MAIMON WORKING PAPERS No. 11 AUGUST 2021

VETERANS AFFAIRS, THE INSTITUTE FOR CLINICAL AND ECONOMIC REVIEW AND GENERIC HEALTH STATE EUGENICS

Paul C Langley, Ph.D., Adjunct Professor, College of Pharmacy, University of Minnesota, MN

Abstract

To link eugenics to health care delivery is by no means a long stretch. The denial of care (and even more disturbing interventions) based upon a perception of the societal 'worth' of a health state experienced by a patient has a long history with interventions to deny care, often enthusiastically endorsed by the medical profession. The Institute for Clinical and Economic Review (ICER) is firmly in this eugenics tradition: the worth of an intervention, based upon generic societal preferences for health states is determined by preference scores to model quality adjusted life year (QALY) simulations that fail the standards of normal science. In common with eugenics, ICER bases its recommendations on a commitment to pseudoscience. The acceptance of ICER pricing and access recommendations by agencies such as Veterans Affairs (VA) raises concerns as whether they recognize the implications of the ICER analytical framework to determine the 'value' or 'worth' of a health state. This is more of a concern given the nature of the VA treating population where a more holistic framework of analysis defining need as a measure of quality of life in health care intervention response can be assessed.

INTRODUCTION

A recent commentary has pointed to the potential for the denial of care when societal preferences for a predetermined set of health symptoms and response levels are applied to health states. In the case of the EQ-5D-3l and EQ-5D-5L which are the most commonly used preference scores for health states, agencies such as the VA factor these into decision making while failing to recognize their manifest and fatal shortcomings. These criticisms cannot be put to one side as a one-off criticism. Unless there is a concerted and continuing effort at re-education, then the current paradigm for the construction of evidence to support formulary decisions will continue. Ongoing and repeated criticism is essential if we are to overturn the imaginary information paradigm; it has to be repeated to convince the various audiences of these issues. Veterans Affairs (VA) is a case in point. They apparently favor the analytical framework proposed by the Institute for Clinical and Economic Review (ICER); indeed they have a long standing collaborative relationship. This sets up a barrier to new ideas; to a paradigm shift in value assessment that abandons the creation of cost-per-QALY imaginary simulations and blanket claims for cost-effectiveness in favor of a more nuanced value assessment and one that recognizes the standards of normal science. The task is made more difficult by an often all too lack of appreciation of the standards of normal science and measurement theory.

The purpose of this commentary is to put the arguments for the VA abandoning the ICER imaginary evidence paradigm in the context of unfortunate parallels between societal preferences for health states as evidenced by pseudoscientific direct and indirect multivariable instruments and the pseudoscience of eugenics.

ICER AND THE VA

The VA is clearly wedded to the standard technology assessment meme; it appears, in two recently released reports on health services research (the QUERI Guidelines: VA Quality Enhancement Research Initiative) and the techniques of cost-effectiveness analysis. There is no apparent understanding of fundamental measurement or, indeed, of the standards of normal science as limitations implicit in the ICER approach to health technology assessments [1] [2]. The VA appears unaware that the ICER analytical framework fails the standards of normal science, in particular the axioms of fundamental measurement. In common with ICER and the wider audience served by the International Society for Pharmacoeconomics and outcomes Research (ISPOR) the commitment by the VA is to the invention of evidence through assumption driven imaginary simulations to plus evidence gaps and support claims for pricing and access. Of course, The VA defends is ongoing collaboration with ICER by making the obvious point that ICER is only one source , in this case of imaginary information, to support decisions by the VA Pharmacy Benefits Management Services (PDM) (a blog jointly written with ICER)[3]. It is difficult to see how imaginary evidence can play any role in formulary evaluations.

It is doubtful if ICER generic preferences supporting imaginary claims for therapy response, pricing and access are of any use to the VA, in fact given their dubious provenance they may be positively misleading. The unique characteristics of the VA population are well known in terms of demographics, social background, disease prevalence and comorbidities[4]. A group, it might be argued, to which a generic assessment of health state societal preferences and their 'worth' is singularly inappropriate when defined by limited clinical symptoms and response levels. If the VA is to best serve the interests of its patients, disease specific outcomes instruments, as long as they meet required measurement standards, are required. Given the apparent cozy relationship with ICER, the VA PBM will no doubt reject this recommendation. After all, ICER has nothing to offer in modeling simulations with outcomes assessments specific and relevant to these various VA target populations. A one size fits all modeling can apparently meet the VA PBM requirements for evaluating imaginary therapy response, irrespective of the lack of coherence in assessing response to therapy.

THE WORTH OF A HEALTH STATE

Preference scores that are an integral part of the creation of quality adjusted life years (QALYs) are generated from algorithms that apply preference weights (sometimes called TTO tariffs) to questionnaire responses. Patients are asked to complete a short questionnaire, in the case of the EQ-5D-3L/5L one with 5 symptoms or attributes (mobility, self-care, usual activity, pain/discomfort, anxiety/depression) which have respectively 3 and 5 response levels (ranging from no problem to extreme problem). Each of the tariffs or preference weights is derived from a community sample who have been asked to value or determine the worth of a set of health state descriptions. These form the basis for the assigned societal preference weights (15 in the case of the EQ-5D-3L and 25 for the EQ-5D-

5L). It should be noted that the societal tariffs are ordinal scores and that the resulting composite score across the attributes is also an ordinal score and one which fails to meet the standards required by the axioms of fundamental measurement. The aggregate societal weight is 'generic' in the sense that the questionnaires are intended to apply to respondents in any health state by stage of disease, yet irrespective of the relevance of particular attributes to that health state. The result is a 'worth' score with implications for society's willingness to allocate resources to health care by disease area and stage of disease. While generic preference scores have received considerable criticism, they continue to be applied widely as there is, for those wedded to the dominant health technology assessment paradigm, nowhere else to go. Presumably ICER and the VA PBM believe that we must have, from a central planning resource allocation perspective within the VA health system, generic preferences and cost-per-QALY estimates of the worth of health states to guide and deny resource use. Under a fixed health care VA budget resources must be directed to where, on the imaginary cost-per-QALY utilitarian calculus, they can maximize the 'worth' of the allocation of limited resources through social pricing and access recommendations. This is simplistic and palpable nonsense, although the VA PBM appears to have bought in to this 'eugenic' view of decision making in health care to prioritize health states.

But we have been here before in the numerous attempts over the past 30 years to create formal criteria for prioritizing (i.e., assessing the worth) of health care interventions. In the US, the ultimately unsuccessful ,Oregon Medicaid experience is the best known while outside of the US the experience to establish criteria for national health systems has been similarly unsuccessful. As Sabik and Lie noted in an early review of the experience in eight countries: ….*there is little evidence that establishment of a framework for priority setting has had any effect on health policy, nor is there any evidence that priority setting exercises have led to the envisaged idea of an open and participatory public involvement in decision making* [5]. In the intervening years little has changed with a more recent review of the Norwegian National Council for Priority Setting concluding that '…while the Council often made use of the official priority-setting criteria they did so in an unsystematic and inconsistent manner' [6] .

Perhaps the VA PBM might acknowledge that in collaborating with ICER in the invention of simulated claims, it is merely chasing a will o'the wisp, as seen in the UK with the National Institute for Health and Care Excellence (NICE). As spider in the National Health Service (NHS) web NICE mandates the construction of imaginary claims with pricing and access driven by imposed cost-per-QALY thresholds. NICE is perfectly aware of the fact it is creating imaginary simulations (supported by academic review centers which have devoted years to the honing of their skills in the assessment of imaginary simulated claims); its defense is that it can think of no other option; a position, of course, echoed by its camp follower ICER. While NICE mandates the creation of imaginary claims, there seems no reason why the VA PBM should emulate this charade by even considering imaginary claims as part of its formulary decision process.

HEALTH STATES WORSE THAN DEATH

As originally intended, the development of both direct and indirect preference instruments was to generate health state scores on a scale anchored with a clinically determined score of unity for 'perfect

health' with no problems with any of the five attributes and 'death' with a worth of zero. The scores were to be created as decrements from the cap of unity, culminating in death at the zero point. Unfortunately, all instruments overshot their target, creating negative scores or 'states worse than death'. From the perspective of determining the 'worth' of a health state, users of this scoring algorithm have a few options with 'states worse than death' being considered of 'no worth' or states worse than a minimum positive score (e.g., a score of 0.10 in the range 0 to 1). In the case of the EQ-5D-5L we would have a range of health states to choose from. A recent attempt to apply societal preference weights for the US with the EQ-5D-5L found that of the possible 3125 health states (5^5) some 624 (20%) had negative scores, assigned by societal preferences to the 'health states worse than death' category. Should these be considered a eugenics target population for denial of health care? How many VA patients would, on completion of the EQ-5D-5L, be categorized as experiencing a state worse than death. This has interesting implications for therapy decisions and healthcare resource allocation within the VA.

EUGENIC RANKINGS OF WORTH

Care has to be taken in interpreting the societal preference scores for the 'worth' of a health state. The key point is that these scores are ordinal; they can be ranked. We have no idea of the 'distance' between the individual ranked scores which means any assessments are restricted to non-parametric statistics. An ordinal score cannot support the standard arithmetic operations of addition, subtraction, multiplication and division. We cannot compute averages and standard deviations, nor can we crosswalk the EQ-5D-5L to the EQ-5D-3L or even create QALYs as these require the ability to multiply time spent in a disease state by a societal weighted preference score for respondents in that health state; a ratio measure.

Even so, the ability to rank health states by societal valuations has problematic implications. First, it is unusual to find studies utilizing a multiattribute instrument to actually present the ranking of respondents, let alone information as to what proportion have negative scores or scores less than a given value. The presence of negative scores is 'hidden' in average score claims. Second, it is even more unusual for those utilizing (or even criticizing) the role of QALYs in health care decisions to ask for (or even be aware of) the distribution of ordinal scores, to include both positive and negative values. Third, the presence of negative scores will depress the overall ordinal score compared to the case where there a no negative values. This may seem a trivial point, but it will also depress the imaginary QALY values, with the potential to inflate the ICER claim for imaginary cost-per-QALY price discounting and more adverse access recommendations. In disease areas where the stage of disease may be associated with an increasing proportion of states worse than death, the average preference score for that stage will be low (even negative) so that the cost- per-QALY will increase leading to a higher probability of societal determined denial of care.

There is no evidence to suggest that the VA has addressed the question of 'states worse than death', asking ICER for a distribution of the ordinal composite preference scores. Although these composite ordinal scores fail the axioms of fundamental measurement (a point typically overlooked) it would be instructive to consider these distributions for the disease states (including comorbid states) present in

the VA patient population. The VA PBM staff might usefully consider the implications of a significant proportion of negative preference scores (or negative 'worth' scores) in target VA populations for formulary decisions. Does this imply a segmentation of target VA patients in disease states by a preference score? Are some less 'worthy' of care than others? Would the VA deny care?

ABANDONING EUGENICS

The VA's long terms association or collaboration with ICER raises a number of uncomfortable questions: to what extent do the VA PBM staff subscribe to the invention of evidence to support formulary decisions? To what extent does the VA PBM staff explicitly reject the standards of normal science and the axioms of fundamental evidence? Are the PBM staff aware of the eugenics undertones in applying societal preference weights to establish the 'worth' of a health state? Have the VA PBM staff ever inquired as to the distribution of preference scores in VA target populations and the prevalence of health states determined as 'worse than death'? Are the VA PBM staff aware that in constructing imaginary assumption driven simulation to establish price and access claims that the standards of normal science, all too evident in product development, are absent?

The perception, unfortunately, is that the ICER analytical framework, the invention of evidence for assumption driven lifetime simulations, is the soft and easily accepted VA PBM option. Rather than focusing on research programs to capture real world, yet provisional evidence, of the impact of therapy interventions in target VA treating populations. In all honesty, rejecting a health technology assessment meme that has dominated therapy evaluations for 30 years is difficult. But we have to start somewhere; there are, as demonstrated in Version 3 of the Minnesota Guidelines, questions a formulary committee must address is a submission for new products is to be realistically evaluated [7]. Perhaps the VA PBM might emulate this?

PATIENT AND CAREGIVER NEED

The VA, from evidence available, seems singularly unconcerned with any attempt to calibrate the need of patients and, where appropriate, caregivers for the VA patient population. This is not to ignore the VA Caregiver Support Program, but to emphasize the need to assess the impact of therapy interventions on the need fulfilment quality of life of target patient groups [8]. To achieve this requires a more holistic and patient centric approach to QoL that puts to one side the societal preference approach to assessing clinically determined health related quality of life, through ordinal preference scores. The focus is on need fulfillment: what is need fulfillment for VA patients and their caregivers (who may also be VA patients) in disease specific target patient groups? To what extent is this need, determined by extensive interviews with the target patient population within disease areas, fulfilled? To what extent can new interventions support a greater need fulfillment? ICER cannot answer this question. It would require a concerted effort to develop need fulfillment instruments for specific disease areas. We are well on the way to achieving this with instruments developed through the application of Rasch Measurement Theory over the last 25 years. It is only recently that the caregiver has received attention with the assessment of need of spousal caregivers in Alzheimer's disease [9]. This model could be applied to the VA population.

At the same time it should be noted more recently, an algorithm has been developed to translate these scores to a bounded ratio scale of need fulfillment: the need or N-QOL scale [10]. This allows a quantitative assessment of the extent to which the need of spousal caregivers are impacted by the introduction of new therapies; a value claim which provides a robust assessment of response to therapy. The transformation algorithm can be applied to over 30 disease states with already developed instruments to assess response to therapy, all of which have relevance to evaluating need fulfillment in VA target patient populations.

CONCLUSIONS

Judged by the standards of normal science and the axioms of fundamental measurement, the VA PBM has a long way to go to meet the required evidence standards for evaluating response to therapy for its membership. It seems wedded to outdated ideas. As a first step it should put ICER to one side. The ICER approach is thoroughly discredited and should not be considered even as one of many inputs to VA PBM decisions. Unfortunately, you are judged by the company you keep and ICER is a liability.

Irrespective of the criteria a formulary committee, such as the VA PBM group, may wish to apply in evaluating and negotiating for product pricing and access, and whether they refer to the clinical characteristics of the product (including comparators), quality of life or resource utilization they must be consistent with the standards of normal science as single attributes that are credible, evaluable and replicable. This means that they should meet required measurement standards for ratio or interval scales, demonstrating dimensional homogeneity or unidimensionality. All claims made for a product should be accompanied by a protocol to show how each claim is to be evaluated within a meaningful timeframe for the target VA population. Obviously, ersatz preference , QALY and cost-per-QALY claims based on assumption driven simulations, with their implications for the 'worth' of health states and product impact are out of the question for obvious reasons. . If a manufacturer wishes to make a claim for 'overall' cost-effectiveness then all inputs to that claim must meet the standards of credibility, empirical evaluation and replication, with a protocol demonstrating how this claim for cost-effectiveness is to be assessed following market entry.

The VA needs to decide whether or not it will continue to stand by and make decisions based on a discredited analytical framework for claiming imaginary cost-effectiveness by inventing evidence from assumption driven lifetime simulation models or propose new guidelines for product assessment that recognize the standards of normal science. There are more robust frameworks that focus on claims for specific clinical, QoL and resource utilization claims. There is no single magic score for cost-effectiveness that can be interpreted to deny access to health care. We need, finally, to put eugenics and estimating the societal worth of health states behind us, denying any attempt to sup with the devil.

Is this likely to occur with the VA PBM? We are dealing with an inventing evidence meme that has been firmly entrenched in the technology assessment literature for over 30 years. The outlook is bleak. Such beliefs are difficult to overturn; particularly when they involve rejecting previous studies and recommendations. Indeed, the more impossible a belief, a mystery, the more support it attracts. It is far easier to follow the ICER mantra and argue that it is what everyone else does; a standard defense of

burning witches in the 17th century. We should not hold our breath. Portugal, as a case in point, continued with the public burning of heretics, or those that do not agree with ICER, well into the mid-18th century.

REFERENCES

[1] US Department of Veterans Affairs. Cost-Effectiveness Analysis, June 2021 https://www.herc.research.va.gov/include/page.asp?id=cost-effectiveness-analysis

[2] US Department of Veterans Affairs. Health Services Research & Development Service QUERI Economic Analysis Guidelines. March 2021. https://www.herc.research.va.gov/include/page.asp?id=queri-economic-analysis-guidelines

[3] Glassman P, Pearson S, Zacher J et al. VA and ICER at three years: Critics concerns answered. *Health Affairs*. Blog. June 15, 2020 https://www.healthaffairs.org/do/10.1377/hblog20200611.662048/full/

[4] RAND Corporation: Current and projected characteristics and unique health care needs of the patient population served by the Department of Veterans Affairs. *RAND Health Quarterly*. 2016;5(4)

[5] Sabik L, Lie R. Priority setting in health care: Lessons from the experience of eight countries. *Int J Equity Heath*. 2008;7(4)

[6] Wester G, Bringedal B. The Norwegian National Council for Priority Detting in Health Care: decisions and justifications. *Health Econ Policy Law*. 2018;13(2):118-136

[7] Langley P. Value Assessment, Real World Evidence and Fundamental Measurement: Version 3.0 of the Minnesota Formulary Submission Guidelines. *InovPharm*. 2020;11(4):No 12 https://pubs.lib.umn.edu/index.php/innovations/article/view/3542/2613

[8] US Department of Veterans Affairs. VA Caregiver Support Program. https://www.caregiver.va.gov/Tips_by_Diagnosis/PTSD.asp

[9] McKenna S, Rouse M, Heaney A et al. International development of the Alzheimer's Patient Partners Life Impact Questionnaire (APPLIQuE). *Am J Alzheimer's Disease & other Dementias*. 2020;35:1-11

[10] Langley P. McKenna S. Fundamental Measurement: The Need Fulfilment Quality of Life (N-QOL) Measure. *InovPharm*.2021;12(2):No. 6 https://pubs.lib.umn.edu/index.php/innovations/article/view/3798