MAIMON WORKING PAPERS No. 2 FEBRUARY 2024

SUBMISSION TO THE AUSTRALIAN HEALTH TECHNOLOGY ASSESSMENT POLICY AND METHODS REVIEW [February 2024]

REJECTING THE PBAC GUIDELINES FOR IMAGINARY COST-EFFECTIVENESS VALUE CLAIMS: A PROPOSED NEW START FOR HEALTH TECHNOLOGY ASSESSMENT IN AUSTRALIA

Paul C Langley, Ph.D., Adjunct Professor, College of Pharmacy University of Minnesota, Minneapolis, MN and School of Pharmacy, University of Wyoming, Laramie, WY

Abstract

The purpose of this submission is to make the case for rejecting outright the current PBAC belief in the gold standard of assumption driven model simulations to create imaginary (false and non-evaluable) cost-effectiveness claims and a continued reliance on ordinal scores to asses cost-effectiveness rather than the Rasch standard of interval measures which have been recognized for over 50 years. The proposed new start in HTA for Australia is to base all value claims whether for clinical physical outcomes, patient or caregiver reported outcomes and drug and resource utilization outcomes (not costs) on three premises:

- *All value claims must refer to single attributes that meet the demarcation standards for normal science: they must be credible, evaluable and replicable*
- *All value claims, notably for patient or caregiver reported outcomes. must be consistent with the limitations imposed by the standards of fundamental measurement: they must be unidimensional with linear, interval and invariance properties*
- *All value claims must be supported by an agreed protocol detailing how they are to be assessed in a meaningful timeframe*

Acceptance of these premises means the rejection of assumption driven, modeled simulation and imaginary (non-evaluable or false) cost-outcomes claims; in particular incremental cost-outcomes ratios. In effect, the PBAC guidelines. This is the solution, if the PBAC (or successor group) is to retain credibility. The focus must be on protocol supported value claims which can be evaluated and reported to health decision makers in a meaningful timeframe, supporting ongoing reappraisals over the lifetime of the product.

Keywords: standards normal science, Rasch measurement, rejecting PBAC modelling, unidimensional value claims, evaluable claim. replicable claim

A NEW START IN HTA

These requirements and their implementation are detailed in a recently released University of Wyoming on-line Certificate Program: A New Start in Health Technology Assessment. This is a 14-module program comprising audio-visual presentations, notes to each module (in total 85,000 words) and short multiple-choice exams for each module. The program is recognized by the US Accreditation Council for Pharmacy Education (ACPE) with credit of 20.5 hours. A Certificate is given by the University of Wyoming to all successfully completing the program.

For further information and a link to the program:

https://www.uwyo.edu/pharmacy/resources/certificate-program-a-new-start-in-health-technology-assessment.html

INTRODUCTION

In common with other reference case guidelines to support health technology assessment (HTA) the PBAC guidelines fail to meet the standards of normal science and fundamental measurement: they are imaginary assumption driven modeled simulation that focus on creating imaginary or non-evaluable (i.e. false) cost-outcomes claims. This failure was made clear some ten years ago, with a critique of the PBAC guidelines presented 2017 [1] [2] . The claims are invented, non-empirically evaluable and entirely imaginary; we have no idea if the value claims are right or wrong and, by design, we will never know. If there is a belief that these imaginary claims are a key input to decision making and resource allocation in health care systems (e.g., denying or limiting access to new therapies) then this belief is entirely misplaced although endorsed by the leaders and professional groups in in health technology assessment (HTA) (e.g., the International Society for Pharmacoeconomics and Outcomes Research [ISPOR]) and HTA groups in academic departments.

It is not just the question of a rejection of the standards of normal science. The failure extends to the question of fundamental measurement and the application of Rasch measurement to patient reported outcomes in HTA [3] [4]. Few attempts have been made to apply Rasch standards in disease specific instrument development; standards which have been agreed for over 50 years. The result, unfortunately, is that not only are all generic multiattribute instruments a failure, but that well over 90% of disease specific instruments also fail; the summation of integer scores from Likert scales is not measurement [5].

Once we admit that to achieve any academic respectability, all value claims in HTA must meet the standards of normal science and fundamental measurement then the current PBAC belief system ceases to have any relevance to health system decision making in Australia; it is entirely redundant and misleading. As detailed in a recent critique of the CHEERS 2022 guidance for submitting imaginary modeled claims to HTA journals, the current standards that support the HTA meme are an analytical dead end [6]. The relevance of HTA must admit to the adoption of a new paradigm.

Following the 2022-2027 strategic agreement between the Commonwealth and Medicines Australia, the present review of HTA in Australia provides an ideal opportunity to make the case for rejecting the PBAC guidelines for imaginary claims. Unfortunately, on the evidence so far

submitted to the HTA Policy and Methods Review Committee, there can be little confidence in the possibility that the present belief system will be challenged; the concern must be that with the entrenched belief in approximate or invented evidence, any 'improvements' will probably be cosmetic aimed at possibly simplifying the process of creating imaginary claims but not addressing the mistake of continuing to live in a cost-outcomes environment, with the mathematically impossible QALY at center stage [7]. After all, as some 30 wasted years have been devoted to promoting imaginary claims, too many have too much to lose from what may be described as a needed paradigm shift. Abandoning a belief system as entrenched as the present one will not be easy although it clearly fails accepted standards of other disciplines.

It is the requirements of this required paradigm shift which are the focus of the University of Wyoming Certificate Program: *A New Start in Health Technology Assessment.* This provides the tools to support a new approach to health technology assessment where the needs of patients in specific disease states takes center stage and not the present reliance on composite generic multiattribute instruments, such as the EQ-5D-3L/5L which in providing only raw ordinal scores are not measures to support value claims for pharmaceutical products and devices.

STANDARDS, MEASUREMENT AND THE PBAC

Since its beginning in the early 1990s the PBAC, in pursuit of blanket cost-effectiveness claims, has relied upon assumption driven modeled simulations (the Markov family) to create imaginary and, by definition, false and non-evaluable claims to support pricing and access decisions. This approach to health technology assessment is not unique to the PBAC and Australia; it is the central belief system n HTA, due in large part not only to the PBAC but to the adoption of the imaginary claim reference case methodology by the NICE in the UK for NHS England, and many other health decision gatekeepers [8]. What makes this belief system unique is not only its denial of the standards of normal science but also a denial of fundamental or Rasch measurement; whether this is just a lack of awareness or a misplaced rejection of the unique contribution of Rasch providing the necessary and sufficient rules for this transformation to an interval measure, as made clear in two seminal papers published in 1989, is not clear although few in HTA seem to be aware of the Rasch contribution to assessing patient and caregiver responses to therapy interventions [9 10].

HTA is unique among the physical and more mature social sciences (e.g., education) in a belief system that rejects the normal standards for evaluation of value claims or hypotheses. It is difficult to grasp why such a belief system or meme has endured; the point is that it has and this puts HTA outside the mainstream in the physical and social sciences in rejecting the standard for demarcation [11]. HTA, in terms of falsification and the commitment to the pursuit of objective knowledge, is no different from intelligent design (which is an apt description of PBAC modelling). The PBAC asks us to take their word for the critical role of assumption driven imaginary claims; a position which was clearly rejected some 350 years ago with the motto adopted by the Royal Society of London *nullius in verba* (take no person's word for it) [12] . The PBAC insists on the impossibility of doing anything other than taking their word for it in cost-effectiveness claims; they are the final imaginary claims arbiter.

In HTA, a failure to apply or even be aware of, Rash standards for patient reported outcomes (PRO) is virtually universal. This is unfortunate because the Rasch model in unique in providing the basis for interval and ratio measurement to support patient or caregiver centric value claims for therapy response. The Rasch model focuses on capturing the manifestation of a latent construct (e., needs fulfillment as a measure of quality of life). The unique contribution of Rasch, presented in the 1950s, as quoted by Bond et al [4] is that:

> *A person having a greater ability than another person should have the greater probability of solving any item of the type in question, and similarly, one item being more difficult than another means that for any person the probability of solving the second item is the greater one (Rasch 1960)* [13]

It should be emphasized that the Rasch measurement model, with its genesis in intelligence and attainment tests, has always focused on the individual. The central premise of the Rasch model is that, in probabilistic or expected response terms, if a respondent with a particular ability encounters a questionnaire item with a particular difficulty, what is the probability that this respondent will get the item correct or respond positively? In other words, instruments must be developed that embody the requirement that the probability of success, meeting a need, depends on the difference between the ability of the person and the difficulty of the item or the difficulty of meeting that need. This means that, as noted above, all Rasch instruments will have the property that they are capturing the manifestation of a latent measure as a single attribute (e.g., needs fulfillment), where the measure of response is linear, interval and invariant. Scores achieved on composite bundles of health state symptoms may bear no resemblance at all to the needs of patients and caregivers, where health may be a key consideration, but not captured as a Rasch single attribute framework. Composite or multiattribute scores fail the standards for Rasch measurement [14].

The concepts underlying the Rasch measure are far removed from the generic and disease specific instruments that characterize HTA. Demonstrating this is one of the principal features of the Wyoming program with the emphasis on the importance of capturing the health-related needs of patients and caregivers in specific disease states. The focus is not on quality of life defined as a pre-set bundle of reported clinical symptoms but a single focus on needs (defined by instrument items) and the extent to which a new therapy enables the target population to better meet their needs. Instruments have been developed for a number of disease states applying the Rasch model over the past 20 years by, in particular, Galen Research, UK with an on-line directory of instruments [15]. Instruments developed over the past 20 plus years by Galen Research include: pulmonary hypertension, atopic dermatitis, psoriasis, growth hormone deficiency (adults), Crohn's disease, herpes, migraine, asthma, COPD, ankylosing spondylitis, osteoarthritis, psoriatic arthritis and rheumatoid arthritis. There is a range of language adaptation version of each instrument.

Unless a disease specific instrument has been developed with the application of Rasch standards then it will not have the desired interval and ratio properties; this sets the manufacturer an interesting challenge to create new instruments and establish PRO disease specific value claims for target patient populations. This assumes that the PBAC (or its successor) adopts a new analytical framework for HTA. Value claims which support replication and reproduction in target Australian populations. Weaning the Australian HTA community off generic composite

instruments and imaginary cost-outcome claims will not be easy. After all, creating imaginary claims has the undeniable benefit that the claim can never be empirically challenged.

A CORNUCOPIA OF INTERVAL UTILITY SCORES

Despite failing to meet Rasch standards for an interval score, utilities or preference scores continue to play center field in HTA. There is no single gold standard generic or multiattribute utility instrument. Rather, we have a range of instruments that represent different systems that vary in the choice, severity, weighting and description of health dimensions and the process applied to create a scoring algorithm for a composite ordinal score. Not surprisingly, for the same patient, different instruments create difference utility scores and, in consequence, different incremental cost-utility ratios and claims for cost-effectiveness. There is, to add further confusion, ongoing debates over the source of preferences for bundles of health as to whether patients or members of the general population should participate. Add, to this the question of whether utilities or preference scores should be treated as having equal value or should varying with the severity of the disease, the presence of disabilities, and life-expectancy of the patient. These are issues which lead to QALYs being banned under Medicare with the *Affordable Care Act* and presently a wider ban proposed under H.R. 485 - *Protecting Health Care for All Patients Act* to prohibit federal health care programs from using the QALY to assess the relative merits of medical interventions. H.R. 485 passed the Houe of Representatives on 7 February and now moves to the Senate. If finally signed off, this will most likely result in the demise of assumption driven imaginary simulations in the US (e.g., models utilized by ICER). The way would then be open for a New Start approach to HTA; but whether this prohibition will have any impact in Australia is an open question.

Ideally, assessment groups should require a common utility algorithm to support imaginary cost-effectiveness claims. NICE, for example, requires all models to utilize the EQ-5D-3L as the base-case. Where this is not available then mapping should be applied to translate, say, results for the EQ-5D-5L to the equivalent EQ-5D-3L score. Indeed, there are studies where disease specific utility or preference scores have been mapped to the EQ-5D-3L. Unfortunately, despite continued advocacy and considerable resources devoted to the effort, mapping is a waste of time. As the utilities or preference are ordinal scores, one ordinal score cannot be mapped to another. It is mathematically impossible as neither are unidimensional with linear, interval and invariant properties

Mention should also be made of another failure to appreciate the role of fundamental measurement: the Tufts University Center for the Evaluation of Value and Risk in Health, Cost-Effectiveness Analysis (CEA) Registry [16]. Developed over the past 47 years, the registry summarizes cost-per QALY models to produce a file of over 46,000 health state preference scores. This can be reviewed to select utility scores that can be applied to assumption driven simulation models (for a small fee). What was never appreciated, resulting in 47 years of wasted effort is that the scores are ordinal, taking both positive and negative values.

The Tufts registry covers both generic and disease specific utility scores. The point to make is that both utility scores are ordinal and fail Rasch standards. In large part, for the disease specific scores, is the aggregation of integer counts from Likert-type question responses. This summation requires

that all questionnaire items are of equal difficulty and the threshold or steps for each item, are equally distant or of equal value; this never considered. If Likert scales are to be the basis for scoring, then we require the application of Rasch rules for dichotomous data: the Rasch Model for Polytomous Data and the Partial Credit Rasch Model. Software packages have been available for over 30 years to support such analyses [3 4 37].

Of course, as the PBAC is committed to generic instruments such as the EQ-5D-3L to create ordinal composite preference scores, then the QALY problem will persist. The PBAC will be subject to the continuing criticism that it insists on utilizing QALYs and cost-per-QALY modelled claims because it has no alternative. It will, alongside other HTA supporters, continue to believe in the HTA gold standard meme. This is supported by leading textbooks whose authors are clearly confused over measurement standards, continuing with QALY models and probabilistic sensitivity analysis to support imaginary cost-effectiveness claims [17]. It is no defense to claim that Australia should continue to use the QALY because other health systems also require assumption driven simulations.

An instrument that certainly should not be considered is the EuroQol successor instrument the EQ-Health and Wellbeing (EQ-HWB) instrument [18]. While considerable time and resources have been devoted to developing an Australian version, it fails Rasch measurement standards [19]. It creates only composite ordinal scores which, due to a similar methodology to the QALY, inevitably lead to negative scores. No thought was given, despite its more recent development, to the standards of Rasch measurement and the need to create single attribute or unidimensional, linear, interval and invariant measures. If there is any thought that it can support a 'new' QALY then this should be put to one side, together with any thought that it can support an 'improved' assumption driven simulation model framework. It will still produce imaginary and false claims. The EQ-HWB also produces negative scores.

RASCH AND INSTRUMENT RESPONSE THEORY

A mistake that is commonly made is to equate Rasch measurement with one-parameter item response theory (IRT). As Bond et al [4] point out the two methodologies have common elements in the emphasis on latent traits and individual item responses to operationalize, the notion of an underlying unidimensional latent trait with local independence of items with estimated probabilities of success for individual items. Application of IRT-theory is exemplified in the PROMIS system for individual item selection and instrument development.

Even so, Rasch and IRT are not equivalent, this has been emphasized over the past few decades. The point is made by Andrich: *To consider that when there is a mismatch between data and a model it might be a problem with the data rather than the model, is in itself a considerable perceptual shift from the traditional perspective on the data-model relationship* [20]. This crucial distinction is further developed by Bond et al in their discussion of why data are selected to meet the rules standards of the Rasch model, with the fit of items and their acceptance the subject of the assessment [4]. Rasch is not concerned with the selection of test-items and fit determined by standard statics which may include supplementary items (two and three parameter IRT) to give a better fit. This is not fundamental measurement; it characterizes the convoluted efforts made to fit generic

utility or preference model to the data (objective a utility or preference range 0 – 1). The common solution, where unity is defined as perfect health, is to apply decrements from unity to create ordinal scores; and inevitably overshooting to give negative ordinal scores (in some cases some 20% of respondents provide negative utility or preference scores) [21]. An issue that does not arise with the Rasch rules to create interval measures and the interval scale is log-odds (a logistic transformation).

With Rasch, as Bond et al emphasize, the emphasis on sound scientific measurement; an estimation method (conditional maximum likelihood) to create instruments with the required interval properties [4]. Rasch is not a collection of items from an item bank; it is a more comprehensive rules-based approach to creating an instrument with items selected that meet Rasch measurement requirements. Rasch differs from IRT and standard statistical assessments where data have primacy; for Rasch the model and its rules have primacy. Rather than a descriptive and exploratory paradigm, the Rasch model is confirmatory and predictive. The data have to fit the model. If the fit to the model is judged satisfactory, with the focus on the size and structure of residuals, then we can claim that the results are an instrument with unidimensional, linear, interval and invariant properties.

This failure to recognize Rasch measurement is seen (in spades) in the almost universal application of the quality adjusted life year (QALY) which, although mathematically impossible, is a centerpiece of the modelled simulation. QALY estimates over the lifetime of the Markov model support incremental cost-utility claims, cost-per-QALY scores and the creation of a cost-effectiveness scale using probabilistic sensitivity analysis.

The problem, or more accurately, the failure of the QALY construct is that it involves multiplying estimated (by the model assumptions) time spent in a disease stage (an interval measure) by a composite, ordinal utility score. This composite ordinal score is generated from generic quality of life instruments which combine patient reporting of symptom severity (an ordinal response) in an algorithm to yield a score. As these scores are ordinal, they cannot support arithmetic or statistical operations. This has been recognized for over a century in measurement theory but not by QALY advocates in HTA.

This failure to recognize, in HTA, the imperative of Rasch measurement is surprising given the widespread acceptance of Rasch measurement in Australia by leading academics and with research funding for Rasch applications over the past 30 or more years. Noteworthy, are the contributions of Professor David Andrich, University of Western Australia and Dr Trevor Bond of James Cook University [3 4].

ENCOURAGING FALSE CLAIMS

Reference case assumption driven modelling is an open invitation to the creation of false value claims [22]. Where a model is proposed that takes a lifetime framework for a hypothetical treating population, the fact that, by design, the value claims are not empirically evaluable encourages model manipulation to create a favorable cost-effectiveness outcome. PBAC models become marketing devices with the sponsors product receiving a modelled favorable cost-effectiveness claim [23]. This does away with any notion of replication or reproduction of value claims (e.g.,

clinical endpoints). This flies in the face of the standards of normal science and the pursuit of what Popper has described as objective knowledge [24] [25]. Claims can never be falsified, which may be considered a bonus, but which mean the patient loses out. There is overwhelming evidence which, put simply, finds the majority of clinical claims failing to be replicated with the original protocol or reproduced in other target populations, including attempts to make the claim more generalizable. A situation which is perpetuated when there is no intention, as with the PBAC guidelines, to provide a framework for a review of clinical claims in terms of both replication and reproduction in target patient populations.

But the PBAC has an additional problem. If it relies on literature reviews to populate and defend other assumption that are required for the assumption driven simulations, what guarantee is there that these assumptions are 'true'? Typically, where there is a search to find and justify assumptions (e.g., utility scores) the number of references is limited; in some cases, only a single claim for utility scores for stages of a target disease. While we know that these, by definition, will be only ordinal composite scores, the presence of 'false' assumptions creates a problem for any faith the PBAC might have in believable simulated imaginary claims. Once a manufacturer challenges PBAC approved assumptions (following a review process) then we are back to base one for discussions over the evidence that the assumptions are not false with a further time and cost penalty that may last months or even years.

Assumptions can be changed and challenged; but this is a singularly useless activity as the assessment and adjudication by, for example the PBAC and its academic assumption monitors, devolves to a choice of assumptions. Parceling out manufacturer modeled submissions to an academic 'assumption police' based in selected universities achieves nothing; they would have to challenge each assumption on grounds of credibility. This overlooks a simple point of logic, Humes's problem of induction regarding claims on the future: the fact that all past futures have resembled past it does not follow that all future futures will resemble future pasts [26]. In these terms creating an assumption driven simulation model, claiming that literature and trial-based assumptions will hold in the future, even if that is hypothetical, is simply a futile endeavor.

There is a more substantive issue which applies across the board in HTA: false claims created by paper mills, consultants and by academics in tweaking and even inventing data to support one more peer reviewed paper [27]. If the PBAC is to have relevance it should direct its attention false claims, not just for imaginary modelling, but for value claims proposed as part of a formulary submission. Hence the important, as detailed in the proposed new start value claim protocols.

PROTOCOLS

The proposed new start in health technology assessment places emphasis on protocols: all value claims submitted must be accompanied by a protocol detailing how the claim will be evaluated and reported to a formulary committee in a meaningful time frame. This has two advantages: first, it makes clear that simulated model claims are finished and, second, it hopefully goes a long way to eliminating concerns over the presence of false claims. Note that it is not the question of the false claims created by assumption driven simulations, but of false claims for clinical and PRO outcomes. In the former case there must be an assessment of the merits of clinical trial claims and

following procedures that the Cochrane Collaboration have in place for problematic studies in systematic reviews [28] . Consideration of a set of standards for the acceptance of pivotal clinical claims must be accommodated in the evaluation of value claims that are claimed to have followed a protocol. At the same time, PRO claims must also be evaluated. This can be achieved by the application of a set of proposed standards that have recently been published [29].

THERAPY CHOICE AND NEED FULFILLMENT

If, following the Rasch standards, we are looking for a measure which may be considered a replacement for discredited the generic composite ordinal instrument scores, then a candidate endorsed by the Wyoming Certificate Program is need fulfillment for patients in specific disease states. This measure, which has received attention in a number of disease states from the mi-1990s, is an interval calibration of patient value [30]. The application of need fulfillment, which follows from an earlier generic measure, the Nottingham Health Profile with the concept that people judge their experiences in relation to expectations, meets the required standards [31].

The needs fulfillment model focuses on the extent to which needs are fulfilled in the presence of disease and its treatment; the value of interventions to patients and caregivers where we can argue that health is a major concern. To achieve this, we need a direct measure of value, applying Rasch model standards, that bypass physical measures of health status and functioning. Health status related quality of life scores may improve following a new therapy intervention but without value to the patient improving. The key is, through extensive interviews, to ask how the patient's life has been impacted by a specific disease. These set the groundwork for initial item selection (statements) to manifest the underlying unidimensional latent needs fulfillment construct. Creating an interval index of the extent to which patient needs are fulfilled, not a profile of health status, is the purview of the application of the Rasch model. A recent example of this is the Galen Research APPLIQue: Alzheimer's Patient Partners Life Impact Questionnaire with a work in progress extension of the Rasch interval scale to a ratio scale [32 33].

PBAC 2016 GUIDELINES: SECTION 3A COST-EFFECTIVENESS

Given the arguments presented above in favor of the standards of normal science and fundamental measurement, it is clear that the submission standards proposed in Section 3A remain, after some 7 years, not fit for purpose. If there is a concerted effort to redraft the PBAC (or successor organization) guidelines for formulary assessment then the entire section should be excised. The starting point should be the focus on rejecting the notion that it is possible to create a valid and empirically evaluable, and non-imaginary, claim for comparator cost-effectiveness. Rather the focus must be on value claims that are empirically evaluable, meet Rasch measurement standards, with emphasis on PROs, and are supported by protocols. At this stage no attempt should be made to bundle particular value claims to attempt an overall measure; the first step must be to assess the individual value claim component of a submission.

The 2016 Section 3A proposed submission standards follow the reference case format that characterizes the modeling framework required by NICE and, in the US, the Institute for Clinical and Economic Review (ICER). In practice, it is a set of instructions to construct imaginary non-evaluable cost-outcomes claims; HTA Lego. In none of these cases is there any concept of the

standards of normal science and Rasch measurement. We are, in effect asked, in contradistinction to the motto of the Royal Society [1662] 'nullius in verba' [take no one's word for it], to take the PBAC's word for it even though it defies the required standards and is only one imaginary cost-effectiveness option among any number, with possible false and invented assumptions for a hypothetical future stream of benefits and costs.

Despite the continued belief in the importance of modelled imaginary cost-effectiveness claims, the fact remains that the entire modelling exercise denies the role of falsification of claims and therapy response expressed in Rasch interval or ratio measure terms. The fact that this imaginary claims approach has been supported for some 30 years must be a concern; particularly if it is used to defend, after some tweaking, a continued commitment to imaginary claims. Sunk costs should not determine future decision frameworks.

Even if we are prepared to put aside any commitment to meeting the demarcation standard for normal science, the PBAC reference model still fails the standards for Rasch fundamental measurement. There is no way this can be defended; unless value claims meet Rasch standards for the creation of instruments with interval or ratio measurement, we are left with unacceptable claims for therapy response, notably the ordinal preference scores supporting the ubiquitous yet false QALY.

The bottom line, unrecognized by the PBAC, is that all value claims for product impact in health technology assessment must be for single attributes (unidimensional) with linear, interval and invariant properties. This is the unique contribution of Rasch measurement, recognized now for some 70 years: observations are ordinal while measures are interval. The fact that presages the needed shift to a new value claim paradigm is that the Rasch measurement model provides the necessary and sufficient means to transform ordinal counts into linear measures; a position recognized for over 50 years. The implications for the PBAC and similar agencies are significant: all value claims (including cost-effectiveness claims) whether they are for purely clinically measured endpoints, patient reported outcomes (PRO's), claims for resource allocation and product switching must be demonstrated to meet Rasch standards. The Australian Assessment of Quality of Life (AQOL) multiattribute utility instruments also fail Rasch measurement standards [34]. In the case of PROs this is easily met with the many Rasch applications that are detailed in the Rasch Measurement Analysis Software Directory, including the Andrich software platform RUMM2030+ [35].

AUSTRALAN HTA REVIEW: CONSULTATION REPORTS

Two consultation reports have been: presented. Consultation I: Report which is described as a thematic summary of information presented [36] and the draft Consultation Report II Options for Reform [37]. Of particular interest is the draft HTA Methods: Economic Evaluation[38] . The singular feature (or rather its absence) is that no one including stakeholders or the leaders in HTA in Australia ask the question: why do we need assumption driven modeled simulations that create imaginary non-evaluable cost-outcome claims when this methodology denies both the standards of normal science for robust and evaluable product value claims and the standards for fundamental measurement in the acceptance of the mathematically impossible QALY as a key input to the

lengthy PBAC decision making? As it stands it is as though the decision has been made to retain the current methodology and to respond to the more egregious failings: time and cost of meeting PBAC requirements, pernicious incentives (deliberate negotiation from a price higher than acceptable) and accommodating rare disease. The issue of the patient perspective and the needs of patients (needs fulfillment) will be put to one side with the continued belief in multiattribute utility or preference instruments to produce composite ordinal scores and with the possibility of introducing the equally flawed multiattribute, composite and ordinal scored EQ-HWB [39] . Given needs fulfillment instrument experience and application, the question of meeting the needs of patients does not have to devolve to a time intensive (years) and costly (millions of dollars) haggling over assumptions to create any number of competing non-evaluable cost-outcomes claims on the future.

CONCLUSIONS

It may seem an unreasonable conclusion, but it is the PBAC earliest guidelines documents that have set the stage for 30 years of modelled, QALY-driven, imaginary claims. A position that makes clear the lack of interest or awareness of the standards of normal science and fundamental measurement. A misplaced focus on the possibility of a blanket cost-outcomes assumption driven modelled claim has set back health technology assessment. Rather than a paradigm that supports ongoing disease and therapeutic class reviews and the needs to which new therapies support target patient groups in needs fulfillment, we have thousands of published of one-off imaginary modelled claims, in large part driven by the impossible QALY. Many of these are just marketing exercises with consultants working to create a positive cost-effectiveness case for the sponsor. Others rest on dubious or invented assumptions with the belief that if data are not available, then it should be invented, possibly with the assistance of paper mills, to provide a cost-effectiveness claim that is imaginary yet beyond reproach.

The Australian health system is in an absurd situation. For 30 years the PBAC has pursued the will o'the wisp of a decision framework that is, and always has been, fatally flawed; an analytical dead end. The focus on cost-outcomes has always been a non-starter, judged by the standards of normal science and fundamental measurement. The result has been the expenditure of probably hundreds of millions of dollars by manufacturers to jump through PBAC hoops. Not only has this delayed the introduction of new therapies for years but has effectively discouraged manufacturers from entering the market place to the detriment of health outcomes. The prospects of effective therapy for those Australians with rare disease is dim. If the HTA review asks for barriers it has to look no further than the PBAC. The PBAC is the principal barrier to a meaningful, more nuanced and comprehensive evaluation framework for therapy decisions. Formulary decisions are not a one-off endeavor promoting imaginary and by-definition non-evaluable comparative product claims which are never revisited. Tinkering with the format of the 2016 guidelines is not sufficient; they are fatally flawed and should play no place in therapy pricing and access decisions. This should be the challenge for the review: should the PBAC be retained or abolished? The prospects are not auspicious.

The Wyoming new start envisages a paradigm shift; a major transition and not one that represents an enhancement of existing standards where existing players retain their place. The proposed new paradigm rejects imaginary claims. Three premises must be respected:

- All value claims must refer to single attributes that meet the demarcation standards for normal science: they must be credible, evaluable and replicable
- All value claims, notably for patient or caregiver reported outcomes. must be consistent with the limitations imposed by the standards of fundamental measurement: they must be unidimensional with linear, interval and invariance properties
- All value claims must be supported by an agreed protocol detailing how they are to be assessed in a meaningful timeframe

These premises represent a major shift; normal science and fundamental measurement are accepted and not rejected. They are in marked contrast to the limited objectives of the current HTA review which suggested that imaginary non-evaluable claims resting on ordinal complex preference scores and mathematically impossible QALYs will continue to be Australia's continuing contribution to HTA. If this is all that is intended to be achieved, then there should be no surprise that the government, given its well-known time-consuming bureaucratic mind-set, is devoting some 5 years and substantial resources to making the pursuit of imaginary claims more palatable. *Plus ça change, plus c'est la meme chose.*

REFERENCES

[1] Langley PC. Validation of modeled pharmacoeconomic claims in formulary submissions, *J Med Econ*. 2015;18(12):993-99

[2] Langley PC. Dreamtime: Version 5.0 of the Australian Guidelines for Preparing Submissions to the Pharmaceutical Benefits Advisory Committee (PBAC). *Inov Pharm.* 2017;8(1): No. 5

[3] Andrich D, Marais I. A Course in Rasch Measurement Theory: Measuring in the Educational and Health Sciences. Singapore: Springer, 2019

[4] Bond T, Yan Z, Heene M. Applying the Rasch Model: Fundamental Measurement in the Human Sciences (4th Ed.). New York: Routledge, 2021

[5] McKenna P, Heaney A. COSMIN reviews: the need to consider measurement theory, modern measurement and a prospective rather than retrospective approach to evaluating patient-based measures. J Med Econ. 2021;24(1):860-61

[6] Langley P. Nothing to Cheer About: Endorsing Imaginary Economic Evaluations and Value Claims with CHEERS 22 [version 1; peer review: 2 approved]. *F1000Research* 2022, 11:248

[7] Langley PC and McKenna SP. Measurement, modeling and QALYs [version 1; peer review: 2 approved]. *F1000Research* 2020, **9**:1048

[8] NICE. Technology Appraisal Guidance. https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-technology-appraisal-guidance

[9] Merbitz C, Morris J, Grip J. Ordinal scales and the foundations of misinference. *Arch Phys Med Rehabil*. 1989;70:308-32

[10] Wright B, Linacre J. Observations are always ordinal; measurements, however, must be interval. *Arch Phys Med Rehabil.* 1989; 70(12):857-60

[11] Pigliucci M. Nonsense on Stilts: How to Tell Science from Bunk. Chicago: University of Chicago Press, 2010

[12] Wootton D. The Invention of Science: A New History of the Scientific Revolution. New York HarperCollins, 2015

[13] Rasch, G. Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests. Copenhagen: Nielsen & Lydiche, 1960.

[14] McKenna S, Heaney A. Composite outcome measurement in clinical research: The triumph of illusion over reality*? J Med Econ*. 2020;23(10):1196-1204

[15] Galen Research, Manchester UK https://www.galen-research.com/measures-database/

[16] Tufts University. CEVR- Center for the Evaluation of Value and Risk in Health. Cost-Effectiveness Analysis (CEA) Registry. https://cevr.tuftsmedicalcenter.org/databases/cea-registry

[17] The classic guide to creating gold standard imaginary ordinal claims is Drummond M, Sculpher M, Claxton K et al. Methods for the Economic Evaluation of Health Care Programmes (4th Ed.).New York: Oxford University Press, 2015

[18] Lee P, Engel L, Lubetkin E, Gao L, Exploring the comparability between EQ-5D and the EQ-HWB in the general Australian population., *Value Health* (2024)

[19] Langley P. After the QALY: Measurement and the road not taken (Part I: The EQ-HWB). *Maimon Working Papers* No. 8, June 2023 https://maimonresearch.com/wp-content/uploads/2023/11/Maimon-Working-Paper-No.-8-Part-1-1.pdf

[20] Andrich D. Understanding resistance to the data-model relationship in Rasch's paradigm: A reflection for the next generation. *J App Measurement*. 2002;3(3):325-359 (Quoted in Bond)

[21] Bernfort L, Gerdle B, Husberg M et al. People in states worse than dead according to the EQ-5D UK value set: would they rather be dead? *Qual Life Res*. 2018;27(7):1827-33

[22] Husereau D, Drummond M, Augustovski F et al. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 22) Statement: Updated reporting guidance for health economic evaluations. *ValueHealth*. 2022;25(1):3-9

[23] Langley P. Facilitating bias in cost-effectiveness analysis: CHEERS 2022 and the creation of assumption-driven imaginary value claims in health technology assessment [version 1; peer review: 3 approved]. *F1000Research* 2022, 11:993

[24] Popper K. Objective Knowledge: An Evolutionary Approach (Rev Ed). London: Clarendon Press, 1979

[25] Thornton, S, "Karl Popper", *The Stanford Encyclopedia of Philosophy* (Winter 2023 Edition), Zalta E & Nodelman U (eds.),

[26] Magee B. Popper. London: Fontana, 1974

[27] Langley P. The Challenge for Health Technology Assessment: Paper mills, false claims and the endorsement of imaginary claims. *Maimon Working Papers No. 14* . August 2023 https://maimonresearch.com/wp-content/uploads/2023/08/Maimon-Working-Paper-No-14-August-2023.pdf

[28] Cochrane Library. Managing potentially problematic studies https://www.cochranelibrary.com/cdsr/editorial-policies#problematic-studies

[29] McKenna S, Heaney A, Langley P. Fundamental Outcome Measurement: Selecting Patient Reported Outcome Instruments and Interpreting the Data they Produce. *InovPharm*. 2021; 12(2): No. 17

[30] McKenna S, Wilburn J. Patient Value: Its nature, measurement, and role in real world evidence studies and outcome-based reimbursement. *J Med Econ*. 2018;23(5): 474-80

[31] Hunt S, McEwen J, McKenna S. Measuring Health Status: a new tool for clinicians and epidemiologists. *J Royal College General Practitioner*s. 1985;35:185-88

[32] Hagell P, Rouse M, McKenna SP. Measuring the impact of caring for a spouse with Alzheimer's disease: Validation of the Alzheimer's Patient Partners Life Impact Questionnaire (APPLIQue). *J Applied Measurement* . 2018; 19(3): 271-282

[33] Langley P. Enhancing the Rasch Response Model for Value Claims: Latent Trait Possession and Formulary Evaluations. *Maimon Working Papers No. 21*. October 2023 https://maimonresearch.com/wp-content/uploads/2024/01/Maimon-Working-Paper-No.-21-October-2023.pdf

[34] Assessment of Quality of life (AQOL) instruments https://www.aqol.com.au/

[35] Rasch Measurement Analysis Software Directory https://www.rasch.org/software.htm

[36] Australian HTA Review Consultation I: Report https://www.health.gov.au/resources/publications/health-technology-assessment-policy-and-methods-review-consultation-1-report

[37] Australian HTA Review Consultation II: Options for Reform (Draft) https://ohta-consultations.health.gov.au/ohta/hta-review-consultation-2/

[38] Australian HTA Review. Draft Paper: HTA Methods – Economic Evaluation https://www.health.gov.au/resources/publications/hta-policy-and-methods-review-draft-paper-hta-methods-economic-evaluation?language=en

[39] Brazier J, Peasgood, Mukuria C et al. The EQ-HWB: Overview of the Development of a Measure of Wellbeing and Key Results. Value Health. 2022;25(4):482-91