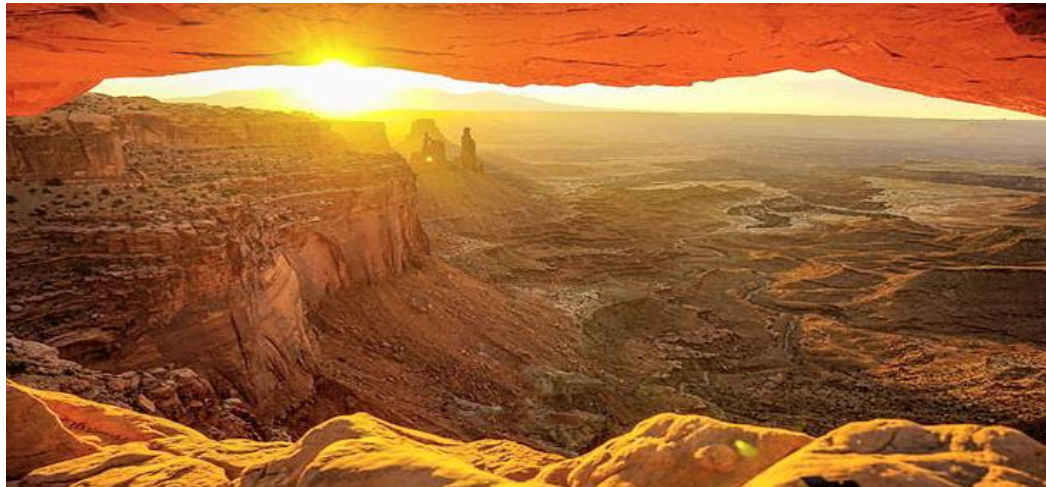


**MAIMON RESEARCH LLC**  
**ARTIFICIAL INTELLIGENCE LARGE LANGUAGE  
MODEL INTERROGATION**



**REPRESENTATIONAL MEASUREMENT FAILURE IN  
HEALTH TECHNOLOGY ASSESSMENT**

**AUSTRALIA: THE END OF THE PBAC REFERENCE  
CASE - RECONSTRUCTING HTA AROUND  
MEASUREMENT**

**Paul C Langley PhD Adjunct Professor, College of Pharmacy, University of  
Minnesota, Minneapolis, MN**

**LOGIT WORKING PAPER No 173 JUNE 2026**

[www.maimonresearch.com](http://www.maimonresearch.com)

**Tucson AZ**

# THE END OF THE PBAC REFERENCE CASE: RECONSTRUCTING HTA AROUND MEASUREMENT

## ABSTRACT

*Australia occupies a unique position in the history of health technology assessment (HTA). Through the Pharmaceutical Benefits Advisory Committee (PBAC) and a network of influential academic research centers, Australia helped establish and disseminate the reference-case framework that came to dominate reimbursement decision making throughout much of the world. Utilities, quality-adjusted life years (QALYs), simulation models and cost-effectiveness analysis became the accepted foundations of therapy assessment, shaping not only policy decisions but also research agendas, professional training and manufacturer submission strategies.*

*This paper argues that the apparent success of the Australian reference case concealed a fundamental methodological problem. The framework standardized arithmetic without first establishing measurement. Drawing upon a series of large language model interrogations of the Australian HTA knowledge base, including the PBAC, leading academic research centres and associated educational programs, the paper demonstrates a consistent pattern of measurement inversion. Foundational principles of representational measurement, ratio measurement, dimensional homogeneity and the distinction between manifest and latent attributes receive weak endorsement, while utilities, QALYs and reference-case simulation models receive strong endorsement. Arithmetic routinely precedes measurement.*

*The consequences of this inversion are examined through a step-by-step analysis of the reference-case framework. Health-state descriptions generate ordinal preference scores, which are transformed into utility values and subsequently combined with time to create QALYs. QALYs then become inputs to simulation models and cost-effectiveness analyses. At each stage, arithmetic operations are undertaken without demonstrating that the entities involved satisfy the requirements imposed by the scales of measurement or the axioms of representational measurement. The result is arithmetic chaos: a sequence of numerical manipulations that produces numbers rather than measures.*

*The paper further examines what is described as the "sound of silence" within Australian HTA. Despite the public availability of interrogation results and the substantial attention they have attracted, neither the PBAC nor leading Australian HTA research centres have offered a substantive response to the measurement critique. This absence of engagement is interpreted as evidence of closure: a condition in which a framework loses the capacity to respond to challenges directed at its foundations. The paper concludes that the future of Australian HTA lies not in refining the reference case but in reconstructing therapy assessment around lawful measurement, linear ratio measures for manifest attributes, Rasch logit ratio measures for latent attributes, protocol-driven evaluation and empirical falsification. Australia helped create the reference-case paradigm; it now has the opportunity to lead its replacement.*

## INTRODUCTION

Australia occupies a unique position in the history of health technology assessment (HTA). For more than three decades, the Pharmaceutical Benefits Advisory Committee (PBAC) has been regarded internationally along with the National Institute for Health and Care Excellence (NICE) in the UK, as one of the most influential agencies in the development and application of cost-utility analysis, reference-case modelling and evidence-based reimbursement decision making. Together with Australia's leading academic research centers the PBAC helped establish a framework that was subsequently adopted, adapted and promoted throughout much of the world. The Australian approach came to be viewed as an exemplar of best practice in HTA, influencing reimbursement systems, academic programs and policy development far beyond its national borders.

The central feature of this framework was the reference case. Manufacturers seeking reimbursement were encouraged to construct simulation models supported by utility scores, quality-adjusted life years (QALYs), projected resource utilization and cost-effectiveness estimates. Over time, these methodologies became embedded not only within reimbursement decision making but also within the teaching, research and professional culture of Australian HTA. Increasingly sophisticated analytical techniques were developed to refine model structures, explore uncertainty and support pricing recommendations. The assumption underlying these efforts was straightforward: better models would lead to better decisions.

Yet a prior question remained unexplored. Do the entities entering these analytical frameworks satisfy the requirements of representational measurement? Before arithmetic can be undertaken, quantitative science requires that the attributes being manipulated possess lawful measurement properties. Measurement must precede arithmetic. This principle is uncontroversial throughout the physical sciences and forms the foundation of modern measurement theory. Surprisingly, it has remained essentially absent from the development and evaluation of HTA methodologies.

Recent large language model interrogations of the Australian HTA knowledge base raise serious concerns in this regard. Assessments of the PBAC, leading Australian HTA research centers and associated academic programs reveal a remarkably consistent pattern. Foundational propositions concerning representational measurement, ratio measurement, dimensional homogeneity and the distinction between manifest and latent attributes receive weak or no endorsement, while propositions supporting utilities, QALYs and reference-case simulation models receive strong endorsement. The result is measurement inversion: arithmetic routinely precedes measurement.

This finding has profound implications. If measurement inversion is the signal, then arithmetic chaos is the consequence. Utilities, QALYs and simulation outputs become numerical constructions detached from lawful measurement. The issue is therefore not whether the reference case can be refined, but whether it can continue to function as a scientifically credible framework for therapy assessment. This paper argues that it cannot. The reference case has reached closure. The future of Australian HTA lies not in defending or modifying the existing framework, but in reconstructing therapy assessment around lawful measurement, empirical evaluation and falsification. Australia helped create the reference-case paradigm. It now has an opportunity to lead its replacement.

## AUSTRALIA AND THE RISE OF THE REFERENCE CASE

Australia played a pivotal role in the emergence of modern health technology assessment. Through the Pharmaceutical Benefits Advisory Committee (PBAC), Australia became one of the first countries to require formal economic evaluation as part of reimbursement decision making. What began as an attempt to bring greater transparency and consistency to decisions regarding pharmaceutical access evolved into a comprehensive framework that would influence HTA practice around the world. By the late 1990s and early 2000s, Australia had established itself as an international leader in the application of cost-utility analysis, QALYs and reference-case methodologies.

The attraction of the reference case was understandable. Health systems face difficult decisions regarding the allocation of scarce resources among competing therapies. A standardized analytical framework appeared to offer a rational solution. Manufacturers could be required to present evidence according to a common structure. Outcomes could be translated into a common metric through utility scores and QALYs. Simulation models could project future consequences of treatment decisions. Cost-effectiveness thresholds could provide a basis for comparing therapies across disease areas. The result appeared to be a coherent and objective approach to reimbursement decision making.

Australian academic research centers played an important role in the development and dissemination of this framework. Institutions such as the Centre for Health Economics Research and Evaluation (CHERE) at the University of Technology Sydney, the Centre for Health Economics at Monash University, the Health Economics Group at the University of Adelaide, HTA at Deakin University and other research centers contributed to the refinement of economic evaluation methods and the training of successive generations of HTA practitioners. Through publications, postgraduate programs, professional networks and advisory roles, these centers reinforced the position of the reference case as the dominant arithmetic precedes measurement paradigm for therapy assessment.

Over time, the reference case became more than a methodological framework. It became an intellectual culture. Utilities, QALYs, Markov models, probabilistic sensitivity analyses and cost-effectiveness thresholds were increasingly treated as the natural language of HTA. Manufacturers organized submission strategies around these requirements. Academic programs trained students in their application. Journals published increasingly sophisticated modelling exercises. The focus of methodological innovation shifted towards improving model structures, refining assumptions and enhancing computational techniques.

What remained unexamined was a prior and more fundamental question that can be traced back to the early experiments of the 17<sup>th</sup> century scientific revolution: did the entities entering these analytical frameworks satisfy the requirements of measurement? The success of the reference case encouraged attention to arithmetic rather than measurement. Health-state descriptions, preference scores, utility algorithms and QALYs were accepted as legitimate analytical inputs with remarkably little scrutiny of their measurement properties. The framework assumed that assigning numbers to observations created quantities suitable for arithmetic manipulation.

This omission would prove critical. The international influence of Australian HTA ensured that the assumptions embedded within the reference case were exported along with the methodology itself. What began as a national approach to reimbursement assessment evolved, along with NICE, into a global framework adopted by agencies, academic centers and professional organizations throughout the world. The rise of the reference case therefore represents one of the most influential developments in modern HTA. It also created the conditions for a problem that would remain largely hidden for decades: the displacement of measurement by arithmetic.

## **MEASUREMENT INVERSION IN AUSTRALIAN HTA**

The concept of measurement inversion provides a useful lens through which to understand the current state of Australian health technology assessment. Measurement inversion occurs when arithmetic precedes measurement; when numerical operations are applied before it has been demonstrated that the underlying attributes possess lawful measurement properties. In a quantitative science, measurement must come first. Arithmetic is permissible only after the measurement status of the entities involved has been established. The evidence suggests that contemporary Australian HTA had never appreciated this sequence and reversed it.

This conclusion emerges from a series of large language model interrogations of the Australian HTA knowledge base. These interrogations examined the published methodological frameworks, educational content, research outputs and policy guidance associated with the PBAC, leading academic research centers and related HTA institutions. The objective was straightforward: determine whether the foundational principles of representational measurement are recognized and applied within the Australian HTA framework.

The results were remarkably consistent. Statements affirming the necessity of measurement before arithmetic, the requirements of ratio measurement, dimensional homogeneity, unidimensionality and the distinction between manifest and latent attributes received weak endorsement. By contrast, propositions supporting utilities, QALYs, cost-utility analysis and reference-case simulation modelling received strong endorsement. The pattern was repeated across agencies, research centers and educational programs. This is the defining characteristic of measurement inversion.

The significance of this finding extends beyond methodological preference. Measurement inversion is not simply another criticism of the reference case. It is a diagnostic signal. It indicates that the analytical framework has become detached from the constraints imposed by scales of measurement and the axioms of representational measurement. Once this occurs, there is no longer a principled basis for determining which arithmetic operations are admissible and which are prohibited. Numerical manipulation begins to substitute for measurement itself.

The consequences are visible throughout Australian HTA. Health-state descriptions are treated as though they were quantitative entities. Ordinal preference scores are transformed into utility values. Time is multiplied by composite ordinal utilities to create QALYs. QALYs are incorporated into simulation models and ultimately combined with composite cost estimates to generate cost-effectiveness ratios. At no point in this sequence is there a demonstration that the entities involved satisfy the requirements necessary for quantitative interpretation. Arithmetic proceeds regardless. The reference case guarantees ordinal scales not ratio measures.

Equally revealing is the near absence of Rasch measurement from the Australian HTA knowledge base. This omission is striking because many of the outcomes of interest to HTA such as symptom burden, treatment satisfaction, functioning and need fulfilment are latent attributes. Yet rather than developing instruments capable of producing lawful measures of latent attribute possession, the dominant framework relies on ordinal scores, utility algorithms and composite indices. The distinction between manifest and latent attributes is ignored, even though it is central to measurement theory.

The result is a system in which arithmetic dominates methodological thinking while measurement remains largely invisible. This is not a criticism of individual researchers, institutions or agencies. It is a criticism of the structure of the knowledge base itself. Over time, the assumptions embedded within the reference case became normalized, reproduced through education and reinforced through policy practice. Measurement inversion became institutionalized.

The importance of recognizing this condition cannot be overstated. Measurement inversion is the signal. It tells us that arithmetic has become detached from measurement. Once this signal is recognized, attention must shift to its consequences.

## **ARITHMETIC CHAOS AND CLOSURE**

If measurement inversion is the signal, arithmetic chaos is the consequence. Australian HTA provides a clear example of how this process unfolds within the reference-case framework. Once arithmetic is allowed to proceed before measurement has been established, there is no longer a principled basis for determining which numerical operations are admissible. Numbers can be assigned, transformed, multiplied, aggregated and divided, but the resulting outputs do not acquire measurement status simply because they are produced by a model.

The process begins with health-state descriptions. These are intended to capture aspects of health, functioning, symptoms and wellbeing that may be affected by disease or treatment. The difficulty is that health-state descriptions are not measures. They are collections of characteristics, frequently multidimensional in nature, which combine manifest and latent elements within a single descriptive structure. They may have descriptive value, but they do not provide a lawful quantitative basis for arithmetic.

The next step is preference elicitation. Techniques such as time trade-off and standard gamble are used to generate preference scores for different health states. These scores indicate relative preference. They do not establish constant units, a true zero or ratio properties. They are ordinal structures. An ordinal score can rank alternatives, but it cannot support multiplication, division or the estimation of meaningful magnitudes. At this point, from the perspective of representational measurement, the reference-case sequence should have stopped. Instead, it continued. Preference scores are then used to generate health-state weights and utility values with a misapplication of econometric modelling. This step is often presented as a transformation from preferences into quantities suitable for economic evaluation. Yet the transformation does not alter the underlying measurement status of the data. Utilities remain composite constructs derived from ordinal preferences and multidimensional health-state descriptions. They are not unidimensional measures. They do not possess demonstrated ratio properties. Many utility systems also permit

negative values for health states considered worse than death, a feature inconsistent with the requirements of ratio measurement, which requires a meaningful non-arbitrary zero.

The creation of the QALY introduces the next stage of arithmetic chaos. Utility values, themselves derived from ordinal preference scores, discount time in an attempt to create a common metric for comparing therapies across disease areas. This operation is inadmissible. Time is a lawful ratio measure, but multiplication requires that both entities entering the operation possess the appropriate measurement properties. Utilities do not meet this requirement. Multiplying a ratio measure by a composite construct derived from ordinal preferences cannot create a lawful measure. The QALY therefore inherits the deficiencies of the utility structure from which it is derived. It has no recognized status within measurement theory and no demonstrated measurement properties. A QALY is not a measure of health status; it is a numerical construct whose quantitative interpretation cannot be justified by the scales of measurement or the axioms of representational measurement.

The process then moves to simulation modelling. QALYs, costs, transition probabilities, treatment persistence assumptions, disease progression estimates and future event rates are incorporated into decision models projected over years or lifetimes. The complexity of these models creates an appearance of rigor, but complexity does not repair measurement failure. Many of the inputs are not lawful measures. Others are assumptions rather than empirical propositions. The model may appear internally coherent, but internal coherence is not the same as scientific validity.

Cost inputs create an additional difficulty. In reference-case modelling, costs often combine heterogeneous resource units: hospital days, physician visits, drug acquisition costs, monitoring costs, adverse event management and other categories. These are frequently aggregated into a single monetary estimate. Although monetary values can be useful for budgeting, the composite cost numerator may conceal the distinct resource attributes that actually matter to health systems. From a measurement perspective, resource claims should focus first on the manifest units themselves before any monetary valuation is applied.

The final output is typically an incremental cost-effectiveness ratio expressed as cost per QALY gained. By this stage the analytical process has travelled a long distance from measurable attributes. The numerator is a composite cost estimate. The denominator is a QALY whose measurement properties remain unresolved. The resulting ratio is presented as evidence for pricing and reimbursement decisions. Yet if the denominator is not a lawful measure, and if the numerator aggregates heterogeneous resource inputs, the ratio cannot be interpreted as a meaningful quantitative measure of therapy value. Again, it cannot be justified by the scales of measurement or the axioms of representational measurement.

This is arithmetic chaos. Health-state descriptions become ordinal preferences; ordinal preferences become utilities; utilities create QALYs; QALYs enter simulation models; simulation outputs become cost-effectiveness ratios. At each stage arithmetic proceeds without first establishing whether the entities involved satisfy the requirements imposed by scales of measurement and the axioms of representational measurement. Each stage inherits the deficiencies of the previous stage. The reference case is just the manipulation of numbers; arithmetic without measurement.

The problem cannot be solved by model refinement. Larger datasets, more detailed transition structures, probabilistic sensitivity analyses, alternative utility algorithms or longer time horizons do not address the prior question of measurement. Sensitivity analysis evaluates uncertainty within a model; it does not establish whether the entities entering the model are lawful measures. More sophisticated arithmetic remains arithmetic. It does not create measurement.

The consequence is closure. Closure does not mean that the PBAC ceases to operate or that Australian researchers stop building models. It means that the reference-case framework no longer possesses the capacity to justify its output according to the standards of quantitative science. A framework reaches closure when it cannot correct its foundational errors from within its own assumptions. The Australian reference case achieved closure with the starting point of valuing health state descriptions; we were just delayed for 40 years in recognizing and reporting it. Its difficulty is not a defective parameter, a weak assumption or an incomplete model structure. Its difficulty is structural: it standardizes arithmetic without establishing measurement.

The recognition of arithmetic chaos therefore changes the nature of the Australian HTA debate. The question is no longer whether the reference case can be improved. The question is whether it can remain the foundation for therapy assessment. If the framework generates numbers rather than measures, then it cannot support credible outcomes regarding therapy impact. The conclusion is difficult to avoid. The reference case has reached closure. Australia now faces the task of reconstructing HTA around measurable attributes, lawful measurement structures, explicit protocols and empirical evaluation. The question is whether or not this is attainable.

## **THE SOUND OF SILENCE: CLOSURE, ABANDONMENT AND RECONSTRUCTION**

One of the most striking features of the Australian HTA experience is not the evidence for measurement inversion itself but the absence of any substantive response to that evidence. We are now in the position that the evidence for measurement inversion in Australia is overwhelming. Australia is a member good standing with the global HTA community who are equal participants; all are in the same boat.

Despite the ready availability of interrogation results on the Maimon Research website in sufficient detail for the PBAC and individually for all academic research groups .in sufficient detail to allow challenge, replication or rebuttal. Yet none of these responses has occurred. This absence of engagement is noteworthy. The PBAC interrogation has attracted hundreds of visits through the Maimon Research website. Similar attention has been given to assessments of Australian research centers and educational programs. Yet neither the PBAC nor the leading university centers have offered a substantive defense of the measurement foundations of the reference case. No attempt has been made to challenge the interrogation methodology. No alternative interpretation of the findings has been proposed. No defense has been mounted for the utility-QALY framework from the perspective of representational measurement. The response has been silence.

Silence, of course, is not proof. Scientific claims are not established by the absence of criticism. Yet silence has diagnostic value when criticism would be expected. If the interrogations were obviously flawed, if the measurement arguments were mistaken, or if the conclusions regarding

utilities, QALYs and simulation modelling were easily dismissed, one might anticipate a vigorous response. Australian HTA has no shortage of articulate researchers, experienced methodologists and influential institutions. The absence of engagement therefore raises an uncomfortable possibility: the issue is not that the arguments are unknown, but that there is no satisfactory response available within the existing framework.

This is precisely the condition that characterizes closure. Closure does not occur because a discipline runs out of practitioners or because institutions cease to function. Closure occurs when a framework loses the capacity to respond to challenges that strike at its foundations. Internal refinements continue. Existing methods are applied with increasing sophistication. New generations are trained in established approaches. Yet the fundamental assumptions of the framework remain insulated from challenge because there is no accepted mechanism through which those assumptions can be re-examined.

The Australian reference-case framework exhibits many of these characteristics. Over four decades, attention has focused on refining simulation models, improving utility algorithms, expanding sensitivity analyses and developing increasingly sophisticated analytical techniques. Yet the prior question of measurement remained absent. The framework has become highly effective at standardizing arithmetic while largely ignoring the requirements of measurement. Consequently, when the measurement question is finally raised, the framework possesses few internal resources with which to respond. The importance of measurement preceding arithmetic appears a novel concept.

This has important implications for the future. Once closure is accepted, then the objective cannot be reform. Reform assumes that the existing structure remains fundamentally sound and requires only adjustment. The present argument is different. If utilities lack lawful measurement properties, if QALYs cannot satisfy the requirements of ratio measurement, and if simulation models merely inherit these deficiencies, then refinement cannot solve the problem. The issue is structural. The foundation of HTA as a viable discipline must be reconsidered.

Reconstruction is required; there is no other answer. Reconstruction begins by abandoning the assumption that assigning numbers to observations creates quantities suitable for arithmetic manipulation. It recognizes that therapy assessment must begin with attributes and that different attributes require different measurement pathways. Manifest attributes require linear ratio measures. Latent attributes require Rasch logit ratio measures. Claims must be linked to protocols, exposed to empirical challenge and subject to possible falsification. The focus shifts from hypothetical futures generated by simulation model numbers to measurable and falsifiable outcomes observed in real patients.

Australia is uniquely placed to lead this transition. It played a central role in the creation and global dissemination of the reference-case paradigm. It possesses internationally respected HTA institutions, experienced researchers and an established policy infrastructure. The same intellectual leadership that helped create the reference case could now be directed toward its closure and replacement. If not the HTA Australian community will just be bystanders to a needed reconstruction not only of HTA but, more practically, the future of the PBAC; if it has one.

Whether this opportunity is taken remains uncertain. What is clear is that the silence surrounding measurement inversion cannot continue indefinitely. The interrogations have placed the made visible what was previously assumed. The measurement question has been placed squarely before Australian HTA. The issue is no longer whether the challenge exists. The issue is whether Australia chooses to respond. The future lies not in defending a framework that has reached closure, but in reconstructing HTA around lawful measurement, empirical evaluation and the continuing evolution of objective knowledge.

## **CONCLUSION**

Australia played a central role in the emergence of modern health technology assessment by endorsing arithmetic before measurement. Through the PBAC, its academic research centers and its influence on international reimbursement practice, Australia helped establish the reference case as the dominant framework for evaluating therapies. For more than four decades this framework has shaped research, education and policy, not only within Australia but throughout the world.

The argument advanced in this paper is that the success of the reference case concealed a fundamental and fatal problem. The framework standardized arithmetic without first establishing measurement. Large language model interrogations of Australian HTA knowledge systems reveal a consistent pattern of measurement inversion in which the requirements of representational measurement receive weak endorsement while utilities, QALYs and simulation modelling receive strong support. Once arithmetic is detached from measurement, arithmetic chaos follows. Health-state descriptions become ordinal preferences, ordinal preferences become utilities, utilities become QALYs and QALYs become cost-effectiveness claims. At no stage are the measurement requirements necessary for quantitative interpretation established.

The significance of this finding extends beyond Australia. Because Australia helped shape the international HTA paradigm, the implications of closure extend far beyond the PBAC. The issue is not whether the reference case can be refined. The issue is whether it can continue to justify its claims according to the standards of quantitative science. The evidence presented here suggests that it cannot. Can Australian research centers respond?

The future therefore lies not in defending the reference case but in replacing it. Reconstruction requires a return to first principles: measurement before arithmetic, explicit attribute claims, linear ratio measures for manifest attributes, Rasch logit ratio measures for latent attributes, protocol-driven evaluation and empirical falsification. Australia now faces a choice. It can remain committed to a closed reference case framework whose measurement foundations have failed, or it can once again assume a leadership role by helping to build a new HTA grounded in lawful measurement and the continuing evolution of objective knowledge.

## **ACKNOWLEDGEMENT**

I acknowledge that I have used OpenAI technologies, including the large language model, to assist in the development of this work. All final decisions, interpretations, and responsibilities for the content rest solely with me.

## REFERENCES

Australia: Pharmaceutical Benefits Advisory Committee (PBAC) – Decisions without measurement, Logit Working Paper No 173. <https://maimonresearch.com/logit-working-paper-no-34/>

Australia: What happens if nothing changes? Questions for the future of health technology assessment. Logit Working Paper No 665. <https://maimonresearch.com/logit-working-paper-no-665-may-2026/>

Australia: The global health technology assessment memplex of measurement failure, Logit Working Paper No 726. <https://maimonresearch.com/logit-working-paper-no-726-february-2026/>