Check for updates

RESEARCH ARTICLE

Assessing fitness-for-purpose and comparing the suitability of COVID-19 multi-country models for local contexts and users

[version 1; peer review: awaiting peer review]

Hannah Clapham[1], Mohamed Gad[2], Adrian Gheorghe[2], Raymond Hutubessy [3], Itamar Megiddo [4], Christopher Painter [5], Francis Ruiz [2], Nejma Cheikh[6], Marelize Gorgens[6], Thomas Wilkinson[6], Marc Brisson[7], Nigel Gay[8], Jane Labadin [9], Jodie McVernon[10], Paula M. Luz[11], Wilfred Ndifon[12], Brooke E. Nichols [13-15], Shankar Prinja[16], Joshua Salomon[17], Akhona Tshangela[18]

[1]National University of Singapore, Singapore, Singapore
[2]Imperial College London, London, UK
[3]World Health Organization, Geneva, Switzerland
[4]University of Strathclyde, Glasgow, UK
[5]Health Intervention and Technology Assessment Program, Nonthaburi, Thailand
[6]World Bank Group, Washington D.C., USA
[7]Laval University, Quebec City, Canada
[8]Independent Researcher, London, UK
[9]Universiti Malaysia Sarawak, Sarawak, Malaysia
[10]The Peter Docherty Institute for Infection and Immunity, Melbourne, Australia
[11]Oswaldo Cruz Foundation, Rio de Janeiro, Brazil
[12]African Institute for Mathematical Sciences, Kigali, Rwanda
[13]Department of Global Health, Boston University School of Public Health, Boston, USA
[14]Health Economics and Epidemiology Research Office, Department of Internal Medicine, School of Clinical Medicine, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa
[15]Department of Medical Microbiology, Amsterdam University Medical Centre, Amsterdam, The Netherlands
[16]Postgraduate Institute of Medical Education and Research, Chandigarh, India
[17]Stanford University, Stanford, USA
[18]Africa CDC, Addis Ababa, Ethiopia

Open Peer Review

Reviewer Status *AWAITING PEER REVIEW*

Any reports and responses or comments on the article can be found at the end of the article.

Abstract

Background: Mathematical models have been used throughout the COVID-19 pandemic to inform policymaking decisions. 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 multi-country COVID-19 models to optimise their usefulness in the COVID-19 response.
Methods: Seven models were identified that satisfied the inclusion criteria for the model comparison and had creators that were willing

to participate in this analysis. A questionnaire, extraction tables and interview structure were developed to be used for each model, these tools had the aim of capturing the model characteristics deemed of greatest importance based on discussions with the Policy Group. The questionnaires were first completed by the CMCC Technical group using publicly available information, before further clarification and verification was obtained during interviews with the model developers. The fitness-for-purpose flow chart for assessing the appropriateness for use of different COVID-19 models was developed jointly by the CMCC Technical Group and Policy Group.

Results: A flow chart of key questions to assess the fitness-for-purpose of commonly used COVID-19 epidemiological models was developed, with focus placed on their use in LMICs. Furthermore, each model was summarised with a description of the main characteristics, as well as the level of engagement and expertise required to use or adapt these models to LMIC settings.

Conclusions: This work formalises a process for engagement with models, which is often done on an ad-hoc basis, with recommendations for both policymakers and model developers and should improve modelling use in policy decision making.

Corresponding author: Raymond Hutubessy (hutubessyr@who.int)

Author roles: Clapham H: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Gad M**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Gheorghe A**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Hutubessy R**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Megiddo I**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Painter C**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Ruiz F**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Cheikh N**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Gorgens M**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Wilkinson T**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Brisson M**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Gay N**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Labadin J**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **McVernon J**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Luz PM**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Ndifon W**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Nichols BE**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Prinja S**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Salomon J**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; **Tshangela A**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Validation, Writing – Original Draft Preparation, Writing – Review & Editing

How to cite this article: Clapham H, Gad M, Gheorghe A *et al*. **Assessing fitness-for-purpose and comparing the suitability of COVID-19 multi-country models for local contexts and users [version 1; peer review: awaiting peer review]** Gates Open Research 2021, **5**:79 https://doi.org/10.12688/gatesopenres.13224.1

First published: 12 May 2021, **5**:79 https://doi.org/10.12688/gatesopenres.13224.1

Introduction

Throughout the COVID-19 pandemic, mathematical models that mimic the natural history and dynamics of COVID-19 have been used in countries across the globe to inform policy-making decisions, such as the allocation of scarce healthcare resources and the implementation of policy measures to curb the spread of COVID-19. Results from COVID-19 models have received significant attention from the media, public and policymakers. However, the expectations of model users and consumers do not always align with the purpose of the models and the modellers' intended use of the models. Producing accurate and robust forecasts of health outcomes due to COVID-19 (which is often the expectation) is extremely difficult[1], particularly beyond a few weeks into the future. There is a common perception that the COVID-19 models produce forecasts of what will happen, when many models aim to be scenario-based tools to inform policymaking – and the precise scenarios explored through modelling rarely materialise as modelled.

The modelling approaches, considerations, data used and purpose of COVID-19 models can vary substantially from model to model[2]. A comprehensive understanding of these characteristics often requires some knowledge of mathematical modelling and concepts related to natural history of disease, infectious disease epidemiology and health economics. These prerequisites can make it difficult for policymakers to assess the fitness-for-purpose of models and to understand which of the many models available are the most appropriate for given policy questions and in the local context, particularly given the time constraints they often work within. Furthermore, as the purpose of COVID-19 models varies, model users may also need to use different models to address different policy questions, and choosing which model to use is not as simple as just attempting to understand which model is the 'best' model. This lack of transparency and accessibility of mathematical modelling carries risks: policymakers may end up misunderstanding, misusing or even ignoring the results of models, which in the worst case could result in harmful policy responses. To add to this challenge, the models are frequently updated with new data and adjustments to their methodologies, so model users need to pay keen attention to these changes too.

Both the science and policy realities of COVID-19 evolve quickly, which constantly raises the issue of potential misalignment between modellers' intentions and policymakers' needs. The novelty of COVID-19 has meant that the scientific understanding of the epidemiology and natural history of the virus has been developing at the same time as models have been built, and modellers face the difficult task of reflecting the differences between imperfect and competing data sources, even when relevant (and local) data are available. Conversely, the local level strategic and programmatic questions that policymakers are considering in response to the pandemic may be different to those that modellers had considered, or there may be differences in their conceptions of how these measures may be implemented that need to be reflected within models.

In response to these issues, 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 multi-country COVID-19 models in use with the aim of facilitating dialogue between model developers and model users. The CMCC is led by members of the World Health Organization (WHO), the World Bank Group, the Bill and Melinda Gates Foundation (BMGF), and the International Decision Support Initiative (iDSI). Further details of the CMCC can be found on the website https://decidehealth.world/CMCC.

Objectives

The CMCC's first phase of work concluded in August 2020. The primary objectives of the first phase of the project were to develop:

1) A flow chart of key questions to assess the fitness-for-purpose of a set of widely used COVID-19 epidemiological models, with additional focus placed on their appropriateness for use in LMICs.

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 these models to LMIC settings.

Furthermore, additional aims of this research included:

1) Producing a description of the applicability of the models to LMIC contexts, identifying particular challenges and opportunities (and key context-specific parameters) for representing the COVID-19 pandemic in resource-constrained health systems and societies;

2) Recommendations for how modelling approaches can be optimised to improve policy engagement and decision-making.

Methods

The methodological process for this project was guided by the CMCC Technical Group members and secretariat following an adaptation of the Guidelines for multi-model comparisons of the impact of infectious diseases[3]. Given that many COVID-19 models have been developed since the start of the pandemic, the model comparison prioritised including only the COVID-19 models most widely used for decision-making in a LMIC context in this process. The models had to satisfy the following inclusion criteria:

o Be applicable, or be capable of adaptation, to multiple countries and/or regions

o Be dynamic and seek to inform the impact of different COVID-19 policy interventions

o Have published results that are available in the public domain

o Have developers that are willing to participate in the comparison study

Seven models that satisfied the inclusion criteria were identified through targeted literature searches and the stakeholders' networks (WHO, World Bank, BMGF, Centre for Global

Development, Health Intervention and Technology Assessment Program [HITAP]) in February 2020. The developers of the selected models were contacted by email and confirmed they were willing to participate in this analysis. The developers of these models are listed below (model name, where applicable):

1. Basel University (Covid-19 Scenarios)

2. COVID-19 International Modelling Consortium (CoMo)

3. Imperial College London (Squire)

4. Institute for Health Metrics and Evaluation

5. London School of Hygiene & Tropical Medicine

6. Institute for Disease Modeling (Covasim and EMOD)

Following the selection process, the Technical Group developed a questionnaire, extraction tables and interview structure, which were used to elicit information about each of the selected models from model developers and model documentation. Face validity and preliminary pilot testing of these documents was performed among members in the management team representing a range of different organisations. No reliability testing was performed but any feedback received during the pilot testing was incorporated and addressed.

The Technical Group also consulted with the Policy Group in developing these instruments. The tables for the model comparison exercise aimed to capture the model characteristics deemed of greatest importance, either due to: 1) their impact on the model results, 2) divergence between the included models, and 3) the priorities of the Policy Group[2]. The questionnaires were preliminarily completed by the CMCC Technical Group using publicly available information prior to interviews with model developers conducted using online teleconferencing software. They were shared with model developers, discussed and completed during interviews, when additional questions were posed regarding the data fitting, model calibration and validation processes. All interviews were attended by MB, AG, RH and IM, with RH and MB leading the interviews. Other technical group members were present in each of the calls but did not attend all interviews. The duration of the interviews ranged from 30 minutes to 1 hour. Written meeting minutes were taken and audio recording of the interviews was conducted. Following finalisation of the model comparison tables, all of the information described in the model comparison tables was verified by the model developers to ensure their accuracy. The fitness-for-purpose flow chart for assessing the appropriateness for use of different COVID-19 models was developed jointly by the CMCC Technical Group and Policy Group.

Results

The main outputs of this study, summarised in the Model Fitness-for-Purpose Assessment report[2], comprised: a fitness-for-purpose flow chart; a comparison of the seven COVID-19 models; and recommendations for modellers and policymakers. A key finding of the exercise was that models are being constantly updated and refined, given the rapidly evolving understanding of COVID-19 transmission and range of possible interventions. As such, any model comparison should be timestamped and, more broadly, regarded as an iterative process rather than as a one-off exercise.

Designed under the premise that a model's fitness-for-purpose is best judged as a function of the setting, the required data and the policy question being asked, the fitness-for-purpose flow chart (Figure 1) includes a flow of questions for policy makers, the analysts supporting them and COVID-19 modellers to consider as part of an ongoing dialogue. The questions expose the trade-offs to be made when selecting a model for a given policy question, context and decision constraints at a point in time. These questions include a consideration of the aim of the model, how it has been adapted to each setting, how it could be used by analysts and policymakers in LMICs and the level of interaction with modellers they would need (and is available). The flow chart offers the foundation for weighing the fundamental policy and technical considerations at play when deciding on a suitable model to use in a given context.

The comparison of the included COVID-19 models (versions as of 31st July 2020) identified both similarities and differences among them. Most models adopted an age-structured SEIR approach and used country-specific demographics as well as country-specific estimated age-stratified contact matrices (e.g. by using Prem et al.[4]). All models used age-stratified rates for severe disease and deaths, but there were key differences between the models in what data or assumptions were used for adjusting this parameter to the LMIC settings. At the same time, most models did not account for particular sub-populations or comorbidities.

The models differed to a greater extent in terms of the policy interventions considered and how these were modelled – some were modelled as altering the contacts for the whole population or in different groups or in different settings (such as home, school or work), and others as altering the risk of transmission given a contact. The models also varied in the way they could be used by those developing policies in LMICs. Some models had interfaces available so people can run their own simulations, whereas others would require more coding skills to run additional simulations or scenarios. A key discussion point was around the calibration of models in different settings. This was currently being handled differently by the groups but was an area of on-going development. The modelling groups highlighted calibration as a challenge given differing reporting across countries, and differences in, and lack of data on, how interventions were implemented in different places. However, this was highlighted as being important for the on-going acceptance and applicability of models, and was therefore identified as a key area requiring country engagement between COVID-19 modellers and local analysts.

The majority of models reviewed did not include the economic impact of COVID-19 or the knock-on impact on other healthcare issues though some groups did highlight separate work considering these impacts, and one group had some costs for
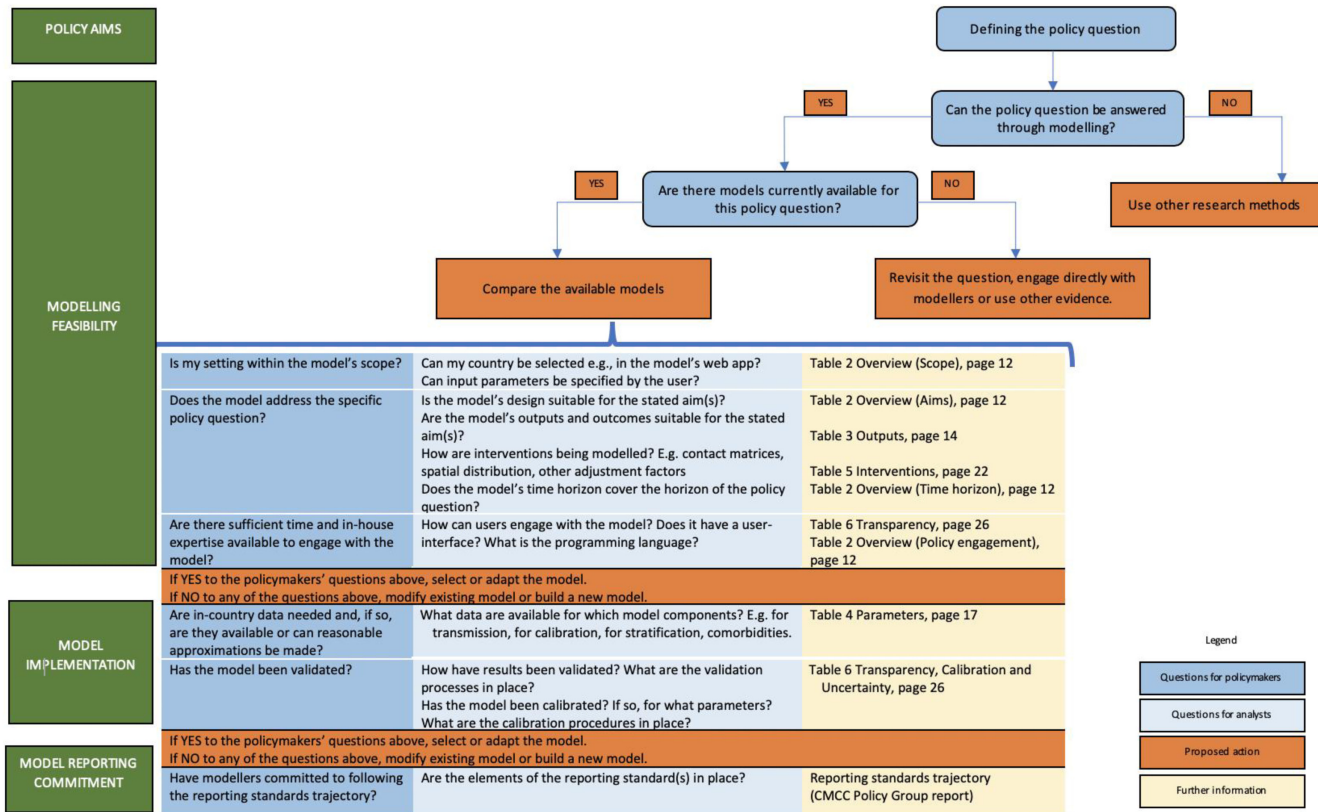
Figure 1. Fitness-for-purpose flow chart. Tables and page numbers cited in this table can be found in the Technical Group Report: Model Fitness-For-Purpose Assessment Report, available online at: https://decidehealth.world/CMCC.

a limited number of settings. This was highlighted clearly by the Policy Group as being a particularly important consideration for policy makers and therefore of great use to be analysed along with the COVID-19 cases and deaths projected.

The proposed recommendations aim to improve the relevance and uptake of COVID-19 models in decision-making, particularly in LMICs (Table 1). Recommendations for policymakers focus on the importance of setting up processes for collecting local data and reviewing evidence as it becomes available. Recommendations for COVID-19 modellers focus on ensuring transparency and clarity, both about the models as a whole and about the influential parameters that should ideally be informed by local data; and on working closely with local partners to validate assumptions and continuously adapt and improve the models.

Discussion

To our knowledge, this research is the first exercise of its kind, bringing together modellers and policymakers to explain, summarise, and evaluate the assumptions of the most well-known COVID-19 models applied to model intervention scenarios in multiple countries. The findings were developed in collaboration with the Policy Group, and the work should inform how to improve modelling use in policy decision making. The results

convey to policymakers, in easily understandable terminology, information on differences and characteristics of COVID-19 models and their fitness for purpose for addressing specific policy questions in their setting. This research also benefits model developers by emphasising the LMIC policymaker perspective and expectations, summarising the key characteristics of other models, as well as discussing the advantages and limitations of models which can help inform future development.

Utilising models for policymaking requires, or at the very least improves with, communication between policymakers, modellers and other technical experts. Policymakers should engage local analysts and modellers, where available, to assess and develop the institutional capacity to engage with models on an ongoing basis as new policy questions requiring evidence arise. Communication between these entities can convey decision constraints such as time and infrastructure and help to assess existing models' fitness-for-purpose (Figure 1), and otherwise, whether adapting existing models or developing new ones is feasible. Policymakers should also plan early for data collection in coordination with local analysts to reduce model uncertainty.

Global health modelling groups should involve policymakers, local modellers and analysts who have an understanding of the

Table 1. Summary of recommendations for policymakers and modellers.

Fit 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, organisational 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[5]. 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[5].
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[5].

local context as much as possible, and transparently present existing models in language that is clear to these groups. Modellers can identify where local context matters in their models to help these groups develop plans for data collection and aid validating model assumptions relevant to the local context. Further, communication between modelling groups and local modellers and analysts can develop local capacity to use and possibly extend models or develop novel ones to answer policy questions as they arise.

During the CMMC Technical Group work, we learnt valuable lessons that can be applied to future model evaluations both globally and within countries. These lessons include ensuring that the exercise is keeping up with the rapidly changing context (e.g. the growing importance of face masks) and evolving models. The collaboration with Policy Group provided the Technical Group insights into what matters most for policymakers in LMICs. The engagement of the modelling groups was key and successive rounds of validation of intermediary comparison results with participating COVID-19 modellers led to insights and course correction and strengthened the exercise. The developed questionnaires are available for use with other models for countries to undertake targeted model comparisons[2]. We hope this is a valuable resource to the modelling community globally.

This work formalises a process for engagement with models, which is often done on an ad-hoc basis, with recommendations for both policymakers and model developers. It provides guidance for future situations where modellers are working to produce locally-relevant models in several regions and countries.

Next possible steps for this work include taking a Structured Decision Making approach such as that developed by Shea *et al.* 2020[6], conducting comparative analyses of model predictions, and informing how to account for information from different COVID-19 models.

Ethics statement
Model developers were invited to take part in the model comparison; completion of the survey was voluntary. In the questionnaire, the project team sought clarification and verification about the details of the models. The questions were not harmful to our subjects. Personal details were not requested as part of this process. Finally, participants were informed that the results of the survey would be presented. For the above reasons, the team did not seek ethical approval for this study. Participants provided written informed consent for participation and publication of resulting data.

Data availability
Underlying data
Figshare: Interviews. https://doi.org/10.6084/m9.figshare.14229260.v1[7].

This project contains the following underlying data:
- LSHTM - Tue, 02 June 2020.docx (Interview transcript for LSHTM model)
- Basel - Thu, 28 May 2020.docx

- CoMo - Fri, 29 May 2020.docx
- ICL - Tue, 09 June 2020.docx
- IDM Covasim - Fri, 05 June 2020.docx
- IDM EMOD - Mon, 08 June 2020.docx
- IHME - Tue, 09 June 2020.docx
- IHME follow-up.docx

Figshare: Questionnaires. https://doi.org/10.6084/m9.figshare.14229242.v1[8].

This project contains the following underlying data:
- CoMo Model description questionnaire V0.9 .docx
- EMOD Model description questionnaire.docx
- CovaSim Model description questionnaire.docx
- Model description questionnaire V0.9_Basel_COVID19_Scenarios.docx
- LSHTM Model description questionnaire_clean.docx
- Model description questionnaire V0.9_squire_v2.docx
- Model description questionnaire V0.9 IHME_for validation.docx

Extended data
Figshare: Interviews. https://doi.org/10.6084/m9.figshare.14229260.v1[7].

This project contains the following extended data:
- Interview guide.docx

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

References

1. Subramanian V, Kattan MW: **Why Is Modeling Coronavirus Disease 2019 So Difficult?** *Chest.* 2020; **158**(5): 1829–1830.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

2. COVID-19 Multi-Model Comparison Collaboration: **Technical Group Report: Model Fitness-for-Purpose Assessment Report**. 2020.
Reference Source

3. den Boon S, Jit M, Brisson M, *et al.*: **Guidelines for multi-model comparisons of the impact of infectious disease interventions.** *BMC Med.* 2019; **17**(1): 163.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

4. Prem K, Cook AR, Jit M: **Projecting social contact matrices in 152 countries using contact surveys and demographic data.** *PLoS Comput Biol.* 2017; **13**(9): e1005697.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

5. COVID-19 Multi-Model Comparison Collaboration. Policy Group Report: **Guidance on Use of Modelling for Policy Responses to COVID-19**. 2020.
Reference Source

6. Shea K, Runge MC, Pannell D, *et al.*: **Harnessing multiple models for outbreak management.** *Science.* 2020; **368**(6491): 577 LP–579.
PubMed Abstract | **Publisher Full Text**

7. Painter C: **Interviews.** *figshare. Journal contribution.* 2021.
http://www.doi.org/10.6084/m9.figshare.14229260.v1

8. Painter C: **Questionnaires.** *figshare. Journal contribution.* 2021.
http://www.doi.org/10.6084/m9.figshare.14229242.v1