Edinburgh Research Explorer

Development of a Critical Appraisal Tool for Models Predicting the Impact of "Test, Trace, and Protect" Programmes on COVID-19 Transmission

OPEN ACCESS

Development of a Critical Appraisal Tool for Models Predicting the Impact of "Test, Trace, and Protect" Programmes on COVID-19 Transmission

ABSTRACT

Objectives: To develop a Critical Appraisal tool for non-computational-specialist public health professionals to assess the quality and relevance of modelling studies about Test and Trace (and Protect – TTP) programmes' impact on COVID-19 transmission.
Study Design: Decision-making tool development
Methods: Using Tugwell et al.'s Health Care Effectiveness equation as a conceptual framework, combined with a purposive search of the relevant early modeling literature, we developed six critical appraisal questions for the rapid assessment of modeling studies related to the evaluation of TTP programmes' effectiveness.
Results: By applying the Critical Appraisal tool to selected recent COVID-19 modeling studies we demonstrate how models can be evaluated using the six questions to evaluate internal and external validity, and relevance.
Conclusions: These six critical appraisal questions are able to discriminate between modeling studies of higher and lower quality and relevance to evaluating TTP programmes' impact. However, these questions require independent validation in a larger and systematic sample of relevant modeling studies which have appeared in later stages of the pandemic.

Key words: COVID-19; disease transmission; mathematical model(s); test and trace; evaluation of contact tracing; critical appraisal

INTRODUCTION

Decision making related to the COVID-19 pandemic has made extensive use of information from studies using complex mathematical models. Specialist technical and contextual knowledge is necessary for detailed "critical appraisal" of such studies. However, public health professionals lacking relevant technical knowledge are often required to evaluate quality and relevance of modelling studies.[1] It would be useful for non-specialists, especially public health professionals with only standard (i.e. MPH-level) training in epidemiology, to be able to quickly assess when to bring new COVID-19 modeling papers (appearing in large numbers since the start of the pandemic) to the attention of modeling specialist colleagues.

Several authors[2-6] have developed approaches to assess internal and external validity for modeling studies. However, these tools are generic and encompass a broad range of models, spanning clinical diagnostic/prognostic decision tools through to burden-of-illness estimates and cost-effectiveness analyses.

We address this gap by developing a "Critical Appraisal" tool, for non-specialists to efficiently screen COVID-19 modeling studies for quality and relevance to COVID-19 test trace and protect (TTP) programmes. TTP programmes test individuals, track or trace potential contacts of positive cases and then protect public health by providing advice regarding isolation or quarantine to both cases and contacts. [We would cite Grantz et al[7] as providing a particularly clear and generalizable pictorial description of precisely how TTP programmes work.] Specifically, we devise a Critical Appraisal question checklist to address the question: "*What are the key indicators of modeling study quality and relevance, for evaluation of TTP programme overall effectiveness in reducing COVID-19 transmission?*"

METHODS

Our objectives were to: 1) identify the key modifiers affecting TTP programme effectiveness in reducing COVID-19 transmission; 2) generate less than ten easy-to-use Critical Appraisal (CA) questions that allow non-modelers, with only basic epidemiological training, to assess the quality and relevance of modelling studies for evaluating such effectiveness; 3) demonstrate application of the proposed CA questions using purposively identified modelling studies.

We applied *Iterative Measurement Loop* methodology (see Tugwell *et al*.[8]), an established critical appraisal (CA) tool for analyzing the population-level effectiveness and efficiency of competing health care interventions, to evaluate TTP programme effectiveness in reducing COVID-19 transmission. This led to a comprehensive list of factors affecting TTP programme effectiveness, based on the "Healthcare Effectiveness Equation" (see Box 1)[8].

We adopt the standard CA tool approach (see CASP and Oxford CEBM websites[9,10]) of identifying a checklist of questions that, in sequence:

1. Screen out studies not directly relevant, i.e. determine whether the study in question *addresses key aspects*, identified through *Iterative Measurement Loop* methods[8], *that co-determine TTP programme overall effectiveness*.
2. Assess *internal validity*, i.e. are study findings logically derived from the data presented and analysed?

3. Assess *external validity*, i.e. are the findings applicable to the reader's *particular decision-making situation?* In this case, the evaluation of a specific COVID-19 TTP programme (e.g. as currently deployed in UK and most HICs.)

To generate specific CA questions, we performed a *purposive* review of modeling papers that assess TTP programme effectiveness, to identify *key shortcomings* with respect to the three criteria above. This was limited to studies of High-Income Countries (HICs), and papers published (or listed on relevant pre-print archives) from early 2020 to May 1, 2021. The review was purposive, rather than systematic or narrative, in that modeling papers fitting the inclusion criteria were sampled until no further generic shortcomings were being identified – so-called "saturation."[11] We were unable to validate against an independent sample of relevant TTP modelling papers, because we exhausted the most widely cited studies published during the study period in developing the CA questions. Such validation, in particular for low to middle income countries (LMICs), has therefore been left to other investigators, who will need to use a representative sample of suitable modelling papers published later in the pandemic.

RESULTS

Critical Appraisal Question Conceptual Framework: How Do COVID-19 TTP Programmes Work, and What are the Key Modifiers of their Effectiveness?

Figure 1 provides a schematic description of the rather complex string of processes involved in TTP programme implementation. These can be distinguished by direct effects ('A' in Figure 1) associated with the positive-tested (index) case and by indirect effects ('B') associated with the contacts of that case. Box 1 shows the key modifiers of any TTP programme's effectiveness that can potentially diminish its overall impact on COVID-19 transmission, as derived from the Iterative Measurement Loop associated with the factors in Figure 1, based on the "Healthcare Effectiveness Equation"[8].

INSERT FIGURE 1 AND BOX1 ABOUT HERE

Purposive Literature Search

The most relevant modelling studies for generating checklist questions were identified through targeted search in Google Scholar and widely used pre-print servers (e.g. bioRxiv, medRxiv), using the keywords "COVID* AND model* AND test* AND trace / tracing AND protect / quarantine / isolate AND effect," and by hand-searching the citations in those studies and published reviews of COVID-19 TTP effectiveness-modelling (sometimes compared with other control measures). The range of identified issues regarding internal or external validity was fully captured by twelve original studies [7,11-22], published between early 2020 (effectively the first such studies after the pandemic began) and May 2021. No additional issues compromising internal or external validity were identified from other

modelling studies published during that time period. As a result, the authors were able to identify six major sorts of shortcoming affecting such modeling, which were then integrated into the Critical Appraisal questions listed below.

Critical Appraisal Questions for Screening Modeling Studies Potentially Relevant To COVID-19 TTP Effectiveness Evaluation

> **QUESTION#1. KEY MODIFIERS: Does the study incorporate or account for the effects, on COVID-19 transmission, of variation in the full set of key modifiers of overall TTP programme effectiveness identified in Text Box 1? [If not, stop here: study not likely to be useful]**

It is important to note that a modeling study may not explicitly mention each individual modifier of effectiveness listed in Box 1, as it may "bundle" several modifiers into one or more model parameters or process. For example, Grantz et al.[7] bundled "coverage" (effectiveness modifier #A1) and "test diagnostic accuracy (i.e. sensitivity)" (#A2) with modifier #A6 "compliance with advice to isolate," into a single parameter -- "isolation completeness" -- representing the probability that an infection in the community is detected and isolated by a TTP programme. This also illustrates that studies may use different terminology for key modifiers. To enable assessment of internal and external validity definition and underlying assumptions for each modifier must be stated.

> **QUESTION #2. STRUCTURE AND SCALE: Are models used in the study employing a structure and scale appropriate for evaluating the impact on COVID-19 transmission of TTP programmes operating at the scale of interest, e.g. national or regional?**

Identifying appropriate model structure and scale to assess COVID-19 TTP programme effectiveness is challenging, and the twelve studies identified were found to be heterogeneous in this respect. In terms of structure, for example one might expect strong dependence of model results on assumed between-individual contact patterns, but some models simply assume homogeneous mixing (e.g. Contreras et al.[18]). Similarly, accounting for asymptomatic or pre-symptomatic carriers of SARS-CoV2[23,24] affects testing coverage of potential transmitters (#A1 in Text Box), but only some in-scope studies do so (e.g. again, not Contreras et al.)[18] Caution is advised when considering models that employ coarse scales or overly simplistic structures for contact patterns. Such models may only be able to provide useful predictions of a qualitative nature (e.g. relative importance of specific modifiers on overall predictions). Internal and external validity of model results should be carefully examined in relation to such scope and scale considerations.

For example, generalising from an early study of the local COVID-19 TTP programme (including a widely downloaded mobile phone app) on the Isle of Wight just off the southern English coast[19] may be problematic; its small study population size, and perhaps even more so its unique geography, surely limit its applicability to large nation states.

> **QUESTION #3. PARAMETERISATION: Are key inputs (e.g. values for COVID-19's key transmission parameters and modifiers of effectiveness of TTP programmes, as listed in Box 1) credibly derived: (i) using models fitted to representative data; or (ii) from suitable peer-reviewed studies, and ideally systematic reviews and meta-analyses?**

This criterion would probably have constituted an unreasonably high bar during the first year of the pandemic, where datasets were just starting to get assembled and modelers were unlikely to be granted full access to raw data. Furthermore, too few primary studies, and certainly systematic reviews of them, had been completed until very recently, with many key studies awaiting final peer-review available only through "pre-print" archives, such as medRxiv. Even as late in the pandemic as the end of 2020, Quilty et al. tally publications relevant to estimating quarantine duration-reduction, under rapid antigen testing, with 59 papers on PubMed and 1934 on medRxiv.[20] However, it is now entirely reasonable to demand critical inputs be derived from high-quality sources and analyses, ideally accounting for multiple sources, appropriately vetted for quality and statistically summarized where appropriate, such as two recent syntheses of incubation period data.[25,26]

QUESTION #4. UNCERTAINTY QUANTIFICATION: Does the study account for a credible range of values for key input parameters, by executing comprehensive *sensitivity analyses,* showing resulting uncertainty, e.g. credible intervals or distributions, for key model outputs?

A key issue is the level of uncertainty associated with best estimates of key parameters. The fewer high-quality primary studies providing suitable data, and the narrower the range of relevant settings in which they were conducted, the more important a comprehensive sensitivity analysis becomes. Both Grantz et al.[7] and Contreras et al.[18] appear to meet this criterion, with sensitivity analyses across a wide range of input parameter values.

QUESTION #5. CONSISTENCY WITH OTHER STUDIES [EXTERNAL VALIDITY]: Are key results arising from the model(s) consistent with other high-quality evidence on impact and performance of TTP programmes?

Assessing external validity is not only a matter of looking explicitly for consistency of results across comparable studies and identifying outliers; it also involves noting entire categories of sub-studies (e.g. estimating key model inputs' distributions in particular settings – see above) where there is virtually no replication available. This a particular problem with COVID-19 research, simply because no study was possible until about February/March 2020. As a specific example of good practice in this regard, we would point to the work of the UK's Modeling Sub-Advisory Group (SPI-M) who have carefully issued consensus statements based on a variety of diverse modeling approaches.[27]

QUESTION #6. SENSE CHECK [EXTERNAL VALIDITY]: What specific questions/settings does the appraiser wish to address? Is the model being appraised credibly applicable to these (e.g. the UK in 2021)?

This final question provides the opportunity to ask: "Do I have any remaining doubts (not covered above) about applicability of this study to the particular TTP programme I want to evaluate?" Potential sources of non-generalisability should be assessed along with issues related to the intended application. For example, the agent-based modelling study of Aleta et al.[22] utilises detailed contact structures, based on pre-pandemic mobility data from Boston, USA, and models effects of applied COVID-19 interventions on these assumptions. This study may provide useful guidelines for developing comparable models, but direct application to

other countries is problematic due to likely differences in the pre-pandemic contact patterns and deployment of social distancing measures.

DISCUSSION (and PRACTICAL LESSONS LEARNED)

Here we describe lessons learned to guide those embarking on a literature (or systematic) review of modeling studies to inform evaluation of TTP programmes:

Relative timing of the modelling study to events. Particularly in the context of CA questions 2 (STRUCTURE AND SCALE) and 3 (PARAMETERISATION), it is important to consider the timing of the study in relation to data and knowledge available at the time of publication, compared to when the Critical Appraisal is conducted. For example, in early studies the proportion asymptomatic cases may be based on purely cross-sectional studies whereas, due latent period, only cohort studies provide a clear picture of the true percentage of cases which are fully asymptomatic[23,24]. Models based on such early estimates of key parameters can therefore be expected to have a "limited shelf life," and must be interpreted with caution.

Demographic context. Key parameters vary within and between settings. For example, the secondary attack rate within a household (or household attack rate) is likely to vary considerably within and across populations, but only some models explicitly account for such heterogeneity. Furthermore, households are not of consistent size, age-sex composition, and crowdedness across societies (let alone comparable with respect to cross-reactive immunocompetence arising from previous exposure to other coronaviruses[28]). Secondary attack rates based on household data will not be fully generalizable from one society -- e.g. China, with low birth rates but many households which include older relatives[29], to another -- e.g. in sub-Saharan Africa, with high birth rates, a very young population overall, and many communities with extremely crowded housing, such as large low-income informal settlements[12].

Geographical, cultural or political features. A further caveat to external validity is that some input parameters may be contextualized by other important but often unstated local geographical, cultural or political features. For example, isolated islands (either physically isolated, such as Iceland, New Zealand and the Faroe Islands) or politically distinct "islands" with historically strong border controls (such as Hong Kong, Singapore and Taiwan) have in some cases introduced strict COVID-19 control measures, including gradations of social distancing through to full "lock-down," while at the same time enforcing draconian inbound-traveler restrictions[14]. The effect of such imported-case exclusion measures can be large[15], and may influence observed impacts of TTP programmes since transmission is rendered entirely internal to the population in question. Such issues are most apparent in studies of closed "institutional/cruise-ship" settings, such as the well-known Diamond Princess outbreak early in the pandemic[16]. Such extreme settings may hold advantages for estimating key transmission parameters, but such estimates may be confounded by atypical features, such as population age-profiles or saturation of air-circulation systems by aerosols, leading to more of a "point (or common) source" epidemic curve, rather than a "person-to-person" transmission curve[30]. Thus, generalizing from "island" settings to societies with more porous borders should be undertaken with extreme caution.

Nuances of TTP programmes. TTP programmes may appear to be similar between jurisdictions, but in fact may be quite different in important respects. For example, TTP programmes with strong legal sanctions against cases or their contacts, who are non-compliant with advice to isolate/quarantine (including mandatory "quarantine hotel" stays under armed guard), would be expected to achieve much higher rates of transmission interruption, compared to more voluntary programmes, relying entirely on "self-isolation at home."[17] There are many such features of TTP programmes that powerfully influence case and contact *compliance* with advice to isolate/quarantine (see Box 1), such as concerns about data security, and they may or may not be fully described in a given published account.

Shortcomings of modelling study reporting. We note, as have other commentators[1-3] that inconsistent and often incomplete reporting was common among the dozen key modelling studies we examined in detail. Standard guidance for such reporting has been published and is constantly being refined.[1,3]

Degree of compliance. When using models to evaluate any TTP programme, a key concern is how that programme is executed on the ground, as well as the full context of other societal behavioural patterns relevant to COVID-19 transmission e.g. compensatory behaviours, and the extent to which the study accounts for such factors, especially via proper reporting practices (see above).

In summary, "the devil is in the details". Anyone reviewing modeling studies which make use of model inputs from settings likely affected by these peculiarities should exercise extreme caution in extrapolating the results to settings which are fundamentally different.

The major strength of this study is that it utilized a purposive sample of about a dozen highly cited early modeling studies of COVID-19 TTP programmes' effectiveness to generate CA questions suitable for use by non-modelers, with only MPH-level training in epidemiology, for screening such studies for more detailed attention by trained modelers.

The major weakness of this study is that it did not attempt a systematic review of this exploding literature (as of spring 2021), but instead relied on the likely saturation of identifiable weaknesses, based on a purposive sample of early studies. This limitation may have resulted in bias, and also limit the applicability of these CA questions to later modelling studies utilizing novel and improved methods and/or higher-quality input data. A second major weakness is that the authors did not attempt to validate the CA questions developed on an independent sample of modeling studies, simply because they had already used all the most highly cited studies of this kind in developing the questions. We leave that important task to others, now that many more pertinent modeling studies have been published.

This study has used a systematic process to develop a brief decision tool – involving creation of a bespoke conceptual framework, a purposive search to identify potential modelling study shortcomings, and the subsequent creation of six CA questions. The tool is intended to allow non-modelers to critically assess modelling studies that aim to address the impact on COVID-19 transmission of TTP programmes, a major global intervention to reduce viral transmission. Only by others' attempts to use these questions can we learn how useful they are. To that end, we invite public health professionals who are involved in evidence reviews on this topic to write to us, in care of the corresponding author, about their experiences with this tool.

REFERENCES

1. Egger M, Johnson L, Althaus C, et al. Developing WHO guidelines: Time to formally include evidence from mathematical modelling studies. *F1000Research.* 2017;6. doi:10.12688/F1000RESEARCH.12367.2

2. Bennett C, Manuel DG. Reporting guidelines for modelling studies. *BMC Med Res Methodol.* 2012; 12: 168. Published 2012 Nov 7. doi:10.1186/1471-2288-12-168

3. 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 Heal.* 2014;17(2):174-182. doi:10.1016/j.jval.2014.01.003

4. Holmdahl I, Buckee C. Wrong but Useful — What COVID-19 Epidemiologic Models Can and Cannot Tell Us. *N Engl J Med*. 2020;383(4):303-305. doi:10.1056/nejmp2016822

5. Brozek JL, Canelo-Aybar C, Akl EA, et al. GRADE Guidelines 30: the GRADE approach to assessing the certainty of modeled evidence—An overview in the context of health decision-making. *J Clin Epidemiol.* 2021;129:138-150. doi:10.1016/j.jclinepi.2020.09.018

6. Zawadzki RS, Gong CL, Cho SK, et al. Where do we go from here? A framework for using Susceptible-Infectious-Recovered Models for policy making in emerging infectious diseases. *Value Heal.* 2021;24(7):917. doi:10.1016/J.JVAL.2021.03.005

7. Grantz KH, Lee EC, D'Agostino McGowan L, Lee KH, Metcalf CJ, Gurley ES, et al. Maximizing and evaluating the impact of test-trace-isolate programs: A modeling study. *PLOS Medicine*. 2021 Apr 30;18(4): e1003585.

8. Tugwell P, Bennett KJ, Sackett DL, Haynes RB. The measurement iterative loop: a framework for the critical appraisal of need, benefits and costs of health interventions. *Journal of Chronic Diseases*. 1985 Jan 1;38(4):339-51.

9. Critical Appraisal Skills Programme. https://casp-uk.net/ [accessed 22 Feb., 2021].

10. Oxford Centre for Evidence-Based Medicine. https://www.cebm.net [accessed 22 Feb., 2021].

11. Gentles SJ, Charles C, Ploeg J, McKibbon KA. Sampling in qualitative research: Insights from an overview of the methods literature. *The Qualitative Report.* 2015 Apr 26;20(11):1772-89.

12. Van Zandvoort K, Jarvis CI, Pearson CA, Davies NG, Ratnayake R, Russell TW, et al. Response strategies for COVID-19 epidemics in African settings: a mathematical modelling study. *BMC Medicine*. 2020 Dec;18(1):1-9.

13. Kretzschmar ME, Rozhnova G, Bootsma MC, van Boven M, van de Wijgert JH, Bonten MJ. Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study. *The Lancet Public Health*. 2020 Aug 1;5(8):e452-9.

14. Strøm M, Kristiansen MF, Christiansen DH, Weihe P, Petersen MS. Elimination of COVID-19 in the Faroe Islands: effectiveness of massive testing and intensive case and contact tracing. *The Lancet Regional Health-Europe.* 2020 December 26, 2020 DOI: https://doi.org/10.1016/j.lanepe.2020.100011

15. Russell TW, Wu JT, Clifford S, Edmunds WJ, Kucharski AJ, Jit M. Effect of internationally imported cases on internal spread of COVID-19: a mathematical modelling study. *The Lancet Public Health.* 2021 Jan 1;6(1): e12-20).

16. Mizumoto K, Chowell G. Transmission potential of the novel coronavirus (COVID-19) onboard the Diamond Princess Cruises Ship, 2020. *Infectious Disease Modelling*. 2020 Jan 1; 5: 264-70.

17. Nussbaumer-Streit B, Mayr V, Dobrescu AI, Chapman A, Persad E, Klerings I, et al. Quarantine alone or in combination with other public health measures to control COVID-19: a rapid review. *Cochrane Database of Systematic Reviews.* 2020 (9); Art. No.: CD013574 DOI: 10.1002/14651858.CD013574.pub2.

18. Contreras S, Dehning J, Loidolt M, Zierenberg J, Spitzner FP, Urrea-Quintero JH, et al. The challenges of containing SARS-CoV-2 via test-trace-and-isolate. *Nature Communications*. 2021 Jan 15;12(1):1-3.19.

19. Kendall M, Milsom L, Abeler-Dörner L, Wymant C, Ferretti L, Briers M, et al. Epidemiological changes on the Isle of Wight after the launch of the NHS Test and Trace programme: a preliminary analysis. *The Lancet Digital Health*. 2020 Dec 1;2(12): e658-66.

20. Quilty BJ, Clifford S, Hellewell J, Russell TW, Kucharski AJ, Flasche S, et al. Quarantine and testing strategies in contact tracing for SARS-CoV-2: a modelling study. *The Lancet Public Health*. 2021 Jan 21.

21. Kucharski AJ, Klepac P, Conlan AJ, Kissler SM, Tang ML, Fry H, et al. Effectiveness of isolation, testing, contact tracing, and physical distancing on reducing transmission of SARS-CoV-2 in different settings: a mathematical modelling study. *The Lancet Infectious Diseases*. 2020 Oct 1;20(10):1151-60.

22. Aleta A, Martin-Corral D, y Piontti AP, Ajelli M, Litvinova M, Chinazzi M, et al. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nature Human Behaviour*. 2020 Sep;4(9):964-71.

23. Byambasuren O, Cardona M, Bell K, Clark J, McLaws ML, Glasziou P. Estimating the extent of asymptomatic COVID-19 and its potential for community transmission: systematic review and meta-analysis. *Official Journal of the Association of Medical Microbiology and Infectious Disease Canada.* 2020 Dec;5(4):223-34.

24. Pollock AM, Lancaster J. Asymptomatic transmission of COVID-19. *BMJ.* 2020; 371:m4851.

25. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Annals of Internal Medicine*. 2020 May 5;172(9):577-82.

26. McAloon C, Collins Á, Hunt K, Barber A, Byrne AW, Butler F et al. Incubation period of COVID-19: a rapid systematic review and meta-analysis of observational research. *BMJ* Open. 2020; 10: e039652.

27. Anderson R, Donnelly C, Hollingsworth D, Keeling M, Vegvari C, Baggaley R. Reproduction number (R0) and growth rate (R) of the COVID-19 epidemic in the UK: Methods of estimation, data sources, causes of heterogeneity, and use as a guide in policy formulation. *Royal Society Rapid Review*. August 24, 2020. https://royalsociety.org/-/media/policy/projects/set-c/set-COVID-19-R-estimates.pdf [accessed 31 March, 2021].

28. Tso FY, Lidenge SJ, Pena PB, Clegg AA, Ngowi JR, Mwaiselage J, et al. High prevalence of pre-existing serological cross-reactivity against severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) in sub-Saharan Africa. *International Journal of Infectious Diseases*. 2021 Jan 1; 102:577-83.

29. Wilder B, Charpignon M, Killian JA, Ou HC, Mate A, Jabbari S, et al. Modeling between-population variation in COVID-19 dynamics in Hubei, Lombardy, and New York City. *Proceedings of the National Academy of Sciences*. 2020 Oct 13; 117(41): 25904-10.

30. Giesecke J. *Modern Infectious Disease Epidemiology.* London: CRC Press; 2017.