Positive tails and normative dogs

I whole heartedly agree with the argument Manski sets forth that that hypothesis testing has done considerable damage to the publication of research, interpretation of clinical findings and how results have informed decisions. It has also contributed to the failure to fully explore sources of heterogeneity. I also strongly agree that there is no reason to value internal over external validity given the likely heterogeneity in treatment effects when there are observed and unobserved differences between study populations and the future target patient population. The list of problems and challenges in making explicit social choices is a long one and includes all those discussed in the paper. There are others, especially the need to compare the alternatives relevant for decision making (whether nor not they are comparators in an RCT). Without it the entire effort may be of little value.

Although I also agree that CPGs have been developed and promoted without adequate consideration of these issues some bodies do understand the difficulties of founding CPGs on the type of RCTs conducted for licensing. The National Institute for Health and Care Excellence (NICE) in the UK is a good example of this. It requires methods of analysis that attempt to address the issues and inform a deliberative decision making process where these problems are key considerations when making the unavoidable judgements required in an explicit and accountable way. For example, decision analytic modelling combined with Bayesian network meta-analysis is commonly used to synthesise evidence and propagate uncertainty, irrespective of bench mark error probabilities or p-values. Similarly attempts are made to address external validity by using baseline event rates more relevant to the target population and where possible exploring whether or not relative effect is likely to be constant with respect to that baseline event rate using study level covariates or individual patient level data (IPD) where that can be acquired. Therefore, explorations of heterogeneity in the magnitude of absolute health benefits and costs are not restricted to statistically significant differences in pre-planned subgroup specific relative effects. Of course this is not to say that NICE or any other central body issuing guidance could not do better, especially if IPD were made available, or that they can observe the same granularity of patient level covariates as practicing clinicians.

Therefore, CPGs can be based on decision analytic methods appropriate to the explicit social objective of the choices they are responsible for. This disputed normative question of what the social objective of health expenditure should be provides the appropriate loss function to apply in the face of uncertainty. Inevitably the explicit characterisation of all sources of uncertainty is partial and disputed. It is conditional on alternative sets of assumptions and judgments required (e.g. for identification and beliefs about causal mechanisms). The important thing is not whether judgments, which are necessarily subjective, are required, but whether decisions made on others’ behalf are made in a way that is accountable to reasoning, evidence and (albeit disputed) social values.[[1]](#footnote-1) This avoids the comforting but false belief that empirics combined with statistical technique with their attendant ‘rules’ might provide access to the ‘truth’. It acknowledges that any analysis no matter how assiduously conducted, even with access to the best data is not ‘true’. Nonetheless it might offer a useful starting point for accountable deliberation of the relative plausibility of different views about what the costs and effects are likely to be when an intervention is made available to future patient populations who appear to be similar to historic patients, who are the only ones we can observe.

The most important advantage of CPGs is addressing the common conflict between the individual (and their clinician) at the point of care and others who have a legitimate claim on resources should they need them. In a collectively funded health care system individual choices by patients and their clinicians will necessarily have an effect on other patients because the resources required to accommodate them will not be available to offer effective care to other patients. Therefore, treatment is never individualistic and decentralised decision making that maximises individual welfare may, even with rational expectations, reduce social welfare. In many respects the primary function of centralised guidance issued by a body like NICE is to overcome this moral hazard. Although NICE does consider subgroups, whether pre-planned or not, it is reluctant to issue guidance conditional on covariates where these might offer opportunities for cost-ineffective use.

This is not only relevant to collectively funded health care systems. Patient and clinician choices in private insurance systems also increase the cost of health care for others. This imposes health opportunity costs as some are unable to access care due to higher co-payments or have their benefits package reduced due to higher premiums. These effects can be disregarded if one believes that the market for health care is undistorted so the health displaced is valued less than the health gained and that individual preferences expressed in this market reflect social welfare. If either one of these conditions fail then treatment is never individualistic and CPG may play an important role in restricting individual choice in the face of moral hazard and imperfect agency to improve social welfare overall.

Another advantage of CPGs is that it is more efficient to review, synthesise and conduct analysis of evidence centrally rather than leave clinicians to assimilate research findings. The empirical findings from psychological research support this but it would also be interesting to consider how an individual clinician might start to form rational expectations and learn from their treatment choices. This is particularly important for the argument that decentralised decision making observes more patient level covariates, which can reasonably be taken for granted. What cannot be taken for granted is how a clinician learns that certain observed covariates are associated with differences in the effect of treatment when they are unable to observe the counterfactual or patients treated by other clinicians. One mechanism would be diversification of treatment choice by individual clinicians which might narrow their bounds over time, although much more slowly than by observing all patients. Observed variation in treatment might reflect this diversification but, as noted, the variation might reflect heterogeneity in prior beliefs across clinicians. If clinicians ‘may not be willing to intentionally randomise treatment’ then it is not clear how they learn or why one should suppose they have a better understanding of the treatment interaction of a covariate even if only they can observe the covariate.

There are two remaining important aspects of the paper: i) how sharp bounds can establish partial identification and ii) what loss function ought to be applied to these bounds. The line of argument seems to run that if the application of Bayesian decision theory is not possible or too difficult, because it requires judgements which are necessarily subjective, a reasonable way to act is to apply minimax regret (MR) to sharp bounds that appear to avoid the type of explicit judgements required by a Bayesian approach. However, we have already seen that explicit or implicit judgments are unavoidable and that CPGs can be founded on Bayesian decision theory with a focus on maximising expected value (e.g., net health benefit). This ‘decision rule’ is founded on a reasonable but disputed and incomplete notion that the purpose of health care is to improve health, which also poses difficult and disputed questions of how to measure health. Some disputes about measurement of health or social values can appeal to empirical evidence of patient and community preferences, but ultimately it remains the responsibility of the decision making body and those that give it legitimacy to identify explicit objectives that offer a reasonable and accountable starting point for deliberation. Once these are ‘agreed’ it implies the appropriate loss function which can be used to consider the (health) value of reducing uncertainty by acquiring more information by conducting research, as well as determining the trade-offs between current patients who might benefit from early access to a new intervention and future patients who would benefit from delayed access so the research can be conducted or conducted more quickly.

The application of Bayesian decision theory has enabled the value of additional evidence to be considered, including: optimal sample size; sequences and portfolios of research; whether or not widespread access to new technologies should be delayed until more research is conducted; how the presence irrecoverable costs might influence decisions; and the link between the need for evidence and the price of a new technology. The distinction between the value of information that identifies sources of heterogeneity and information that resolves uncertainty given what is known about heterogeneity has been explored and applied. Coupled with this have been greater efforts to reflect all sources of uncertainty (e.g., ‘structural’ uncertainty), including the use of formal elicitation methods. Therefore, Bayesian decision theory and the implications of uncertainty for decision making has been a rich and useful seam of work that avoids the arbitrary bench marks of hypothesis testing, critical error probabilities, the effect size of the power calculation, or the maximum regret recommended by Manski. It ought not to be dismissed lightly.[[2]](#footnote-2)

Nevertheless, there is real value in the greater appreciation and use of sharp bounds when considering how distributions might be assigned to uncertain parameters. The choice is whether to assign them explicitly or implicitly. For example, the MR choice is the same as maximising expected utility using the midpoint of the interval, i.e., the bounds have implicitly assigned a uniform distribution to the unknown parameter once we apply the objective function used to construct the MR rule. The combination of sharp bounds and MR implies a judgement about distribution – it is unavoidable. If a more reasonable view of probability within the bounds was not uniform then the mean may differ leading to the ‘wrong’ decision, i.e., not consistent with an accountable and legitimate judgement. But even if the distribution was symmetrical there would be too much weight in the tails so the cost of uncertainty and the value of acquiring more evidence would be overestimated risking poor research decisions. The inevitable need for judgement is also illustrated when assumptions are used to tighten bounds. It implies two things: i) probability was not uniform within the original bounds and ii) that these assumptions themselves are uncertain unknown parameters which require a prior distribution to be assigned.

Choosing bounds and MR over maximising expected value of an explicit objective function would avoid placing prior distributions on parameters, i.e., the latest posterior distribution from the accumulating evidence. However, why, on normative grounds, maximising an agreed explicit social objective function (which could include elements of risk aversion), should be rejected in favour of an MR decision rule which implies a different objective function is unclear. I fear that positive issues (the usefulness of making greater use of sharp bounds) combined with using MR to avoid the logic of Bayes is a ‘positive tail’ that is in danger of wagging the ‘normative dog’.[[3]](#footnote-3) For social choices made on others behalf the normative must come first and wag the tail of appropriate analysis.

1. See: Chapter 2 of Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes. 4th ed. Oxford: Oxford: Oxford University Press, 2015 for a discussion of the disputed social objectives of health expenditure and the role that economics ought play in decisions about health care; Chapter 4 for how these normative positions determine the type of analysis that would be appropriate; and Chapter 11 for an overview of how clarity about social objectives enables an appropriate loss function to be applied to our necessarily incomplete attempts to characterise all sources of uncertainty when informing decisions which requires making predictions about an unknown future rather than describing the past (the observed sample). [↑](#footnote-ref-1)
2. See Chapter 11 of Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes. 4th ed. Oxford: Oxford: Oxford University Press, 2015, which provides an overview of these efforts and points to a number of the important published contributions in this area. [↑](#footnote-ref-2)
3. This is a reference to an excellent paper; Culyer, A.J., Evans, R.G., 1996. Mark Pauly on welfare economics: normative rabbits from positive hats, Journal of Health Economics 15, 243-25, which identifies a similar problem in a different context. This seminal paper includes the instruction from Keynes in 1932 to inform rather than prescribe social choice; “If economists could manage to get themselves thought of as humble, competent people, on a level with dentists, that would be splendid! This and many other splendid papers can also be found in Chapter 4 of Cookson RA, (ed.), Claxton KP, (ed.), Culyer T. The Humble Economist: Tony Culyer on Health, Health Care and Social Decision Making. York and London: University of York and Office of Health Economics, 2012, which is a free down load at <https://www.york.ac.uk/che/publications/books/the-humble-economist/> [↑](#footnote-ref-3)