MAIMON WORKING PAPERS

MAIMON WORKING PAPER No. 6 FEBRUARY 2022

CORE VALUE CLAIMS, PROTOCOLS AND FORMULARY EVALUATIONS

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

ABSTRACT

In fall 2020, Version 3.0 of the Minnesota Formulary guidelines was published, followed by a review commentary in Innovations in Pharmacy. The hallmark of these guidelines was the insistence that value claims submitted to formulary committees to support new products or ongoing disease areas and therapeutic class reviews should meet the standards of normal science, notably the axioms of fundamental measurement. This meant that all value claims should refer to single attributes with either interval or ratio measurement properties. At the same time it was proposed that a protocol should accompany each value claim detailing how the manufacturer proposed to evaluate the claim and report back to the formulary committee. What was missing from this scenario was the question of core value claims. These are value claims that should be addressed specific to the needs of a target patient population in a disease area, where the product is indicated for members of that target group. This may reflect what the health system and formulary committee consider an unmet need, yet to be addressed by manufacturers, or dissatisfaction with previous formulary submissions and their emphasis on modeled imaginary value claims. The purpose of this brief commentary is to point to: (i) the benefits of protocols and (ii) the structure within which evaluable core value claims can be formulated.

Keywords: value claims, protocols, ratio measurement, imaginary modeling

INTRODUCTION

Therapy choice in health care systems should rely on value claims that are credible, evaluable and replicable. This has been established as part of the standards of normal science in the discovery of new, yet always provisional, new facts regarding therapy benefits in target patient populations since the scientific revolution of the 17th century. Unfortunate, in health technology assessment the focus for the past 30 years has been on inventing evidence

presumably to support formulary decisions [1]. It is not clear whether formulary committees are in agreement. It is of interest to note that this commitment to assumption driven modeled simulations, stretching decades into the future, has been reaffirmed by the publication of the CHEERS 22 guidance for economic evaluations [2]. This guidance, with the endorsement of some 15 journals and multiple authors, is apparently intended to support the structure of imaginary economic evaluations for submission to journals. Interestingly, no mention of the evidence requirements of formulary committees is made. There is no recognition that the guidance opens the door to a multitude of competing models, probably funded by manufacturers, which may be submitted to, probably somewhat bemused, formulary committees. Mention should also be made, in the US, of the Institute for Clinical and Economic Review (ICER) and their promotion of imaginary outcome and cost-effectiveness claims from assumption driven models; they will without doubt endorse the CHEERS 22 position [3] . Unlike CHEERS 22 which puts to one side implications of cost-per-quality adjusted (QALY) thresholds to support pricing and access recommendations for health systems, the ICER evidence reports have no such limitation. To be frank: inventing evidence is quick, easy and low cost, in particular where consulting and academic groups have off-the-shelf assumption driven modeling package. Rather than address real world issue of therapy benefit, the reporting on core value claims, there is a refuge in the continuous replication of modeled imaginary claims from one ICER report to the next. It becomes a production line for imaginary value claims.

It is not the intention here to propose that the various journals in receipt of a multitude of competing disease specific modeled claims should reconsider their imaginary claims policies; apart from the obvious point that is seems a waste of time. Rather, the focus, following from Version 3.9 of the Minnesota formulary guidelines, is on the evidence needs of formulary committees who may be in the unenviable position of having to deflect imaginary claims that will, presumably, contradict each other while recognizing evidence gaps that can be met through the empirical evaluation of core value claims [4].

FORMULATING CORE VALUE CLAIMS

A core value claim would typically be presented as a question: the anticipated impact or benefit that a new product contributes to the target patient population. A benefit expressed in comparative terms given the existing distribution of outcomes associated with therapies that the new therapy is intended to complement or replace. This may seem a somewhat trivial exercise, but once we consider how it must be evaluated quantitatively the exercise becomes difficult. If a manufacturer attempts to convince a formulary committee on the basis of the pivotal trial results that have convinced the FDA and expert advisors to grant

marketing approval, the committee is likely to complain that the evidence base is thin. Historically, health technology assessment has responded to this by creating, or more accurately, inventing evidence from an assumption driven lifetime modelling simulation. All this accomplishes, even with its pretense of analytical sophistication, are a bundle of non-evaluable claims for costs, QALYs and statements of the likelihood of being deemed cost-effective. The commitment to the evaluation of core value claims aims to turn this around; to return to the milieu of the discovery of new, yet provisional facts, through the process of conjecture and refutation.

In the case of clinical claims for a new product, the core value claims will pose a series of questions asking the manufacturer to develop and implement agreed evaluation protocols to assess quantitatively the claimed benefits in the target patient population. The focus will be on translating the claims from pivotal clinical trials and associated RCTs for the product and its actual and potential comparators into the real world treating environment of the target patient population. Replicating claims from clinical trials is not new, but has presented mixed results with a substantial proportion of trial claims proving impossible to replicate. A recent report from the RCT DUPLICATE initiative (funded by the FDA) seeking to emulate primary endpoints of RCTs (n=10) with non-randomized real world data found agreement between 60% and 90% depending on the agreement metric [5]. The purpose here, however, is not to emulate, which seems a pointless exercise, but to seek protocols that support the benefit assessment of claimed primary and secondary endpoints in a target population which may differ considerably from the selected population (inclusion and exclusion criteria) of the RCT. The question to address is to evaluate, in the event of a lack of correspondence (which is likely) is to identify those patient characteristics contributing not to just a failure to emulate but the wider failure of the pivotal (and if required) secondary outcome claims to translate the claimed RCT benefits to a target patient population. Further issues for claimed endpoints which may complicate primary claims include the practice of 'averaging' outcomes from two pivotal trials (which seems sleight of hand), or attempting to sieve the data to find a subgroup that can be presented as likely to benefit.

Before proceeding there is an important caveat: the so-called "efficiency-effectiveness gap" [6]. Among clinicians, the RCT is the gold standard, the result is the RCT is often applied as the standard for evaluating real world data claims. This is absurd as we would always, given the limited ambit of RCTs in the selection of patients from a target patient population, expect discrepancies. We are not attempting to assess efficacy from observational data only effectiveness. We have the tools in econometric modeling to undertake evaluations of the determinants of primary (and secondary) outcomes claims and to support protocol specifications for the assessment of core value 'efficacy' claims. In an important sense we put

clinical research behind us (the claims are in place) and focus in core value claims in real world research; the distinction is fundamental. Certainly, data quality is a concern, but we have the ability to address this issue as part of the submitted protocol.

The initial focus should be on the primary outcome claim from pivotal clinical trials, modified possibly by feedback from trials undertaken by the manufacturer to complement the pivotal trial claim. Other value claims may be suggested from trial secondary trial endpoints as these are typically underpowered and speculative. In all cases the question of fundamental measurement is pervasive: the only core value claims of interest are those with ratio or interval properties.

PROTOCOL BENEFITS

Insisting on protocols to evaluate value claims solves a problem that has bedeviled formulary evaluations: the elimination of imaginary modeled claims that reject the standards of normal science. If a health system or formulary committee insists that all value claims are to be accompanied by an assessment protocol that meets required evidence standards, then the CHEERS 22/ICER meme is redundant. As detailed in previous commentaries, the term meme or belief system is employed, rather than paradigm as the commitment to imaginary non-valuable claims is a barren exercise; an analytical dead end [7] .

The fact that protocols and standards for protocol design are common in drug development and evaluation as the development process proceeds through Phase I to Phase III, and even post-marketing Phase IV, makes the application of protocols to support core value claims evaluation reasonably straightforward. The key difference to the standard protocol design to support, for example, pivotal Phase III trials, is the requirement that the protocol endorse the axioms of fundamental measurement. That is, all claims that the protocol support must have either ratio or interval properties. This is not a standard that we find for protocols to support RCTs as ordinal outcome measures are found, specifically PRO measures, that fail to meet the required standards.

A further benefit of a protocol is that is commits a manufacturer to support a core value claim and to report the results of an evaluation to a formulary committee in an agreed timeframe. It is of little practical benefit for therapy choice to have an open ended timeframe; all too often this is a recipe for doing nothing. A final point is whether agreement to implement a protocol should be a basis for provisional pricing or a rejection of a submission subject to resolution of the core value claims.

It must be emphasized that there is no concept of a product being cost-effective; such claims reflect a focus on a single gold standard metric, the impossible QALY with the also impossible incremental-cost-per-QALY calculations, and the application of disallowed cost-per-QALY thresholds and attempts to capture uncertainty and the likelihood of different prices being cost-effective from probabilistic sensitivity analysis [8]. Rather the focus should be on core value claims that are single attributes. There should be no attempt to combine these in a single metric, which would make no sense, but rather to present the value claims and the protocols to the formulary committee to factor into their decisions.

As a final point, manufacturers who have successfully completed Phase I or Phase II assessments for a new compound would be advised to review the core value claims requirements before deciding on a final Phase III protocol for their product. A major concern is that all too many trial protocols report results for PROs that fail to meet the required measurement standards. Given the resources required to support a Phase III trial and the timeframe for reporting, failure to include a meaningful PRO could be a disaster. This is seen, for example, where a multiattribute preference instrument such as the EQ-5D-5L is included as an outcome measure with a view to making modeled QALY claims. The entire exercise would be invalid as the algorithm for the E!-5D-5L produce only ordinal scores. These cannot support QALYs.

DEFINING THE TARGET POPULATION

A target population must be identified by the International Classification of Disease clinical modification (ICD-10-CM) code to support protocol evaluations. It is the responsibility of the formulary committee to determine the diagnostic code appropriate to the target population and ensure that is part of the protocol request. Unless an ICD-10-CM code (or codes) can be assigned the attempt evaluate core value claims will fail. All core value claims must relate to a target patient population that can be quantitatively defined. This is of particular relevance to core value claims for resource utilization, including drug utilization and compliance with therapy, and mapping National Drug Codes (NDC) to ICD-10-CM codes could be critical.

CLASSIFYING CORE VALUE CLAIMS

The core value claims may be usefully categorized in terms of their focus and required measurement and response status. Three categories are proposed: (i) clinical claims that meet ratio or interval measurement standards; (ii) PROs that have to meet ratio or interval status; and (iii) resource utilization claims that meet ratio standards.

Clinical Claims

Whether placebo-controlled phase III clinical claims that meet ratio standards are of interest to the formulary committee is an open question. This may satisfy the FDA in the process of drug approval, but may be of limited interest as they typically fail to make a comparative case for therapeutic benefit. A core clinical value claim should, therefore, focus on a requirement for a comparative assessment for clinical benefit; this should not include a modeled indirect comparison but on a proposal for one or more either randomized clinical trials or observational study outcomes to be evaluated and reported. It is not the intent to justify an RCT as the gold standards, but to assess whether or not respondents have experienced the proposed RCT benefit. While it is possible to apply the Phase III protocols to attempt to replicate the pivotal claims, this has all too often failed to reproduce the original claims accepted by the FDA. It is at the discretion of the formulary committee whether they want to replicate these trials.

The protocol for clinical core value claims should not stand alone. Manufacturers should indicate, given the typical commitment to a number of Phase III and Phase IV trials, to indicate whether any of these meet the requirements of the core clinical value claim (or claims). Of particular interest is the relaxation of pivotal Phase III trial protocols, given a profile of the target patient population. One question that could be addressed is the extent to which the pivotal Phase III protocols are representative of the target patient population and the implications for the determinants of net benefit as proposed by the RCTs.

Patient Reported Outcomes

The CATCH-22 in PRO core value claims is the fact that the majority of both generic and disease specific PRO measures fail the standards for fundamental evidence; an issue that is not addressed in CHEERS-22. Joseph Heller's *Catch 22*, published in 1961, has been described as both a dark comedy and absurdist fiction, and as one of the iconic novels of the 20th century. The term Catch 22, is defined by the Collins English Dictionary as follows: "*If you describe a situation as a catch-22, you mean it is an impossible situation because you cannot do one thing until you do another thing, but you cannot do the second thing until you do the first thing.*" As such, it has application to the recently published CHEERS 22 guidance for constructing imaginary pharmaceutical value claims and the belief system that supports it; indeed CHEERS 22 might, somewhat uncharitably perhaps, be described as absurdist fiction but not a dark comedy, unless one is unduly cynical.

The CATCH-22 dilemma is straightforward: if there is agreement that there needs to be a radical rewriting and overhaul of disease specific PROs then this would require recognition of the standards of normal science and the role of fundamental measurement. You cannot attempt to achieve the first without acknowledging the second, but you cannot embrace the standards of normal science and fundamental evidence without rejecting the current PRO situation even though you might view it as a major and unwelcome undertaking.

You either embrace the apparent information meme or you reject it; there is no halfway house. You either buy into it lock stock and barrel or you do not. This is why the role of protocols in formulary decision making is potentially groundbreaking in forcing attention to the need for empirically evaluable core value claims. Much as the approximate information meme is focused, following CHEERS 22 on imaginary modeled claims, it fails, as ICER demonstrates, at providing needed guidance and evidence to formulary committees. CHEERS 22 makes no reference to formulary committees or how the imaginary claims should be presented to committees. It is perhaps doubly unfortunate that both CHEERS 22 and CATCH 22 share the same numeric suffix.

If there is agreement that both generic and disease specific PRO instruments and claims to criteria for ratio or interval measurement properties need to be addressed, they seldom reflect or capture the patient (and caregiver) voice [9] .While users treat them as ratio measures this is because few consider the axioms of fundamental measurement either in their development or in their role in evaluating therapy response. ICER, for example, after the axioms of fundamental measurement have been explained, together with limitations, insists that health economists have confidence that the claimed ordinal multiattribute generic instruments are actually ratio measures in disguise [10]. ICER is not alone, this lack of understanding of measurement theory can be traced back over 30 years with the inevitable result that the overwhelming majority of PROs have to be rejected. This provides manufacturers with two obstacles to overcome: first, in responding to a core value claim requirement they will have to demonstrate that the selected PROs have the required properties to support either interval or ratio responses and, second, where there is no PRO instrument that can be justified as proving the basis for empirical assessment, they have to consider whether or not to invest in developing an appropriate measure.

Once a formulary committee has decided on a PRO core value claim (e.g., need fulfillment quality of life) then it should submit a check list to the manufacturer to ascertain whether or not the manufacturer believes its PRO choice has the required measurement properties. The list of questions has been determined in a recent publication [11]. The key elements are:

- **What do you intend to measure?**
- **What does the PRO instrument measure?**
- **Is the instrument patient or clinician-centric?**
- **How were the items in the instrument generated?**
- **Are the items specific to the disease being measured or are they generic?**
- **Has the instrument been tested with relevant respondents?**
- **Was the instrument's reproducibility reported?**
- **How strong is the evidence for construct validity?**
- **Does the instrument measure ant the ordinal, interval or ratio level**
- **Did the developers apply modern measurement techniques – preferably Rasch Measurement Theory?**
- **Did the authors report evidence of internal validity?**
- **Did the authors report the effectiveness of the response format?**
- **Did the authors report item fit?**
- **Did authors report the evaluation of local item dependency?**
- **Did the authors report assessment of differential item functioning?**
- **Did the authors report overall assessment of fit to the Rasch model?**
- **What assessments of responsiveness did authors report?**

Manufacturers would be expected to address and report on each of these questions for the PROs selected for evaluating core value claims; to include PROs already reported on from clinical trials if these are intended to capture the core value claims and support replication of those claims. Reporting to formulary committees other trial-based PRO measures even though they are not intended to capture a core value assessment should be discouraged unless these questions are also addressed.

If there are concerns that these questions are too demanding, then the formulary committee should refer the manufacturer to the Rasch standard instruments that have been developed over the past 25 years by Galen Research. In addition, if there is concern regarding a core value that requests need-fulfillment quality of life claims it might then be pointed out that there is, for the first time, a bounded ratio scale for need-fulfillment quality of life claims relevant to some 30 disease states [12].

Resource Utilization

Core value claims for resource utilization present few problems from the perspective of fundamental measurement as the default would be a ratio scale. The question is one of

categorization: irrespective of the target impact of a new therapy, whether it is on physician visits, emergency room visits, hospitalizations or drug utilization, the manufacturer has to provide a protocol that defines how the claimed impact for a new therapy is to be assessed. The formulary committee may request resource utilization impacts within broad categories (e.g., impact on physician visits) but this would then have to be translated into the type of physician visit as defined by Current Procedure Terminology (CPT) Codes. Blanket statements regarding 'saving' on physician visits would be unacceptable. Within broad categories for urgent care and hospitalization the same issue arises in identifying the relevant codes for subsequent tracking. In addition, as noted above, the protocol would have to define the target patient population by ICD-10 code to facilitate data extracts for resource utilization.

As well as reporting on core value claims for specific resource utilization categories, it may be of interest to a formulary committee to propose core value claims for compliance with therapy. Clearly, if compliance is expected to be low then the merits of introducing a target population to a product may be muted, notably if the indication is addressing a chronic condition. Nevertheless, despite technical issues in developing models to predict compliance behavior, it is a necessary corrective to overly optimistic claims for product uptake and compliance. The core value request can be quite general: asking manufacturers to provide estimates of compliance for the first 12 months after product launch together with other core value claims for drug uptake and utilization, and the implications of this, including therapy switching, for resource utilization, defined by the relevant CPT and NDC codes.

CONCLUSIONS

Identifying the core value claims relevant to a target patient group in a disease area creates a number of hurdles for manufacturers; the most important being the choice of PRO to reflect one or more core value claims that the manufacturer intends to undertake and report to the formulary committee. The formulary committee or health system decision makers determine the core value claims specific to target populations. The protocol provides the link between the core value claims and is assessment by the manufacturer. This is a key step in ensuring that all value claims meet the standards of normal science. With protocols, imaginary claims are, by design, impossible and must not burden formulary committees with analytical nonsense.

REFERENCES

[1] Neumann P, Willke R, Garrison L. A Health Economics Approach to US Value Assessment Frameworks – Introduction: An ISPOR Special Task Force Report. *Value Health*. 2018;21:119-123

[2] Husereau D, Drummond MN, Augustovski A. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) 2022 Explanation and Elaboration: A Report of the ISPOR CHEERS II Good Practices Task Force. *Value Health*. 2022; 25(1):10-31

[3] Langley P. Let a Thousand Models Bloom: ICER Analytics Opens the Floodgates to Cloud Pseudoscience. *Inov Pharm*. 2021;12(1):No. 5

[4] 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

[5] Franklin J, Patorno E, Desai R et al. Emulating randomized clinical trials with nonrandomized real-world evidence studies. *Circulation*. 2021;1`43:1002-13

[6] Thompson D. Replication of randomized controlled trials using real-world data: what could go wrong?. *ValueHeath*. 2021;24(1):112-115

[7] Langley P. Nonsense on Stilts – Part 1: The ICER 2020-2023 value assessment framework for constructing imaginary worlds. *Inov Pharm*. 2020;11(1):No. 12

[8] Langley P. The Great I-QALY Disaster. *InovPharm*. 2020; 11(3): No 7

[9] McKenna S, Wilburn J. Patient Value: Its nature, measurement, and role in real world evidence studies and outcomes-based reimbursement. *J Med Econ*. 2018;21(5):474-870

[10] Langley P. To Dream the Impossible Dream: The Commitment by the Institute for Clinical and Economic Review to Rewrite the Axioms of Fundamental Measurement for Hemophilia A and Bladder Cancer Value Claims. *InovPharm*. 2020;11(4):No. 22

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

[12] Langley P. McKenna S. Fundamental Measurement: The Need Fulfilment Quality of Life (N-QOL) Measure. *InovPharm*.2021;12(2):No. 6