MAIMON WORKING PAPERS No. 1 JANUARY 2024

IMPLEMENTING THE RARE DISEASE TEN COMMANDMENTS FOR VALUE CLAIMS, PRICING AND REIMBURSEMENT: THE WYOMING NEW START CERTIFICATE PROGRAM IN HEALTH TECHNOLOGY ASSESSMENT

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

Background

A Maimon Working Paper published in May 2022 put forward ten commandments for standards to be applied in preparing formulary and other submissions for rare disease products. The standards made clear the importance of meeting the standards for normal science and fundamental measurement. The commandments have received considerable attention with well over 500 views on the Maimon Research website. These commandments, which are briefly revisited here, are the only basis for making valid product claims or ensuring, through protocols, that prospective valid claims can be evaluated with outcomes-based contracting. The purpose of this Working Paper is to make clear the argument for rejecting the analytically nonsensical quality adjusted life year (QALY) which has been proposed and applied as a pricing criterion. Attempts to impose cost-per-QALY caps for rare disease are an analytical dead end. The analytical standards proposed in the Ten Commandments and the recently released University of Wyoming Certificate Program (ACPE accredited) A New Start in Health Technology Assessment are those that be applied in rare disease. Adopting these standards means the long overdue rejected of assumption driven modeled simulations that create imaginary cost-per-quality adjusted life year (QALY) claims and the equally irrelevant imaginary claims for cost-effectiveness. While evidence a product launch for rare disease products is typically limited, this is no excuse to develop assumption driven modelled non-evaluable claims such as those pursued by the Institute for Clinical and Economic Review. If there is one lesson to be learned: supporting rare disease value claims is a long-term endeavor focused on reporting on further value claims and supporting pricing and access negotiations over the lifetime of any rare disease product.

INTRODUCTION

The current belief system in health technology assessment with the focus on assumption driven model simulations is a charade; it fails all standards for normal science and fundamental evidence [1]. It has no role to play, despite protestations from interest groups such as the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) and the Institute for Clinical and Economic Review (ICER), in to the establishment of value claims. The focus is on costs and outcomes, where both fail the standards required, and, the *pièce de résistance*, non-empirically evaluable claims for cost-effectiveness. The central place of simulation models in health technology assessment is a complete failure. Claiming that the appropriate evidence base is the construction of approximate, modeled cost-outcome claims is nonsensical. This is a disservice, not only to the range of pharmaceutical products and devices but also to rare disease where the evidence base, sparse though it is a product assessment and market entry, must initiate through interactions between health systems and manufacturers, a process of claims assessment to meet these evidence gaps. These requirements are basis for a proposed Ten Commandments published a few years ago and, more recently, the release of an ACPE accredited certificate program by the

School of Pharmacy, University of Wyoming: *A New Start in Health Technology Assessment* [23]. An accompanying Maimon Working Paper provides detail on the structure and content of the program to include audio-visual presentations, extensive notes for each module and assessment standards required to complete the program [4]. A key and fundamental element of the New Start is to recognize, for value claims, the unique contribution of Rasch measurement [5] [6].

BASIS FOR THE TEN COMMANDMENTS

The proposed Ten Commandments to support the entry of rare disease products into a global market place, rest upon three key elements. These, as detailed in the Wyoming Certificate Program are (i) the imperative of recognizing and following the standards of normal science for value claims assessment; (ii) the need to focus on patient reported outcome (PRO) and other claims on measures which are unidimensional and single attribute scored on an interval or ratio measure with invariant properties; and (iii) to recognize the role of models, where assumptions may be present, but which must have claims that are empirically evaluable. These are standards accepted in the physical and more mature social sciences but, either by accident or design, absent from health technology assessment guidelines and pharmacy programs such as the PharmD. Instead, there is a continuing belief in assumption driven simulation modelled unevaluable imaginary claims despite the manifest deficiencies.

The purpose of this brief note is to emphasize once again the importance of standards, described as the Ten Commandments, for product development and support over its lifetime. At the same time, attention is drawn to the Wyoming New Start program with its provision of the necessary tools to ensure uptake of value claims that meet the required (non-imaginary) standards. The New Start. is the basis for rejecting the major barrier to formulary approval, the ICER simulated imaginary cost-per-QALY claims and ISPOR sponsored guidelines such as the CHEERS 2022 guidance for submitting imaginary or false modelled claims to journals [7] .

THE TEN COMMANDMENTS

The Ten Commandments or decalogue are a set of principles, as evidenced by their biblical precursor, that should guide decision making and evidence creation in health technology assessment. They should form the framework for formulary decisions with negotiations over pricing and access as well as principles that underpin ongoing assessment and reporting of therapy impact in rare disease populations. The Ten Commandments are not unique to rare disease interventions, but have a critical role to pay in supporting submissions for negotiations with rare disease interventions where, unless addressed by the manufacturer, imaginary modeled cost-effectiveness value claims may be given an undeserved prominence with their false claims.

The Ten Commandments for value claims whether clinical, PROs or for resource utilization are:

- VALUE CLAIMS: All value claims for rare disease products must conform to the standards of normal science and fundamental measurement
- SINBGLE ATTRIBUTES: All value claims must be for single attributes
- MEASUREMENT: Unless value claims are expressed as single attributes with ratio or interval properties they should be rejected

- FILTER: All value claims must be filtered and assessed to reject nominal or ordinal scales
- VALUE CLAIMS PROTOCOLS Each value claim should be accompanied by a protocol detailing how that value claim is to be empirically assessed and reported in a meaningful timeframe
- PATIENT REPORTED OUTCOME CLAIMS: All PRO claims must be disease or target patient population specific and capture the patient and caregiver voice
- PREPARATION: At the initiation of phase 3, with protocols for pivotal claims, manufacturers must have determined the value claims proposed for their product
- PRO INVESTMENT: Manufacturers should be prepared to commit to investing in PRO value claims that meet measurement standards, capturing the patient and caregiver voice
- ABANDONING MODELS: Unless a model can create evaluable claims that can be captured by a protocol for assessment and reporting, the model must be rejected
- NEGOTIATION: All value claims must be presented with contractually agreed timelines to support evaluation and replication as well as meeting measurement standards

THE WYOMING NEW START

The principal focus on the New Start Wyoming program is on value claims; it operationalizes the Ten Commandments. The program details the standards that value claims should meet and how they should be evaluated. Value claims are classified into those that are clinical, those that capture patient and caregiver outcomes for need fulfillment quality of life and claims for the impact of a new therapy on drug utilization, including compliance, and other health care resources. The value claims are framed as ones that should be revisited over the lifetime of the product, consistent with the earlier Ten Commandments.

From the perspective of manufacturers and other developing potential rare disease therapies the New Start provides a framework or blueprint for initial market access and formulary submissions. The emphasis is on evidence gaps, which can in the case of PROs be addressed as part of the product development process as well as protocol supported value claims to address issues of replication and reproduction of clinical claims and claims for resource utilization to support estimates of resource utilization.

The New Start Certificate Program is presented in three sections: (i) a review of standards essential for evidentiary claims for product and therapy assessment; (ii) a detailed assessment of the failure of simulated modeled information for therapy decisions; and (iii) the required standards for protocol-supported value claims. The sections are contained within a 14-module package including extensive notes (overall 85,000 words), audiovisual presentations, and a short true-false and multiple-choice assessment for each module. The modules are designed for those who are looking for a new framework for the evaluation of pharmaceutical products and devices, to support robust value claims, ongoing disease areas, therapeutic class reviews, and if required, outcomes-based contracting.

A University of Wyoming Certificate will be given to all those who successfully complete the program. In the case of registered pharmacists the Certificate Program provides 20.5 hours of ACPE approved pharmacy continuing education. The link to the program, as noted above, is:

https://pharmacyshare.catalog.instructure.com/courses/certificate-program---a-new-start-in-health-technology

THE SOTATERCEPT FIASCO: IMAGINARY CLAIMS

The recently released ICER report for sotatercept (Merck) an activin signaling inhibitor in pulmonary arterial hypertension is an ideal example of what to avoid in making the case for value claims in rare disease [8]. In common with the standard ICER assessment framework, the lifetime simulation modeled value claims are driven by assumption, which in this case rest upon limited information from the literature relying on only one or two studies or expert opinion. None of the value claims are intended to be empirically evaluable, the Markov design ensures this is the case, and all fail the standards of fundamental measurement. This applies in particular to the utilities range as quality of life inputs claimed to capture the disutility decrement associated with states of disease (initial WHO-FC I stage 0.729 to final WHO-FC IV stage 0.515: Table 4.2); none should be taken seriously as they are composite ordinal scores where the decrements for each progressive stage of the disease are meaningless; we might just as well label them A > B >C >D. This lack of understanding of fundamental measurement, where the Rasch standard is to create single attribute, linear, interval and invariant measures, is illustrated in the claims . The result is claims over the transition model lifetime for the hypothetical population of QALY gain, compared to background therapy alone of 2.03 which, at the assumed direct medical costs, yields a cost per QALY claim of $2,380,000, assuming a placeholder annual price of $400,000.

The notion of health benefit price benchmarks for the annual cost of treatment for sotatercept, yields a price range that would be consistent with modelled incremental cost-effectiveness ratios between $100,000 and $150,000 per QALY gained. This price range is $17,900 to $26,900; a far cry from $400,000 which is assumed. Not surprisingly, at this assumed benchmark price sotatercept is not claimed, falsely, to be cost effective.

If this is the analysis to be applied to all rare disease interventions then it is most unlikely that any investor would consider their support; a conclusion that would hold even the price range was increased substantially. Fortunately, the entire analysis is nothing more than smoke and mirrors, resting on untenable assumptions which should be rejected out of hand as a guide to cost-effectiveness and pricing for any product, notably in this case for a rare disease . As detailed in the Ten Commandments and in more detail in the New Start, there are two fundamental reasons for this rejection: (i) the model fails the standards of normal science, the demarcation between science and non-science, in failing, by design, to produce any value claim that is empirically evaluable and (ii) the model fails to meet the standards of Rasch or fundamental measurement in focusing the analysis, if that is the right word, on claims for quality of life which are mathematically indefensible. The utility values fail because they are composite ordinal scores; a feature which characterizes all generic quality of life instruments, including, in this case the SF-36. Of course, applying scores from other generic quality of life instruments would create equally bizarre and meaningless results. PRO measures that meet the Rasch standard are critical for any value claim.

If we continue to accept ICER-type modelling to support pricing and access, which would deny virtually all rare disease proposals, then we are defending a ‘discipline’ (if that is the right word)

which is unique in claiming that valid formulary decisions can be based upon assumption driven simulations that explicitly reject the standards of normal science and fundamental evidence. ICER in its insistence on assumption driven claims for validation overlooks an elementary point of logic: given a fact that all past futures have resembled past pasts does not mean that all future futures will resemble future pasts or, more simply put, the problem of induction [9]. There is no way one can claim, as the basis for creating imaginary claims, that one set of assumptions is more 'realistic' or preferred to another. The conclusion is obvious: any one ICER or CHEERS 2022 model is just one of (potentially) an indefinite number of models for a particular product in a target patient population each resting on a different set of assumptions, each yielding, in ICER's terms, its own imaginary health benefit price benchmark.

It is precisely these failures which underpin the Ten Commandments and are the key rationale for the New Start. Accepting these means that the modelled rare disease fiasco that characterizes the ICER report on sotatercept would not be repeated. A situation that could have been avoided if those developing these instruments from the mid-1980s had an appreciation of the importance of Rasch or fundamental measurement; Rasch measurement had had been recognized since the 1950s and supported by off-the-shelf software from the late 1970s.

PATIENT AND CAREGIVER NEEDS

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, needs fulfillment, where the measure of response is linear, interval and invariant.

The concepts underlying the Rasch measure are far removed from the generic and disease specific instruments that characterize health technology assessment. Demonstrating this is one of the principal features of the New Start with the emphasis on the importance of capturing the needs for patients and caregivers in specific rare diseases. A handful of instruments have been designed that follow the standards of Rasch measurement, none of which however have focused on a rare disease.

Potentially, an assessment of the extent to which a new intervention can increase the extent to which patient and caregiver needs are met should be a key part of any negotiation for pricing and access. Again, in making the case for disease specific measures, it is important to stress that the new start in health technology assessment, as summarized in the Ten Commandments, is not an alternative to constructing imaginary claims. Failure to consider unmet needs is a major failing in rare disease development. It puts manufacturers, in their negotiations with health systems and government agencies on the back foot. It is no good falling back on simulated modelled claims

applying generic scores, which are all too often marketing devices, because for the reasons outlined above, such claims will be dismissed out of hand by astute negotiators.

What is overlooked is that for the pat 60 years we have the tools to create disease specific measures which meet Rasch or fundamental measurement standards. Rasch modelling is unique in that it is the only basis for creating measures with unidimensional, linear, interval and invariant properties.. The Ten Commandments make quite clear the imperative of evaluating the impact of a new therapy on the needs of patients are caregivers. The task is not difficult as an instrument can be readily developed in a matter of months. All that is required is the willingness of manufacturers to underwrite development in time for an acceptable instrument to be introduced at phase 3 in product development.

It is important to differentiate the Rasch framework from item response theory (IRT). IRT is not fundamental measurement; it is not designed to create single attribute, linear, interval and invariant measures. This is due to a critical difference: The data, for Rasch application, do not have primacy. The intent is not to fit a model, such as an IRT model to data, but to fit the data (i.e. instrument items) to the requirements of the Rasch model. In other words, while IRT and true score theory (TST) instruments are exploratory and descriptive of the data (i.e., item responses) the Rasch framework requires the data to fit the model. Response items are selected to fit the model which means that Rasch is confirmatory and predictive. The item selection and fitting process is relatively straightforward with a number of statistical packages available to accomplish this. If the fitting criteria are sufficiently realized then we are justified in making the case that we have a unidimensional, linear, interval and invariant measure. Just as we should reject simulated imaginary models so we must reject claims for the manifestation of a latent construct such as needs fulfillment that have been created by instruments supported by the PROMIS item selection system. Claims based on the PROMIS instrument development framework should be rejected.

PROTOCOLS, REPLICATION AND REPRODUCTION

The New Start program makes clear that for any product or therapy, gaining market access and an acceptable price is not a one-off or once and for all activity. Rather, and this applies in spades to rare disease, ownership of a new product or therapy must recognize a long-term requirement to support that product in the target treating populations. This is evidenced, as noted in the Ten Commandments, with the requirement for value claim assessment protocols. Whether the value claim is couched simply in clinical terms, with a Rasch standard measure to support those claims, in terms of patient or caregiver outcomes or in terms of resource utilization, there must be a commitment to evaluating and reporting on those claims, not just initially but to support ongoing disease area and therapeutic class reviews.

There are two main reasons for this: (i) potential doubts over the validity of claims from one or two pivotal phase 3 trials given the difficulty of replication (applying the same protocol) puts a premium on the manufacturers to commitment to implementing protocols for replication and then reproduction with more flexible protocols (i.e., a hypothesis with a greater information content) to capture specific characteristics of the patient population and (ii) to eliminate false claims. Putting to one side the obvious opportunity to create false non-evaluable claims from modelled

simulations, the presence of false claims where there is an intent to deceive is a key issue, particularly where the financial returns from rare disease therapies can be substantial [10]. There are a number of examples in the social sciences (notably psychology) where small sample data, which is true of many rare diseases, have been invented. This may be an entirely false set of data or the 'tweaking' of an existing data set, responses from patients or caregivers, to meet the magical confidence level statistical standards There is, of course, a continuum of malfeasance, all too often dismissed by colleagues and institutions, which is never questioned in the peer review process. Hence protocols, where a detailed description of how a manufacturer proposes, to substantiate a value claim has to be addressed and results reported in a meaningful time frame. Making this requirement known in advance should act as a check in false claims or 'tweaking' of responses from patients and caregivers. Even so, there is always the possibility of claims that a new instrument meets Rasch standards. This is relatively easily evaluated as the available software packages can provide a check if a sample response is assessed for the required Rasch criteria, apart from the developer providing audited evidence for instrument creation. Such an assessment would, of course, be part of a protocol where the proposal is to validate the Rasch instrument in a target treating population to support claims for invariance in response.

Manufacturers in rare disease should be prepared to work with target health systems to track the experience of patients and caregivers over the life of the product. Value claims in rare disease are more than likely to be subject to outcomes-based contracting. This puts a premium on working with the target patient population with value claim feedback to support ongoing disease area and therapeutic class reviews.

It is of passing interest that there is no apparent effort by ICER in evaluating whether or not study based assumptions are false as opposed being invented or proposed by experts or the model developers. This is, of course, no reason to do this as the claims are imaginary, although all too many assumptions are based upon single studies (e.g., utilities). All ICER can offer is a promise that if new data to support assumptions becomes available, they might revisit the model to assess impact. This would seem a fruitless endeavor.

The picture is complete where the health system decision makers, including formulary committee members, are trained to apply standard questions to both assess the merits of a protocol as well as the merits of an initial formulary submission which makes value claims based on prior clinical trials with protocols to substantiate and extend those claims. This is an issue captured in the New Start program where a framework is presented detailing questions a formulary committee should ask of all value claims, including assumptions made to support these value claims (e.g., protocol design for target patient inclusion).

CONCLUSIONS

The principal challenge facing manufacturers developing novel products and therapies in rare disease is to overcome the insistence by groups such as ICER and by leaders in health technology assessment that the value claim standards must be in imaginary modelled cost-effectiveness terms. Such value claims are essentially meaningless yet can lead to unnecessary conflict between manufacturers and health system decision makers in pricing and access negotiations. This is not a

situation unique to rare diseases with small target populations but is one that is easily exploited given the limited evidence at product launch leading to an unwarranted insistence on imaginary modeled lifetime simulations. The New Start proposals for health technology assessment provide, as evidenced by the Ten Commandments, a way of escaping from the focus on cost-per-QALY simulations and ersatz recommendations for pricing.

The resolution of this commitment to false or imaginary information is for manufacturers to take the initiative and recognize how the New Start proposals provide the basis for a commitment to value claims supporting marketing approval and market entry as well as an ongoing commitment to supporting protocol to create value claims. In the former case the focus is on integrating value claims (notable quality of life or needs fulfillment claims) in support of phase 3 pivotal trials, where the measured response in needs fulfillment terms is a primary protocol endpoint. The decision to assess the impact of value claims should be made at early phase 2 in product development.

These decisions are entirely the prerogative of the manufacturer. Certainly, discussions over the importance that might be attached to specific value claims can be held with health system decision makers prior to a formal formulary presentation. The concern must be of course, that both parties may have a limited understanding of the need to move to claims that meet the standards of normal science and fundamental measurement; hence the importance of the Wyoming certificate program.

It is not a question of limited evidence to support claims at product launch. The manufacturer is tasked to identify areas where, as part of product development addition evidence can be created. This should be a complement to identifying clinical claims. These are first steps: value claims then must be continually challenged over the lifecycle of the product or therapy. If not then the status quo with its emphasis on imaginary cost-per-QALY and effectiveness claims will continue to perform at center stage.

REFERENCES

[1] 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 (https://doi.org/10.12688/f1000research.109389.1).

[2] Langley P. Rejecting Pseudoscience: Ten Commandments for Rare Disease Pricing and Reimbursement. *Maimon Working Paper No. 15* May 2022 https://maimonresearch.com/wp-content/uploads/2022/05/Working-Paper-No.-15.pdf

[3] School of Pharmacy. University of Wyoming. Certificate Program: A New Start in Health Technology Assessment. https://pharmacyshare.catalog.instructure.com/courses/certificate-program---a-new-start-in-health-technology

[4] Langley P. The University of Wyoming Certificate Program: A New Start in Health Technology Assessment. *Maimon Working Paper No. 22* November 2023 https://maimonresearch.com/wp-content/uploads/2023/11/Maimon-Working-Paper-No-22-V2.pdf

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

[6] Wright B, Linacre J. Observations are always ordinal; measurements, however, must be interval. *Arch Phys Med Rehabil.* 1989; 70(12):857-60 https://www.researchgate.net/publication/20338407_Observations_are_always_ordinal_measurements_however_must_be_interval/link/5563b02408ae9963a11ef326/download

[7] 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 https://www.valueinhealthjournal.com/article/S1098-3015(21)03146-6/pdf

[8] ICER. Sotatercept for Pulmonary Arterial Hypertension. Final Evidence Report. January 8, 2024 https://icer.org/wp-content/uploads/2023/05/PAH_Final-Evidence-Report_For-Publication_01082024.pdf

[9] Russell B. Problems of Philosophy, 1912.

[10] 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