

Another Rush to Judgment: The Imaginary Worlds of ICER and Recommendations in Duchenne Muscular Dystrophy

Paul C Langley, PhD

Adjunct Professor, College of Pharmacy, University of Minnesota

Abstract

Previous commentaries in the Formulary Evaluation section of INNOVATIONS in Pharmacy have pointed to the lack of credibility in modeled claims for cost-effectiveness and associated recommendations for pricing by the Institute for Clinical and Economic Review (ICER). The principal objection to ICER reports has been that their modeled claims fail the standards of normal science: they are best seen as pseudoscience. The purpose of this latest commentary is to consider the recently released ICER report for Duchenne muscular dystrophy (DMD). As ICER has continued in the case of DMD to apply its modeled cost utility framework with consequent recommendations for pricing adjustments, these recommendations also lack credibility. This commentary emphasizes again not only why the ICER methodology fails to meet the standards of normal science but to point to the importance in rare diseases for accelerated approval, while recognizing that evidence generation will continue. While this assessment of the ICER DMD model does not imply any support for this methodology, a key point is the application of quality of life measures which fail to capture the experience of patients with DMD and, importantly, the interests of both patients and caregivers. While ICER would argue that even with a limited evidence base it is important to address issues of pricing and access for new products, their reports are used as justification for coverage and reimbursement by insurers and health system decision makers without recognition of their lack of scientific merit. This rush to judgement by ICER must raise concerns about potentially adverse formulary decisions that result in access restrictions on new products. If ICER is to make a contribution to the entry of new products in the health market place then it should consider an alternative methodology that generates claims that are empirically evaluable in a timeframe relevant to health decision makers. As it stands, ICER's recommendation should be rejected. This is not a research program that meets accepted scientific standards but one that relies on the willingness of an audience to accept the proposition that evidence is constructed not discovered.

Keywords: Duchenne muscular dystrophy (DMD), ICER pseudoscience, patient voice, caregiver voice

Introduction

On 15 August 2019 the Institute for Clinical and Economic Review (ICER), a not-for profit health technology assessment group, released its final report and meeting summary on Duchenne muscular dystrophy (DMD) ¹. The report focused on an assessment of the comparative clinical effectiveness and value of two first-of-their-kind novel exon-skipping therapies, eteplirsen (Exondys) and golodirsen (under FDA review), together with the corticosteroid, deflazacort (Emflaza). The base case comparator was prednisone, a mainstay of corticosteroid therapy for DMD.

The ICER report concluded that, based on a comparative clinical assessment and then a modeled cost-per-QALY evaluation that 'there is no persuasive evidence that these exon-skipping therapies improve outcomes that matter to patients, including functional status, quality of life, or length of life. Eteplirsen has been on the market for three years and yet we still found notably inadequate data on patient outcomes. Corticosteroid therapy does improve outcomes in DMD, and

there is some evidence that deflazacort may have larger benefits and smaller harms than prednisone'.

The purpose of this commentary, in common with previous commentaries on ICER reports published in *INNOVATIONS in Pharmacy* is to review the value claims for these products, in particular eteplirsen ^{2 3 4 5 6}. The principal question is whether or not the techniques employed by ICER in making value claims meet the standards of normal science ^{7 8 9}. If they do not then the value claims should be put to one side in favor of an alternative value assessment framework that generates value claims that are credible, evaluable and replicable.

Duchenne Muscular Dystrophy

DMD is classified as a rare disease. It is the most common and most severe of the nine types of muscular dystrophy. It is a genetic disorder caused by the absence of the protein dystrophin. It is characterized by progressive skeletal muscle degeneration and weakness that control movement. Onset is usually in early childhood, typically between the ages 3 and 5 years. As it progresses, muscle weakness and atrophy impact the trunk and forearms and eventually other muscles. DMD results from mutations of the DMD gene on the short arm (p) of the X chromosome (Xp21.2). As this gene regulates production of the protein dystrophin that is associated with maintaining the integrity of the membrane of muscle cells, the absence of dystrophin is associated with this most severe form of muscular dystrophies (with a less severe form, Becker

Corresponding author: Paul C Langley, PhD, Adjunct Professor
College of Pharmacy, University of Minnesota
Minneapolis MN
Director, Maimon Research LLC; Tucson, AZ
Email: langley@maimonresearch.com

muscular dystrophy associated with decreased dystrophin protein)

DMD affects approximately 1 in 3,500 live male births with an estimated 20,000 children diagnosed with DMD each year. It is rapidly progressive with children exhibiting developmental delays, toe walking, a waddling gait, difficulty climbing stairs, rising from a sitting position, fractures and repeated falling. As the disease progresses other abnormalities may appear to include progressive spinal curvature, wasting of thigh and pectoral muscles and joint contractures. In the absence of interventions, leg braces may be needed by ages 8-9 and a wheelchair by ages 10 to 12.

If a boy survives to late teens complications may appear such as cardiomyopathy, increased susceptibility to respiratory infections, coughing and respiratory failure. Poor respiratory function can include headaches, difficulty concentrating and poor sleep patterns. Dysmotility of the gastrointestinal tract may also develop, with associated constipation and diarrhea. A recent descriptive study of the health profile of a cohort of DMD patients tracked from 1982 to 2012 from the Muscular Dystrophy Tracking and Research Network (MD STARnet) found that median years for critical milestones were: development of cardiomyopathy 16.7 years, initiation of non-invasive ventilation 18.0 years, gastroenterology 19.0 years and death 21.8 years¹⁰.

To add further to the burden of disease is the cognitive and neurobehavioral profile of DMD patients. A retrospective study of 59 boys found 27% had a full scale IQ of < 70, 44% with a learning disability 19% with intellectual disability, attention deficit hyperactivity disorder (ADHD) in 32%, autism spectrum disorders in 15% and anxiety in 27%¹¹. Further evidence of neuropsychological comorbidities from a self-report group (n=228) found 7.9% were considered to have epilepsy together with other brain-related comorbidities, to include ADHD, obsessive compulsive disorder, anxiety disorders and sleep disorders¹². Overall, approximately one third of DMD patients may have various elements of cognitive impairment.

Chronic pain is also present in DMD patients, while often overlooked. A recent review of pain experience concluded that pain is pervasive in the DMD population, beginning in adolescence with an estimated two thirds reporting experience of pain¹³.

At the same time, life expectancy has increased in DMD populations with the advent of corticosteroids and the application of better interventional techniques such as non-invasive positive pressure ventilation. Current estimates are that on average life expectancy has increased from 19 to 26 years, with some surviving into their 30s and 40s. At the same time, while the evidence from long-term data sets such as the TREAT-NMD Global Database indicate that while average life

expectancy has increased there is associated greater disease variability in terms of progression¹⁴.

Value Claims in DMD

Understanding the complexity of DMD in its presentation and associated comorbidities is critical in any appraisal of its impact on patients, caregivers and the family group. If, for example, quality of life (QoL) or health related quality of life (HRQoL) are considered relevant outcomes in establishing the value case for interventions such as exon-skipping therapies, then care must be taken to ensure that measures for both patients and caregivers are appropriate. Value claims that rest on incomplete or inappropriate measures are not acceptable. At the same time, if value claims are expressed in QoL or HRQoL terms then these must be credible and evaluable.

The same arguments apply to claims for intervention costs. If value proposition focus on comparative costs of, for example, a corticosteroid such as prednisone and the additional costs of an exon-skipping therapy, then these costs must be expressed in evaluable terms. Making claims for lifetime costs that are, by definition, non-evaluable hardly seems a useful endeavor.

The requirement for testable hypotheses in the evaluation and provisional acceptance of claims made for products and devices is unexceptional. Since the 17th century, it has been accepted that if a research agenda is to advance, if there is to be an accretion of knowledge, there has to be a process of discovering new facts. Indeed, as early as the 16th century Leonardo da Vinci (1452 – 1519) in notes that appeared posthumously in 1540 for his *Treatise on Painting* (published in 1641) clearly anticipated the standards for the scientific method which were widely embraced a century later in rejecting thought experiments that fail the test of experience. By the 1660s, the scientific method, following the seminal contributions of Bacon, Galileo, Huygens and Boyle, had been clearly articulated by associations such as the Academia del Cimento in Florence (1657) and the Royal Society in England (founded 1660; Royal Charter 1662) with their respective mottos *Provando e Riprovando* (prove and again prove) and *nullius in verba* (take no man's word for it)¹⁵.

By the early 20th century, standards for empirical assessment were put on a sound methodological basis by Popper in his advocacy of a process of 'conjecture and refutation'^{16 17}. Hypotheses or claims must be capable of falsification; indeed, they should be framed in such a way that makes falsification likely. Life becomes more interesting if claims are falsified because this forces us to reconsider our models and the assumptions built into those models. This leads, then, to the obvious point that claims or models should not be judged on the realism or reasonableness of assumptions as representing an 'acceptable' imaginary perception of an unknown and unknowable future 'reality'. We should not 'take their word for it'; rather, we should reject it as pseudoscience.

Although Popper's view on what demarcates science (e.g., natural selection) from pseudoscience (e.g., intelligent design) is now seen as an oversimplification involving more than just the criteria of falsification, the demarcation problem remains¹⁸. Certainly, there are different ways of doing science but what all scientific inquiry has in common is the 'construction of empirically verifiable theories and hypotheses'. Empirical testability is 'one major characteristic distinguishing science from pseudoscience'; theories must be tested against data. Indeed paradoxically, while the development of pharmaceutical products and the evidence standards required by the Food and Drug Administration (FDA) for product evaluation and marketing approval are driven by adherence to the scientific method, once a product is launched and claims made for cost-effectiveness and, in the case of ICER, pricing and access recommendations, the scientific method is put to one side.

The rejection of a research program that meets the standards of normal science by groups such as ICER is best exemplified by the latest version of the Canadian health technology guidelines where it is stated: *Economic evaluations are designed to inform decisions. As such, they are distinct from conventional research activities, which are designed to test hypotheses*¹⁹. While this position puts modeled health technology assessment in the category of pseudoscience, it is also what may be described as a relativist position. Rather than subscribing to the position that the standards of normal science are the only standards to apply in health care decisions and value claims, the relativist believes that all perspectives are equally valid. Health care decisions are to be understood sociologically. No one body of evidence is superior to another. Results of a lifetime modeled simulation are on an equal basis with those of a pivotal Phase 3 randomized clinical trial. For the relativist, the success of a scientific research program, in this case one built on hypothetical models and simulations, rests not on its ability to generate new knowledge but on its ability to mobilize the support of the community. Basing decisions on models and simulations underpins the consensus view that evidence is constructed, never discovered. Instead of coming to grips with reality science is about rhetoric, persuasion and authority. Truth is consensus.

The ICER Reference Case

Central to the ICER construction of value claims is the reference case. Standards for model building, the construction of imaginary worlds, are clearly stated with the preference for models that take a long-term or lifetime perspective with value propositions expressed in cost-per QALY terms. Once a cost-per-QALY estimate (or estimates under different scenarios) has been constructed, the acceptability of a proposed product price is then assessed against cost-per-QALY willingness to pay thresholds (typically \$50,000, \$100,000 and \$150,000 per QALY with exceptions for higher cut-offs for rare diseases). Whether a product 'adds value' is then determined in terms of its impact on estimated lifetime modeled QALYs set against a proposed

lifetime product cost where both are driven by hypothetical constructed evidence. .

Unfortunately, the position taken by the Canadian guidelines reflects the consensus view in health technology assessment. Over the past 30 years, literally thousands of modeled claims have been presented in the literature, including leading health technology assessment journals. Annual reviews of the status of cost-effectiveness or modeled claims in the three journals, *Pharmacoeconomics*, *Value in Health* and the *Journal of Medical Economics* found that the majority of models presented non-evaluable claims (typically lifetime cost-per-QALY)^{20 21 22 23 24}. Where models were funded by a manufacturer a high proportion supported, in their modeled cost-per-QALY assessment, the manufacturer's product. While subscribing to modeling standards proposed by, for example, the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) and the Academy of Managed Care Pharmacy (AMCP) all too many of the papers were essentially marketing exercises^{25 26}.

At the same time with the emergence of the National Institute of Health and Care Excellence (NICE) in the UK in the late 1990s, the creation of imaginary worlds has received widespread support²⁷. Versions of the NICE reference case are found in the majority of single payer health systems.

The point, of course, is that in advocating a reference case to support value propositions and the case for price discounts, it was never intended that the value claims expressed in cost-per-QALY terms would be empirically evaluable. They were, as made clear by the CADTH guidelines, 'for information only'. The only basis on which a reference case value claim could be challenged was by using the same reference standards to construct a different model. A model whose claims would similarly be non-evaluable. The debate would then center on the choice of assumptions with each side arguing that theirs is a more realistic yet imaginary construct.

To be fair, however, if an agency such as NICE or the Pharmaceutical Benefits Advisory Committee in Australia (PBAC) require submissions for product reviews to subscribe to a reference case structure then, irrespective of concerns as to scientific status, manufacturers are obliged to create models that meet those standards²⁸. This does not mean that the US has to follow suit. The fact that ICER has taken upon itself a standards setting role is beside the point. There is no reason why formulary committees and health decision makers should accept the ICER reference case as relevant to health care decisions, particularly if it is seen as failing the standards of normal science.

The ICER Reference Case specifies the approach that ICER and its collaborators follow for cost-effectiveness analyses in constructing cost-per-QALY imaginary worlds²⁹. This is little different from the initial reference case standards established

by NICE. However, unlike agencies such as NICE and the PBAC, where the reference specifications are to inform manufacturers of the standards required for modeled submissions to the review agency, the ICER reference case is to establish the framework for the development of the ICER in-house model. Manufacturers are not asked to submit their own model, although this would be relatively easy if they have already developed models for other jurisdictions, but to merely have limited access to the ICER model as it is developed, with the opportunity to suggest modifications but without the possibility of major modifications or competing models. Although unlikely, it would be of interest to ask an academic center in the UK or Australia to undertake a review of ICER models with the results in the public domain.

The options open for competing models within a reference case framework is readily seen in the response and transparency of NICE and its reviewers to manufacturer's models. The manufacturer's model may be accepted without change, accepted with modifications or rejected and an alternative model proposed. Critiques of the models proposed by manufacturers are not, of course, in terms of the empirical status of claims. This would be impossible given the reference case timeframes.

The ICER DMD Model

The objective of the ICER model was to 'assess the cost-effectiveness of deflazacort, eteplirsen and golodirsen for patients diagnosed with DMD in the US'. This translates to a model to judge the appropriateness of current pricing. In the case of deflazacort the comparison was between deflazacort and supportive care and prednisone and supportive care. Eteplirsen and golodirsen were assessed as add-on therapy to corticosteroids and supportive care and corticosteroids and supportive care alone. However, in the case of eteplirsen and golodirsen 'there was insufficient evidence to model specific treatment effects'.

The model proposed by ICER in the construction of the imaginary DMD world is a multi-state partitioned survival model as it was considered to be the best fit to data available on DMD progression. The model was applied to a hypothetical cohort of patients with DMD initiating treatment at 5 years of age. Patients were partitioned to progress through four health states (early ambulatory, late ambulatory, early non-ambulatory and later non-ambulatory) followed by death. Costs and QALYs were assigned to each stage with aggregate QALYs evaluated for the cohort as its members progressed by assumption through these stages. Costs and QALYs were discounted.

The ICER model is only one of a number of alternative DMD imaginary worlds that have been published. Landfeldt et al proposed a number of cost-effectiveness model frameworks for DMD, all of which were lifetime models utilizing a Markov state-transition approach³⁰. These presented the case for a

hypothetical intervention versus standard of care. None met the standards required for credible and evaluable claims.

Also worth noting is the NICE assessment of the role of altaluren in DMD³¹. Although this assessment takes a UK perspective and resulted in an agreement for discounting and other financial incentives, it is of interest because of the models proposed and reviewed by the NIICE expert review group all of which is in the public domain. Again, as the models presented follow the requirements of the NICE reference case, the claims fail to meet standards for credibility and evaluation.

ICER: Evidence Base

If we accept, for the purposes of argument, a role for imaginary lifetime cost-per-QALY models, then a reasonable question is how robust must the evidence base be to go forward with modeling? ICER acknowledges that in modeling DMD the evidence base was sparse and insufficient to model specific treatment effects for eteplirsen and golodirsen, including in the latter case an analysis comparing it to corticosteroids and supportive care alone. In addition, the ICER report makes the following points in listing the key model assumptions:

- There was no evidence based mechanism for relating treatment effects to the proportion of years spent in early and late ambulation or early and late non-ambulation
- Early and late ambulatory stages are not explicitly defined in the international trial data so the proportions by age reported in the original published model were applied
- There is insufficient evidence to establish treatment effects on time in ambulation beyond rightward shifts in the survival curves. In the absence of long-term data, rightward shifts in the ambulation curve along with equivalent shifts in the mortality curve offered a reasonable upward bound approximation of a treatment effect.

If there is, as ICER notes, insufficient evidence to model specific treatment effects, the question is why ICER continued to press on with its analysis and recommendations. However, as needs must, ICER pressed forward applying their model to see whether there were (as noted above), by assumption, 'potential' yet imaginary treatment effects. This was achieved by shifting the ambulatory and mortality curves for 10, 20, 40 years, and applying this imaginary framework to explore 'resulting incremental cost-effectiveness ratios given the eteplirsen current price'. Whether this meets any criteria of common sense is an open question. To assume that current prices would be maintained over these timeframes appears odd, let alone the neglect of the potential market approval and entry of new therapies targeted at specific DMD gene mutations, together with market entry and subsequent pricing decisions by their manufacturers.

The FDA Perspective

It is worth noting in the case of eteplirsen, where the target gene mutation is present in approximately 13% of the DMD population, the conclusions by the FDA Director for Drug Evaluation and Research in finding that the data submitted by the company met the standard for the accelerated approval of the drug³². In disagreeing with a number of the findings from a committee review, the Director found that based on the surrogate endpoint of increased dystrophin protein production there was a reasonable likelihood of predicting clinical benefit. This conclusion on the surrogate endpoint, given the deficiencies recognized in the development program by the FDA 'represents the greatest flexibility possible for FDA while remaining within its statutory framework'. This flexibility, the Director continued, 'is warranted because: (i) the life-threatening nature of the disease; (ii) the lack of available therapy; (iii) the fact that the intended population is a small subset of an otherwise rare disease; and (iv) the fact that this is a fatal disease in children'. The FDA, in allowing for the use of the accelerated approval pathway recognized that DMD is a difficult disease to study and develop treatments. It is both rare, presents in heterogeneous forms and lacks well established endpoints that span the full spectrum of the disease.

ICER Model: Measuring and Valuing Health

The ICER reference case stipulates that in measuring and valuing health outcomes, the instrument used for assessing health preferences should (i) reflect those of the general US population and (ii) describe those preferences usually from an indirect measurement based on a generic classification system.

In the case of DMD, there are two groups of interest: (i) the DMD patient and (ii) the caregiver(s) and family. While utility estimates for those two groups have been proposed and applied in the construction of the ICER DMD construct, the issue as noted by Uttley et al, is whether in the case of DMD we are focusing on quality of life (QoL) or health related quality of life (HRQoL)³³. The two constructs are different, although the terms are often used interchangeably. QoL is a multidimensional construct capturing physical health, personal, including psychological, health, beliefs, social relations and environmental influences. HRQoL is narrower in focusing on the patient's (and caregiver(s) in pediatric and young adolescent populations) assessment of the impact of illness on physical, psychological and social interactions. In DMD, given the complexity of the disease state, the broader QoL construct may be the more relevant measure, in its application to both patients and caregivers. In the case of the ICER model, the utility measures adopted are focused on HRQoL, not the broader QoL. There was apparently no choice as these QALY measures were the only ones available. Irrespective, therefore, of potential shortcomings, ICER adopted them to model their QALY claims.

Even so, there is no agreement on what the broader QoL construct should embrace. Bann et al, in proposing a comprehensive model of QoL in muscular dystrophy point out

that in their review of QoL scales none captured all of the domains identified in their model³⁴. Ryder et al, in an evidence review of DMD noted that while the Pediatric Quality of Life Inventory (PedsQL) is the most frequently used tool for measuring HRQoL it does not correlate well with disease progression while a recent review of the range of muscular dystrophy conditions in adults point to the need for a greater understanding of mental wellbeing, independence and the management of fatigue and pain^{35 36}.

In constructing the modeled DMD imaginary world, the ICER options were limited to the application of utilities from a single paper presenting utility estimates for the stage of disease scored from the Health Utilities Instrument Mk 3 (HUI Mk3) for the hypothetical DMD patient population³⁷. This instrument utilizes preferences from a Canadian population, parents with schoolchildren, and not a representative US population. The HUI Mk 3 (and its earlier Mk 2 version) includes a health status classification system scoring system based on standard gamble utilities with the scores on the conventional 0 – 1 scale. The scoring formula is a multiplicative multi-attribute utility function. A visual analogue scoring is also available, though seldom used. A time-trade off technique with econometric modeling to generate the scoring function the scoring function

The ICER report bases its estimates of caregiver burden on a single study by Landfeldt et al³⁸. This analysis is designed to assess subjective burden associated with DMD. It utilizes a range of instruments supported by supplementary question to give an overall picture of caregiver burden. Specifically the EQ-5D-3L, a visual analogue scale and the SF-12 Health Survey supported by the Zarit Caregiver Burden Interview together with supplementary caregiver questions to link patient and caregiver HRQoL (caregiver assessment of their son's current health and mental status). ICER only utilizes the EQ-5D-3L mean utility scores from the Landfeldt et al study to support the reference case model for its caregiver scenario analysis.

If we are concerned with measuring the QoL or HRQoL of DMD patients and their caregivers then, if the HUI Mk3 is the centerpiece to ICER assumptions regarding patient utilities and the EQ-5D-3L caregiver utilities then it is appropriate to consider four questions³⁹:

- Does the instrument proposed, its classification system, cover the attributes and the level of those attributes that are likely to be important to the patient population and caregivers under study?
- Has the proposed instrument been used in similar patients and caregiver populations and was in response to therapy interventions and the natural course of the disease?
- Is the instrument likely to be responsive to the changes expected in the study population and caregivers in constituent population groups?

- Does the instrument have ceiling effects or floor effects that will reduce its sensitivity for the patients and caregivers under study?

While the HUI Mk3 has been used widely, the question is whether or not it is appropriate in capturing the range of factors that may impact QoL, as opposed to HRQoL in a DMD population. If the claim is for HRQoL then the case needs to be made for excluding factors that may be appropriate in QoL. Importantly, and this point is emphasized by Uttley et al, a condition specific instrument for DMD does not currently exist. While it is possible to identify QoL domains and sub-domains from DMD studies for both patients and caregivers, the key point is that these domains are unlikely, following Uttley et al, to be addressed by generic preference based instruments such as the HUI Mk3. Indeed, where alternative generic HRQoL instruments have been used the results are either inconclusive or contradictory. The authors conclude: *“This research highlights that factors relevant to QoL in DMD spans all domains of the included HRQoL instruments and beyond. Many of the nuances specific to DMD such as the impact on the carer and wider family, are not currently captured by standard HRQoL instruments. The breadth and diversity of themes and emerging trends in this review (e.g., impact of age within DMD) highlight the need to consult with patients and their families about what aspects of QoL need to be covered by the QoL instruments used.*

Although ICER, in referencing the Uttley paper, was aware of the criticisms of the application of generic HRQoL instruments to evaluating QoL in DMD patients (and the application of the EQ-5D-3L generic instrument in representing by assumption caregiver preferences) they nevertheless went ahead in applying the preference scores from the HUI Mk3 and the EQ-5D in their model building. These scores are fundamental to their cost-per-QALY and willingness to pay threshold claims. Rather than ICER stepping back and allowing the development of a DMD instrument capturing the preferences of both patients, caregivers and providers the need to generate a report and recommendations based upon these instruments were, apparently, seen as paramount.

Of course, if we reject the ICER modelling approach on the grounds that it fails to meet the standards of normal science, then this point becomes moot, What is important, however, is not the development of competing imaginary worlds, but the development of instruments that capture the QoL of patients, caregivers providers and one which can be used to evaluate the impact of competing therapies.

Family Impact of DMD

The impact of DMD is on the patient and the patient’s family, to include those family members as caregivers who are most involved with supporting the patient. If there is a QoL impact then, presumably, it should be evaluated for the family unit as a whole ⁴⁰. It seems odd to focus on the QoL of the patient to the exclusion of the caregiver(s) and the caregiver(s) to the

exclusion of the patient. The ICER modeling of HRQoL (not QoL) ignores this possibility and fails to offers a measure that combines the measurement of family QoL in assessing how increased survival may impact the aggregate of QoL experience in the family unit. To argue that the utility scores assumed relevant for the ICER model are the only ones available and should therefore be used is not acceptable.

Imaginary Voting

Following the final evidence review ICER convenes a public meeting with a panel to present a summary of the claims made. At the end of the meeting the panel is asked to vote on a series of questions. For DMD the ICER-convened New England Comparative Effectiveness Public Advisory Council (CPAC) was held in Boston on July 2019. The final question (of 9 asked) was an overview for eteplirsen “Given the available evidence on comparative effectiveness and incremental cost-effectiveness and considering other benefits, disadvantages and contextual considerations, what is the long term value for money of treatment with eteplirsen versus supportive care and corticosteroids alone”. Responses were ‘low’, ‘intermediate’ and ‘high’; there was no option to abstain.

As argued in this commentary not only is the evidence base limited for DMD as a rare disease which has received accelerated approval, but any claims for incremental cost-effectiveness and claims for value for money are created by a construct that does not meet the standards of normal science. The panel was asked, therefore, to vote for an imaginary world with no credible and evaluable claims. Whether panel members realized this is an open question. The result: 16 voted for ‘low’ imaginary value and 1 for an ‘intermediate’ imaginary value for the 17 panel members.

Project Hercules

The rush to judgement by ICER may be contrasted to Project Hercules in the UK at the School of Health and Related Research, University of Sheffield⁴¹. This is a multinational collaboration initiated in 2017 by Duchenne UK to develop tools and evidence to support health technology assessments and reimbursement decisions for new treatments in DMD.

Deliverables include:

- A critique of commonly used quality of life metrics (completed: see reference)
- A natural history model of DMD
- A core economic model
- A bespoke quality of Life metric
- A guide to health technology Assessment
- A burden of illness study

Although not detailed by ICER in their report, Project Hercules is of particular interest in the recognition that there is no condition specific preference-based measure of DMD. They propose to develop the first condition specific measure for DMD to use in cost-effectiveness studies⁴². This is intended to

overcome deficiencies in standard generic measures, including the HUI Mk3 and the EQ-5D-3L that are a cornerstone of the ICER QALY estimates. For example, neither the HUI Mk3 nor the EQ-5D adequately capture fatigue, social participation or dignity. The Hercules study is designed (i) to evaluate the content validity of these generic measures including the EQ-5D-Y for youths and the Child Health Utility 9D while (ii) generating new content for either a condition specific instrument or a bolt on addition to an existing instrument.

Stifling Innovation

Basing recommendations for pricing of innovative products, particularly those targeting rare disease populations, on the constructed evidence of an imaginary world must raise concerns about the impact of ICER's hypothetical claims on the willingness to invest in these products. Unlike the FDA who in the case of eteplirsen granted accelerated approval status based on a surrogate endpoint following an exhaustive evidence review, the ICER recommendations are based on constructed evidence. As emphasized above, constructing evidence to support pricing recommendations is, from the standpoint of the methods of normal science, unacceptable. A show of hands to vote on whether or not the ICER constructed modeled evidence demonstrates cost-effectiveness is hardly satisfactory.

ICER notes in the preamble to the DMD report that 'Readers should be aware that new evidence may emerge following the publication of this report that could potentially influence the results. ICER *may* (emphasis added) revise its analysis in a formal update to this report in the future'. While this is laudable it is unnecessary, the fact remains that in its rush to judgement ICER has not only produced recommendations from a constructed imaginary world while also recognizing that the evidence base is sparse. Promising further DMD imaginary world iterations is hardly reassuring. A more appropriate approach might have been to follow the initiative of Project Hercules and support the establishment of a more comprehensive evidence base to support credible and evaluable claims.

Conclusions

As detailed in this commentary and in previous commentaries on ICER's reference case methodology, the approach taken by ICER in constructing imaginary worlds to support pricing recommendations fails to meet the standards of normal science. The approach should be seen as pseudoscience and any conclusions reached subject to this caveat.

If we accept the position taken by the CADTH guidelines that the role of model building in health technology assessment is to inform decisions, a reasonable question is how a health system

decision maker is to value the information provided. How is the decision maker to judge the weight to be given to a non-evaluable cost-per-QALY claim? How does the decision maker judge the 'take our word for it' approach of ICER if there are concerns regarding differing frameworks and outcome scenarios offered by competing models?

To argue that the role of hypothetical lifetime models is to provide information to formulary committees and others is beside the point. Any model can be challenged on its structure and its supporting assumptions. Even, if as seems likely, the proposed Project Hercules preference based measure calibrated for measuring HRQoL (QoL?) in DMD patients and caregivers is judged adequate, this does not mean that its presence justifies its application in reference case models. The model, unless the claims are credible and evaluable, would still fail to meet the standards of normal science. The same argument would apply to 'improved' direct medical cost measures even if they include costs borne by caregivers and the family. It seems appropriate to abandon the phrase 'for information' in favor of 'knowledge', accepting a framework in health technology assessment that focuses on technology adoption and credible, evaluable and replicable knowledge accumulation.

This ICER position is made that more questionable by the likelihood that no one on a formulary committee has any skills in model building; the ICER model is a 'black box'. It would be probably unwise to fall back on the ICER 'voting' procedures as those voting, as noted, are probably also unfamiliar with model building. As it stands, health care decision makers have a number of DMD models to choose from in addition to the ICER model, particularly with reference to competing estimates of QALYs. Additional models are in production, notably model(s) associated with the DMD Project Hercules. All would be rejected if they were focused on reference case scenarios.

Constructing imaginary worlds to justify product pricing and access decisions should be seen as an analytical dead end. As noted, we have no idea whether the claims and their supporting models are 'right' or if they are 'wrong' both in their structure, assumptions and conclusions, and we will never know. Indeed, with a hypothetical cohort, in this case of 5-year old DMD patients tracked by assumption over the next 10, 20 or 30 years we were never intended to make a judgement. All ICER offers are recommendations based on the ratio of assumed lifetime QALYs to assumed lifetime costs.

Conflicts of Interest

PCL is a consultant to Patients Rising, a public interest advocacy group.

References

- ¹ Institute for Clinical and Economic Review. Deflazacort, Eteplirsen, and Golodirsen for Duchenne Muscular Dystrophy: Effectiveness and Value, Evidence Report August 15, 2019 https://icer-review.org/wp-content/uploads/2018/12/ICER_DMD-Final-Report_081519-1.pdf
- ² Langley PC. Cost-Effectiveness and Formulary Evaluation: Imaginary Worlds and Entresto Claims in Heart Failure. *Inov Pharm.* 2016;7(3): No. 6 <https://pubs.lib.umn.edu/index.php/innovations/article/view/449>
- ³ Langley PC. Multiple Sclerosis and the Comparative Value Disease Modifying Therapy Report of the Institute for Clinical and Economic Review (ICER). *Inov Pharm.* 2017;8(1): No. 12 <https://pubs.lib.umn.edu/index.php/innovations/article/view/492>
- ⁴ Langley PC. Imaginary Worlds and the Institute for Clinical and Economic Review (ICER) Evidence Report: Targeted Immune Modulators for Rheumatoid Arthritis. *Inov Pharm.* 2017;8(2): No. 10. <https://pubs.lib.umn.edu/index.php/innovations/article/view/515>
- ⁵ Langley PC. Rush to Judgement: Imaginary Worlds and Cost-Outcomes Claims for PCSK9 Inhibitors. *Inov Pharm.* 2017;8(2): No. 11 <https://pubs.lib.umn.edu/index.php/innovations/article/view/516>
- ⁶ Langley PC. Another Imaginary World: The ICER Claims for the Long-Term Cost-Effectiveness and Pricing of Vesicular Monoamine Transporter 2 (VMAT2) Inhibitors in Tardive Dyskinesia. *Inov Pharm.* 2017;8(4): No 12 <https://pubs.lib.umn.edu/index.php/innovations/article/view/927>
- ⁷ Langley PC. Resolving Lingering Problems or Continued Support for Pseudoscience? The ICER Value Assessment Update. *Inov Pharm.* 2017;8(4): No 7 <https://pubs.lib.umn.edu/index.php/innovations/article/view/933>
- ⁸ Langley PC. Transparency, Imaginary Worlds and ICER Value Assessments. *Inov Pharm.* 2017;8(4): No 11 <https://pubs.lib.umn.edu/index.php/innovations/article/view/926>
- ⁹ Langley PC. Alternative Facts and the ICER Proposed Policy on Access to Imaginary Pharmacoeconomic Worlds. *Inov Pharm.* 2018;9(2): No. 10 <https://pubs.lib.umn.edu/index.php/innovations/article/view/1300>
- ¹⁰ Pandya S, James K, Westfield C et al. Health profile of a cohort of adults with Duchenne muscular dystrophy. *Muscle Nerve.* 2018;58(2);219-23
- ¹¹ Banihani S, Smile S, Yoon G et al. Cognitive and neurobehavioral profile in boys with Duchenne muscular dystrophy. *J Child Neurol.* 2015;30(11):1472-82
- ¹² Hendriksen R, Vles J, Aalbers M et al. Brain-related comorbidities in boys and men with Duchenne muscular dystrophy: A descriptive study. *Eur J Paediatr Neurol.* 2018;22(3):488-97
- ¹³ Jacques M, Stockley R, Bostock E et al. Frequency of pain in Adult males with muscular dystrophy. *PLoS One.* 2019;24(2):e0212437
- ¹⁴ Koeks Z, Bladen C, Salgado D et al. Clinical Outcomes in Duchenne MuscularSDystrophy: A study of 5345 patients from the TREAT-NMD DMD global database. *J Neuromuscular Diseases.* 2017;295-306
- ¹⁵ Wootton D. *The Invention of Science: A new history of the scientific revolution.* New York: Harper Collins, 2015
- ¹⁶ Popper KR., *The logic of scientific discovery* .New York: Harper, 1959.
- ¹⁷ Lakatos I, Musgrave A (eds.). *Criticism and the growth of knowledge.* Cambridge: University Press, 1970.
- ¹⁸ Piglucci M. *Nonsense on Stilts: How to tell science from bunk.* Chicago: University of Chicago Press, 2010)

- ¹⁹ Canadian Agency for Drugs and Technologies in Health (CADTH). Guidelines for the economic evaluation of health technologies: Canada. Ottawa: CADTH, 2017
- ²⁰ Langley PC. Imaginary worlds: Modeled claims for cost-effectiveness published in *PharmacoEconomics* January 2015 to December 2015. *Inov Pharm* . 2016;7(2): No. 9.
<https://pubs.lib.umn.edu/index.php/innovations/article/view/432>
- ²¹ Langley PC, Rhee TG. Imaginary worlds: A systematic review of the status of modeled cost-effectiveness claims published in the *Journal of Medical Economics* from January 2015 to December 2015. *Inov Pharm* . 2016;7(2): No. 16.
<https://pubs.lib.umn.edu/index.php/innovations/article/view/439>
- ²² Langley PC, Rhee TG. Imaginary worlds: The status of modeled economic evaluation claims published in *Value in Health* January 2015 to December 2015. *Inov Pharm* . 2016;7(2): No. 18.
<https://pubs.lib.umn.edu/index.php/innovations/article/view/441>
- ²³ Langley PC, Rhee TG. Imaginary worlds: A systematic review of the status of modeled cost-effectiveness claims published in the *Journal of Medical Economics* from January 2015 to December 2015. *Inov Pharm* . 2016;7(2): No. 16.
<https://pubs.lib.umn.edu/index.php/innovations/article/view/538>
- ²⁴ Langley PC, Rhee T. The Imaginary Worlds of ISPOR: Modeled Cost-Effectiveness Claims Published in *Value in Health* from January 2016 to December 2016. *Inov Pharm* . 2017;8(2): No. 14
<https://pubs.lib.umn.edu/index.php/innovations/article/view/538>
- ²⁵ Langley PC. Validation of modeled pharmacoeconomic claims in formulary submissions. *J Med Econ* . 2015;18(12):993-99
- ²⁶ Langley PC. Modeling imaginary worlds: Version 4 of the AMCP Format for Formulary Submissions. *Inov Pharm* . 2016;7(2): No. 11
<https://pubs.lib.umn.edu/index.php/innovations/article/view/434>
- ²⁷ Langley PC. Sunlit uplands: the genius of the NICE reference case. *Inov Pharm* . 2016;7(2): No.12.
<https://pubs.lib.umn.edu/index.php/innovations/article/view/435>
- ²⁸ 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 <https://pubs.lib.umn.edu/index.php/innovations/article/view/485>
- ²⁹ ICER. ICER's Reference Case for Economic Evaluations: Principles and Rationale. July 2018
- ³⁰ Landfeldt E, Alfredsson L, Straub V et al. Economic evaluation in Duchenne Muscular Dystrophy: Model frameworks for cost-effectiveness analysis. *Pharmacoeconomics* . 2017;35:249-58
- ³¹ National Institute for Health and Care Excellence. Ataluren for treating Duchenne muscular dystrophy with a nonsense mutation in the dystrophin gene. Highly specialised technologies guidance [HST3] Published date: 20 July 2016
- ³² Food and Drug Administration. Center Director Decisional Memo. https://parentprojectmd.org/wp-content/uploads/2018/04/Exondys51_FDA_Approval_Decision_Memo.pdf
- ³³ Uttley L, Carlton J, Woods HB et al. A review of quality of life these in Duchenne muscular dystrophy for patients and carers. *Health Qual Life Outcomes*, 2018;16:237
- ³⁴ Bann CM, Abresch RT, Biesecker B et al. Measuring quality of life in muscular dystrophy. *Neurology* . 2015;84:1034-42
- ³⁵ Ryder S, Leadley RM, Armstrong N et al. The burden, epidemiology, costs and treatment for Duchenne muscular dystrophy. *Orphanet J Rare Diseases* . 2017;12:79
- ³⁶ Jacques MF, Stockley RC, Onambele-Pearson GL et al. Quality of life in adults with muscular dystrophy. *Health Qual Life Outcomes* . 2019;17:121

³⁷ Landfeldt E, Lindgren P, Bell CF, et al. Health-related quality of life in patients with Duchenne muscular dystrophy: a multinational, cross-section study. *Dev Med Child Neurol*. 2015;58:508-15

³⁸ Landfeldt E, Lindgren P, Bell C et al. Quantifying the burden of caregiving in Duchenne muscular dystrophy. *J Neurol*. 2016;263:905-15

³⁹ Drummond M, Sculpher M, Torrance G et al. *Methods for the Economic Evaluation of Health Care Programmes*. 3rd Ed. New York: Oxford University Press, 2005

⁴⁰ Summers J, Poston D, Turnbull A et al. Conceptualizing and measuring family quality of life. *J Intellectual Disab Res*. 2005;49(10):777-83

⁴¹ <https://hercules.duchenneuk.org/about-project-hercules/>

⁴² Powell P, Carlton J, Rowen D et al. Producing a preference-based quality of life measure for people with Duchenne muscular dystrophy: a mixed-methods study protocol. *BMJ Open* 2019;9e023685