MAIMON WORKING PAPERS No. 17 SEPTEMBER 2023

REPLICATION: SCIENTIFIC OBJECTIVITY AND FALSE CLAIMS IN HEALTH TECHNOLOGY ASSESSMENT

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

Abstract

Replication of value claims is a cornerstone for progress and the discovery of new facts. Unfortunately, Health Technology Assessment (HTA) falls short of this standard in its endorsement of the creation of evidence to support formulary decisions with assumption driven modeled simulations. The concept of replication as a necessary component of the standards for normal science is absent. In short, rather than promoting these standards to support a belief in the process of discovery, the mind-independent reality of objective knowledge, HTA focus on the mind-dependent fantasy of invented claims. This rejects discovery in favor of one-off modeled claims that are entirely imaginary. If we are to have faith in the claims made for pharmaceutical products this must be addressed, not by assumption driven simulations embodying clinical claims, but a commitment to the evaluation of single attribute clinical claims, where these claims are presented in terms of fundamental measurement. This allows us to assess directly the merits of the clinical claim utilizing real world evidence where submissions to formulary committees include agreed protocols to undertake short term yet meaningful assessments. Forcing manufacturers to propose and underwrite replication studies has advantages in minimizing the impact of publication bias and the proliferation of false claims in HTA. We are in a situation where claims from individual trials, meta-analyses and systematic reviews should all be treated with caution; no claim should be taken at face value. They must be considered provisional until reaffirmed by one or more replication studies. This brings us back to the standards for discovery and objectivity in therapeutic intervention which should be, as far as can be determined, value free; a view from nowhere. The purpose of this commentary is: (i) to make clear what is meant by scientific objectivity and objective knowledge; (ii) to make clear that false claims have badly undermined our faith in clinical claims; and (ii) point to a commitment to replication as an effective counterweigh.

INTRODUCTION

Health technology assessment (HTA) has given scant attention to the replication of value claims for pharmaceutical products [1]. This is unfortunate because the degree of 'faith' or 'belief' that we have in value claims for a product rest in large part on our ability to replicate and confirm those claims in real world treatment situations [2] . To support a continued belief in clinical claims we require a structured reporting of a possible proliferation of positive confirmations in treating situations; a reporting which assumes, unwisely, of a balance in the opportunities provided to report positive and negative findings. While this is no doubt a laudable and essential activity to create evidence in support of a product, the literature over the past 50 years has put the lie to this objective. There are few pharmaceutical products which have been assessed in a balanced reporting of clinical value [3]. Typically, only positive value claim assessments are reported in the literature. The balance is therefore towards an ongoing cumulative positive confirmation of the value claims. A key issue is the extent to which these claims can be trusted when false claims may achieve equal billing with those that meet the required professional standards [4]

This is not a concern in HTA; the focus on assumption driven simulations with non-falsifiable claims has effectively expedited and insulated HTA value claims from any commitment to replication [1]. With simulated modeled claims it is impossible to assess the claim; this first level of defense is reinforced by an analytical framework that is deliberately designed to fend off attempts to assess claims.

At the same time, the criticism has been made on a number of occasions that HTA has turned its back on the standards of normal science and fundamental measurement in the pursuit of assumption driven simulated modelled imaginary claims [1]. This is not just a repudiation of the pursuit of what may be described as ontological objective knowledge but of the role of evidence in evaluating value claims. Clearly, imaginary modelled claims for cost-effectiveness claims are hardly a subject for meta-analysis or systematic reviews; not only because they are non-falsifiable but as one-off claims which are seldom, if ever, replicated due the assumption driven status.

In answer to this pursuit of imaginary value claims it has been proposed that we look to a new start in HTA, one that recognizes the standards of normal science and the evaluation of single attribute value claims that meet Rasch measurement standards [5]. While this framework has always been accepted in the process of product development, once the results of pivotal trials become available these are not only factored into simulated models but in the presentation of results are subject to bias in selective release and interpretation. The latest example of a guidance is the CHEERS 2022 guide to creating what we will describe as assumption driven simulated fantasy claims to HTA journals which, surprisingly, have committed themselves to accepting or providing an open door to assumption driven modeled non-falsifiable claims [6].

The purpose of this note is to make the case that if we are to embrace the standards of normal science and fundamental measurement, the watch keepers must also be formulary committees. Standards for formulary submission should be in place that ensure that the clinical evidence presented is robust, reporting all trials or observational study results for new products and devices with clinical value claims supported by protocols for replication. In short, with the present management in place of clinical study results, there can be a loss of faith in publications, systematic reviews and meta-analyses to justify value claims. This is unacceptable; if we are to focus on assessment in real world treating environments then a more rigorous review and tracking of protocols is required.

SCIENTIFIC OBJECTIVITY AND OBJECTIVE KNOWLEDGE

Scientific objectivity, often considered an ideal for scientific inquiry, manifests the idea that scientific claims, methods, results, including the scientists themselves, are not, or should not be, influenced by particular perspectives, value judgments, community bias or personal interests [7]. As a value, objectivity is considered to come in degrees; there is no absolute standard of objectivity. The notion that science provides a ‘view from nowhere’ is unattainable. Even so, the notion that there is an unattainable yet agreed reference point lies at the basis of the commitment to scientific realism and the resolution of factual disagreements. From a Kuhnian perspective science evolves within a paradigm, a successive series of attempts to approach an actual truth through puzzle solving [8]. Observations can undermine a paradigm which, in Kuhn’s terms means that in overturning a paradigm we overturn the meaning of observational concepts and the perceptions of the scientists working within that paradigm. This raises the question of whether one paradigm is more faithful to the facts when the same issues are being addressed. It is only within one paradigm or world view that we can agree on the meaning of the question of objectivity.

Concern with the objective nature of scientific enquiry raises a major issue: can any form of scientific inquiry be value free? The notion of the value free ideal, which aims to reduce as far as possible contextual issues as they relate to two activities: (i) the gathering of evidence in relation to the chosen research problem and (ii) the provisional acceptance of a scientific hypothesis or theory as a tentative adequate answer and (iii) feedback through a process of error elimination and replication is a key factor in Popper's view on objective knowledge [9] . Expressed in terms of what is described as the value neutrality thesis, the claim is that we can—at least in principle—gather evidence and assess/accept theories without making contextual value judgments. Following Popper, the two arguments that have had the most weight are: (i) science focuses on the acceptance or rejection of hypotheses; and (ii) no hypothesis is ever confirmed beyond reasonable doubt so that the decision to accept or reject a hypothesis reflects an implicit value judgement. The fact that within a community of scholars there may be agreement that the acceptance or rejection of a hypothesis should recognize the $p<5\%$ rule, this may be of little relevance to the decisions made by individual scientists.

What is all too often overlooked in arguments for values and contexts is that the acceptance of scientific theories is only one of several places for values to enter scientific reasoning. This includes implicit value judgements, experimental design, implementation of the study protocol, data characterization, choice of analytical standards and interpretation. All are presumably subject to the desired or required end product and publication. It is, if you like, a movable feast that can be justified to select items from a flexible menu.

Not surprisingly, value judgements can also have negative effects to the extent that they favor the dissemination of false claims through the suppression of null or negative findings and, in all too many cases, support the active manipulation or creation of data to support a preconceived belief in the strength of favored hypotheses or simply create an imaginary data set by reverse engineering from a required statistical endorsement . False claims are not going to go away. Certainly, we can mount a concerted effort to minimize their impact on clinical decision making but all too many people have too much to lose (or gain) from the promotion of false facts, particularly where power structures in universities and other organizations make whistleblowing an invitation to career suicide.

SCIENTIFIC REALISM

The commitment in HTA to assumption driven modelled simulation that create non-falsifiable (i.e., false) value claims for therapy choice, has been described as a meme that lacks any commitment to discovery and the standards of normal science [1]. It fails the demarcation test and must be seen, from an epistemological perspective, as equivalent to creationism or intelligent design; a mind-dependent fantasy [10]. It fails, in other words any link or correspondence to scientific realism or to the nature of scientific knowledge. In these terms scientific realism is *a positive epistemic attitude toward the content of our best theories and models, recommending belief in both observable and unobservable aspects of the world described by the sciences* [11]. Accepting the concept of scientific realism endorses positions that believe in the reality of something; a belief totally at variance to the imaginary worlds of HTA and false facts.

It is useful to consider scientific realism in terms of three dimensions of belief: a metaphysical or ontological dimension, a semantic dimension and an epistemological dimension. The metaphysical dimension is focused on realism: the mind-independent existence of a reality that is the subject of inquiry. A pseudo-reality that is not created in the terms of the practice of HTA simulation

modeling, but one that exists to be discovered. The semantic dimension is focused on the truth value of scientific claims; whether those claims are true or false, or observable or unobservable. The epistemological dimension embraces theoretical claims regarding a mind-independent reality as knowledge of the world; a correspondence theory of truth where theories provide approximate descriptions of observable and unobservable realities. There approximate descriptions, which are always provisional, stand in contrast to the HTA commitment to the nebulous term 'approximate information' or the created evidence from assumption driven simulations. HTA in other words is not committed to a concept of a mind-independent reality but to the opposite concept of a mind-dependent fantasy; discovery of new, yet provisional, facts is replaced by the creation of evidence to support truth claims.

SCIENTIFIC RIGOR

Our knowledge of a mind-independent reality is, by definition, imprecise. It is limited, for both observable and non-observable claims by the structure, content and application of our most accepted theories and by the recognition that for any applied theoretical framework truth, a provisional claim, is only approximate even if defended as converging in its modifications and applications to a more precise validation. In the case of drug development with the application of RCTs to support value claims the compounds themselves are part of the physical and chemical properties of fundamental particles at the quantum level. These objects, properties and interactions exist independently of the observer. They occur due to specific molecular structures and mechanisms of how drugs work in the body to treat or manage disease.

However, its essential to note that while the physical and chemical aspects of drug interactions are mind-independent, the translation and interpretation of their significance relies on human decisions in the design of instrumentation or measurement protocols such as RCTs. In other words, to establish, even provisionally, the truth status of a theory, we need to propose how the theory is to be tested. It is at this juncture that we face the issue of rigor in establishing that a theoretical claim does not produce true claims for false results; the question of underdetermination in science. This point was made over a century ago by Duhem (now Duhem-Quine thesis) who pointed out that a hypothesis cannot be used to derive testable predictions in isolation as auxiliary assumptions are required such as background theories and hypotheses about instruments and measurement [12] . If the hypothesis fails it is then important to review the auxiliary assumptions. A classic example of this is the so-called base rate fallacy where the lower the incidence of the disease at large, the lower the probability that a positive result signals the presence of the disease; background assumptions need to be clarified over the incidence or prevalence of the disease.

REPLICATION AND CORROBORATION

The standards of normal science are often summarized as the process of (i) credible hypothesis development; (ii) empirical evaluation; and (iii) replication of the initial study. Replication is typically considered the gold standard in establishing confidence in initial claims for therapy impact where a scientific experiment is repeated with the same methods, procedures and conditions to determine whether or not the original results can be reproduced. The primary aim is to test the reliability and validity of initial finding, ensuring that observed effects are not due to chance or special conditions present in the original study. Corroboration complements replication as it involves seeking additional evidence or support for a particular hypothesis by conducting different types of studies or various other measures to strengthen the overall body of evidence.

Replication of randomized clinical trial (RCT) results and other studies supporting value claims presents two challenges: (i) the traditional replication where there is duplication and application of an original protocol to assess whether or not the original results and consequent value claims can be duplicated; and (ii) the attempt to replicate the original results with real world data or evaluate a value claim that derives from pivotal trials. The latter presupposes that attempts to duplicate an original protocol have been demonstrated. The problem or crisis of replication arises when there is doubt that a duplication can yield the original results; replicating false claims raises questions. The objective is what we may describe as canonization , a transition from the original claims to a 'fact'. Where the original claim is taken for granted rather than as an open hypothesis. The acceptance is taken to be epistemological rather than ontological; it is not taken as descriptive or representation of an underlying physical reality. The problem, hence the term crisis, arises because the RCT replication often fails; the conclusions are ontologically false. This is not necessarily straightforward as multiple studies may be required to assess subsidiary hypotheses and not just a direct replication. Unlike ontological claims, epistemological claims are those accepted as facts by a relevant community reflecting in all too many cases the failure to publish negative results, publication bias, and the acceptance of false claims by journal editors.

There is, however, the vexed question of whether or not we want to reproduce or replicate the results of a study where the design of the study is considered inappropriate. In a paper published almost 20 years ago, the case was presented that most published research findings supporting clinical claims are false . This is not a novel proposition as refutation of research findings is a commonplace in RCTs, epidemiological studies and molecular research. In the subsequent 20 years little has changed [13]. It is challenging because the claim is made that the high rate of non-replication comes from the strategy of claiming conclusive research findings on a single study assessed by statistical significance (the p test). The situation is more damaging when the object of the study is to minimize costs and time to complete an RCT with the view of supporting market entry, following FDA approval, by a $p < 5\%$ value, rather than a focus on the positive predictive value or post-study probability that it is true. The issue is one of possible bias, not chance variability, where the combination of design, data characteristics and presentation produce research findings when they should not be produced.

The conclusions or corollaries regarding the probability that a research finding is less likely to be true include:

- The smaller the study sample size
- The smaller the study effect size
- The greater the number and the lesser the selection of tested relationships
- The greater the flexibility in designs and analytic modes
- The greater the financial and other interests and prejudices
- The hotter the scientific field (with more scientific teams involved)

The concern that a research claim from one or two Phase 3 pivotal studies prior to FDA approval and product launch, has a significant probability of being a false claim has to be the default position for formulary evaluations. Certainly, with the catalyst of the 21st Century Cures Act of 2016, the last few years have witnessed increased attention being given to the design and assessment of clinical trials, including standards for trial design and implementation [14]. Master protocols have been proposed to include umbrella studies, basket or bucket trials, platform studies and master

observational trials (MOT) in precision medicine. Even with these innovations there are two questions which are unresolved: (i) what are the implications of these new protocol designs for the creation of false claims; and (ii) are the reviewers for claims based on these designs equipped to both review them with the appropriate skills and assess them for inclusion in meta-analyses and systematic reviews.

The position that we should assume by default that all clinical claims are false claims unless they meet the review and evaluation standards for presumptive claims. The position can be usefully summarized by the distinction, following the demarcation criteria for intent, between two types of HTA, including purely clinical studies: (i) studies that focus on discovery where the analyst faces a mind-independent reality and (ii) studies where there is no intent or interest in discovery to confront an objective reality, but rather the intent is focused on creating a mind-dependent fantasy. For those focused on the standards of normal science and measurement it is just common sense to think in terms of a scientific or metaphysical reality where *the world as it is independent of how humans or other agents take it to be* [15]. Objects, fix the world's nature and exist independently of our ability to discover they do. Einstein made this point in his famous paper with Podolsky and Rosen in 1935, although we don't have to put this in terms of certainty [see quantum mechanics], saying: *If, without in any way disturbing a system, we can predict with certainty (i.e., with probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity* [16].

PUBLICATION BIAS: NULLIUS IN VERBA

It would be surprising if the founding members of the Royal Society in 1662 had foreseen the extent to which their commitment the role of evidence to support belief, even provisionally, had been subverted by publication bias and the creation of evidence in the physical and social sciences. Publication bias, the suppression of negative and embarrassing results, reflects both the failure to submit results by research groups (the 'file draw problem') and pharmaceutical manufacturers but also by the unwillingness of journal editors to publish negative findings. The result is what has been described as the canonization of false facts; claims, with repeated positive findings published and negative findings suppressed, become false facts.

It is difficult to come up with an estimate of the extent of publication bias. Some 10 years ago an estimate of publication bias found that, from a small sample, that 20.8% of null findings were published compared to 61.5% of positive findings [17]. There is no evidence that this is likely to change even with the increasing emphasis on false claims by leading publishers, the handful of journals focused on null results publication and, particularly noteworthy the tracking of journal retractions by Retraction Watch (https://retractionwatch.com/).

To those that, possibly naively, subscribe to the belief in the discovery of new facts and the rejection or modification of existing paradigms, the suppression of unwanted negative findings is a major stumbling block with researchers probing a range of provisional hypotheses which have long since been discredited but the previous findings ignored. Even though there have been attempts to report negative findings in selected journals together with red-flag warnings on possible bias or false facts in meta-analyses and systematic reviews (e.g. PRISMA-P [18]), it is an up-hill battle with the added effect of the proliferation of predatory journals, paper mills, the active involvement of academic researchers in creative falsification and the often difficult task of

retraction of published papers with academic institutions proving remarkably resistant to remedial interventions.

FALSE CLAIMS AND FALSIFIABILITY OF CLAIMS

In HTA the publication of false facts takes on a unique characteristic: assumption driven modelled simulations producing non-falsifiable claims for imaginary incremental cost-per-QALY and cost-effectiveness outcomes. This is not a biased selection of modelled claims that are 'significant' but the publication of one-off unique claims; canonization within a disease area is not at issue, but rather the canonization of the methodology of a discipline. In the area of the clinical sciences the impact of publication bias is to obscure what we may describe, as noted above, as objective reality or a mind independent reality. In HTA this is rejected; modelled imaginary claims are manifestations of a belief in a mind-dependent fantasy. This does not deny the presence of publication bias in HTA but sees it as complementing the creation of false evidence for claims with paper mills and others by encouraging the creation of evidence to meet well defined yet false methodological standards to support formulary acceptance, pricing and access.

A hypothesis or theory may be considered credible or valid it is falsifiable; these are not false claims which fail to reach the intent of this standard. In the case of fraud, the intent involves deception, misrepresentation or the withholding of information, often for personal gain. It is a legal and ethical issue not a scientific one. The presence of absence of falsifiability does not determine whether a claim constitutes fraud per se; but a false claim can be presented and demonstrated to be falsifiable.

Clearly, if the intent to deceive or provide a non-falsifiable modeled manufactured claim to support a particular hypothesis is absent then there should be no implication of fraud. Unfortunately, if there is intent to make a case for a product by a judicious choice of model and assumptions, then we can consider fraud and in an odd juxtaposition, the creation of false yet non-falsifiable claims; perhaps categorizing them as fantasy claims is more accurate. Claims that are false in the sense of failing to meet the standards of normal science and fundamental measurement, but not false in the sense of being manufactured to support a false claim. If this is the case, then it should be made clear to the prospective audience that the exercise involves hypothetical scenarios and assumptions that create fantasies which the more gullible believe should have a real impact on pricing and formulary decisions.

Intent has, presumably be inferred. While this does not arise in the case of agencies such as the Institute for Clinical and Economic Review (ICER) in the US which applies a significant degree of transparency in its fantasy model development, there remains a menagerie of independent consultants, paper mills and academics where intent is impossible to determine. In many cases it is a free go to first base where the journal editor has neither the time or resources to evaluate the modelled claim together with limited, if any, understanding of fundamental measurement. This stands in contrast to employment of academic reviewers in a number of single payer systems. If we add to this the possibility that the model will only be presented if it yields positive results for the client and the further possibility of bias by the journal editor to favor positive claims for cost-effectiveness, then we move further away from a commitment to discovery. Even if academic or similar reviewers are employed to evaluate assumption driven simulations, it does raise the question of why an academic group would devote its time to shuffling assumptions to provide a competing set of imaginary claims for cost-effectiveness.

Fortunately, rather than trying to disentangle the various modeled simulation, the saving grace is that they fail the standards for normal science and should be rejected by journal editors. The first step, for leading journal editors in HTA is to rescind their commitment to the fantasy of CHEERS 2022 and adopt a policy that automatically excludes assumption driven simulations and their non-falsifiable claims for cost-effectiveness from publication; they could still appear in second-line and predatory submission fee and free access journals.

There is one further criterion we should factor into the rejection of false claims: measurement. The standards for modern or Rasch measurement are quite clear in the transformation of observations or counts to measurement. The Rasch model provides the necessary and sufficient means to transform observations or counts to single attribute or unidimensional linear, interval and invariant measures [19]. Rasch provides the only basis for credible measurement. In HTA the implication is straightforward: claims that do not rest on a Rasch measure are supporting false claims based on an ordinal scale. This applies in particular for claims based on multiattribute generic instruments as well as those that are based on integer summation. Even if claimed to be falsifiable these scales, typically composite and ordinal, do not reach the standard of credible measures; they must be rejected, including where this extends to quality adjusted life years (QALYs).

It is interesting to note that with the various standards proposed for evaluating claims the question of measurement is never raised; the focus is on falsification but without recognizing that you can only apply statistical tests an interval or ratio measure that captures a single attribute. In HTA the COSMIN checklist is the obvious candidate although it shows no awareness of Rasch measurement; the same lack of awareness characterizes the various CONSORT standards, let alone the GRADE quality check and the CHEERS 2022 guidance for creating mind-dependent fantasy simulations [20] [21] [22]. CHEERS 2022 is not alone; many single payer health systems have proposed fantasy assumption driven simulation frameworks, often described as reference cases, to guide manufacturers and others in the creation of non-falsifiable cist-effectiveness claims. In the US there are the ICER guidelines, in the UK the National Institute for Health and Care Excellence (NICE) reference case guidance and in Australian the guidance prepared by the Pharmaceutical Benefits Advisory Committee [23] [24] [25]. At the risk of repetition, none seem to have an awareness to meet the required normal science standards for creating falsifiable and replicable value claims with required interval measurement properties.

In the case of CONSORT, which submits to the COSMIN standards, with the 2022 outcomes extension there are a further 10 outcome standard items added to the 2010 statement [26] [27] . There appears to be no concept whatsoever of the standards for Rasch or fundamental measurement where all outcome claims must be for a unidimensional or single attributes with linear, interval and invariant properties; items defining a single attribute manifestation of a latent construct must fit to the Rasch model. This has been a standard for measurement in the social sciences for over 60 years. This failure is compounded where an as an example of the 5 core elements of a defined outcome, the case study is depression, the CONSORT example is for the Montgomery-Asberg Depression Rating Scale (MADRS) [28]. This is an unfortunate choice as it fails completely Rasch standards as a simple integer summation to produce a composite ordinal scale and not a required interval or ratio measure.

FORCING A RESOLUTION

Whether intentional or not, the HTA meme has effectively distanced itself from any commitment to the replication of value claims; fantasy claims are not indicative of belief in a mind-independent

external reality and attempts to make sense of that reality. HTA is committed to a mind-dependent fantasy as the driver for formulary decisions and pricing. Obviously, only avenue open is, to vary the assumptions of a simulated modelled claim. This might involve the application of a Tornado diagram to identify the most impactful or troublesome assumptions and adjust these to assess the impact on the base line imaginary claims. This seems a singularly fruitless exercise as there is no basis for assessing the merits of the baseline claim in terms of the standards of normal science and measurement; just a variation on the original fantasy. This may appeal to a more credulous audience with a firm belief in imaginary claims and imaginary scenarios, but achieves nothing in terms of the assessed credibility of the claims for cost-effectiveness.

But there is one possible way forward to mitigate, from the perspective of a duty of care, claimed benefits, through health systems putting a brake on false claims and what is, by any standard, fraud. In proposals for a new start in HTA it has been proposed that, as health systems are the ultimate consumers of pharmaceutical products, dodgy or otherwise, then as part of a formulary submission clinical value claims which meet the required standards for fundamental measurement, should be supported by an agreed protocol to demonstrate how the manufacturer intends to evaluate the merits of the value claim [1]. Required standards and protocols for both single and parallel arms applied to real world evidence are readily available [29] . As gatekeeper, a formulary committee or similar group is in a pivotal position to demand, in a relatively short yet meaningful time frame, a degree of confirmation or otherwise of the product in a real- world treating environment. This is in clear contradistinction to the reference case scenarios where claims are based on imaginary assumption driven simulations and which proliferate in single payer health systems. In other words, a commitment to a mind-independent reality as the reference point replaces a mind-dependent fantasy as the decision criterion.

At the same time the downside is that with the opportunities to utilize expanding access to data bases to support value claims, there is the opening for the entry of paper mills, predatory journals and academics to provide false claims. The restriction, from the perspective of the formulary committee, is that that with a protocol requirement there is tighter control over the process of value assessment. Even so, there will no doubt be an incentive to present ‘validated’ claims for products which rely on false claims. Formulary committees should avoid the offer to accept existing value claims unless their provenance can be established. The preferred path forward is for protocols to support value claims that are specific to a target patient population that may be itself specific to a health system. Again, the formulary committee is the gatekeeper.

CONCLUSIONS

If we accept the proposition that in evaluating the clinical impact of competing therapies our perspective is the acceptance of a mind-independent reality not a mind-dependent fantasy, then our starting point must be one or more protocol driven clinical value claims for outcomes in a real-world treating environment. Certainly, clinical trials and systematic reviews can be the starting point, but claims from these sources must always be treated with caution. They cannot be accepted at face value; the default must always be that they support false positive claims. This may seem a harsh judgement, but the evidence for replication, corroboration and putative real-world impact for target patient populations leads to no other conclusion. Indeed, we may go even further and argue that if we apply the logic of induction to existing claims, then rather that assume fact has become truth we must take the position that prior claims, even if supported by an accumulation of so-called evidence, are logically no guide to future claims for outcomes. The fact that past futures

have resembled past pasts does not mean that future futures will resemble future pasts. If the truth is out there, we have no basis for assuming that we can converge to that truth when all we have are isolated and individual studies that may yield coherent provisional support for value claims but we can go no further than simply report results. The truth may be out there, but so are lies.

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] Franklin J, Patorno E, Desai R et al. Emulating randomized clinical trials with nonrandomized real-world evidence studies: First results from the RCT DUPLICATE initiative. *Circulation*. 2021;143(10):1002-1013

[3] Nissen S, Magidson T, Gross K et al. Publication Bias and the canonization of false facts. *eLife*. 2016;5:e21451

[4] Noorden R. How many clinical trials can't be trusted? *Nature*. 2023;619:455-458

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

[6]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

[7] Reiss, J, Sprenger J. Scientific Objectivity. *The Stanford Encyclopedia of Philosophy* (Winter 2020 Edition), E. Zalta (ed.), https://plato.stanford.edu/archives/win2020/entries/scientific-objectivity/

[8] Kuhn T. The Structure of Scientific Revolutions. Chicago: Chicago University Press, 1962

[9] Popper K. Objective Knowledge: An Evolutionary Approach. Oxford: Oxford University Press, 1972

[10] Pigliucci M. Nonsense on Stilts: How to tell science from bunk. Chicago: Chicago University Press, 2010

[11] Chakravartty, A. Scientific Realism, *The Stanford Encyclopedia of Philosophy* (Summer 2017 Edition), E Zalta (ed.), https://plato.stanford.edu/archives/sum2017/entries/scientific-realism/

[12] Gillies, D). "The Duhem Thesis and the Quine Thesis". In Curd M, Cover J. (eds.). Philosophy of Science: The Central Issues. New York: Norton, 1998.

[13] Ioannidis J. Why most published research findings are false. *PLoS Med*. 2005;2(8):e124

[14] Subbiah V. The next generation of evidence-based medicine. *Nature Medicine*. 2023;29:January 49-58

[15] Khlentzos, D. Challenges to Metaphysical Realism, *The Stanford Encyclopedia of Philosophy* (Spring 2021 Edition), E Zalta (ed.), https://plato.stanford.edu/archives/spr2021/entries/realism-sem-challenge/

[16]Einstein A., Podolsky B, and Rosen, N. "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?" *Physical Review.* 1935; 47(10): 777–780.

[17] Franco, A., Malhotra, N., Simonovits, G. Publication bias in the social sciences: Unlocking the file drawer. *Science*, 2014; *345*(6203), 1502–1505. https://doi.org/10.1126/science.1255484

[18] Shamseer L, Moher D, Clarke M et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ*. 2015;349:g7647 https://www.bmj.com/content/bmj/349/bmj.g7647.full.pdf

[19] 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

[20] McKenna SP, Heaney A. Setting and maintaining standards for patient-reported outcome measures: can we rely on the COSMIN checklists? *J Med Econ*. 2021;24(1):502–511 https://www.tandfonline.com/doi/epdf/10.1080/13696998.2021.1907092?needAccess=true&role=button

[21] McKenna S, Heaney A. COSMIN reviews: the need to consider measurement theory, modern measurement and a prospective rather than retrospective approach to evaluating patient-based measures. *J Med Econ*. 2021;24(1):860-61 https://www.tandfonline.com/doi/epdf/10.1080/13696998.2021.1948232?needAccess=true&role=button

[22] Meader N, King K, Llewellyn A. A checklist designed to aid consistency and reproducibility of GRADE assessments: development and pilot validation. *Systematic Reviews*. 2014;3:82

[23] Institute for Clinical and Economic Review. ICER's Reference Case for Economic Evaluations: Principles and Rationale. https://icer.org/wp-content/uploads/2020/10/ICER_Reference_Case_013120.pdf

[24] National Institute for Health and Care Excellence. NICE health technology evaluations: the manual. Process and methods 31 January 2022 https://www.nice.org.uk/process/pmg36/resources/nice-health-technology-evaluations-the-manual-pdf-72286779244741

[25] Pharmaceutical Benefits Advisory Committee. Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee (PBAC). Version 5.0. September 2016. https://pbac.pbs.gov.au/

[26] Moher D, Hopewell S, Schulz K et al. CONSORT 2010 explanation and updated guidelines for reporting parallel group randomized trials. *BMJ*. 2010;340:c332

[27] Butcher N, Monsour A, Mew E et al. Guidelines for reporting outcomes in trial reports. The CONSORT Outcomes 2022 extension. *JAMA*. 2022;328(22):2252-2264

[28] Montgomery S, Asberg M . A new depression scale designed to be sensitive to change". *Br J of Psychiatry*. 1979; 134 (4): 382–89.

[29] Food and Drug Administration. Considerations for the Design and Conduct of Externally Controlled Trials for Drug and Biological Products: Guidance for Industry (Draft). February, 2023. https://www.fda.gov/media/164960/download