F1000Research

Check for updates

OPINION ARTICLE

Developing WHO guidelines: Time to formally include evidence from mathematical modelling studies [version 1; referees: 1 approved]

Matthias Egger[1,2], Leigh Johnson[2], Christian Althaus[1], Anna Schöni[1], Georgia Salanti[1], Nicola Low[1], Susan L. Norris [ID][3]

[1]Institute of Social and Preventive Medicine (ISPM), University of Bern, Bern, 3012, Switzerland
[2]Centre for Infectious Disease Epidemiology and Research (CIDER), University of Cape Town, Cape Town, 7925, South Africa
[3]World Health Organization, Geneva, Switzerland

Abstract

In recent years, the number of mathematical modelling studies has increased steeply. Many of the questions addressed in these studies are relevant to the development of World Health Organization (WHO) guidelines, but modelling studies are rarely formally included as part of the body of evidence. An expert consultation hosted by WHO, a survey of modellers and users of modelling studies, and literature reviews informed the development of recommendations on when and how to incorporate the results of modelling studies into WHO guidelines. In this article, we argue that modelling studies should routinely be considered in the process of developing WHO guidelines, but particularly in the evaluation of public health programmes, long-term effectiveness or comparative effectiveness. There should be a systematic and transparent approach to identifying relevant published models, and to commissioning new models. We believe that the inclusion of evidence from modelling studies into the Grading of Recommendations Assessment, Development and Evaluation (GRADE) process is possible and desirable, with relatively few adaptations. No single "one-size-fits-all" approach is appropriate to assess the quality of modelling studies. The concept of the 'credibility' of the model, which takes the conceptualization of the problem, model structure, input data, different dimensions of uncertainty, as well as transparency and validation into account, is more appropriate than 'risk of bias'.

Open Peer Review

Referee Status: ✓

	Invited Referees
	1
version 1 published 29 Aug 2017	✓ report

1 Wilma A. Stolk [ID], University Medical Center Rotterdam, Netherlands

Discuss this article

Comments (0)

Corresponding author: Susan L. Norris (norriss@who.int)

Author roles: Egger M: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Software, Supervision, Writing – Original Draft Preparation; **Johnson L**: Conceptualization, Funding Acquisition, Writing – Review & Editing; **Althaus C**: Conceptualization, Funding Acquisition, Investigation, Methodology, Resources, Software, Supervision, Writing – Review & Editing; **Schöni A**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Software, Visualization, Writing – Review & Editing; **Salanti G**: Conceptualization, Funding Acquisition, Writing – Review & Editing; **Low N**: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Resources, Software, Supervision, Visualization, Writing – Review & Editing; **Norris SL**: Conceptualization, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

How to cite this article: Egger M, Johnson L, Althaus C *et al.* **Developing WHO guidelines: Time to formally include evidence from mathematical modelling studies [version 1; referees: 1 approved]** *F1000Research* 2017, **6**:1584 (doi: 10.12688/f1000research.12367.1)

First published: 29 Aug 2017, **6**:1584 (doi: 10.12688/f1000research.12367.1)

Introduction

Mathematical models have a long history in public health[1]. In 1760, Daniel Bernoulli developed a model of smallpox transmission and control. William Hamer published a measles transmission model in 1906 and Ronald Ross a model of malaria transmission in 1908. In recent years, the number of publications related to mathematical modelling has increased steeply. Today, mathematical modelling studies are not restricted to infectious diseases but address a wide range of questions.

The World Health Organization (WHO) provides recommendations on many public health, health system and clinical topics. WHO guidelines are developed using processes and methods that ensure the publication of high-quality recommendations, as outlined in the *WHO Handbook for Guideline Development*[2]. WHO uses the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach to rate the certainty of a body of evidence and to produce information that is used by guideline panels to formulate recommendations, based on the balance of benefits and harms and other considerations[3].

Many of the questions addressed in mathematical modelling studies are relevant to the development of guidelines. Increasingly, WHO and other guideline developers need to decide whether and how the results of mathematical modelling studies should be included in the evidence base used to develop recommendations. We reviewed the 185 WHO guidelines that were approved by the Guidelines Review Committee from 2007 to 2015: 42 (23%) referred to mathematical modelling studies. However, these studies were rarely formally assessed as part of the body of evidence, and quality criteria for modelling studies were often lacking. A major barrier to the incorporation of evidence from mathematical modelling studies into guidelines is the perceived complexity of the methods used to construct and analyse these studies. At present, there are no widely agreed methods for, or approaches to, the evaluation of the results of mathematical modelling studies, and to their integration with primary data to inform guidelines and recommendations. In April 2016 WHO organized a workshop in Geneva, Switzerland to discuss when and how to incorporate the results of modelling studies into WHO guidelines (see *Acknowledgements* for names of participants; see the Meeting Report). Specifically, the following three questions were discussed at the workshop:

(1) When is it appropriate to consider modelling studies as part of the evidence that supports a guideline?

(2) How should the quality and risk of bias in mathematical modelling studies be assessed?

(3) How can the GRADE approach be adapted to assess the certainty of a body of evidence that includes the results of modelling and to formulate recommendations?

The role of modelling in economic evaluation is well recognised in guideline development and at WHO, and was therefore excluded from discussions. At the workshop, we considered the results of a survey of experts (see Box 1) and a rapid literature review (see below). In this paper, which reflects the opinions of the authors but not necessarily that of all workshop participants, we first define

---

Box 1. Web-based expert survey on the role of mathematical modelling in guideline development

The survey was conducted between March 17 and April 4, 2016. It consisted of 10 questions: four on the characteristics of the respondents, three on the role of mathematical models in guideline development, two questions on quality criteria for mathematical models and one on the challenges in using mathematical modelling in guideline development (see Figure S1). Using snowball sampling, mathematical modellers, epidemiologists, guideline developers and other experts were invited to participate in the survey. A total of 151 individuals from 28 countries and 87 different institutions responded. About half of respondents were modellers, and the other half users of the results from modelling studies. The majority of respondents (58%) had been part of a guideline development group in the past.

Ninety-five percent of respondents answered yes to the question "Should mathematical modelling inform guidance for public health interventions?" and 60% indicated that findings of mathematical modelling studies can sometimes provide the same level of evidence as those of empirical research studies. When asked to list situations in which mathematical modelling could be particularly useful for the development of guidelines, the absence of empirical data on the effectiveness, cost-effectiveness and impact of an intervention, and on the comparative effectiveness of different interventions was most frequently mentioned. We also asked about situations where mathematical modelling studies should not be used or have been inappropriately used in the development of guidelines. Respondents reported that modelling should not be used "to cover up" for the lack of evidence from empirical research, and due emphasis should be given to the uncertainty of model predictions. When asked about the five most important criteria for the quality of reporting of modelling studies, respondents mentioned that the model structure should be clearly described and justified, the important sources of uncertainty reported, and model validity addressed. Assumptions should be clearly stated, justified and discussed and the sources of parameter estimates described. Finally, respondents identified the interpretation of results from modelling studies, the evaluation of their quality and the communication of uncertainty as major challenges in using mathematical modelling in guideline development. These challenges would be best addressed by including at least one modelling expert in guideline development groups.

---

models and modelling studies. We then address the three questions outlined above and conclude with some recommendations on the use of evidence from modelling studies in guidelines development.

What is a mathematical modelling study?

Using a common terminology across different disciplines, for example infectious disease modelling and modelling in chronic disease, will facilitate the assessment, evaluation and comparison of mathematical modelling studies. We define a *mathematical model* as a "mathematical framework representing variables and their interrelationships to describe observed phenomena or predict future events"[4]. *Mathematical modelling studies* are studies that address defined research questions using mathematical modelling, for example the potential of HIV testing with immediate antiretroviral therapy to reduce HIV transmission[5], or the likely impact of different screening practices on the incidence of cervical cancer[6]. In contrast, *statistical modelling* is typically concerned with associations between variables assessed in empirical studies, rather than

an understanding of wider phenomena or systems. The results from statistical analyses of empirical data often inform mathematical models. Mathematical modelling studies also increasingly integrate statistical models into complex models to relate the model output to data.

Role of mathematical modelling studies in guideline development

Mathematical models typically address questions that cannot easily be answered with randomized controlled trials (RCTs) or observational studies. Table 1 lists specific situations and examples where the results of mathematical modelling are particularly relevant to guideline development, based on the survey, published examples and the Geneva workshop. Mathematical modelling can overcome some of the limitations of results obtained from the carefully controlled settings in which RCTs are typically conducted. First, the main trial results provide an average effect estimate that applies to a specific intervention and study population. Mathematical modelling studies can be used to extrapolate from the results of RCTs to different target groups and settings, to long term outcomes, and to bridge the gap between efficacy and (long-term) effectiveness[7]. Second, interventions to prevent and control infectious diseases have non-linear effects. RCTs that address short term effects at the individual level might not be suitable for estimating the longer term effects of introducing an intervention, say a vaccine, in a whole population if indirect herd effects influence the incidence of infection and hence the impact of the intervention[8,9]. Third, rapid guidance is often needed early in outbreaks or public health emergencies when relevant interventions for prevention or management might simply not have been evaluated. The results of mathematical modelling studies can be used to draft emergency guidelines or to assess the epidemic potential of new outbreaks[10].

The findings of mathematical modelling studies are only as good as the data and assumptions that inform them. Guideline recommendations should therefore not be based on the outputs of models when uncertainty in the empirical data has not been appropriately quantified, when the model makes implausible assumptions or has not been validated adequately, or when the model predictions vary widely over a plausible range of parameter estimates.

Assessing the quality of a mathematical modelling study: Rapid review

We performed a rapid review of the methodological literature to identify criteria that are proposed to assess the "quality" of mathematical modelling studies (see Table S1 for the detailed search strategy). Specifically, we aimed to identify criteria proposed to assess the quality of single mathematical modelling studies, including best practice standards or criteria for assessing risk of bias or reporting quality and criteria proposed to assess the quality of a body of evidence from mathematical modelling studies. We were also interested in identifying checklists or other instruments developed to assess the quality of mathematical modelling studies.

Table 1. Situations in which mathematical modelling studies may be useful for guideline development.

Situation	Examples of relevant mathematical modelling studies
The long-term effectiveness or cost-effectiveness of an intervention is unclear.	Life time effect on decompensated cirrhosis of obeticholic acid as second-line treatment in primary biliary cholangitis[37]. Outcomes and costs over 10 years of donepezil treatment in mild to moderately severe Alzheimer's Disease[38]. Long-term clinical outcomes, costs and cost-effectiveness of interventions in diabetes mellitus (types 1 and 2)[39].
The outcomes of an intervention in real world, routine care settings are unclear.	Outcomes of medical management of asymptomatic patients with carotid artery stenosis who were excluded from clinical trials[40]. Effects on blood pressure and cardiovascular risk of variations in patients' adherence to prescribed antihypertensive drugs[41].
The comparative (relative) effectiveness of different interventions overall or in subgroups of patients is unclear.	Comparative effectiveness of different statins and statin doses in patient groups with varying baseline cardiovascular risk[42]. Relative effect of different strategies of incorporating bevacizumab into platinum-based treatment on survival in ovarian cancer[43]. Relative real-world drug effectiveness of disease modifying anti-rheumatic drugs (DMARDs)[44].
The overall effects of an intervention at the population level, including direct and indirect effects, are unknown.	Effects of different vaccination strategies with serogroup C meningococcal conjugate vaccines on meningococcal carriage and disease[45]. Public health impact of vaccinating boys and men with a quadrivalent HPV vaccine[46]. Impact of expanding access to antiretroviral therapy ("treatment as prevention") on new HIV infections[30].
The population burden of a disease or condition is unknown.	Estimate of the global burden of latent tuberculosis infection[47]. Burden of healthcare-associated infections on European population health[48]. Global variation in stroke burden and mortality[49].

Source: WHO expert survey and consultation.

We identified 20 relevant articles (see Figure 1 for a flow chart of the identification of eligible articles)[9,11–28]. Most gave recommendations for good modelling practice and were compiled by a task force in a consensus process or based on a systematic or narrative review of the literature. The widely cited 2003 paper by Weinstein and colleagues organized 28 recommendations under the headings "structure", "data", and "validation"[15]. A questionnaire or checklist was not included. A subsequent series of seven articles[9,22–26,28] by the joint International Society for Pharmacoeconomics and Outcomes Research (ISPOR) and Society for Medical Decision Making (SMDM) task force elaborated upon these recommendations, providing detailed advice on conceptualizing the model, state transition models, discrete event simulations, dynamic transmission models, parameter estimation and uncertainty, and transparency and validation. The 79 recommendations are summarized in the first article of the series[28].

We identified four articles[16,18,21,27] that present comprehensive frameworks of good modelling practice, with detailed justifications of the items covered and attributes of good practice. They include signalling or helper questions to facilitate the critical appraisal of published modelling studies: the number of questions ranges from 38 in Caro *et al.*[16] to 66 questions in Bennett and Manuel[21]. The four frameworks cover similar territory, including items related to the problem concept, model structure, data sources and synthesis of the evidence, model uncertainty, consistency, transparency and validation (Table 2). Two of the frameworks include sponsorship and conflicts of interest[16,21].

In a qualitative study Chilcot *et al.*[11] performed in-depth interviews with 12 modellers from academic and commercial sectors, and model credibility emerged as the central concern of decision-makers using models. Respondents agreed that developing an understanding of the clinical situation or disease process being investigated is paramount in ensuring model credibility, highlighting the importance of clinical input during the model development process[11].

Model comparisons and modelling consortia

Published mathematical models addressing the same issue may reach contrasting conclusions. In this situation, careful comparison of the models may lead to a deeper understanding of the factors that drive outputs and conclusions. Ideally, the different modelling groups come together to explore the importance of differences in the type and structure of their models, and of the data used to parameterize them[29–31]. For example, several groups of modellers have investigated the impact of expanding access to antiretroviral therapy (ART) on new HIV infections. The HIV Modelling Consortium compared the predictions of several mathematical models simulating the same ART intervention programs to determine the extent to which models agree on the epidemiological impact of expanded ART[30]. The consortium concluded that although models vary substantially in structure, complexity, and parameter choices, all suggested that ART, at high levels of access and with high adherence, has the potential to substantially reduce new HIV infections in the population[30]. There was broad agreement regarding the short-term epidemiologic impact of ART scale-up, but more variation in longer-term projections and in the efficiency with which
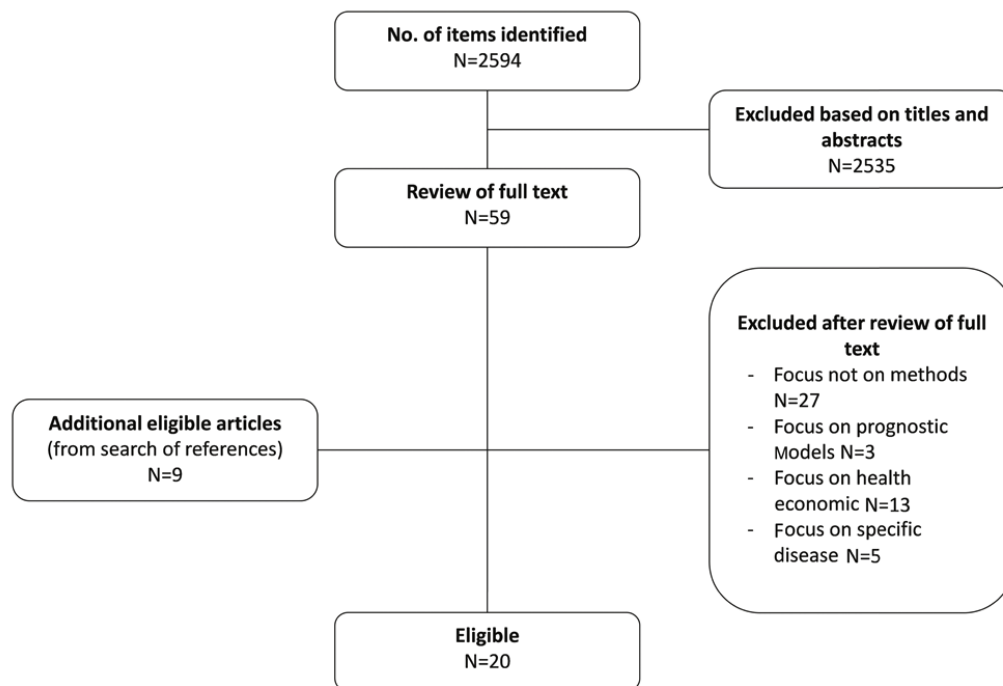

Figure 1. Rapid review of literature on good practice in mathematical modelling: flow of identification of eligible studies.

Table 2. Items covered by four published frameworks developed to assess good modelling practice.

Philips 2006[18]	Bennett 2012[21]	Caro 2014[16]	Peñaloza Ramos 2015[27]
Structure	**Structure**	RELEVANCE	**Problem concept**
Decision problem/objective	Decision problem/objective	Population	Decision problem
Scope/perspective	Scope/perspective	Interventions	Analytical perspective
Rationale for structure	Rationale for structure	Outcomes	Target population
Structural assumptions	Structural assumptions	Context	Health outcomes
Strategies/comparators	Strategies/comparators		Comparators
Model type	Model type	**CREDIBILITY**	Time horizon
Time horizon	Time horizon	**Validation**	
Disease states/pathways	Disease states/pathways	External validation	**Model concept**
Cycle length	Cycle length	Internal verification	Choice of model type
	Parsimony	Face validity	Model structure
Data			
Data identification	**Data**	**Design**	**Synthesis of evidence**
Pre-model data analysis	Data identification	Problem concept	Data sources
Baseline data	Data modelling	Model concept and structure	Utilities
Treatment effects	Baseline data		Cycle length and half-cycle correction
Utilities	Treatment effects	**Data**	Resources/costs
Data incorporation	Risk factors	Process of obtaining and	Patient heterogeneity
Assessment of uncertainty	Data incorporation	values of inputs	Parameter precision
Methodological	Assessment of uncertainty		
Structural	Methodological	**Analysis**	**Model uncertainty**
Heterogeneity	Structural	Adequacy	Analyses of uncertainty related to the
Parameter	Heterogeneity	Uncertainly	decision problem
	Parameter		Parameter estimation
Consistency		**Reporting**	Structural uncertainty
Internal consistency	**Consistency**	Adequacy	Other analyses of uncertainty
External consistency	Internal consistency		
	External consistency	**Interpretation**	**Model transparency and validation**
		Balance	Transparency
	Validity		Validation
	Output plausibility	**Conflict of interest**	Face validity
	Predictive validity	Potential conflicts and steps	Internal validity
		taken to address them	Cross-validation
	Computer implementation		External validity
	Transparency		Predictive validity
	Sponsorship		

treatment can reduce new infections. The impact of ART on HIV incidence long-term is expected to be lower if models: (i) allow for heterogeneity in sexual risk behaviour; (ii) are age-structured; (iii) estimate a low proportion of HIV transmission from individuals not on ART with advanced disease (at low CD4 counts); (iv) are compared to what would be expected in the presence of HIV counselling and testing (compared to no counselling and testing); (v) assume relatively high infectiousness on ART; and (vi) consider drug resistance[30,32,33].

Assessing mathematical modelling studies using the GRADE approach

GRADE was conceived with the intention of creating a uniform system to assess a body of evidence to support guideline development in response to a confusing array of different systems in use at that time[34]. It has since been adopted by over 90 organisations, including WHO. GRADE addresses clinical management questions, including the impact of therapies and diagnostic strategies, diagnostic accuracy questions (i.e., the accuracy of a single diagnostic or screening test), the (cost-) effectiveness and safety of public health interventions, and questions about prognosis.

The GRADE approach encompasses two main considerations: the *degree of certainty* in the evidence used to support a decision and *the strength of the recommendation*. The degree of certainty, i.e., the confidence in or quality of a body of evidence, is rated as "high", "moderate", "low", or "very low" based on an assessment of five dimensions: study limitations (risk of bias), imprecision, inconsistency, indirectness, and publication bias. The initial assessment is based on the study design: RCTs start as high certainty and observational studies as low certainty. Based on the assessments of the five dimensions, RCTs may be down-rated and

observational studies up- or down-rated. Judgment is required when assessing the certainty of the evidence, taking into account the number of studies of higher and lower quality and the relative importance of the different dimensions in a given context. The second consideration is the strength of the recommendation, which can be "strong" or "conditional", for or against an intervention or test, based on the balance of benefits and harms, certainty of the evidence, the relative values of persons affected by the intervention, resource considerations, acceptability and feasibility, among others[35].

We believe that evidence from mathematical modelling studies could be assessed within the GRADE framework and included in the guideline development process. Specifically, guideline groups might include mathematical modelling studies as an additional study category, in addition to the categories of RCTs and observational studies currently defined in GRADE. The dimensions of indirectness, inconsistency, imprecision and publication bias are applicable to mathematical modelling studies, but criteria may need to be adapted. The concept of bias relates to results or inferences from empirical studies[36]. "Lack of credibility" may therefore be a more appropriate term for modelling studies than "risk of bias". The assessment of the credibility of a model is informed by a comprehensive quality framework and should cover the conceptualization of the problem, model structure, input data, different dimensions of uncertainty, as well as transparency and validation (Table 2). The framework should be tailored to each set of modelling studies by adding or omitting questions and developing review-specific guidance on how to assess each criterion. The certainty of the body of evidence from modelling studies can then be classified as high, moderate, low, or very low. In the evidence-to-decision framework a distinction should be made between *observed outcomes* from empirical studies and *modelled outcomes* from modelling studies (see an example).

Conclusions and recommendations
Based on the discussions and presentations at the workshop in Geneva, the survey and rapid systematic review, we believe a number of conclusions can be formulated.

When is it appropriate to consider modelling studies as part of the evidence that supports a guideline?
1. The use of modelling studies should routinely be considered in the process of developing WHO guidelines. Findings of mathematical modelling studies can provide important evidence that may be highly relevant. Evidence from modelling studies should be considered specifically in the absence of empirical data directly addressing the question of interest, when modelling based on appropriate indirect evidence may be indicated. Examples for such situations include the evaluation of long-term effectiveness, and the impact of one or several interventions (comparative effectiveness), for example in the context

of public health programmes where RCTs are rarely available.

2. Modelling may be more acceptable and more influential in situations where immediate action is called for, but little direct empirical evidence is available, and may arguably be more acceptable in public health than in clinical decision making. In these situations (for example the HIV, Ebola, or Zika epidemics) funding is also likely to become available to support dedicated modelling studies.

3. The use of evidence from mathematical models should be carefully considered and there should be a systematic and transparent approach to identifying existing models that may be relevant, and to commissioning new models.

How should the credibility of mathematical modelling studies be assessed?
4. No single "one-size-fits-all" approach is appropriate to assess the quality of modelling studies. Existing frameworks and checklists may be adapted to a set of modelling studies by adding or omitting questions. In some situations, the approach will need to be developed *de novo*.

5. Additional expertise will typically be required in the systematic review groups or guideline development groups to appropriately assess the credibility of modelling studies and interpret their results.

6. The credibility of the models should not be evaluated only by modellers, and not only by modellers involved in the development of these models.

How can the GRADE approach be adapted to assess a body of evidence that includes the results of modelling and to formulate recommendations?
7. The inclusion of evidence from modelling studies into the GRADE process is possible and desirable, with relatively few adaptations. GRADE is simply rating the certainty of evidence to support a decision and any type of evidence can in principle be included.

8. The certainty of the evidence for modelling studies should be assessed and presented separately in summaries of the evidence (GRADE evidence profiles), and classified as high, moderate, low, or very low certainty.

9. The GRADE dimensions of certainty (imprecision, indirectness, inconsistency and publication bias) and the criteria defined for their assessment are also relevant to modelling studies.

10. For modelling studies, the concept of the 'credibility' of the model, which takes the structure of the model, input data, dimensions of uncertainty, as well as transparency and

validation into account, is more appropriate than 'study limitations' or 'risk of bias'.

11. When summarizing the evidence, a distinction should be made between observed and modelled outcomes.

We look forward to discussing these recommendations with experts and stakeholders and to developing exact procedures and criteria for the assessment of modelling studies and their inclusion in the GRADE process.

Competing interests
Susan L. Norris is a member of the GRADE working group. No other competing interests were disclosed.

Supplementary material
Table S1. Search strategy in MEDLINE from inception to January 2016 without language restrictions, combining terms for mathematical models with terms for quality assessment and health care decision-making.

Click here to access the data.

Figure S1. Questionnaire of the online survey on the use of mathematical modelling in guidelines for public health decision making.

Click here to access the data.

References

1. Smith DL, Battle KE, Hay SI, *et al.*: **Ross, macdonald, and a theory for the dynamics and control of mosquito-transmitted pathogens.** *PLoS Pathog.* 2012; **8**(4): e1002588.
   PubMed Abstract | Publisher Full Text | Free Full Text

2. World Health Organization: **WHO Handbook for Guideline Development.** Geneva, 2014.
   Reference Source

3. Guyatt GH, Oxman AD, Vist GE, *et al.*: **GRADE: An emerging consensus on rating quality of evidence and strength of recommendations.** *BMJ.* 2008; **336**(7650): 924–6.
   PubMed Abstract | Publisher Full Text | Free Full Text

4. Eykhoff P: **System identification; parameter and state estimation.** Chester: John Wiley & Sons Ltd., 1974.
   Reference Source

5. Granich RM, Gilks CF, Dye C, *et al.*: **Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: a mathematical model.** *Lancet.* 2009; **373**(9657): 48–57.
   PubMed Abstract | Publisher Full Text

6. Canfell K, Barnabas R, Patnick J, *et al.*: **The predicted effect of changes in cervical screening practice in the UK: results from a modelling study.** *Br J Cancer.* 2004; **91**(3): 530–6.
   PubMed Abstract | Publisher Full Text | Free Full Text

7. Egger M, Moons KG, Fletcher C, *et al.*: **GetReal: from efficacy in clinical trials to relative effectiveness in the real world.** *Res Synth Methods.* 2016; **7**(3): 278–81.
   PubMed Abstract | Publisher Full Text

8. Weinstein MC: **Recent developments in decision-analytic modelling for economic evaluation.** *Pharmacoeconomics.* 2006; **24**(11): 1043–53.
   PubMed Abstract | Publisher Full Text

9. Pitman R, Fisman D, Zaric GS, *et al.*: **Dynamic transmission modeling: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--5.** *Value Health.* 2012; **15**(6): 828–34.
   PubMed Abstract | Publisher Full Text

10. Camacho A, Kucharski AJ, Funk S, *et al.*: **Potential for large outbreaks of Ebola virus disease.** *Epidemics.* 2014; **9**: 70–8.
    PubMed Abstract | Publisher Full Text | Free Full Text

11. Chilcott J, Tappenden P, Rawdin A, *et al.*: **Avoiding and identifying errors in health technology assessment models: qualitative study and methodological review.** *Health Technol Assess.* 2010; **14**(25): iii–iv, ix–xii, 1–107.
    PubMed Abstract | Publisher Full Text

12. Weinstein MC, Toy EL, Sandberg EA, *et al.*: **Modeling for health care and other policy decisions: uses, roles, and validity.** *Value Health.* 2001; **4**(5): 348–61.
    PubMed Abstract | Publisher Full Text

13. Philips Z, Ginnelly L, Sculpher M, *et al.*: **Review of guidelines for good practice in decision-analytic modelling in health technology assessment.** *Health Technol Assess.* 2004; **8**(36): iii–iv, ix–xi, 1–158.
    PubMed Abstract | Publisher Full Text

14. Goldhaber-Fiebert JD, Stout NK, Goldie SJ: **Empirically evaluating decision-analytic models.** *Value Health.* 2010; **13**(5): 667–74.
    PubMed Abstract | Publisher Full Text

15. Weinstein MC, O'Brien B, Hornberger J, *et al.*: **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 Health.* 2003; **6**(1): 9–17.
    PubMed Abstract | Publisher Full Text

16. Jaime Caro J, Eddy DM, Kan H, *et al.*: **Questionnaire to assess relevance and credibility of modeling studies for informing health care decision making: An**

ISPOR-AMCP-NPC good practice task force report. *Value Health.* 2014; **17**(2): 174–82.
PubMed Abstract | **Publisher Full Text**

17. Bilcke J, Beutels P, Brisson M, *et al.*: **Accounting for Methodological, Structural, and Parameter Uncertainty in Decision-Analytic Models: A Practical Guide.** *Med Decis Mak.* 2011; **31**(4): 675–92.
PubMed Abstract | **Publisher Full Text**

18. Philips Z, Bojke L, Sculpher M, *et al.*: **Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment.** *Pharmacoeconomics.* 2006; **24**(4): 355–71.
PubMed Abstract | **Publisher Full Text**

19. Chilcott J, Brennan A, Booth A, *et al.*: **The role of modelling in prioritising and planning clinical trials.** *Health Technol Assess.* 2003; **7**(23): iii, 1–125.
PubMed Abstract | **Publisher Full Text**

20. Forsberg HH, Aronsson H, Keller C, *et al.*: **Managing health care decisions and improvement through simulation modeling.** *Qual Manag Health Care.* 2011; **20**(1): 15–29.
PubMed Abstract | **Publisher Full Text**

21. Bennett C, Manuel DG: **Reporting guidelines for modelling studies.** *BMC Med Res Methodol.* 2012; **12**: 168.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

22. Siebert U, Alagoz O, Bayoumi AM, *et al.*: **State-transition modeling: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--3.** *Value Health.* 2012; **15**(6): 812–20.
PubMed Abstract | **Publisher Full Text**

23. Briggs AH, Weinstein MC, Fenwick EA, *et al.*: **Model parameter estimation and uncertainty: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--6.** *Value Health.* 2012; **15**(6): 835–42.
PubMed Abstract | **Publisher Full Text**

24. Karnon J, Stahl J, Brennan A, *et al.*: **Modeling using Discrete Event Simulation: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force--4.** *Value Health.* 2012; **15**(6): 821–7.
PubMed Abstract | **Publisher Full Text**

25. Roberts M, Russell LB, Paltiel AD, *et al.*: **Conceptualizing a model: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--2.** *Value Health.* 2012; **15**(6): 804–11.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

26. Eddy DM, Hollingworth W, Caro JJ, *et al.*: **Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--7.** *Value Health.* 2012; **15**(6): 843–50.
PubMed Abstract | **Publisher Full Text**

27. Ramos MC, Barton P, Jowett S, *et al.*: **A Systematic Review of Research Guidelines in Decision-Analytic Modeling.** *Value Health.* 2015; **18**(4): 512–29.
PubMed Abstract | **Publisher Full Text**

28. Caro JJ, Briggs AH, Siebert U, *et al.*: **Modeling good research practices--overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--1.** *Value Health.* 2012; **15**(6): 796–803.
PubMed Abstract | **Publisher Full Text**

29. Mandelblatt JS, Cronin KA, Bailey S, *et al.*: **Effects of mammography screening under different screening schedules: model estimates of potential benefits and harms.** *Ann Intern Med.* 2009; **151**(10): 738–47.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

30. Eaton JW, Johnson LF, Salomon JA, *et al.*: **HIV treatment as prevention: systematic comparison of mathematical models of the potential impact of antiretroviral therapy on HIV incidence in South Africa.** *PLoS Med.* 2012; **9**(7): e1001245.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

31. Althaus CL, Turner KM, Schmid BV, *et al.*: **Transmission of *Chlamydia trachomatis* through sexual partnerships: a comparison between three individual-based models and empirical data.** *J R Soc Interface.* 2012; **9**(66): 136–46.
PubMed Abstract | **Publisher Full Text**

32. Hontelez JA, Lurie MN, Bärnighausen T, *et al.*: **Elimination of HIV in South Africa through expanded access to antiretroviral therapy: a model comparison study.** *PLoS Med.* 2013; **10**(10): e1001534.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

33. Law MG, Prestage G, Grulich A, *et al.*: **Modelling the effect of combination antiretroviral treatments on HIV incidence.** *AIDS.* 2001; **15**(10): 1287–94.
PubMed Abstract | **Publisher Full Text**

34. Atkins D, Best D, Briss PA, *et al.*: **Grading quality of evidence and strength of recommendations.** *BMJ.* 2004; **328**(7454): 1490.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

35. Andrews JC, Schünemann HJ, Oxman AD, *et al.*: **GRADE guidelines: 15. Going from evidence to recommendation-determinants of a recommendation's direction and strength.** *J Clin Epidemiol.* 2013; **66**(7): 726–35.
PubMed Abstract | **Publisher Full Text**

36. Guyatt GH, Oxman AD, Vist G, *et al.*: **GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias).** *J Clin Epidemiol.* 2011; **64**(4): 407–15.
PubMed Abstract | **Publisher Full Text**

37. Samur S, Klebanoff M, Banken R, *et al.*: **Long-term clinical impact and cost-effectiveness of obeticholic acid for the treatment of primary biliary cholangitis.** *Hepatology.* 2017; **65**(3): 920–928.
PubMed Abstract | **Publisher Full Text**

38. Getsios D, Blume S, Ishak KJ, *et al.*: **Cost effectiveness of donepezil in the treatment of mild to moderate Alzheimer's disease: a UK evaluation using discrete-event simulation.** *Pharmacoeconomics.* 2010; **28**(5): 411–27.
PubMed Abstract | **Publisher Full Text**

39. Palmer AJ, Roze S, Valentine WJ, *et al.*: **The CORE Diabetes Model: Projecting long-term clinical outcomes, costs and cost-effectiveness of interventions in diabetes mellitus (types 1 and 2) to support clinical and reimbursement decision-making.** *Curr Med Res Opin.* 2004; **20 Suppl 1**: S5–26.
PubMed Abstract | **Publisher Full Text**

40. Smolen HJ, Cohen DJ, Samsa GP, *et al.*: **Development, validation, and application of a microsimulation model to predict stroke and mortality in medically managed asymptomatic patients with significant carotid artery stenosis.** *Value Health.* 2007; **10**(6): 489–97.
PubMed Abstract | **Publisher Full Text**

41. Lowy A, Munk VC, Ong SH, *et al.*: **Effects on blood pressure and cardiovascular risk of variations in patients' adherence to prescribed antihypertensive drugs: role of duration of drug action.** *Int J Clin Pract.* 2011; **65**(1): 41–53.
PubMed Abstract | **Publisher Full Text**

42. Schuetz CA, van Herick A, Alperin P, *et al.*: **Comparing the effectiveness of rosuvastatin and atorvastatin in preventing cardiovascular outcomes: estimates using the Archimedes model.** *J Med Econ.* 2012; **15**(6): 1118–29.
PubMed Abstract | **Publisher Full Text**

43. Barnett JC, Alvarez Secord A, Cohn DE, *et al.*: **Cost effectiveness of alternative strategies for incorporating bevacizumab into the primary treatment of ovarian cancer.** *Cancer.* 2013; **119**(20): 3653–61.
PubMed Abstract | **Publisher Full Text**

44. Didden E, Ruffieux Y, Hummel N, *et al.*: **Prediction of Real-World Drug Effectiveness Pre-Launch: Case study in Rheumatoid Arthritis.** *Value Health.* 2017; submitted.

45. Trotter CL, Gay NJ, Edmunds WJ: **Dynamic models of meningococcal carriage, disease, and the impact of serogroup C conjugate vaccination.** *Am J Epidemiol.* 2005; **162**(1): 89–100.
PubMed Abstract | **Publisher Full Text**

46. Elbasha EH, Dasbach EJ: **Impact of vaccinating boys and men against HPV in the United States.** *Vaccine.* 2010; **28**(42): 6858–67.
PubMed Abstract | **Publisher Full Text**

47. Houben RM, Dodd PJ: **The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling.** *PLOS Med.* 2016; **13**(10): e1002152.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

48. Cassini A, Plachouras D, Eckmanns T, *et al.*: **Burden of Six Healthcare-Associated Infections on European Population Health: Estimating Incidence-Based Disability-Adjusted Life Years through a Population Prevalence-Based Modelling Study.** *PLoS Med.* 2016; **13**(10): e1002150.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

49. Johnston SC, Mendis S, Mathers CD: **Global variation in stroke burden and mortality: estimates from monitoring, surveillance, and modelling.** *Lancet Neurol.* 2009; **8**(4): 345–54.
PubMed Abstract | **Publisher Full Text**

Open Peer Review

Current Referee Status: ✔

✔ **Wilma A. Stolk** (iD)

Erasmus MC, Department of Public Health, University Medical Center Rotterdam, Rotterdam,
Netherlands

In this opinion article, the authors discuss when and how to incorporate the results of modelling studies into WHO guidelines, by addressing three questions: (1) When is it appropriate to consider modelling studies as part of the evidence that supports a guideline? (2) How should the quality and risk of bias in mathematical modelling studies be assessed? (3) How can the GRADE approach be adapted to assess the certainty of a body of evidence that includes the results of modelling and to formulate recommendations? Based on findings from a web-based expert survey, a rapid literature review to identify criteria for assessing the "quality" of mathematical modelling studies, and on discussions and presentations at a workshop on the topic that was held April 2016 in Geneva, the authors conclude that modelling studies should indeed routinely be considered in the process of developing WHO guidelines, particularly in the evaluation of public health programmes, long-term effectiveness or comparative effectiveness. As for other types of evidence taken into consideration, there should be a systematic and transparent approach to identifying existing models that may be relevant and the quality and credibility of models should be systematically assessed. Relatively few adaptations are needed in the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach to rate the certainty of a body of evidence and to produce information that is used by guideline panels to formulate recommendations, based on the balance of benefits and harms and other considerations.

MINOR COMMENTS:
1. Recommendation 4 is "No single 'one-size-fits-all' approach is appropriate to assess the quality of modelling studies. Existing frameworks and checklists may be adapted to a set of modelling studies by adding or omitting questions. In some situations, the approach will need to be developed *de novo*." I'd prefer to turn it around: based on existing frameworks and checklists, generic criteria can be developed to assess the quality of modelling studies, although – depending on the situation –questions may have to be added or omitted. I am not convinced that in some situations a completely new approach is needed, and this would also not be advisable. The authors should either delete the last statement, or explain under which circumstances such a new approach is needed, ideally illustrated with an example.

2. Recommendation 8 is "The certainty of the evidence for modelling studies should be assessed and presented separately in summaries of the evidence (GRADE evidence profiles), and classified as high, moderate, low, or very low certainty." In the text, the authors state that RCTs start as high certainty and observational studies as low certainty, although this certainty score may be up- or down-rated based on detailed assessment of five dimensions. Is it possible to give an indication of

where modelling studies would start, with a justification? If not, can the authors describe factors to be considered when determining the start class?

3. The questionnaire of the online survey on the use of mathematical modelling in guidelines for public health decision making is included as Figure S1, which combines a series of screen shots. The quality of this figure is poor and I recommend to include the questionnaire as a text document.

Is the topic of the opinion article discussed accurately in the context of the current literature?
Yes

Are all factual statements correct and adequately supported by citations?
Yes

Are arguments sufficiently supported by evidence from the published literature?
Yes

Are the conclusions drawn balanced and justified on the basis of the presented arguments?
Yes

Competing Interests: No competing interests were disclosed.

Referee Expertise: Epidemiology / mathematical modelling, with focus on neglected tropical diseases

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.