps.*

● *The models' natural history are variants of SEIR (susceptible-exposed-infected-recovered). These include additional natural history states such as asymptomatic, symptomatic and infections requiring hospitalisation.*

Table 3 - Model outputs

Outcomes	LSHTM	Basel *COVID-19 Scenarios*	Imperial *Squire*	Oxford CoMo	IHME	IDM *Covasim*	IDM *EMOD*
Symptomatic Cases	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reported cases	No	No	Yes	Yes	Yes	Yes	No
Hospital beds needed	Yes	Yes	Yes	Yes	Yes	Yes	No
ICU beds needed	Yes	Yes	Yes	Yes	Yes	Yes	No
Ventilators needed	No	No	Yes	Yes	Yes	No	No
Oxygen needs	No	No	Yes	No	No	No	No
Deaths	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Generic health outcome measure (DALY or QALY)	Yes*	No	No	No	No	No	No
Economic outcomes	Yes*	No	No	No	No	No	No

** Not included in the web-app presently. QALYs, health sector costs and macroeconomic (GDP) impact only for the UK.*
Health sector and household medical costs for a limited number of LMIC settings.

Important context specific model parameters

Our synthesis of model parameters does not focus on the general parameters that are important for natural history modelling, but on the context-specific ones that are particularly relevant for adapting models focused on high-income settings to LMICs. Though several modelling groups have conducted internal ad hoc **sensitivity analyses**, none of the groups have conducted an extensive sensitivity or robustness analysis and published the results at the time of the interviews; some have reported to be currently working on this aspect. Consequently, our understanding of the most important parameters needed in modelling and of which parameters are critical in terms of using localised data is incomplete. In interviews with the modelling groups we identified parameters that were thought to be particularly important for predicting outcomes, and these are presented in Table 4. These should be considered when answering the question: “Are in country data needed and, if so, are they available or can reasonable approximations be made?”

The **infection fatality ratio** (IFR) affects models’ predicted deaths and their distribution across sub-populations. Most models account for the age distribution, which is important because of the observed relationship between age and mortality, but do not include other sub-populations, such as those with comorbidities. Most models use data from early in the pandemic in China, and in particular data from the Diamond Princess cruise ship as well as other sources where testing coverage was/is high. They take this approach since IFR is difficult to estimate in settings with unknown testing coverage. However, the implicit assumption is that IFR by age-group is consistent across countries. This is a strong assumption that may not be true in LMICs, where comorbidities occur earlier in life and are less likely to be treated than in high-income countries. A couple of models attempt to adjust for comorbidities, though in an informal manner—shifting the IFR ten years down from that observed so far or providing the user an option for input multiplier on IFR. The Imperial College’s Squire model has adjusted for comorbidities in bespoke analyses. Several models also adjust the IFR for the health system’s capacity to treat infections, though data on this effect are also limited. These findings are similar for the parameter on the proportion of infections that are severe, which models often describe as ones requiring hospitalisation or ventilators and often use data on these outcomes to describe severity. More work is required to understand variability in age specific IFR by country/setting to assess the impact of generalising IFR from one setting to another.

The transmission and contact rates are important parameters for projecting the spread of infectious diseases. Some models describe these separately as i) the rate of contact and ii) probability of infection per contact, while others abstract from the more explicit modelling of transmission events and combine these into one value called the transmission rate. Contact rates are age-structured and country-specific in most models; however, they are typically based on contact matrices (e.g. Prem et al., 2017) that have been inferred using statistical models that either extrapolate from contact surveys in other countries, or from DHS data on demographics such as household size, and not from collection of primary data on contact rates in each country. The LSHTM group recently validated synthetic contact matrices against out-of-sample empirical data and found that they generated similar findings in all but a few age groups and recommended that “modellers should consider using multiple contact matrices constructed using different methods for sensitivity analyses” (Prem et al., 2020).

For countries where the data are unavailable, models often use values for countries that are deemed similar. Several of the models differentiate between home, school, work, and community settings, but only one model—IDM’s spatial ABM EMOD—considers urban versus rural settings. Models that more explicitly model contact and infection typically adjust the probability or rate of infection in their calibration procedure (if occurring), so transmission rates produce case patterns that match observed data.

It will be important to get information on the IFR from populations with different underlying morbidity profiles, with the understanding that this is often difficult data to collect, interpret and therefore use in a model given the differences in testing and reporting systems between places. Information on the age of cases and deaths over time will be particularly useful here, with ideally also information on who is tested by age. The age-specific contact rates, and how these vary by setting (e.g. home, work, or school) in each country are important data to obtain from local contexts.

Table 4 - Parameters of interest: assumptions and data used

	LSHTM	Basel *COVID-19 Scenarios*	Imperial *Squire*	Oxford CoMo	IHME	IDM *Covasim*	IDM *EMOD*
INFECTION FATALITY RATIO							
Age-stratified	Yes	Yes	Yes	Yes	Yes	Yes	Yes[◆]
Co-morbidities accounted for	IFR shifted 10 years down	No	Included in bespoke country-specific analyses by modifying IFR	No	No	User input modifies proportion severe, which adjusts IFR	User input modifies proportion severe, which adjusts IFR
Impacted by healthcare capacity	No	Yes	Yes	Yes	No	Yes	Yes
IFR data source (not demography)	China	China or user input	China; impact of health care capacity from UK expert elicitation; impact of comorbidities from UK and international data and prevalence from Global Burden of Disease data	China and USA; or user input	Prioritize locations with high testing capacity (e.g. New Zealand, Germany, and Diamond Princess)	China; or user input overrides default	China; or user input overrides default
PROPORTION of INFECTIONS THAT ARE SEVERE							
Age-stratified	Yes	Yes	Yes	Yes	Yes	Yes	No
Co-morbidities included	Calculated from data source below and then shifted down 10 years	No	No	No	No	User input modifies (custom code needed for targeting subpopulations)	No
Impacted by healthcare capacity	No	No	No	No	No	No	No
Severity data source	China and multicountry analysis (Davies et al)	China, UNSD database API	China and UK	US and China	Prioritize locations with high testing capacity	China	NA

** EMOD's IFR is applied statistically as a post dynamic simulation multiplier of the number of cases*

	LSHTM	**Basel *COVID-19 Scenarios***	**Imperial *Squire***	**Oxford CoMo**	**IHME**	**IDM *Covasim***	**IDM *EMOD***
TRANSMISSION RATE / PROBABILITY*◆							
Country specific?	No	Can be estimated fitting to data in online tool (country data pre-sets)	Estimated by fitting to data in online tool (country data pre-sets)	Estimated fitting to data in online tool visual calibration	Yes, by combining country specific covariates with global coefficients	Estimated in calibration procedure	Estimated in calibration procedure
Data source	Own meta-analysis	Fitted to ECDC, US COVID, other government portals	Fitted to death data from ECDC	Fitted to user specified data on cases or deaths	Fitted to country-specific data collated by IHME	Fitted to user specified data on any time-series observation (e.g. deaths and hospitalization)	Fitted to user specified data on cases
CONTACT RATES							
Country specific?	Yes	No	Yes	Yes	Combined with transmission rate	Yes	Yes
Age-structured?	Yes	No	Yes	Yes	No	Yes	Yes
Setting	Home, work, school, and community	No	Home, work, school and community	Home, work, school, and community	Partially captured by mobility covariates (e.g. school and business closure)	Home, work, school, and community	Home, work, school, and community
Urban / rural variation	Yes◆	No	No	No	Partially captured by population density covariate	No	Yes
Data source	A combination of Prem et al. 2017 and Mossong et al. 2008	NA	Representative contact matrix as described in Walker et al., 2020 and stratified by World Bank income status	Prem et al. (or country thought to be closest by user)	Global data collated by IHME	Prem et al.	Prem et al.

* *EMOD's IFR is applied statistically as a post dynamic simulation multiplier of the number of cases*

◆ *Not included in the web-app presently.*

▲ *Models vary in how they include transmission and this row applies to all cases. Some models include contact rates and multiply these by parameters related to transmission probability (e.g. infectiousness or susceptibility) while others combine contact rates and transmission probability into one transmission rate parameter.*

Note : NA - not applicable

Modelling projected interventions

We previously described that models may differ in their aim, whether to project case numbers or the impact of interventions. Additionally, the models differ in which interventions they include and how they implement them. Different model structures are more conducive for explicitly modelling particular interventions.

Table 5 describes how interventions are implemented in the participating models. Those that explicitly model transmission, that is, by considering age-structured contact matrices and infection probabilities, typically model interventions that reduce transmission opportunities by altering **contact matrices**; in the case of ABMs, contact networks are altered. Importantly, in so doing, the mechanism by which multiple interventions are modelled (e.g. social distancing, stay at home advice, school closing, case isolation and shielding) is essentially the same, changing the frequency or occurrence of contact in the relevant groups or the whole population or the overall probability of infection upon contact for the relevant groups or the whole population. Interventions that affect the likelihood of transmitting or acquiring disease typically affect the transmission rate in compartmental models and individual level parameters describing transmission in ABMs. Interventions that affect disease progression change the rate at or duration after which subgroups (compartmental models) or individuals (ABMs) move from one health state to another.

ABMs allow for greater granularity of intervention specification than compartmental models. Compartmental models model interventions that target population subsets by adding subpopulation compartments or altering age-dependent contact matrices to represent changing contacts in subpopulations (e.g. in school). ABMs can typically target interventions at the level of individuals or subsets of individuals that have similar characteristics. The modelling of contact tracing is importantly different between the models and the modelling of contact tracing compartmental models requires strong assumptions. Evidence on how well these assumptions match observed data compared to more explicit modelling, using ABMs, does not exist yet for COVID-19. The incorporation of interventions within the IHME model is significantly different from all other approaches. It uses multiple covariates taken to represent the occurrence of the intervention in a location, such as testing per capita and use of mask covariates, to predict the infection rate later used in the mechanistic model.

In addition to the ability to incorporate interventions, there are considerations about how these interventions will differ by setting. Important parameters in modelling interventions are coverage, or the degree to which an intervention is implemented in the target population (or the efficacy of implementation), and adherence, or the degree to which individuals are able to comply with interventions. These are generally unknown in different settings, and assessment using models fit to data, or estimating of transmission rates (Rt) when different interventions are in place in different settings may be of help here. This information can then be used to project the impact forward in time. It may also be possible to gain some understanding of the extent to which interventions have been or can be put in place from qualitative studies or assessments in these settings. For example, in settings where many individuals are living together in dormitory or camp type accommodation, the ability to comply with social distancing measures has been seen to be low. Consideration of these populations within countries will also be important for projecting the impact of interventions using models.

Interactions between important parameters and modelling interventions

There are additional important interactions between the parameters of the model, the model structures and how reliably the different interventions can be modelled for different countries. Important interactions are particularly apparent for the targeted interventions. For example, in estimating the impact of shielding the vulnerable population, which population group is at most risk of severe outcomes and how this relates to who is modelled as shielded in the model is important. In many LMIC settings shielding strategies will not be implementable. Similarly, the age-structured contacts for this group and whether they are in the home or outside setting will be very important for modelling the impact of in-home shielding.

There are also important considerations about the level of physical space in the home and the amount of care people require that will impact the number of contacts that can be reduced by shielding. Such factors may vary greatly between settings. This highlights the need for information on these two parameters in different settings.

Table 5 - Projecting the Impact of Interventions: Inclusion and Implementation

Intervention	**LSHTM**	**Basel *COVID-19 Scenarios***	**Imperial *Squire***	**Oxford CoMo**	**IHME**	**IDM *Covasim***	**IDM *EMOD***
Social distancing	Alter contact matrices	Alter transmission rate	Alter transmission rate	Alter contact matrices	Alter social distancing related covariates for two-stage prediction of mobility and then infection rate	Alter network contacts and/or infection probability	Alter transmission matrix**
Stay at Home advice	Alter contact matrices	No	No	Alter contact matrices	Alter stay at home advice covariate for predicting infection rate	Alter network contacts	Alter transmission matrix**
School closing	Alter contact matrices	No	No	Alter contact matrices	Alter school closure covariate for predicting infection rate	Alter network contacts	Alter transmission matrix**
Case isolation	Alter contact matrices	No	No	Alter contact matrices	No	Remove contacts or reduce infection probability	Alter transmission matrix**
Shielding / cocooning	Alter contact matrices	Alter transmission rate	No	Alter contact matricess	No	No	Alter transmission matrix**
Testing	No	No	No	Symptom-based	Alter testing per capita covariate for predicting infection rate	Explicitly model individual testing	Alter transmission matrix**

Intervention	**LSHTM**	**Basel *COVID-19 Scenarios***	**Imperial *Squire***	**Oxford CoMo**	**IHME**	**IDM *Covasim***	**IDM *EMOD***
Contact tracing / quarantine	No	No	No	Population-level calculation based on overdispersion of infected	No	Explicitly model tracing of contacts in different settings	Explicitly model tracing of contacts in different settings
International travel restrictions	Use meta-population structure	Reduce imported cases	No	Reduce imported cases	No	No	Reduce imported cases and alter migration parameters
Healthcare worker prophylaxis	No	No	No	No	No	No	Reduce transmission and susceptibility
Masks	No	Alter transmission rate (applies to all settings)	No	Alter infection probability (applied to all settings)	Alter mask use covariate for predicting infection rate	Alter transmission probability	Reduce transmission and susceptibility (applied to all settings)
Novel treatment	Alter IFR	No	No	Alter severity and risk of death parameters	No	Reduce infection probability or natural history parameters	Alter individual’s state
Vaccination	Yes ; separate compartments	No	No	Alter infection probability	No	Alter infection probability or natural history parameters	Alter transmission matrix or natural history parameters

[*] *We explicitly note that masks are applied across settings since there is variation in how this intervention is applied across models. However, in the Basel COVID-19 Scenarios model all interventions are applied across all settings since contact rates are not specific to settings.*

[◆] *EMOD alters a transmission matrix (which incorporates multipliers for susceptibility and infectiousness) rather than the contact matrix.*

Model transparency, calibration, and reporting

Table 6 summarises how each of the included models addresses issues of transparency, calibration and reporting. Models without clear comprehensive documentation detailing model's specifics (structure and assumptions) and how to use them are less accessible to users. In part due to the need for rapid response to produce model outputs, documentation availability varies among the participating COVID-19 models. Moreover, in cases where documentation is available, most are currently incomplete or outdated, and efforts are being made to expand and update the available information. Several of the models provide open-source code, and thus, provide accessibility for other modellers with sufficient expertise (e.g. in a specific programming language) and time to understand the models in more detail. Models that provide a user-facing app (Table 2) provide information on how to use the models. Time constraints and the urgency of the COVID-19 pandemic partially justifies the lack of peer-reviewed publications for most models though ideally these will eventually be available.

As suggested in the report "Guidance on use of modelling for policy responses to COVID-19" from the CMCC's Policy Group, undergoing a peer-review process that includes the evaluation of the model's structure, use and analysis by the wider scientific community based on the source code and comprehensive written documentation would constitute the ideal reporting standard. Modellers are advised to commit to meeting this standard within a 6-9 months' timeframe.

In terms of validation, Weinstein et al. (2003) recommended that, for all models, an overall assessment of the models' internal validation is expected; that is, models should be thoroughly tested internally and debugged, and calculations should be checked to ensure accuracy. In Phase 2 of the CMCC's work, we will aim to assess models' "convergent validity" or "between model validation" – the extent to which models' predictions converge to each other and also to observed data. As mentioned previously, models should not be dismissed if they fail to converge or predict but such exercise should call for explanations and/or recalibration or respecification, as needed.

Model calibration is attempted by all groups, with the frequency varying across the models. The models use different data as targets for model calibration. Reported deaths is used by multiple groups and may constitute an appropriate data source due to its lower likelihood of under-reporting, however there is marked variability in coding and even requirement for cause of death certification across LMIC settings; in some cases excess burials (if recorded), may be more informative.

The IHME model adopts a particular, two-stage approach to calibration. First, infections are back calculated in each country (area/setting) from the data on cases and deaths, and an underlying daily transmission rate calculated. Then, a global regression analysis estimates the effect of a range of possible "drivers" on daily transmission rates. The global regression coefficients are then used to predict future transmission rates in each country (area/setting). Thus, in the IHME model, the transmission parameters used to make future projections in any country (area/setting) may differ significantly from those implicit in calculating the past incidence. Countries using this approach should verify the extent to which the regression fits their observed data closely.

Regarding uncertainty, it is important to understand what type of uncertainty models report in order to interpret their results. All seven models capture parameter uncertainty, though to a varying degree, and a number of models (stochastic ones) also capture stochastic uncertainty. Parameter uncertainty describes our epistemological uncertainty. Models vary parameter values with a range, or drawing from distributions, that represent our knowledge of true parameter values. The reported uncertainty reflects that uncertainty. Models that report stochastic uncertainty also capture the inherent randomness in real world events, and thus, they can describe the distribution of possible outcomes given that randomness.

All models are in constant development. Moreover, it must be acknowledged that they have all been developed relatively quickly given the evolution of the pandemic. The modelling groups highlighted incorporating additional interventions, and calibrating transmission parameters to case data as future additions.

Table 6 - Parameters of interest: assumptions and data used

	LSHTM	**Basel *COVID-19 Scenarios***	**Imperial *Squire***	**Oxford CoMo**	**IHME**	**IDM *Covasim***	**IDM *EMOD***
TRANSPARENCY							
Documentation availability	Peer-reviewed manuscripts, pre-published reports, and GitHub documentation.	medRxiv preprint manuscript posted May 12, 2020	Website, peer-reviewed manuscripts, pre-published reports, and GitHub documentation.	Not available on the model's website or GitHub repository	Complete documentation; on IHME's website including pre-published manuscripts and GitHub documentation.	Pre-published manuscript, web-app, and GitHub documentation	Non-COVID-19-specific documentation on web-app and GitHub; Incomplete COVID-19 specific documentation in pre-published manuscript
Open-source code	Yes	Yes	Yes	Yes	Yes	Yes	Non-COVID-19-specific code available on GitHub
CALIBRATION							
Frequency	Yes, for the UK; calibration to LMICs in process	No, but update underlying data every 2-3 days*	Daily	Project based (by app user)	Weekly	Project based (coder needed)	Project based (coder needed)
Fitted to	Multiple sources including contact rates, reported cases, ICU capacity, deaths, serology and virology	No	Deaths	Cases and deaths	Disease prevalence, cases, hospitalisations, and deaths and a range of covariates	Any combination of time-series observations (with model outputs)	Any combination of time-series observations (with model outputs)
Data	UK data	Official data ECDC and others; not validated	ECDC	UK data by the modelling group and national datasets by consortium members	IHME collated data across countries to draw general conclusions on transmission drivers that are assumed globally	Project based; open-access data and local data	Project based; open-access-data and local data
UNCERTAINTY							
Describes parameter/ stochastic/model uncertainty	Parameter, stochastic, and scenario based	Parameter	Parameter in paper and dashboard; none in online tool	Parameter	Parameter	Parameter and stochastic	Parameter (few variables) and stochastic
SENSITIVITY							
Type and parameters	R0, severity, susceptibility	No	Sensitivity analyses performed using Squire package◆	In progress	No	One-way sensitivity analysis for a number and specific model applications	Yes, but not done extensively

** The Basel COVID-19 Scenarios model outputs are not continuously calibrated to fit data, though the underlying data are updated every 2-3 days.*

◆ The Squire package is an R package.

CONCLUSION AND RECOMMENDATIONS

Understanding and responding to infectious disease dynamics requires interdisciplinary cooperation, often involving policymakers, epidemiologists, computer scientists, mathematicians, and ecologists, among others. The depth and breadth of data required to model specific situations, and the strength of interaction between stakeholders during model development, are key to the effectiveness and policy relevance of any modelling effort. A policy-facing model comparison exercise can narrow the gap that naturally emerges in crisis management among various stakeholders and stimulate the formation of efficient feedback loops potentially leading to, ultimately, better policy responses.

The COVID-19 multi-model comparison study is, to our knowledge, a unique exercise to date. It has brought together both infectious disease modelers, policymakers in LMICs and development partners to better connect policy design, analysis and implementation in the global fight against the COVID-19 pandemic.

The study has shown that the current versions of the seven models included in the comparison share commonalities and differences that impact on their relative fitness-for-purpose. The majority of the models are mechanistic, incorporate age-distribution in transmission and mortality parameters, and base the IFRs on strong assumptions from data collected early in the pandemic. At the same time, most models do not account for particular sub-populations or comorbidities, do not include indirect COVID-19 effects on other diseases and do not model economic outcomes. There were also differences across the models, such as in their purposes and aims, in how COVID-19 transmission and contact patterns have been implemented, in the breadth of interventions considered, and in how these interventions were constructed.

Phase 1 of the multi-model comparison study produced the following deliverables: the model description questionnaire (available as an online appendix), which can be used by researchers and analysts as a basis for other similar exercises, related to COVID-19 or not; and the fitness-for-purpose flow chart to support judgements on model selection given a specific policy issue. The proposed flow chart may be useful for decision-makers and analysts relying on modelled evidence in general (not only for COVID-19) both at global and country level. It is hoped that its application can foster and guide an ongoing dialogue between the two types of audiences – policy and technical – through a series of fundamental questions that expose the trade-offs that have to be made when selecting a model for a given policy question, context and decision constraints at a point in time. Rarely will there ever be a perfect model waiting to be used, yet decisions will still need to be made. The flow chart offers the foundation for weighing the fundamental policy and technical considerations; how these considerations are to be weighed in practice will depend on the specific context of the policy decision.

A number of lessons relevant to multi-model comparison exercises also emerged from this process. Keeping up with the rapidly changing policy and behavioural context (e.g. the growing importance of face masks) and correspondingly evolving models has required maintaining an intense level of engagement with the participating modelling groups throughout data collection while trying to minimise demands on their time. Working alongside the Policy Group of the CMCC from the outset has provided the Technical Group with invaluable insights into what matters most for policymakers in LMICs at this time; these insights are reflected in the design of key deliverables such as the model description questionnaire and the fit-for-purpose flow chart. Finally, discussing intermediary outputs of the comparison with participating COVID-19 modellers has offered insights and opportunities for course correction and has undoubtedly strengthened the exercise. The modelling groups willingness to collaborate in full openness in disclosing both model characteristics and development processes was fundamental to this model comparison study.

We also propose a number of recommendations for optimising modelling approaches with a view to improving the engagement between policymakers and modellers, ultimately leading to better policy decisions (Table 7). We hope that our recommendations will also be useful for funders of modelling groups when defining the scope and terms of references in using models for policymaking in LMICs e.g. encouraging modelling groups to engage with decision-makers early on in the process and/or stimulate modelling groups in participating in future multi-model comparison exercises to improve models.

In Phase 2 (currently being planned), CMCC's work will focus on performing direct quantitative comparisons among participating models for a selection of policy questions, using standardised parameter sets within the elicitation framework proposed by Shea et al. (2020).

Table 7 - Summary of recommendations

Fitness-for-purpose flow chart domain	For policymakers	For modellers
Modelling feasibility	Question whether modelling tools can address the policy question and if so, whether to use/adapt existing models or to develop completely new ones (if necessary and feasible). Engage local analysts and researchers early on in assessing the individual, organizational and institutional capacity for engaging with models and in the fitness-for-purpose assessment. Make clear to local analysts and modellers the types of decision constraints being faced e.g. time, stakeholder coordination, infrastructure, budget and how the results will be used and disseminated. Conduct the fitness-for-purpose assessment iteratively and not as a one-off exercise because i) models evolve quickly; and ii) depending on context and the specific policy question, the fitness-for-purpose of a given model may also change.	Identify concrete approaches to involve policymakers and analysts in LMICs in developing or adapting the model and user interfaces to maximise relevance for such settings; consider following the collaborative modelling approach in the Policy Group report (Figure 8). For models that are already developed: Clarify in all available model documentation the types of policy questions that the current model version can be used for, and its main limitations; consider using the policy question typology in the Policy Group report (Table 3).
Model implementation	Consider and plan early on for rapid data collection which may minimise uncertainty in model results. Set up a consultative process involving local analysts and researchers for defining, reviewing and validating model assumptions where data are not available. Prefer models that have been calibrated and validated for your setting using appropriate methods given the policy question.	Identify clearly a minimum set of model parameters that should be ideally informed by local data in order for the model to be applied credibly in a given context; refine it continuously based on context-specific uncertainty and sensitivity analyses; support local analysts and policymakers in identifying appropriate data sources and in collecting additional data. Wherever possible engage with local partners and experts to validate key assumptions and the quality of the sources of the setting specific parameters. Ideally, train local analysts on how to use the model themselves._
Model reporting	Seek commitment from modellers to adhere to the recommended reporting trajectories	Commit to the recommended reporting trajectories proposed in the Policy Group report (Figure 6)

GLOSSARY

- **Agent-based models:** Agent-based models simulate each individual in the population separately (in terms of characteristics, behaviours and outcomes), as opposed to cohort models which model a homogeneous population and simulate average characteristics, behaviours and outcomes.
- **Calibration:** Calibration is a method for estimating uncertain parameters, by altering a parameter or set of parameters so that the model results can best match empirical data.
- **Contact matrix:** A matrix that quantitatively describes the heterogeneous social mixing patterns between population subgroups, typically implemented for age-groups as age captures a significant amount of the variation in contacts. Each matrix row describes an age-group and each column similarly describes an age-group. Each cell describes the contact rate between individuals in the age-groups described by the row and column. Different contact matrices are often implemented for different settings such as home, schools, work, and the community.
- **Dynamic model:** Dynamic models are frequently used for modelling communicable diseases as they reflect changing risks of infection to the uninfected population based on the size of the infected population. They differ from static models that assume a constant risk of infection.
- **Infection fatality ratio (IFR):** The total number of deaths from COVID-19 divided by the total number of COVID-19 infections (including both confirmed and undetected infections). The IFR is distinct from a case-fatality ratio which is the total number of deaths from COVID-19 divided by the total number of confirmed COVID-19 infections.
- **Mechanistic models:** Models that make explicit hypotheses about the biological mechanisms that drive infection dynamics. Such hypotheses range from simple representations of the time it takes to complete some part of the disease process to attempts to explicitly represent social interactions of people in an entire country or even the world.
- **Sensitivity analysis:** Sensitivity analysis is a technique that is used to understand the robustness or validity of results produced by a model. Sensitivity analysis is performed by altering key parameters in the model, individually or simultaneously, as well as by varying the underlying assumptions or structure of a model.
- **Uncertainty:** Uncertainty in modelling refers to the fact that projections made by models are not certain and are subject to the conditions, assumptions and methods used in the model being accurate and appropriate representations of the real world. There are several types of uncertainty in modelling, including (Bilcke et al., 2011):
    1. Methodological uncertainty – which normative modelling approach should be used? For example, which perspective or discounting approach to adopt in the analysis.
    2. Structural uncertainty — What structural aspects should be incorporated to capture the relevant characteristics of the disease and intervention being investigated? For example, which disease states to include or which function to use for extrapolating data into the future.
    3. Parameter uncertainty — What is the true value of each model parameter?
- **Validation:** There are many different types of validation (Weinstein et al., 2013). Internal Validation refers to ensuring that the mathematical calculations are accurate and consistent with the specifications of the model (internal validity), that model inputs and outputs are consistent with available data (calibration), and that their results make sense and can be explained at an intuitive level (face validity). Cross-validation (between-model validation) refers to explaining sources of differences when models of the same decision come to different conclusions. External and predictive validation refers to the ability of the model to make accurate predictions of future events.

REFERENCES

Bilcke, J., Beutels, P., Brisson, M., & Jit, M. (2011). Accounting for Methodological, Structural, and Parameter Uncertainty in Decision-Analytic Models: A Practical Guide. Medical Decision Making, 31(4): 675-692. https://doi.org/10.1177/0272989X11409240

den Boon, S., Jit, M., Brisson, M., Medley, G., Beuttels, P. et al. (2019) Guidelines for multi-model comparisons of the impact of infectious disease interventions. BMC Medicine 17, 163. https://doi.org/10.1186/s12916-019-1403-9

Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S., Scalia Tomba, G., Wallinga, J., Heijne, J., Sadkowska-Todys, M., Rosinska, M. & Edmunds, W.J. (2008). Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. PLOS Medicine, 5(3): e74. https://doi.org/10.1371/journal.pmed.0050074

Prem, K., Cook, A.R. & Jit, M. (2017). Projecting social contact matrices in 152 countries using contact surveys and demographic data. PLOS Computational Biology, 13(9): e1005697. https://doi.org/10.1371/journal.pcbi.1005697

Prem, K., van Zandvoort K, Klepac, P., Eggo, R.M., Davies, N.G., Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Cook, A.R. & Jit, M. (2020). Projecting contact matrices in 177 geographical regions: an update and comparison with empirical data for the COVID-19 era. medRxiv, 2020.07.22.20159772; doi: https://doi.org/10.1101/2020.07.22.20159772

Shea, K., Runge, M.C., Pannell, D., Probert, W.J.M., Li, S.-L., Tildesley, M. & Ferrari, M. (2020). Harnessing multiple models for outbreak management. Science, 08 May 2020: Vol. 368, Issue 6491, pp. 577-579. DOI: 10.1126/science.abb9934

Walker, P.G.T., Whittaker, C., Watson, O.J., Baguelin, M., Winskill, P. et al. (2020). The impact of COVID-19 and strategies for mitigation and suppression in low- and middle-income countries. Science, 369 (6502): 413-422. DOI: 10.1126/science.abc0035

Weinstein, M.C., O'Brien, B., Hornberger, J., Jackson, J., Johannesson, M., McCabe, C. and Luce, B.R. (2003). Principles of Good Practice for Decision Analytic Modeling in Health Care Evaluation: Report of the ISPOR Task Force on Good Research Practices—Modeling Studies. Value in Health, 6: 9-17. doi:10.1046/j.1524-4733.2003.00234.x

APPENDIX 1

Number of LMICs included in participating models' web-user interfaces

Model web user interface	**Number of LMICs**
Basel (*Covid-19 scenarios*)	123
CoMo	131
IHME	101
Imperial College London (*COVID-19 Scenario Analysis Tool*)	131
LSHTM	132
Total LMICs per World Bank (*July 2019 release*)	138

Notes : The right-hand column denotes the number of LMICs that can be selected nominally from the web-user interfaces as of 03 August 2020. The web user interfaces of Basel and CoMo also allow the user to define bespoke characteristics of a target setting (e.g. population size).

© World Health Organization 2020.

Some rights reserved. This work is available under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 IGO licence (CC BY-NC-SA 3.0 IGO; *https://creativecommons.org/licenses/by-nc-sa/3.0/igo*).

Under the terms of this licence, you may copy, redistribute and adapt the work for non-commercial purposes, provided the work is appropriately cited, as indicated below. In any use of this work, there should be no suggestion that WHO endorses any specific organization, products or services. The use of the WHO logo is not permitted. If you adapt the work, then you must license your work under the same or equivalent Creative Commons licence. If you create a translation of this work, you should add the following disclaimer along with the suggested citation: “This translation was not created by the World Health Organization (WHO). WHO is not responsible for the content or accuracy of this translation. The original English edition shall be the binding and authentic edition”.

Any mediation relating to disputes arising under the licence shall be conducted in accordance with the mediation rules of the World Intellectual Property Organization. (*http://www.wipo.int/amc/en/mediation/rules/*)

Suggested citation. COVID-19 Multi-Model Comparison Collaboration (CMCC) technical group report : Model fitness-for-purpose assessment report. Geneva : World Health Organization ; 2020 (WHO/2019-nCoV/Multi-Model_Comparison_Collaboration/2020.1). Licence : *CC BY-NC-SA 3.0 IGO*.

Cataloguing-in-Publication (CIP) data. CIP data are available at *http://apps.who.int/iris*.

Sales, rights and licensing. To purchase WHO publications, see http://apps.who.int/bookorders. To submit requests for commercial use and queries on rights and licensing, see *http://www.who.int/about/licensing*.

Third-party materials. If you wish to reuse material from this work that is attributed to a third party, such as tables, figures or images, it is your responsibility to determine whether permission is needed for that reuse and to obtain permission from the copyright holder. The risk of claims resulting from infringement of any third-party-owned component in the work rests solely with the user.

General disclaimers. The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of WHO concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate border lines for which there may not yet be full agreement.

The mention of specific companies or of certain manufacturers’ products does not imply that they are endorsed or recommended by WHO in preference to others of a similar nature that are not mentioned. Errors and omissions excepted, the names of proprietary products are distinguished by initial capital letters.

All reasonable precautions have been taken by WHO to verify the information contained in this publication. However, the published material is being distributed without warranty of any kind, either expressed or implied. The responsibility for the interpretation and use of the material lies with the reader. In no event shall WHO be liable for damages arising from its use.

This publication contains the collective views of the Technical Group of the COVID-19 Multi-Model Comparison Collaboration (the CMCC) and does not necessarily represent the decisions or the policies of WHO.

Ref. WHO/2019-nCoV/Multi-Model_Comparison_Collaboration/2020.1