UNIVERSITY *of* York

This is a repository copy of *Enhancing COVID-19 decision making by creating an assurance case for epidemiological models*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/165293/

Version: Published Version

Article:

White Rose
university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Enhancing COVID-19 decision making by creating an assurance case for epidemiological models

Ibrahim Habli,[1] Rob Alexander,[1] Richard Hawkins,[1] Mark Sujan,[2] John McDermid,[1] Chiara Picardi,[1] Tom Lawton[3]

[1]Department of Computer Science, University of York, York, UK
[2]Department of Social Science and Systems in Health, Warwick Medical School, University of Warwick, Coventry, UK
[3]Department of Anaesthesia and Critical Care, Bradford Teaching Hospitals NHS Foundation Trust, Bradford, UK

Correspondence to
Dr Ibrahim Habli;
ibrahim.habli@york.ac.uk

INTRODUCTION

When the UK government was first confronted with the very real threat of a COVID-19 pandemic, policy-makers turned quickly, and initially almost exclusively, to scientific data provided by epidemiological models. These models have had a direct and significant influence on the policies and decisions, such as social distancing and closure of schools, which aim to reduce the risk posed by COVID-19 to public health.[1] The models suggested that depending on the strategies chosen, the number of deaths could vary by hundreds of thousands. From a safety engineering perspective, it is clear that the data generated by epidemiological models are safety critical, and that, therefore, the models themselves should be regarded as safety-critical systems.

With safety-critical systems, we typically associate large installations such as nuclear power plants and oil refineries, but also software that is used in cockpit flight management systems, as well as medical devices, such as ventilators. Common to these systems is the requirement that they have robust and rigorous assurance so that we can justifiably place trust in their operation. If we accept that epidemiological models fall into this category of safety-critical systems, that is, that failures in their design, operation and use can kill people, then we should expect that they hold up to this requirement of assured performance.

However, there are significant gaps in the understanding of COVID-19 and this introduces uncertainty into epidemiological models, such as whether individuals who have recovered have enduring protective immunity, and what the extent of asymptomatic infection is.[2] In addition, mitigation strategies such as contact tracing rely on people's behaviour (eg, whether or not people download and use a contact tracing app), which

can be hard to predict and model reliably.[3] Epidemiological models that are executed as simulations on computers can also require fairly complex software code, and this needs to be dependable—but from experience, we know that this cannot be taken for granted.[4] It is, therefore, vital that decision makers are made aware of the quality of the models and the assumptions underlying them and that they can reflect on the limitations of the models in relation to practical decisions about the management of the pandemic.

Again, looking across at safety-critical systems, it is best practice and frequently an explicit regulatory requirement to produce a safety or assurance case—a structured, explicit argument supported by evidence.[5][6] Assurance cases are a primary means by which confidence in the safety of the system is communicated to and scrutinised by the diverse stakeholders, including regulators and policy-makers.

In this commentary, we put forward the suggestion that developers of epidemiological models consider complementing their models with an assurance case that explains to users how, and to what extent, the resulting evidence can support and substantiate policy decisions. We argue that such an assurance case has the potential to enable a wider understanding, and a critical review, of the expected benefits, limitations and assumptions that underpin the development of the epidemiological models and the extent to which these issues, including the different sources of uncertainty, are considered in the policy decision-making process.

ASSURANCE CASES IN UK SAFETY-CRITICAL SYSTEMS

An assurance case may consider different critical dependability properties of a system,
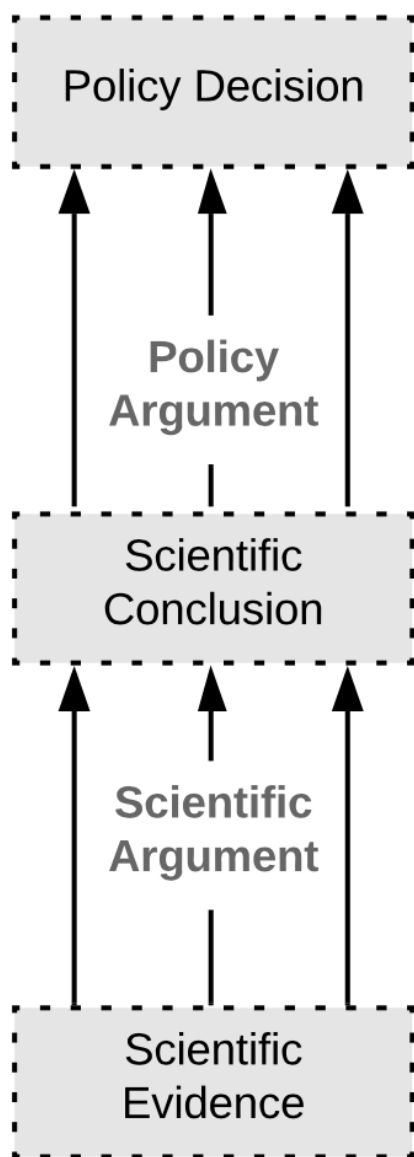
Figure 1 Overall assurance case structure for a COVID-19 policy.

such as safety, security, availability and maintainability. With respect to safety, an assurance case is primarily used to communicate and critically evaluate a safety argument about risk-based decisions to commission a system or a service. Put simply, an assurance case helps us determine whether we can be confident that a system is safe for use. We expect the developers and operators of systems to convince us through the argument that they have considered all relevant risks and that they have dealt with them satisfactorily. We want to be shown the evidence for that and we want to know about any gaps in the argument or the evidence. Data from modelling, simulation, testing and in-service usage provide the evidence base for claims about safety. However, this evidence is rarely conclusive. It entails different sources of uncertainty and hinges on technical, organisational and social assumptions. The argument should make these issues and the way in which

they relate to each other explicitly to enable the different stakeholders to critically review, modify, accept or reject the claims.

The use of assurance cases is a long-established practice in the engineering domain of safety-critical systems. Particularly in the UK, the development of an assurance case is a mandatory requirement in key sectors such as defence, nuclear and rail.[7] More pertinently, in the National Health Service (NHS), compliance with the NHS Digital clinical safety standards DCB0129 and DCB0160 requires an assurance case for Health IT systems.[8] Examples of assurance cases submitted to NHS Digital include adult and children's social care case management applications, as well as prominent and well-known systems such as Babylon Health's GP at Hand suite of applications (which includes the AI-based symptom checker).

Although, thus, medical devices (through EU regulations) and, to a certain extent, Health IT systems are regulated and have to follow in their development established risk-management and assurance standards, this does not yet seem to apply to epidemiological models. As epidemiological models are intended to inform decision making at the policy level, although safety-critical, they fall outside the scope of regulation and assurance standards, because they are not used in the provision of clinical services.

ASSURANCE CASES FOR EVIDENCE-BASED COVID-19 POLICY

As a highly salient example, we consider Report 9 by the Imperial College COVID-19 Response Team ('the impact of non-pharmaceutical interventions (NPIs) on the reduction of COVID-19 mortality and healthcare demand').[9] This is an example of epidemiological modelling based on microsimulation to provide primary evidence that can have significant policy implications. We can view the structure of a COVID-19 policy assurance case as an integration of the following, as illustrated in figure 1:

    A. Scientific evidence, such as data from epidemiological modelling (in this case: microsimulation).

    B. *Scientific conclusion*, often referred to as 'scientific advice', concerning the effect of the different public health strategies based on the scientific evidence.

    C. *Policy decisions* concerning the chosen public health strategy based on the scientific conclusion, but also considering national values, policy goals and so on.

Distinguishing between scientific conclusions and policy decisions is important because policy decisions about how to manage the risk of COVID-19 are *risk-informed* rather than *risk-based*.[10] This means that policy decisions involve broader considerations than just scientific conclusions about risk—examples include ethical, economic and societal concerns and tradeoffs (eg, the debate about reopening of schools). This perspective also helps explain why—given the same scientific evidence and conclusions—different countries might rationally and justifiably adopt different policies to manage COVID-19 risk.[10]

It is important that the relationship of these structural elements of the assurance case (scientific evidence, scientific conclusions and policy decisions) is explained through well-reasoned and sound arguments:

D. *Scientific argument* explaining the extent to which the scientific evidence (A) supports the scientific conclusions (B).

E. *Policy argument* explaining the extent to which the scientific conclusions (B) are sufficient to support the policy decisions (C).

F. *Confidence argument* explaining the trustworthiness of the scientific evidence (A), for example, the trustworthiness of the epidemiological model.

Scientific evidence

An important part of the scientific evidence comes from epidemiological models, even though there will be other sources of scientific evidence, such as literature reviews. Epidemiological models are engineering artefacts. For use in safety-critical decision making, they should be systematically specified, implemented and tested. The rigour with which this is performed should be proportionate to the criticality of these models to the decision-making process. Given the prominence and importance of epidemiological models in determining a response to COVID-19, the criticality of the models is extremely high.

The COVID-19 model used by the Imperial team is based on a modified individual-based simulation that was developed to support pandemic influenza planning. Models can be intended for a specific purpose, and therefore, a confidence argument would need to justify the suitability of the model for the new context, including the continued validity of the original parameters. This is important because ad-hoc reuse and modification of designs have been associated repeatedly with catastrophic accidents in other safety-critical domains (eg, the recent Boeing 737 Max accidents[11]).

The quality of the software design and the code of the simulator is an important factor, particularly its amenability to inspection and testing.[4] For instance, Neil Ferguson, the lead author of the Imperial report, stated the following: '*For me the code is not a mess, but it's all in my head, completely undocumented. Nobody would be able to use it… and I don't have the bandwidth to support individual users*'.[12] In a safety-critical context, this would significantly undermine confidence in the simulation results. It is actually common practice in high-risk software engineering to employ different software teams to produce different versions of the same software programme to guard against mistakes. That indicates how important—and how difficult—it is to get the software design and coding done without errors.

The validity of the simulation results hinges on large uncertainties and many societal assumptions, for example, about population behavioural changes. In large part, this is because COVID-19 is a novel virus, which is still relatively poorly understood.[2] The developers of the model made many of their assumptions explicit by listing

the corresponding parameters and where data exist to support the chosen parameter values. This is useful and enables an independent assessment and evolution of the model. However, decision makers also need to know how confident they can be that these parameters and assumptions are adequate. For example, the report states an assumption that 30% of patients with COVID-19 who are hospitalised will require critical care (invasive mechanical ventilation) based on early reports from cases in the UK, China and Italy. We now know that this was a significant overestimate due to a combination of miscommunication ('critical care' in many other countries includes non-invasive measures such as continuous positive airway pressure devices) and the effects of the initial official UK advice to 'intubate early'.

Scientific conclusion and argument

Given the novelty of COVID-19 and the large uncertainties around the design of the model and its underpinning data, the transition from the scientific evidence to the overall scientific conclusion is not straightforward.[13] There are usually multiple sources of evidence, neither of which fully supports a conclusion by itself, and each of which has associated uncertainty. It is important, then, to explain through an argument why the evidence provides sufficient support to the conclusions and how confident we can be.

Figure 2 illustrates how the scientific argument can be captured and represented in a structured way through identifying the claims that are made, the evidence that supports those claims and the relationships between them.

In figure 2, only a very simplified structured argument is shown as an example (adapted based on[9]). The results of modelling for different NPIs are used as evidence to support claims about the impact these interventions will have on the number of deaths. This is accompanied by evidence about the suitability for repurposing the model for COVID-19, evidence coming from independent inspection of the software code and evidence about the reliability of the model outputs. Such evidence is used to support an epidemiological claim that a specific combination of NPIs will result in a certain number of deaths with reasonable confidence.

The full argument would be much more comprehensive and draw on further evidence. For example, while we have referred to practices for increasing confidence from the high-risk software engineering domain, these could be complemented with reference to existing best practices for simulation model building and validation.[14 15] The assurance case provides a structure for representing the diverse evidence but does not prescribe what evidence is provided. This is for the developers and assessors of the assurance case to reflect on.

To represent the scientific argument in a structured way, we have used a graphical notation, the Goal Structuring Notation,[16] which is widely used in safety-critical
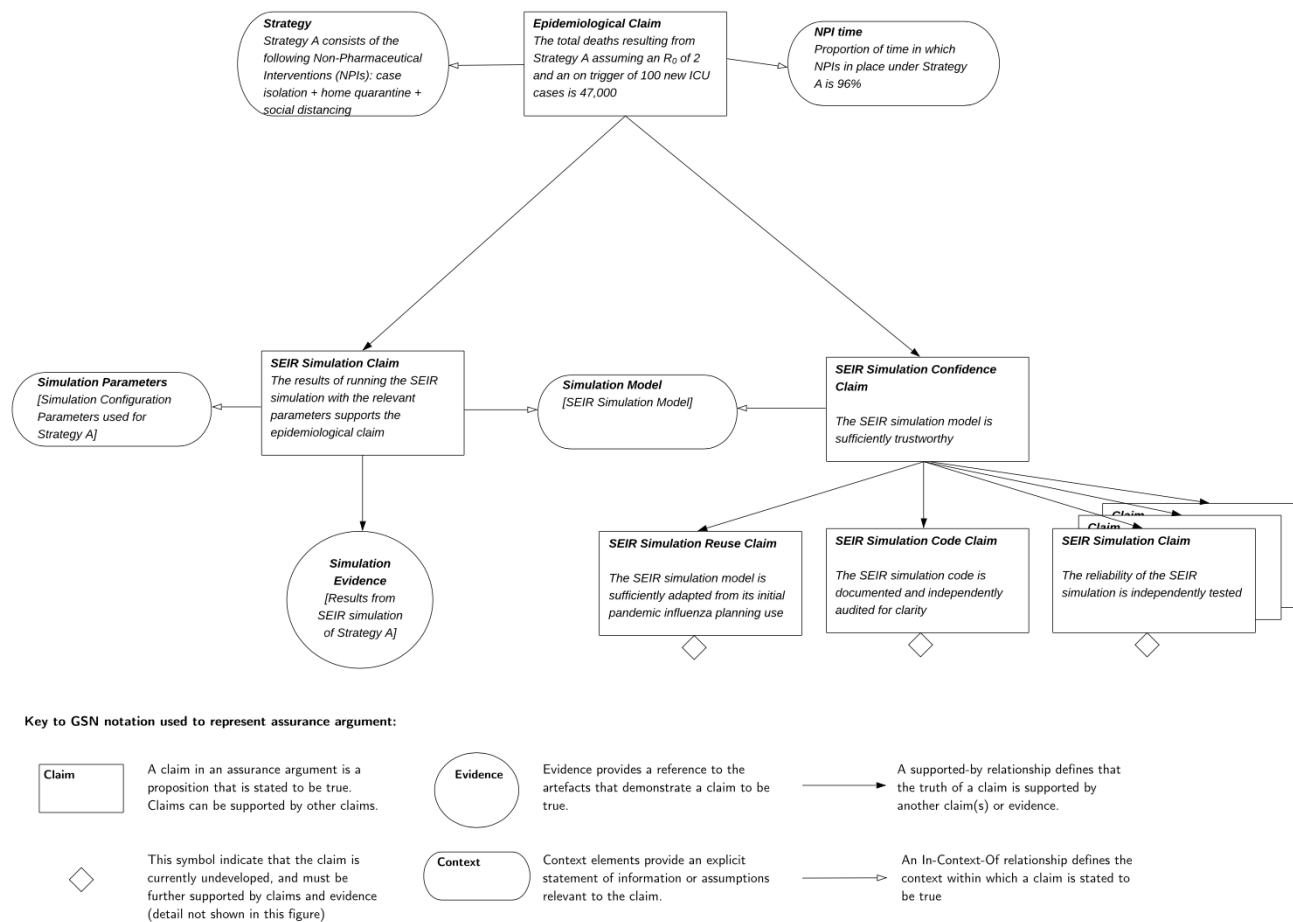
Figure 2 An example of part of a structured scientific argument for COVID-19 simulation. ICU, intensive care unit.

domains for creating and communicating structured assurance arguments.

Policy decisions and argument

Moving from scientific advice and evidence to a policy decision requires that policy-makers consider assumptions, risk acceptance beliefs and tradeoffs (such as between economic and medical impact) that are not often direct and amenable to rigorous scientific examination.[13] The transition from scientific conclusions to a policy decision should therefore involve a complex and diverse policy argument that builds on the scientific conclusions, but also brings to bear these additional considerations.[17] Imperial College Report 9 contains some explicit suggestions for policy (decisions), but it does not contain a policy argument.[9]

A good policy argument should justify the reliance on particular sources of scientific advice and models and acknowledge the extent to which the underlying sources of uncertainty in the evidence were considered. The policy argument should make clear how tradeoffs were made and how evidence concerning the economic, legal and ethical implications of the chosen policy was generated and appraised. In the COVID-19 context, such evidence should also incorporate estimation of non-COVID-19 health harms, for example, potential delays in cancer diagnosis and treatment.

CONCLUSIONS AND RECOMMENDATIONS

Our society is currently placing great weight on epidemiological models of COVID-19 effects. Although such models are essential for dealing with the pandemic, it is hard to know which models we should trust, to what extent and under what conditions. Therefore, we need to make an interdisciplinary effort to create transparency around these models.

The use of assurance cases can help with creating such transparency. Scientists developing and using epidemiological models can document systematically in the assurance case the assumptions they have made, the confidence they have in their respective assumptions and the steps they have taken to ensure that the outputs of their models are valid. This is not dissimilar in nature to the limitations section of a research paper. However, for epidemiological models it would also include additional considerations such as what steps have been taken to ensure that the code that executes the models is correct. Making these assumptions, safeguards and remaining uncertainties explicit increases transparency and can help

policy-makers place justified confidence in the scientific conclusions. This principle also applies to the policy level. Transparency in policy decisions, along with a description of the assumptions and uncertainties allows for public scrutiny of decisions taken.

In such an effort, epidemiologists and health data scientists will have a central role, but they will need support from software engineers, including those with safety-critical software experience. Working together, such collaborations will be able to create standards for developing, testing and maintaining these models in a consistent, rigorous and auditable manner. They will be able to build assurance cases that communicate the uncertainty, assumptions and tradeoffs to a wide variety of stakeholders. This knowledge will then aid policy-makers in using pandemic models in exactly the ways that they are useful and not in the ways that they are not.

Is it realistic to expect these changes to happen? It might be cynical to reason about the level of transparency at the policy level, but as far as the scientific level is concerned, we can be more optimistic. The computer code for some epidemiological models has been made available for scrutiny and we have seen similar efforts for transparency with regards to the code for the (now abandoned) NHSX contact tracing app, which was subjected to independent security analysis. Even so, there is still a need to raise awareness about the contributions disciplines such as safety engineering can make and the onus is as much on engineering bodies to demonstrate the benefits their techniques can bring, as it is on healthcare regulators to require the application of such techniques through regulations.

REFERENCES

1 Adam D. Special report: the simulations driving the world's response to COVID-19. *Nature* 2020;580:316–8.
2 Holmdahl I, Buckee C. Wrong but useful - what Covid-19 epidemiologic models can and cannot tell us. *N Engl J Med* 2020;383:303–5.
3 Peek N, Sujan M, Scott P. Digital health and care in pandemic times: impact of COVID-19. *BMJ Health Care Inform* 2020;27:e100166.
4 Thimbleby H. A call for professional software engineering to be used in epidemic modelling. Available: http://www.harold.thimbleby.net/reliable-models.pdf [Accessed 28 Apr 2020].
5 Sujan MA, Habli I, Kelly TP, *et al*. Should healthcare providers do safety cases? lessons from a cross-industry review of safety case practices. *Saf Sci* 2016;84:181–9.
6 Cleland GM, Sujan MA, Habli I, *et al. Evidence: using safety cases in industry and healthcare*. The Health Foundation, 2012.
7 Bloomfield R, Bishop P. Safety and assurance cases: Past, present and possible future–an Adelard perspective. In: *Making systems safer*. London: Springer, 2010: 51–67.
8 Habli I, White S, Sujan M, *et al*. What is the safety case for health it? A study of assurance practices in England. *Saf Sci* 2018;110:324–35.
9 Ferguson N, Laydon D, Nedjati Gilani G, *et al*. Report 9: impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. Available: https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf [Accessed 27 Apr 2020].
10 Aven T, Bouder F. The COVID-19 pandemic: how can risk science help? *J Risk Res* 2020:1–6.
11 Daniels D. The Boeing 737 max accidents. *28th safety-critical systems symposium*, York, UK, 2020.
12 The Times. Neil Ferguson interview: no 10's infection guru recruits game developers to build coronavirus pandemic model, 2020. Available: https://www.thetimes.co.uk/article/neil-ferguson-interview-no-10s-infection-guru-recruits-game-developers-to-build-coronavirus-pandemic-model-zl5rdtjq5 [Accessed 29 Mar 2020].
13 Donnelly CA, Boyd I, Campbell P, *et al*. Four principles to make evidence synthesis more useful for policy. *Nature* 2018;558:361–4.
14 Hammond RA. Considerations and best practices in agent-based modeling to inform policy. In: *Assessing the use of agent-based models for tobacco regulation*. National Academies Press (US), 2015.
15 Collins A, Petty M, Vernon-Bido D, *et al*. A call to arms: standards for agent-based modeling and simulation. *J Artif Soc Soc Simul* 2015;18.
16 SCSC. Goal structuring notation community standard. Available: https://scsc.uk/scsc-141B [Accessed 28 Apr 2020].
17 Aven T. Risk assessment and risk management: review of recent advances on their foundation. *Eur J Oper Res* 2016;253:1–13.