Morton, Alec (2017) Treacle and smallpox : two tests for multi-criteria decision analysis models in health technology assessment. Value in Health, 30 (3). pp. 512-515. ISSN 1524-4733 ,

Treacle and smallpox: Two tests for Multi-Criteria Decision Analysis models in Health Technology Assessment

Alexander David Morton
University of Strathclyde Department of Management Science

Treacle and smallpox: Two tests for Multi-Criteria Decision Analysis models in Health Technology Assessment

Abstract

Multicriteria Decision Analysis (MCDA) is, rightly, receiving increasing attention in Health Technology Assessment. However, a distinguishing feature of the health domain is that technologies must actually improve health, and good performance on other criteria cannot compensate for failure to do so. We argue for two reasonable tests for MCDA models: the treacle test (can a winning intervention be incompletely ineffective?) and the smallpox test (can a winning intervention be for a disease which noone suffers from?). We explore why models might fail such tests (as the models of some existing published studies would do) and offer some suggestions as to how practice should be improved.

Introduction

As Health Technology Assessment (HTA) becomes more widely used in jurisdictions around the world, and for a wider range of medical technologies and services, several voices have called for the discipline to give greater recognition to the range of values which decision makers bring to bear in making decisions, alongside maximising health. In particular, there has been increased interest in Multicriteria Decision Analysis (MCDA) as a framework for making decisions, as a complement or even an alternative to the standard techniques of economic evaluation. This growth in interest can be seen both in the growth of publications on this topic in leading HTA journals over the last few years (see Figure 1 for a count of articles mentioning MCDA in the title or abstract according to Web of Science), as well as the recent publication of the ISPOR task force on MCDA[1,2] and recent reviews and advocacy [3,4].

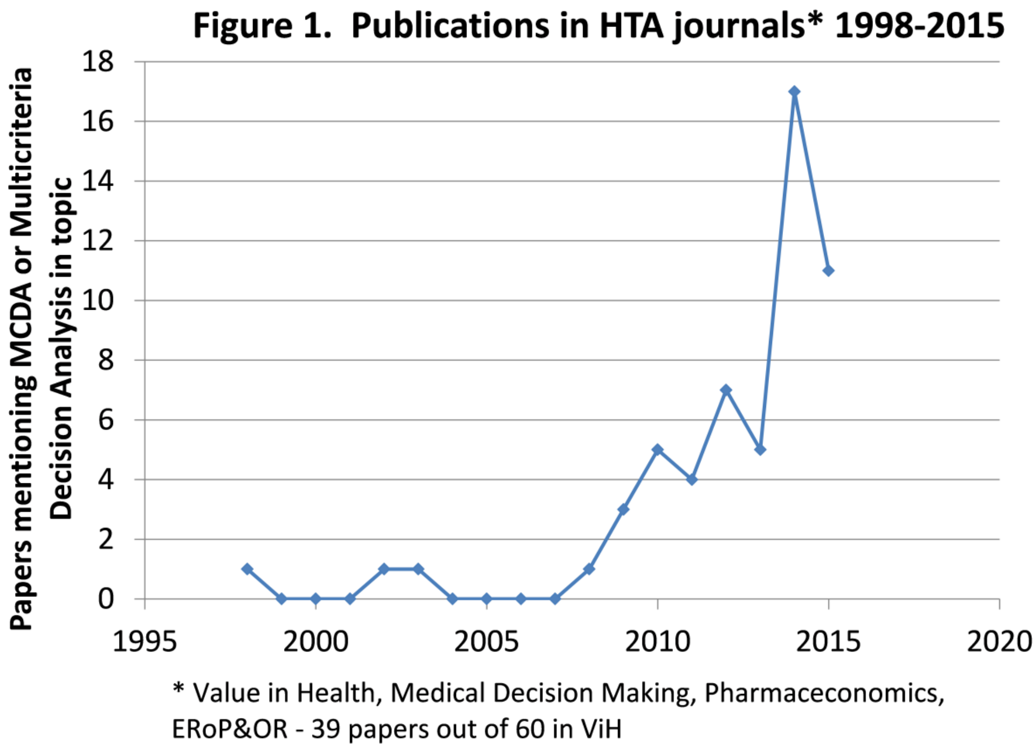

Figure 1

The attraction of MCDA is that it offers a way to include many of the values which decision makers care about beyond immediate population health gains (measured for example through QALYs) and money. It is incontrovertible that such values exist. For example: (i) enormous amounts of ink have been spilled on the existence and implications of distributional preferences for health, according to which different sorts of people are given priority, such as the young . (ii) The DCP3 project which aims to produce guides to cost effectiveness of healthcare in low- and middle-income countries uses an "Extended Cost Effectiveness" approach in which the insurance benefits of coverage are calculated [6]. This recognises the important financial protection role of the public healthcare system in many poorer countries, where many people are currently exposed to substantial financial risk in the event that they fall ill. (iii) Many stakeholders have argued that current HTA decision rules are failing to fully account for the different sources of value of new antibiotics, which should be rewarded for their innovativeness relative to the current available armamentarium, as drugs with novel mechanisms of action can be expected to hold their own against resistance [7].

Such concerns have filtered through to discussions about how HTA agencies should make decisions or decision recommendations. The Scottish Medicines Consortium, for example, considers whether certain "decision modifiers" apply in making its decisions in addition to cost-effectiveness [8]; the HTA agency for England, NICE, recently considered (and ultimately rejected) wide-ranging changes to its processes in order to formally account for non-cost-effectiveness related considerations in its decisions [9]; and an equity concept, that of "proportional shortfall" features prominently in the Dutch decision framework [10].

The core idea of MCDA is to treat these additional values in a systematic manner. Thus, roughly following the ISPOR taskforce, the main stages of the MCDA process are: defining the decision problem; identifying and structuring the criteria; measuring performance of alternatives on the different scoring; scoring the alternatives (or perhaps more accurately, scoring the performance levels of the alternatives); weighting criteria; calculating overall scores; performing sensitivity analysis; and reporting analysis to decision makers.

Of course, the existence of such values does not necessarily imply that MCDA (understood as a formal body of techniques) should necessarily have a role in the HTA process. It may be perfectly satisfactory to take these values into account informally, as part of the deliberative process which surrounds a formal analysis based on traditional CEA. While my view is that MCDA can add real value and insight, I believe that the standard CEA paradigm contains an important insight which is relevant for the structuring of MCDA models for HTA. This insight is that health improvement is not one among a number of criteria in the assessment of a medical technology: it is at the absolute core of our reasons for developing and providing such technologies.

Two tests

In this commentary, I outline two tests which I believe that any MCDA model should pass in order to qualify as fit for purpose for its use in HTA processes. These tests are as follows:

- the treacle test: it should not be possible for a completely clinically ineffective intervention (such as Venice treacle, a popular 17th century treatment for bubonic plague[11]) to get a higher overall score than an intervention which does patients some positive good, eg because the ineffective intervention is intended for severely (perhaps terminally) ill patients, children, or disadvantaged population groups.

- the smallpox test: it should not be possible for a treatment for a disease with zero global prevalence to get a higher overall score than a treatment for a disease which actually affects some people in the population (for example because the treatment is cost effective, easy to administer, or there is a convincing evidence base).

I do not believe these tests are unreasonably demanding. Yet many published MCDA models in the HTA literature would potentially fail them. For example, many MCDA models have a criterion "disease severity" which is assigned a score to be summed up with other criterion scores [12] [13] [14]. Smallpox, although declared by the UN to be eradicated in 1980, is a serious disease, with a high case fatality rate, and so on this principle, smallpox vaccination could get several points on this criterion: say the score which smallpox gets on the disease severity criterion is s and the weight associated with this criterion is w. Now consider another intervention targeted at a troublesome but non-life threatening disease which has not yet been eradicated, which receives a score on the severity criterion of $t<s$. Now as long as the aggregate weighted scores on the other criteria total less than $w(s-t)$, smallpox vaccination will get a higher score than our second intervention, which, although perhaps unattractive in many ways, is nevertheless effective against a disease which some people actually have. Similarly, Venice treacle, though now no longer regarded as an effective treatment for bubonic plague, would get points simply for being targeted at severely ill patients, and these points could balance out its therapeutic ineffectiveness.

This is not a hypothetical argument. For example, in the model of Wilson et al [15], the four interventions W, X, Y and Z have overall scores of 6.21, 5.45, 7.87 and 3.05, respectively, based on weighted scores from seven criteria, including criteria "need" (operationalised through questions including "What is the prevalence / incidence of the disease or condition this proposal is intended to treat?") and "effectiveness" (operationalised through questions including "Is the proposal proven to work?"). Thus with the original numbers of the model, option Y seems the most attractive alternative.

If the score for "need" for Y is set to zero (keeping everything else constant), then its overall value is reduced to 6.09, which is lower than the overall score of W, but still higher than the overall score of X and Z. If additional to setting the score for "need" to zero, we also set the "effectiveness" score to zero, option Y now has an overall score of 5.04, which is still higher than that of option Z. Thus the model may prioritise an intervention which is a completely ineffective treatment for a disease which noone has over a treatment which stands to yield some clinical benefit to patients.

One objection to this line of reasoning is that in practical evaluation settings, Venice treacle and smallpox vaccination are not interventions which would be realistically considered. This is to misunderstand the key point here. A decision rule which produces absurd results in cases where the answer is obvious should not be trusted as guide to wise decision making in situations which are harder to call.

A short detour into theory

MCDA applications use what I will call an "evaluation model". An evaluation model is a mathematical formula which describes how the criterion-level evaluations translate into an overall score for the objects of assessment. The evaluation model may be written down as a mathematical function such as $\sum_j w_j v_j(\cdot)$ where j indices criteria and w is a vector of weights and $v_j(\cdot)$ is a set of criterion-level scoring functions. Alternatively (but equivalently) the evaluation model may be described as set of procedures ("score options on the different criteria, weight the scores and add them up"). In principle there are an infinite number of different evaluation models (as many as there are mathematical functions). The main purpose of this note is to critique the use of the additive evaluation model $\sum_j w_j v_j(\cdot)$ in a HTA setting where the set of criteria includes criteria such as population need and clinical effectiveness, and also concepts such as equity or severity.

The additive evaluation model is a widely used and familiar evaluation model in many domains of life. To understand why the additive model produces results which seem odd in this setting we need to delve back in the underpinning theory of MCDA. As recognised by the ISPOR task force in their second paper (p 128), the additive model embeds an important preferential independence assumption

– namely how much I care about performance on one criterion does not depend on performance on the other criteria. This is surely not true for health technologies: a health technology which does not actually improve the health of some actual people is worthless, irrespective of whether it is intended for people with rare diseases, at the end of life, whether it is deployed in a particularly accessible way, or whether the quality of the evidence for its (non-) effectiveness is particularly strong.

In a properly conducted MCDA which uses a theoretically grounded weighting technique, such as swing weighting (as recommended by the ISPOR task force), in a genuinely deliberative way, failures of this independence assumption should become evident. Swing weighting, to recall, involves asking the question: "How much do you care about a worst-to-best improvement on a given criterion, all other things being equal?" If I was asked such a question about a criterion like "disease severity", my response would be "I cannot tell you how important a change in the disease severity of the target patient group would be to my evaluation of the treatment: it depends on how effective the treatment is *and how many people can benefit"*. I believe that many thoughtful individuals would respond in a similar way if they are given time to properly understand the question. Of course, it is always possible to get people to give numbers by asking vague questions [16], and/ or by simply bulldozing through respondents' doubts about what they are being asked. But this misses the point of decision analysis, which is to aid reflection and deepen understanding of a decision problem, not to produce numbers.

Possible remedies

Standard MCDA texts [17] [18] suggests that there are two approaches to situations where there is a failure of additive independence: restructure the model or choose a different functional form for the evaluation model.

- Restructuring the model. If I have an evaluation model for selection of a place to live and I have two criteria which capture closeness to park and closeness to swimming pool, I might subsequently discover a failure of preferential independence (I like to either go for a run or a swim in the morning: if my house is close to the park, I no longer care about closeness to the swimming pool). In this case, an appropriate action would be to restructure the model, replacing these two criteria with a single criterion: closeness to exercise facilities. In the HTA context, however, an example of such a restructuring might be replacing an effectiveness and a disease severity criterion with a criterion which is concerned with minimising the population shortfall relative to some reference level of health (for example one might aspire to raise the life expectancy of those suffering from some disease to within 90% of the level of the general population).

- Choosing a different functional form for the evaluation model. While there are a range of different functional forms available, beyond the additive model, one simple and practical functional form is a multiplicative one, whereby scores on different criteria are multiplied through to come up with an overall index of value (e.g. the ISafE model used in the selection of the Thai essential medicines list [19]). This seems an intuitive functional form in the health technology assessment context. The arithmetic of the multiplicative model is no harder than the arithmetic of the additive model (which, after all, also involves multiplication) and will be familiar to anyone with at least primary school education. (Indeed, the concept of the QALY as a measure of time-integrated health related well-being requires high school mathematical background but this is now well-understood and accepted.) Indeed, the practice of applying less stringent cost effectiveness thresholds for particular classes of technology can be considered as implying a multiplicative value model. For example, applying a cost-effectiveness threshold of £50,000 rather than £30,000 for life-extending treatments for patients at the end of life amounts to "upweighting" the QALYs of patients at end of life by a multiplier of 5/3.

Both these responses preserve the important intuition that there are things which do and should matter in decision making alongside health and money, but recognises that these additional considerations complement rather than substitute for health gain, which must be the central aim of any health technology.

Conclusion

The philosophy of this commentary is that decision rules should reflect and embody sound decision making principles. A sure sign that a decision rule is faulty is if one applies it to a situation where the answer is obvious and it produces the wrong result. Analysts who recommend such rules weaken their credibility – and the credibility of the entire discipline – in the eyes of decision makers.

Rather than brainstorming criteria and using scoring and weighting to construct an overall value score, MCDA practice in HTA should try to ensure that value is modelled in a way which makes sense. My key point is that clinical effectiveness and population need are not criteria on an equal footing with others: they are the foundational, in the sense that without them, there can be no value.

As a way to test out whether a particular modelling approach makes sense, two tests are offered: the treacle test (can a winning intervention be incompletely ineffective?) and the smallpox test (can a winning intervention be for a disease which noone suffers from?). An evaluation model which fails these natural and reasonable tests should be treated with great suspicion.

References

1. Thokala P, Devlin N, Marsh K, et al. Multiple Criteria Decision Analysis for Health Care Decision Making—An Introduction: Report 1 of the ISPOR MCDA Emerging Good Practices Task Force. Value Health. 2016;19(1):1-13. doi:10.1016/j.jval.2015.12.003.

2. Marsh K, IJzerman M, Thokala P, et al. Multiple Criteria Decision Analysis for Health Care Decision Making—Emerging Good Practices: Report 2 of the ISPOR MCDA Emerging Good Practices Task Force. Value Health. 2016;19(2):125-137. doi:10.1016/j.jval.2015.12.016.

3. Marsh K, Lanitis T, Neasham D, Orfanos P, Caro J. Assessing the Value of Healthcare Interventions Using Multi-Criteria Decision Analysis: A Review of the Literature. PharmacoEconomics. 2014;32(4):345-365. doi:10.1007/s40273-014-0135-0.

4. Baltussen R, Jansen MP, Mikkelsen E, et al. Priority Setting for Universal Health Coverage: We Need Evidence-Informed Deliberative Processes, Not Just More Evidence on Cost-Effectiveness. Int J Health Policy Manag. forthcoming.

5. A Williams. Intergenerational equity: an exploration of the “fair innings” argument. Health Econ. 1997;6:117-132.

6. Stéphane Verguet, Ramanan Laxminarayan, Dean T. Jamison. Universal public finance of tuberculosis treatment in India: an extended cost-effectiveness analysis. Health Econ. 2015;24:318-332.

7. Rex JH, Outterson K. Antibiotic reimbursement in a model delinked from sales: a benchmark-based worldwide approach. Lancet Infect Dis. 2016;16(4):500-505. doi:10.1016/S1473-3099(15)00500-9.

8. Scottish Medicines Consortium. SMC Modifiers used in Appraising New Medicines. http://www.scottishmedicines.org.uk/About_SMC/Policy_statements/SMC_Modifiers_used_in_Appraising_New_Medicines. Accessed 26/09/2016.

9. Kusel J. Why Has Value Based Assessment Been Abandoned by NICE in the UK? Value Outcomes Spotlight J Int Soc Pharmacoeconomics Outcomes Res. 2015;1(5):22-25.

10. van de Wetering EJ, Stolk EA, van Exel NJA, Brouwer WBF. Balancing equity and efficiency in the Dutch basic benefits package using the principle of proportional shortfall. Eur J Health Econ HEPAC Health Econ Prev Care. 2013;14(1):107-115. doi:10.1007/s10198-011-0346-7.

11. Griffin JP. Venetian treacle and the foundation of medicines regulation. Br J Clin Pharmacol. 2004;58(3):317-325. doi:10.1111/j.1365-2125.2004.02147.x.

12. Sussex J, Rollet P, Garau M, Schmitt C, Kent A, Hutchings A. A Pilot Study of Multicriteria Decision Analysis for Valuing Orphan Medicines. Value Health. 2013;16(8):1163-1169. doi:10.1016/j.jval.2013.10.002.

13. Goetghebeur MM, Wagner M, Khoury H, Levitt RJ, Erickson LJ, Rindress D. Bridging Health Technology Assessment (HTA) and Efficient Health Care Decision Making with Multicriteria Decision Analysis (MCDA) Applying the EVIDEM Framework to Medicines Appraisal. Med Decis Making. 2012;32(2):376-388. doi:10.1177/0272989X11416870.

14. Baltussen R, Niessen L. Priority setting of health interventions: the need for multi-criteria decision analysis. Cost Eff Resour Alloc. 2006;4:14. doi:10.1186/1478-7547-4-14.

15. Wilson EC, Rees J, Fordham RJ. Developing a prioritisation framework in an English Primary Care Trust. Cost Eff Resour Alloc. 2006;4:3. doi:10.1186/1478-7547-4-3.

16. A Morton, B Fasolo. Behavioural Decision Theory for Multi-Criteria Decision Analysis: a guided tour. J Oper Res Soc. 2009;60:268-275.

17. R L Keeney, H Raiffa. Decisions with Multiple Objectives: Preferences and Value Tradeoffs. Chichester: Wiley; 1976.

18. Belton V, Stewart TJ. Multiple Criteria Decision Analysis: An Integrated Approach. Boston: Kluwer; 2002.

19. RP Chongtrakul, N Sumpradit, W Yoongthong. ISafE and the evidence-based approach for essential medicines selection in Thailand. Essent Drugs Monit. 2005;34:18-19.