Assumptions tested

From: Key Concepts for assessing claims about treatment effects and making well-informed treatment choices (Version 2022)

2.2d Consider whether important assumptions were tested.

Explanation

Sometimes treatment claims are based on chains of evidence, or models. For example, the effects of using a diagnostic test may depend on how accurate the test is, assumptions about what will be done based on the test results, and evidence of the effects of what is done. Similarly, evidence of the effects of public health and health system policies sometimes comes from models that combine different types of studies and assumptions; and assumptions are sometimes made when fair comparisons are combined in systematic reviews. When treatment comparisons depend on assumptions, it is important to consider their basis and to test how sensitive the results are to plausible changes in the assumptions made. For example, a model used to compare the effects of using different diagnostic tests on outcomes that are important to patients might require assumptions about what actions doctors or patients will take, based on test results. If that is uncertain, it is important to consider whether changing the assumptions has a substantial impact on the estimated difference in outcomes important to patients.

During and prior to the Covid-19 pandemic there have been few randomized trials of public health measures used to control spread of infections, such as school closures *[Glasziou 2021]*. As a result, estimates of the effects of those interventions have frequently been based on models and non-randomized studies. The modelling studies make many different assumptions and often suggest different effects. For example, some modelling studies have suggested that school closures can reduce community transmission of the coronavirus, while others disagree *[Walsh 2021 (SR)]*. These models depend on many assumptions, and changes in these assumptions can change the results. Different models make different assumptions about per-contact transmission probabilities, how many parents go to work or work at home when schools are closed or opened, changes in contacts outside of home because of schools closing or opening, what other protective measures are in place, what happens during holidays, what proportion of infected people have symptoms, how long they are infected before they have symptoms and are tested, how long the symptoms last, contact tracing, how many people without symptoms are tested, the accuracy of testing, delays in getting test results, and compliance with and effects of isolation and quarantine. Because of all these assumptions and important uncertainty about many of them, the results of these modelling studies are very uncertain.

Early in the pandemic, some assumptions were empirically informed, such as how populations are distributed spatially. However, other assumptions were seemingly anecdotal, such as an assumption that children were twice as likely as adults to transmit the coronavirus. That assumption helped justify school closures. However, subsequent epidemiological studies suggested, if anything, children may be less likely to transmit the virus *[Reddy 2020]*. In addition, some models did not consider health consequences beyond deaths from coronavirus or how social and economic consequences might affect health. Models can be helpful when there is extreme uncertainty, but it is important to recognise their limitations and uncertainty.

Basis for this concept

Many different types of models are used to estimate treatment effects. One type is marginal structural models, which are increasingly used in analyses of routinely collected data. These models take account of confounders arising during follow-up when patients switch or stop treatments, as well as baseline differences. Like all non-randomized study designs, the underlying assumption is that all relevant confounders are known, measured, and correctly integrated in the analyses. A systematic review compared treatment effects found in marginal structural model studies with those found in randomized trials for mortality and other outcomes *[Ewald 2020 (SR)]*. The review found important differences, including effects going in the opposite direction for eight of the 19 included comparisons.

New medicines are normally approved for marketing based on the results of randomized trials. A systematic review of medicines that were approved for marketing without randomized trials found that the majority of models that were used to estimate effects were based on "historical controls" (how patients were treated in the past) without any adjustment for differences in patient population (see Concept 1.2e), with a high risk of bias *[Hatswell 2017 (SR)]*.

Modelling studies combine information from a variety of sources to compare treatments. Expert judgement is often used when there is limited or conflicting evidence about a variable or "parameter" included in a model. Systematic reviews of the use of expert judgements in modelling studies in health research and in health technology assessments found extensive use of expert judgement, but most modelling studies did not provide adequate details of how expert judgements were elicited *[Cadham 2021 (SR), Grigore 2013 (SR)]*. This makes it difficult to assess the reliability of those judgements and the findings of the modelling studies. Expert judgements may be misleading due to cognitive biases, overconfidence, and the choice of experts *[Morgan 2014]*. To reduce the risk of misleading judgements, there should be a protocol for selecting experts, helping them make systematic and transparent judgements, and combining (or not combining) judgements from different experts *[Morgan 2014 , Schunemann 2019]* (See Concept 1.4c).

When direct evidence is lacking, models can be used to link together evidence of the effects of screening (see Concept 1.3e) or diagnostic tests on outcomes that are important to people *[Petitti 2018]*. However, these models also can be misleading *[Koleva-Kolarova 2015 (SR)]*. Overall, the certainty of these models corresponds to the certainty of the weakest link in the chain of evidence *[Schünemann 2019]*.

Modelling is unavoidable in evaluations of the cost-effectiveness of treatments and decision analyses [*Buxton 1997*]. However, these models can be misleading. For example, a systematic review of models assessing the cost-effectiveness of antipsychotic medication for schizophrenia found 60 models *[Jin 2020 (SR)]*. The models varied greatly, and the quality of the models was generally low due to failure to capture the health and cost impact of adverse effects and input data from the best available source.

Challenges with modelling studies include choosing which technique to use (and not making an arbitrary or biased choice), avoiding arbitrary (or biased) ranges for variables (parameters) when examining the impact of uncertainty, and making details of the model available when that is in conflict with the "intellectual property" generated by a substantial investment in developing a model *[Caro 2012]*. The trustworthiness of a model depends on transparency and validation *[Eddy 2012]*. Unfortunately, both are often lacking, making it difficult to judge how much confidence can be placed in the findings of a model. Sensitivity analyses can be used to assess the uncertainty of a model from the assumptions that are made. Sources of uncertainty include uncertainty about the

values or data used as input for each variable (parameter) in the model, uncertainty about the model (how the variables are combined), and uncertainty about how the model compares to other models using different methods). A systematic review of 406 cost-effectiveness analyses found that most analyses only addressed one of those sources of uncertainty (most often uncertainty about the variables) and that sensitivity analyses were often poorly reported *[Jain 2011 (SR)]*.

In summary, modelling studies can provide valuable information about the effects of treatments and treatment choices, but when they are used to assess the effects of treatments or to inform decisions, their reliability and uncertainty need to be carefully assessed and reported *[Briggs 2012 , Brozek 2021 , Egger 2017]*.

Implications

Whenever treatment comparisons depend on assumptions, consider whether the assumptions are well-founded and how sensitive the results are to plausible changes in the assumptions that are made.

References

Systematic reviews

Cadham CJ, Knoll M, Sánchez-Romero LM, Cummings KM, Douglas CE, Liber A, et al. The use of expert elicitation among computational modeling studies in health research: a systematic review. Med Decis Making. 2021:272989x211053794. https://doi.org/10.1177/0272989x211053794

Ewald H, Ioannidis JPA, Ladanie A, Mc Cord K, Bucher HC, Hemkens LG. Nonrandomized studies using causal-modeling may give different answers than RCTs: a meta-epidemiological study. J Clin Epidemiol. 2020;118:29-41. https://doi.org/10.1016/j.jclinepi.2019.10.012

Grigore B, Peters J, Hyde C, Stein K. Methods to elicit probability distributions from experts: a systematic review of reported practice in health technology assessment. Pharmacoeconomics. 2013;31(11):991-1003. https://doi.org/10.1007/s40273-013-0092-z

Hatswell AJ, Freemantle N, Baio G. Economic Evaluations of Pharmaceuticals Granted a Marketing Authorisation Without the Results of Randomised Trials: A Systematic Review and Taxonomy. Pharmacoeconomics. 2017;35(2):163-76. https://doi.org/10.1007/s40273-016-0460-6

Jain R, Grabner M, Onukwugha E. Sensitivity analysis in cost-effectiveness studies: from guidelines to practice. Pharmacoeconomics. 2011;29(4):297-314. https://doi.org/10.2165/11584630-000000000-00000

Jin H, Tappenden P, Robinson S, Achilla E, Aceituno D, Byford S. Systematic review of the methods of health economic models assessing antipsychotic medication for schizophrenia. PLoS One. 2020;15(7):e0234996. https://doi.org/10.1371/journal.pone.0234996

Koleva-Kolarova RG, Zhan Z, Greuter MJ, Feenstra TL, De Bock GH. Simulation models in population breast cancer screening: A systematic review. Breast. 2015;24(4):354-63. https://doi.org/10.1016/j.breast.2015.03.013

Walsh S, Chowdhury A, Braithwaite V, Russell S, Birch JM, Ward JL, et al. Do school closures and school reopenings affect community transmission of COVID-19? A systematic review of observational studies. BMJ Open. 2021;11(8):e053371. https://doi.org/10.1136/bmjopen-2021-053371

Other references

Briggs AH, Weinstein MC, Fenwick EA, Karnon J, Sculpher MJ, Paltiel AD. 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. https://doi.org/10.1016/j.jval.2012.04.014

Brozek JL, Canelo-Aybar C, Akl EA, Bowen JM, Bucher J, Chiu WA, 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;129:138-50. https://doi.org/10.1016/j.jclinepi.2020.09.018

Buxton MJ, Drummond MF, Van Hout BA, Prince RL, Sheldon TA, Szucs T, et al. Modelling in economic evaluation: an unavoidable fact of life. Health Econ. 1997;6(3):217-27. https://doi.org/10.1002/(sici)1099-1050(199705)6:3%3C217::aid-hec267%3E3.0.co;2-w

Caro JJ, Briggs AH, Siebert U, Kuntz KM. 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. https://doi.org/10.1016/j.jval.2012.06.012

Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force--7. Value Health. 2012;15(6):843-50. https://doi.org/10.1016/j.jval.2012.04.012

Egger M, Johnson L, Althaus C, Schöni A, Salanti G, Low N, et al. Developing WHO guidelines: Time to formally include evidence from mathematical modelling studies. F1000Res. 2017;6:1584. https://doi.org/10.12688/f1000research.12367.2

Glasziou PP, Michie S, Fretheim A. Public health measures for covid-19. BMJ. 2021;375:n2729. https://doi.org/10.1136/bmj.n2729

Morgan MG. Use (and abuse) of expert elicitation in support of decision making for public policy. Proc Natl Acad Sci U S A. 2014;111(20):7176-84. https://doi.org/10.1073/pnas.1319946111

Petitti DB, Lin JS, Owens DK, Croswell JM, Feuer EJ. Collaborative Modeling: Experience of the U.S. Preventive Services Task Force. Am J Prev Med. 2018;54(1s1):S53-s62. https://doi.org/10.1016/j.amepre.2017.07.003

Reddy S. How epidemiological models fooled us into trusting bad assumptions. Barrons. April 29, 2020. https://www.barrons.com/articles/the-danger-of-overreliance-on-epidemiological-models-51588179008

Schünemann HJ, Mustafa RA, Brozek J, Santesso N, Bossuyt PM, Steingart KR, et al. GRADE guidelines: 22. The GRADE approach for tests and strategies-from test accuracy to patient-important outcomes and recommendations. J Clin Epidemiol. 2019;111:69-82. https://doi.org/10.1016/j.jclinepi.2019.02.003

Schunemann HJ, Zhang Y, Oxman AD. Distinguishing opinion from evidence in guidelines. BMJ. 2019;366:l4606. https://doi.org/10.1136/bmj.l4606