MAIMON WORKING PAPERS No 28 DECEMBER 2025

THE REFERENCE CASE IS EPISTEMIC BULLSHIT

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

ABSTRACT

In 1985, Harry Frankfurt introduced the epistemic category of "bullshit" to describe forms of discourse that are neither lies nor simple errors, but are instead characterized by indifference to truth conditions. Bullshit, in Frankfurt's precise sense, is produced without regard to whether its claims are true or false, focusing instead on whether those claims function acceptably within a given institutional context. This paper applies Frankfurt's framework to the reference case simulation modelling that sits at the core of contemporary health technology assessment (HTA).

The reference case defines admissible methods and outputs for reimbursement and coverage decisions, mandating cost–utility analysis, quality-adjusted life years (QALYs), incremental cost-effectiveness ratios (ICERs), and probabilistic simulation. Although widely critiqued on ethical, distributive, and procedural grounds, these critiques have consistently failed to address a logically prior issue: whether the numerical quantities manipulated in the reference case meet the axioms required for arithmetic to be meaningful under representational measurement theory. This paper argues that they do not.

Drawing on the axioms of representational measurement—unidimensionality, scale-type admissibility, dimensional homogeneity, and invariance—the paper demonstrates that the reference case violates the conditions required for numbers to support truth-apt quantitative claims. Utilities are ordinal yet treated as interval; interval quantities are multiplied and aggregated as if they were ratio measures; heterogeneous attributes are combined without justification; and simulation outputs are presented as evidence despite lacking falsifiability. These are not technical imperfections but category errors that render the resulting outputs epistemically null.

The persistence of the reference case is explained not by fraud, deception, or incompetence, but by structural indifference to truth conditions. The framework functions as a system of procedural acceptance rather than measurement-based inquiry, sustained by institutional consensus and conformity. Defences of the reference case increasingly rely on an implicit redefinition of truth as sociological agreement rather than correspondence constrained by measurement.

The paper concludes that the reference case fits Frankfurt's definition of epistemic bullshit precisely. It survives not despite its lack of measurement validity, but because that lack is never made relevant. The implications are stark: HTA must either abandon its claim to quantitative evidence or reconstitute itself around falsifiable, measurement-valid claims that allow objective knowledge to evolve.

INTRODUCTION

In 1985, the philosopher Harry Frankfurt introduced a new epistemic category that filled a conceptual gap left open by traditional accounts of error, deception, and pseudoscience [1]. In his essay *On Bullshit* (Reprinted as a monograph by Princeton University Press), Frankfurt argued that there exists a form of discourse that is neither lying nor simple falsehood, and not adequately captured by concepts such as fraud, ignorance, or methodological mistake. Bullshit, in Frankfurt's precise sense, is defined by indifference to truth conditions. The bullshitter does not aim to conceal the truth, as the liar does, nor does the bullshitter necessarily assert claims believed to be false. Rather, the bullshitter is unconcerned with whether what is said is true or false at all. What matters is whether the discourse functions appropriately within a given context, satisfies expectations, or achieves acceptance.

This was a novel and important contribution because it identified a mode of reasoning that could persist indefinitely within professional and institutional settings without triggering the usual safeguards against error. Bullshit is particularly resilient because it does not depend on false premises that can be corrected, nor on deliberate deception that can be exposed. Instead, it thrives where outputs are required, procedures are routinized, and success is defined by conformity rather than correspondence to reality. Frankfurt's concern was not primarily moral but epistemic: bullshit erodes the very distinction between truth and non-truth by making that distinction irrelevant to the production of claims.

The relevance of this category extends far beyond casual speech. Frankfurt explicitly noted that bureaucratic, managerial, and expert systems are especially fertile environments for bullshit, precisely because they reward plausibility, technical appearance, and procedural compliance over truth. In such systems, numerical outputs can acquire authority simply by existing, even when the conditions that would make them meaningful are never examined. It is this institutional form of bullshit that is the focus of the present note.

The purpose of this brief paper is to consider whether Frankfurt's epistemic category provides an appropriate and illuminating description of reference case simulation modelling in health technology assessment (HTA). The reference case sits at the core of contemporary HTA practice. It defines admissible methods, mandates specific modelling architectures, and produces the quantitative outputs, cost-effectiveness ratios, threshold comparisons, probabilistic acceptability curves, that are treated as evidence for coverage and reimbursement decisions. Over more than four decades, this framework has become the dominant belief system of HTA, replicated across jurisdictions and embedded in guidelines, journals, and professional training. It functions as a memeplex: a self-reinforcing constellation of methods, assumptions, and norms that reproduces itself largely without interrogation.

This paper does not argue that reference case modelling is fraudulent, dishonest, or intentionally misleading. Nor does it seek to classify it as pseudoscience in the conventional sense. Instead, it asks a more fundamental question: is reference case modelling indifferent to the truth conditions of its numerical claims? Specifically, does it proceed without regard to whether the quantities it manipulates meet the axioms required for numbers to function as measures under representational

measurement theory? If so, then Frankfurt's category of bullshit is not a rhetorical flourish but an exact epistemic diagnosis.

By applying Frankfurt's framework to the reference case, this paper aims to shift the debate away from incremental methodological refinement and toward a prior and unavoidable issue: whether the quantitative core of HTA is capable, even in principle, of producing claims that can be true or false. If that capacity is absent, then no amount of ethical adjustment, transparency, or procedural reform can rescue the reference case from its epistemic status.

MISSING THE EPISTEMIC ARGUMENT

Criticism of the HTA reference case has been extensive, sustained, and often sophisticated. Over four decades, commentators have raised concerns about equity, fairness, severity weighting, end-of-life adjustments, discounting conventions, choice of perspective, the legitimacy of thresholds, and the political implications of cost-effectiveness analysis. Others have questioned the realism of lifetime models, the plausibility of extrapolations, the handling of uncertainty, and the transparency of assumptions embedded in simulation structures. These critiques are not trivial. Many have led to procedural refinements, supplementary analyses, and layers of ethical or deliberative adjustment. Yet despite their volume and variety, they share a common and decisive limitation: they never confront the epistemic status of the numbers themselves [2].

What unites almost all existing critiques is an implicit acceptance that the reference case is, at bottom, a quantitative exercise whose outputs are at least capable of being true or false. The debates assume that the arithmetic is meaningful and that the quantities being manipulated are measures, even if imperfect ones. Critics argue about how QALYs should be weighted, not whether they can be aggregated. They argue about thresholds, not whether ratios constructed from utilities and costs are admissible in principle. They worry about bias and uncertainty, not about whether the underlying quantities possess the scale properties required for arithmetic. In doing so, they remain entirely within the belief system or memeplex of the reference case [3].

This is the sense in which the critics miss the epistemic argument. The decisive question is not whether the reference case is ethically defensible, socially just, or procedurally transparent. It is whether the numerical outputs it generates have meaning at all. Representational measurement theory establishes that arithmetic operations are admissible only when numbers represent empirical attributes under specific axioms: unidimensionality, scale-type admissibility, dimensional homogeneity, and invariance. If these conditions are not met, then addition, multiplication, aggregation, and ratio construction are not merely approximate or controversial; they are undefined. No amount of ethical sophistication can rescue arithmetic performed on non-quantities.

The persistence of critique without epistemic penetration has had an unintended consequence. By focusing on refinements rather than foundations, critics have helped stabilize the reference case. Each new adjustment, severity modifiers, equity weights, alternative thresholds, scenario analyses, creates the impression of progress while leaving the core untouched. The system appears responsive, adaptive, and self-critical, even as it continues to operate on numbers whose measurement status is never examined. Debate becomes a mechanism of reproduction rather than disruption.

This pattern also explains why attacks on the QALY have repeatedly failed to dislodge it. Ethical objections, however forceful, do not threaten a system whose authority rests on the appearance of quantification rather than on measurement validity. As long as critics argue that QALYs are unfair rather than incoherent, the reference case can respond with procedural accommodations. What it cannot accommodate is the claim that its arithmetic is meaningless. That claim lies outside the vocabulary of the debate as it has been conducted.

In missing the epistemic argument, critics have treated the reference case as a flawed measuring instrument rather than as a non-measuring system. This misclassification matters. A flawed instrument can be recalibrated, supplemented, or corrected. A system that never measured anything in the first place cannot. The failure to recognize this distinction has allowed the reference case to persist unchallenged at its foundations, insulated by a continuous cycle of critique that never asks the logically prior question: are the quantities being manipulated capable of bearing the numerical meanings that the reference case assigns to them?

THE EPISTEMIC BULLSHIT ARGUMENT

Any claim that relies on arithmetic must first satisfy the conditions that make arithmetic meaningful. This is not a methodological preference or a disciplinary convention; it is a logical requirement. Representational measurement theory specifies the axioms under which numbers can represent empirical attributes and under which numerical operations are admissible [4]. At a minimum, these axioms include unidimensionality, so that a single attribute is being represented; scale-type admissibility, so that the permitted arithmetic matches the properties of the scale; dimensional homogeneity, so that like is combined with like; and invariance, so that numerical representations are stable across permissible transformations. When these conditions are met, arithmetic can support claims that are, in principle, true or false. When they are not met, arithmetic ceases to have meaning.

The reference case systematically violates these axioms. Utilities derived from preference instruments are ordinal, yet treated as interval. Interval quantities are multiplied by time to create QALYs, which are then treated as ratio measures despite lacking a true zero and dimensional coherence. Heterogeneous attributes are aggregated into single indices without justification. Simulation models manipulate these quantities over hypothetical lifetimes, producing ratios, thresholds, and acceptability curves that presuppose precisely the scale properties that are absent. These are not marginal technical errors. They are category errors. Once they occur, the exercise is no longer approximate measurement; it is non-measurement.

It is important to be clear about the implications of this failure. When measurement axioms are violated, the problem is not that results are biased, uncertain, or context-dependent. The problem is that the numerical outputs have no defined truth conditions. They cannot be approximately correct or approximately incorrect because there is no underlying quantity to which they correspond. In such cases, appeals to approximation are misplaced. Approximation presupposes a target that can, in principle, be approximated. Where arithmetic is undefined, approximation is meaningless.

This brings the argument squarely into Frankfurt's epistemic territory. Bullshit, as Frankfurt defined it, is discourse produced without regard to whether its claims are true or false. The reference case qualifies because it proceeds exactly as it would even if the measurement conditions were known to be violated. The arithmetic does not pause for scale-type justification. Aggregation does not depend on dimensional homogeneity. Simulation outputs are presented as claims without any requirement of falsifiability. Truth is not denied; it is simply irrelevant to the production of the numbers.

The persistence of this framework is not mysterious. Organizations seeking formulary approval or reimbursement operate in an environment where conformity to the reference case is a condition of participation. The task is not to demonstrate measurement validity but to follow prescribed rules and produce sanctioned outputs. The reference case offers a procedural pathway that is predictable, replicable, and administratively legible. Whether the quantities manipulated are measures is not a question that needs to be answered in order to succeed. Indeed, asking it would be counterproductive, because it would call into question the very framework that grants legitimacy to the submission.

This dynamic does not require bad faith, deception, or even awareness of the underlying problem. It requires only indifference. Indifference to scale properties. Indifference to truth conditions. Indifference to falsification. What matters is that a number is produced in the correct format, using the correct template, and justified according to accepted conventions. In this sense, the reference case is not defended because it is believed to be true, but because it works as a system of acceptance. It allows decisions to proceed without confronting the epistemic void at their core.

The result is a belief system in which numerical outputs carry authority precisely because the conditions of truth are never examined. This is the defining feature of epistemic bullshit. The reference case does not fail despite its lack of measurement; it survives because that lack is never made relevant.

A DIFFERENT CONCEPT OF TRUTH

Once the epistemic void at the heart of the reference case is exposed, its defense can no longer rest on claims of technical adequacy or methodological refinement. Instead, a different line of argument emerges, sometimes explicitly but more often implicitly: that the reference case operates under an alternative conception of truth. In this view, truth is not a matter of correspondence between numerical claims and empirical attributes, but of acceptance within a professional community [5]. What counts as "true enough" is determined sociologically, through consensus, convention, and institutional endorsement. The reference case is defended not because its numbers measure anything, but because they are recognized, reproducible, and authoritative within the HTA ecosystem.

This move is subtle but decisive. It replaces epistemic truth with procedural legitimacy. A result is treated as valid because it has been generated according to agreed rules, reviewed by peers, and accepted by committees. The question of whether the arithmetic is meaningful is displaced by the question of whether the process was followed correctly. Truth becomes a property of compliance

rather than correspondence. In effect, the reference case is insulated from epistemic critique by redefining what it means for a claim to be acceptable.

Such a conception of truth is deeply at odds with the logic of measurement. Representational measurement theory does not allow truth by consensus. Whether a scale supports addition or multiplication is not a matter of agreement but of logical necessity. Whether quantities are dimensionally homogeneous is not settled by committee vote. Whether a model generates falsifiable claims does not depend on how many agencies require it. These are binary conditions: either they are satisfied or they are not. To suggest that truth can be socially negotiated in this context is to abandon the very idea that numbers constrain belief.

Yet this relativized conception of truth is precisely what allows the reference case to persist. By treating truth as something that emerges from shared practice, HTA institutions can remain indifferent to measurement validity while still presenting their outputs as "evidence-based." The authority of the reference case derives not from its relation to reality, but from its embeddedness in guidelines, journals, training programs, and regulatory routines. Once enough actors behave as if the outputs mean something, their meaning is taken for granted.

This is not a neutral philosophical stance. It is a functional necessity for a system that cannot survive epistemic scrutiny. If truth were understood in the classical sense, as correspondence constrained by measurement axioms, the reference case would collapse immediately. Its quantities could not be defended as approximations, because there is nothing to approximate. Its simulations could not be treated as provisional truths, because they lack falsifiability. The only way to preserve the framework is to redefine truth so that these questions no longer arise.

Here the connection to Frankfurt's analysis becomes unavoidable. Bullshit thrives in environments where truth is decoupled from meaning and replaced by social performance. The bullshitter is not concerned with whether a statement is true, only with whether it is accepted as appropriate in context. The reference case exemplifies this condition. Its defenders do not argue that its quantities satisfy measurement axioms; they argue that everyone uses them, that decisions require them, and that abandoning them would be disruptive. These are sociological justifications, not epistemic ones.

To invoke a different conception of truth is therefore not an escape from the bullshit diagnosis; it is its confirmation. A system that must redefine truth in order to function has already conceded that truth, in the ordinary sense required for quantitative claims, is no longer relevant. The reference case does not offer a competing account of measurement-based truth. It offers a substitute for truth altogether.

CONCLUSION: WILL THE EPISTEMIC CHARADE CONTINUE?

After more than forty years, the persistence of the reference case raises a question that can no longer be avoided: is this epistemic charade simply too entrenched to abandon? Entire professional identities, institutional routines, journals, guidelines, and careers have been built around cost-utility analysis, QALYs, ICERs, and the procedural rituals that sustain them. The sunk costs are immense, not only in financial and organizational terms, but in intellectual commitment. To

concede that the reference case never rested on valid measurement would be to acknowledge that decades of "evidence-based" decision making were, in epistemic terms, ungrounded. It is unsurprising, then, that direct attacks on the framework, however precise, risk simply bouncing off a system that lacks the conceptual vocabulary to recognize what is being said.

From a technical standpoint, the case is already settled. The measurement errors embedded in the reference case are neither subtle nor controversial. Ordinal utilities cannot support arithmetic. QALYs lack dimensional homogeneity and a true zero. ICERs constructed from such quantities are undefined. No refinement of modelling practice, no expansion of sensitivity analysis, and no revision of reporting standards can rescue claims built on non-measures. One can, with ease, demonstrate that cost per QALY results have no truth conditions and therefore cannot be approximately correct, misleading, or informative. They are simply null. On this basis, initiatives such as CHEERS 2022, which seek to standardize and legitimize reporting of these outputs, do not advance science; they entrench non-measurement behind ever more elaborate procedural language [6].

The question, however, is not whether the reference case can be demolished. It can. The more difficult question is what follows. If epistemic exposure merely hardens institutional resistance, then critique alone may be insufficient. In that case, the path forward may not lie in arguing endlessly about QALYs, thresholds, or modelling conventions, but in changing what decision makers ask for. Rather than attempting to reform the reference case, decision makers could simply require something it cannot provide: falsifiable claims grounded in valid measurement. Claims about manifest outcomes must be expressed on ratio scales and evaluated prospectively. Claims about latent constructs must be supported by Rasch-calibrated measures that satisfy invariance and unidimensionality. Claims must be replicable, open to refutation, and embedded in a research strategy rather than offered as one-off modelling artefacts.

This shift would render much of the existing debate irrelevant. There would be no need to argue about the ethics of QALYs or the plausibility of lifetime simulations. Non-falsifiable claims would simply fail to qualify. What remains would be a cumulative process of inquiry, where knowledge evolves through testing, replication, and correction. This is not a radical proposal; it is a return to the ordinary logic of scientific progress. The alternative is to continue treating numerical storytelling as evidence, indefinitely.

The reference case can persist only by insulating itself from the evolution of objective knowledge. Whether it does so is ultimately a choice. But the epistemic stakes are now clear. The question is no longer whether the reference case is imperfect, but whether health technology assessment wishes to remain a domain of bullshit, or to rejoin the disciplines in which numbers mean something because they are constrained by measurement and answerable to truth.

ACKNOWLEDGEMENT

The author acknowledges the use of ChatGPT (version 5, OpenAI) in drafting and editing portions of this paper. All AI-assisted text was reviewed, verified, and substantively revised by the author, who assumes full responsibility for the final content and interpretation.

A NEW START IN MEASUREMENT FOR HEALTH TECHNOLOGY ASSESSMENT

For readers who are looking for an introduction to measurement that meets the required standards, Maimon Research has just released two distance education programs. These are:

- Program 1: Numerical Storytelling – Systematic Measurement Failure in HTA.
- Program 2: A New Start in Measurement for HTA, with recommendations for protocol-supported claims for specific objective measures as well as latent constructs and manifested traits.
-

Each program consists of five modules (approx. 5,500 words each), with extensive questions and answers. Each program is priced at US$65.00. Invitations to participate in these programs will be distributed in the first instance to 8,700 HTA professionals in 40 countries.

More detail on program content and access, including registration and on-line payment, is provided with this link: https://maimonresearch.com/distance-education-programs/

REFERENCES

[1] Frankfurt H. On Bullshit. Princeton: Princeton University Press, 2005

[2] Drummond M, Sculpher M, Claxton K et al. Methods for the Economic Evaluation of Health Care Programmes (4th Ed). New York: Oxford University Press, 2015

[3] Dawkins R. The Selfish Gene. 30th Anniversary Ed. Oxford: Oxford University Press, 2006

[4] Krant D, Luce R ,Suppes P ,Tversky A. (Eds.) *Foundations of Measurement, Volume I: Additive and Polynomial Representations*. New York: Academic Press, 1971

[5] Barnes B. *Scientific Knowledge and Sociological Theory.* London: Routledge & Kegan Paul, 1974.

[6] Husereau D, Drummond M, Stavros P et al. Consolidated Health Economic Evaluation Reporting Standards 2022 (CHEERS 2022) Statement: Updated Reporting Guidance for Health Economic Evaluations.*Value in Health* 25, no. 1 (2022): 3–9.