Bright line or lottery? On significance and value in medical decision making

Jörg Mahlich and Srirangan Dheban

*Miltenyi Biomedicine, Bergisch Gladbach, Germany & Dice - University of Düsseldorf, Düsseldorf, GermanySrirangan Dheban, Zurich, Switzerland

ABSTRACT

It is widely acknowledged that using p-value thresholds as the basis for making decision on health care spending is not appropriate. In the context of medical decision making, we argue that patient preferences need to be a stronger factor. Depending on attitudes to risk, patients might prefer a medical treatment that performs on average worse than a comparator but offers a small probability of a large gain such as a cure. However, what has been labeled ‘value of hope’ is not yet fully reflected in the decision-making process of drug approval and health technology assessment (HTA). Therefore, patient risk preferences should be formally incorporated within the decision-making framework for regulatory and reimbursement decisions.

Introduction

In this short communication, we propose to combine two ongoing academic discussions that impact medical decision making: one circle around the role of statistical significance in policy decision making and the other is about how to determine value of medical treatments. It is widely acknowledged that using a certain p-value to determine whether a new treatment works better than a defined standard of care is highly misleading. In 2016, the American Statistical Association published a white paper on the use of p-value and argued that “Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold. Practices that reduce data analysis or scientific inference to mechanical ‘bright-line’ rules (such as ‘p < 0.05’) for justifying scientific claims or conclusions can lead to erroneous beliefs and poor decision making. A conclusion does not immediately become ‘true’ on one side of the divide and ‘false’ on the other. Researchers should bring many contextual factors into play to derive scientific inferences, including the design of a study, the quality of the measurements, the external evidence for the phenomenon under study, and the validity of assumptions that underlie the data analysis. Pragmatic considerations often require binary, ‘yes-no’ decisions, but this does not mean that p-values alone can ensure that a decision is correct or incorrect. The widespread use of ‘statistical significance’ (generally interpreted as ‘p = 0.05’) as a license for making a claim of a scientific finding (or implied truth) leads to considerable distortion of the scientific process” [1]. Following a general acceptance of this view amongst statisticians, it has since been argued ‘the p-value (should) be demoted from its threshold screening role and instead, treated continuously, be considered along with currently subordinate factors’ [2].

When it comes to regulatory and reimbursement decisions, one factor that is currently considered as inferior to statistical significance is patient preference, although authorities such as the FDA are currently developing guidelines for patient-focused drug development (PFDD) that should ensure that data relating to patient and caregiver experience are heard in regulatory decision making [3]. The European counterpart of PFDD is PREFER (Patient Preferences in Benefit-Risk Assessments during the Drug Life Cycle), a public-private collaborative research project under the Innovative Medicines Initiative (IMI) with an involvement of the EMA and other relevant stakeholder [4].

Attitudes towards risk and medical decision making

One excellent example why looking at statistical significance alone is misleading and results in poor policy decisions is the field of drug approval and HTA. Here, appraisals often serve as the foundation of pricing decisions and, hence, are used to determine value. However, saying that drug A performs significantly ‘better’ in a set of endpoints than drug B does not necessarily mean that drug A would bring higher value for patients or society at large. To make
any inference about the value of one treatment relative to another, it is important to consider patient preference. Using Neumann–Morgenstern’s rigorous axiom-based economic utility theory [5], decisions can be analyzed that involve risky (probabilistic) outcomes. Individuals can be classified as risk averse, risk neutral, and risk seeking depending on their preferences to take or avoid risks: A risk averse individual for instance would prefer to receive €2 for certain instead of receiving a lottery ticket that pays out €1 and €3 with equal probability of 50% (i.e., expected value of €2), indicating a concave utility function. Concave utility functions imply risk aversion, e.g., to be indifferent between the lottery and the certain outcome, the expected value of the lottery needs to be higher than the certain outcome to compensate for the risk, with the difference between the two called risk premium. A risk neutral individual is indifferent between the two options and his or her utility function is linear. A risk seeking individual would prefer the lottery over the certain outcome. The risk premium is negative (convex utility function), indicating that he or she would be willing to pay a price to participate in the lottery even though the expected net outcome (expected value minus lottery price) is negative. This is because the individual places a high value on the opportunity of receiving a high gain.

In health economics, a preference for risk results in what is termed ‘value of hope’ [6,7]. Business models for lotteries exploit the risk proclivity of consumers. With annual growth rates of 10% and projected sales of 353 bn USD by 2026 [8], the global lottery market indicates the presence of a high number of risk-seeking individuals.

Risk attitudes differ by age, gender, socioeconomic status, cultural factors, and even body height. Male gender, higher parental education, and increasing height are positively correlated with willingness to take risks [9]. As lottery games such as standard gamble (SG) are used to elicit utility values (which are the basis of any cost utility analysis), utilities are influenced by the same set of variables [10–13]. The SG is a method of valuing peoples’ preferences. It has its theoretical basis in the von Neumann–Morgenstern axioms of expected utility theory and aims at measuring the ‘disutility’ of a health state by observing an individual’s willingness to accept a certain risk of death in order to avoid the health state. The term ‘skewness preferences’ describes the fact that most people prefer right-skewed risks (which feature a small probability of a severe positive impact) and avoid left-skewed risks (which feature a small probability of a severe negative impact). This can explain why people at the same time pay for insurance against left-skewed risks that yield a rather large loss with a small probability and seek right-skewed risks such lotteries that are defined by large gain with a small probability [14].

In Figure 1, the concept of the ‘value of hope’ is illustrated. The expected outcome – for instance in terms of median overall survival – is 28 months for the active treatment (A) and 32 months for placebo (P). Still, a risk-taking patient prefers the active treatment over placebo although the expected average outcome is significantly worse. This is because the ‘fat tail distribution’ (i.e., greater probabilities of events producing outsized gains and losses) of the active treatment offers sufficient opportunities (even though small) that are of value for patients. This value is reflected in a positive willingness to pay for treatment A over P. Such type of decision making applied to patient preferences can be very well underpinned by Bayesian analysis to make probabilistic statements about how likely it is to achieve an improvement of a medical condition larger than a certain threshold [15]. The active treatment distribution in Figure 1 for example offers a chance of surviving at least 40 months of 21%, while the respective probability is only 2% for the placebo arm. This could be especially interesting for patients failing to respond to therapies and being left with no or less attractive alternatives as all these alternatives perform similarly to the previous treatment on average. In such cases, a probabilistic comparison of the less attractive choices could aid to choose the one that maximizes the probability to improve the medical condition dramatically relative to the other alternatives. However, most regulatory agencies would probably not approve an active treatment if the evidence suggests it is not statistically superior to placebo, a decision that would violate preferences of risk-taking patients. Empirical studies concluded that patients with chronic diseases are risk averse [16], while incurable patients are risk seeking [17] and therefore would benefit most from fat tail distributions. The policy implication of this argument is that is that decision makers should not allocate medical resources in a way that maximizes the number of aggregated live years but aggregated utilities. In his social aggregation theorem, Harsanyi (1955) showed that individual von Neumann–Morgenstern utilities can be aggregated to a social welfare function if the ex-ante Pareto principle is met [18]. This assumption asserts that if all agents prefer one allocation of prospects to the other; then, the social planner also prefers it [19].

**Conclusion**

Patient preferences should be systematically considered in regulatory and HTA decision making. Statistical significance tests of relative clinical benefits should be reported along with a range of measures to value the additional benefit of a treatment, but not serve as a sole basis for making a decision on whether to adopt, or how to
reimburse, a technology. Among many attributes that are relevant for determining addition ‘value’, those attributes can be holistically viewed and summarized using multi criteria decision making techniques [20]. Our proposal has far reaching consequences. As risk seeking patients might prefer drugs that perform on average worse than a comparator but offer some small probability of a huge gain, those preferences need to be reflected in pricing decision to make sure that the pharmaceutical industry channels R&D investments into those projects that maximize patients’ value. Our suggestions relate to a broader debate around including patient perspectives in decision making [21,22].

Figure 1. Illustration of ‘value of hope’.

Disclosure statement
JM may own stock/stock options of JNJ. SD does not report any conflict of interest.

ORCID
Jörg Mahlich @ http://orcid.org/0000-0003-1110-2793

References
[1] Wasserstein R, Lazar N. The ASA statement on p-values: context, process, and purpose. Am Stat. 2016;70 (2):129–133.
[2] McShane B, Gal D, Gelman A, et al. Abandon statistical significance. Am Stat. 2019;73(sup1):235–245.

[3] Food and Drug Administration. Patient-focused drug development: collecting comprehensive and representative input. Silver Spring: FDA; 2020. Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders. Available from: https://www.fda.gov/media/139088/download

[4] Patient Preferences in Benefit-Risk Assessments during the Drug Life Cycle (PREFER). Including the patient perspective. [cited 2021 Aug 20]. Available from: https://www.imi-prefer.eu/about/

[5] von Neumann J, Morgenstern O. Theory of games and economic behavior. London: Wiley; 1944.

[6] Lakdawalla DN, Romley JA, Sanchez Y, et al. How cancer patients value hope and the implications for cost-effectiveness assessments of high-cost cancer therapies. Health Aff (Millwood). 2012;31(4):676–682.

[7] Garrison L, Jansen J, Devlin N, et al. Novel Approaches to Value Assessment Within the Cost-Effectiveness Framework. Value Health. 2019;22(6):S12–S17.

[8] Researchdive 2020. Lottery market report. Available from: https://www.researchdive.com/154/lottery-market

[9] Dohmen T, Falk A, Huffman D, et al. Individual risk attitudes: measurement, determinants, and behavioral consequences. J Eur Econ Assoc. 2011;9(3):522–550.

[10] Sellers S, Cherepanov D, Hanmer J, et al. Interpersonal discrimination and health-related quality of life among black and white men and women in the USA. Qual Life Res. 2013;22:1307–1312.

[11] Obaidi LA, Mahlich J. A potential gender bias in assessing quality of life - a standard gamble experiment among university students. Clinicoecon Outcomes Res. 2015;7:227–233.

[12] Robert SA, Cherepanov D, Palta M, et al. Socioeconomic status and age variations in health-related quality of life: results from the national health measurement study. J Gerontol B Psychol Sci Soc Sci. 2009;64:378–389.

[13] Mahlich J, Dilokthornsakul P, Sruamsiri R, et al. Cultural beliefs, utility values, and health technology assessment. Cost Eff Resour Alloc. 2018;16:19.

[14] Dertwinkel-Kalt M, Köster M. Salience and skewness preferences. J Eur Econ Assoc. 2020;18(5):2057–2107.

[15] Fisch R, Jones I, Jones J, et al. Bayesian design of proof-of-concept trials. Ther Innov Regul Sci. 2015;49:155–162.

[16] Lakdawalla D, Malani A, Reif J. The insurance value of medical innovation. J Public Econ. 2017;145:94–102.

[17] Shafrin J, Schwartz TT, Okoro T, et al. Patient versus physician valuation of durable survival gains: implications for value framework assessments. Value Health. 2017;20(2):217–223.

[18] Harsanyi J. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. J Political Econ. 1955;63(4):309–321.

[19] Miyagishima K. Efficiency, equity, and social rationality under uncertainty. Econ Theory. 2021. DOI:10.1007/s00199-020-01335-4.

[20] Angelas A, Kanavos P. Multiple Criteria Decision Analysis (MCDA) for evaluating new medicines in health technology assessment and beyond: the advance value framework. Soc Sci Med. 2017;188:137–156.

[21] Wale J, Scott A, Hofmann B, et al. Why patients should be involved in health technology assessment. Int J Technol Assess Health Care. 2017;33(1):1–4.

[22] Hunter A, Facey K, Thomas V, et al. Guidance for patient involvement in medicines research and development: health technology assessment. Front Med (Lausanne). 2018;5:231.