

Beyond the Memoryless Cohort: A Critical Re-evaluation of Markov Assumptions, QALY Frameworks, and a Need for Next-Generation Health Economic Modelling *A Scoping Review with Conceptual Synthesis*

Anuj Gupta* 

SME – Public Health System and Policy, Lean Business Services, Saudi Arabia

*Corresponding Author

Anuj Gupta, SME – Public Health System and Policy, Lean Business Services, Saudi Arabia.

Submitted: 2026, Aug 17; **Accepted:** 2026, Aug 26; **Published:** Under Process

Citation: Gupta, A. (2026). Beyond the Memoryless Cohort: A Critical Re-evaluation of Markov Assumptions, QALY Frameworks, and a Need for Next-Generation Health Economic Modelling A Scoping Review with Conceptual Synthesis. *J Eco Res & Rev*, 6(2), 01-11.

Abstract

Objectives: The scoping review aims to understand the structural, epistemic, and ethical boundaries of the commonly used and popular health economic modelling tools and methodologies—especially the cohort Markov model and the quality-adjusted life-year (QALY). The aim is to further outline a scope and research agenda for next-generation modelling approaches, which can inform policymakers, leading to an equity-sensitive resource allocation in an age of targeted medicine, digital health, and real-world evidence (RWE).

Methods: The structured scoping review leverages PRISMA-ScR methodology to identify and establish structural, knowledge-based, and ethical limits of presently used health economic models using evidence from related literature between 2015 and 2025, identified through systematic searches of Scopus and PubMed/MEDLINE. Further, the research for the paper also includes public information on ISPOR-SMDM Good Research Practices, NICE Decision Support Unit guidance, and grey literature from HTA bodies. For analysis, multiple variables, such as model architecture, structural-uncertainty management and handling, validation approach, equity sensitivity, RWE integration, and AI/ML augmentation, were considered. Thematic synthesis was augmented through systems-mapping analysis and a conceptual framework proposal.

Results: A total of 847 records were retrieved from systematic research of published and grey literature. 62 of the 847 records met the inclusion criteria. Following an in-depth analysis of the included research, the Markov assumption of memoryless gives rise to three critical systematic challenges:

- Ahistorical transition estimation
- State-space explosion under heterogeneous treatment sequencing
- Suppression of patient-level variance

Further, while the utilitarian aggregation underlying QALYs is distributionally neutral, it is increasingly contested by approaches such as distributional cost-effectiveness analysis (DCEA) and the application of severity modifiers, although uptake remains limited. Additionally, many models leveraging these methodologies were found to rely on assumptions which do not fully reflect health and health system operations. Therefore, to address these challenges, a conceptual 'Strategic House' framework—which integrates causal systems, flexible model engines, equity analysis, and real-time decision support is proposed.

Conclusions: Advancing HTA requires a transition to adaptable, routinely updated modelling approaches that explicitly prioritize equity. While the Strategic House framework provides a robust foundation for reforming national HTA systems, real-world implementation and evaluation are critical next steps.

Keywords: Markov Model; QALY, Structural Uncertainty, Equity, Systems Thinking, Real-World Evidence, PRISMA-ScR, Health Technology Assessment

1. Introduction

The cohort Markov state-transition model has shown impressive staying power in institutions. First developed for medical decision-making by Beck and Pauker in 1983 and refined with a key practical guide by Sonnenberg and Beck in 1993, it became the gold standard for health technology assessment (HTA) thanks to the ISPOR-SMDM Modelling Good Research Practices Task Force [1-4]. Major national HTA agencies—like the UK's National Institute for Health and Care Excellence (NICE), Canada's Agency for Drugs and Technologies in Health (CADTH), and Australia's Pharmaceutical Benefits Advisory Committee (PBAC)—have shaped their submission and review processes around these state-transition models. The quality-adjusted life-year (QALY), as the main outcome measure in these setups, has likewise become a cornerstone [5,6].

This institutional success stems from real strengths: ease of use, clarity for non-expert reviewers, low computational demands, and a robust body of methods for estimating parameters, running probabilistic sensitivity analysis (PSA), and validating models [7-9]. For many common decision problems—like chronic conditions with clear clinical stages, straightforward treatment paths, and fairly uniform patient groups—the cohort Markov approach is still a practical and reliable option.

- Critical challenges faced by HTA bodies are decision problems that go beyond the usual issues, arising from three structural shifts that have exposed the limitations of the standard paradigm.
- Precision medicine and multi-line treatment sequencing create path-dependent patient journey and treatment history, undermining the memoryless nature of these models [10,11].
- Population-level feedback loops generated by digital health interventions, behavioural programmes, and complex public health strategies are not captured by static, closed-population models [12,13].
- The equity discourse has matured beyond rhetorical acknowledgement: distributional cost-effectiveness analysis and extended cost-effectiveness analysis show that the QALY's distributional neutrality is not sufficient and faces methodological challenges [14-16].

The Markov models are not incorrect; within their scope, they are mathematically sound. However, with the growing complexity of HTA decisions, a critical gap is created and needs to be addressed immediately. The review aims to identify these gaps, establish methodical tools to resolve the gaps identified and outline a research plan and a practical operational framework for next-generation health economic modelling. By leveraging the systematic scoping methods, this review moves beyond the scope of the literature and proposes an integrated future-state architecture for national-level health economic decision support.

2. Methods

2.1. Review Framework and Reporting Standard

This review follows the PRISMA Extension for Scoping Reviews PRISMA-ScR; Tricco et al., 2018 and the Arksey-O'Malley framework (2005), as refined by Levac, Colquhoun, and O'Brien (2010) [17-19]. The methodology helps to map the conceptual edges and limitations of well-established modelling methods, leading to a thematic blending over and beyond pooled effect estimation. The review lays the foundation for changes to existing models by combining systematic evidence with an inductive, author-developed conceptual framework (the Strategic House).

2.2. Eligibility Criteria

Inclusion criteria:

- Studies addressing structural, epistemic, or ethical limitations of state-transition models, QALY frameworks, or related decision-analytic architectures in health economic evaluation.
- Methodological innovations including microsimulation, system dynamics, agent-based modelling, Bayesian structural-uncertainty handling, DCEA/ECEA, RWE integration, and AI/ML augmentation of health economic models.
- Peer-reviewed articles (systematic reviews, scoping reviews, methodological guidance, applied exemplars) and grey literature from HTA bodies (NICE DSU, ISPOR, EUnetHTA).
- English language, published January 2015–October 2025.
- Any human population, any geography, any HTA or health-system context. Exclusion criteria:
- Pure clinical prediction models without an economic evaluation component.
- Burden-of-disease estimations without policy comparison or resource-allocation framing.
- Editorials, conference abstracts without methodological detail, opinion pieces without analytic content.
- Non-English publications.
- Publications before 2015, except foundational methodological references (e.g. Sonnenberg and Beck, 1993; Beck and Pauker, 1983) cited in included good-practice guidance [1.2].

The post-2015 period was selected to capture the accelerating methodological evolution following the ISPOR-SMDM 2012 Task Force series, including the emergence of PALISADE (2022), CHEERS-AI (2024), DCEA operationalization, and generative-AI-augmented modelling.

2.3. Search Strategy and Information Sources

Systematic database searches was done in Scopus and PubMed/MEDLINE on 18 March 2025. Targeted hand-searching of the ISPOR Value in Health archive, NICE DSU Technical Support Documents, and grey literature from CADTH, IQWiG, PBAC, and EUnetHTA was done. Forward and backward citation from anchor papers (ISPOR-SMDM 2012 series, PALISADE, CHEERS-AI) was used to identify additional relevant sources.

2.3.1. Scopus search strategy (Title, Abstract, Keywords):

TITLE-ABS-KEY (("health economic evaluation" OR "cost-effectiveness analysis" OR "health technology assessment") AND ("Markov model" OR "microsimulation" OR "system dynamics" OR "causal loop" OR "discrete event simulation") AND (limitation* OR assumption* OR memoryless OR "state-space explosion" OR heterogeneity) AND (QALY OR "distributional cost-effectiveness" OR DCEA OR "value assessment")) AND PUBYEAR > 2014 AND DOCTYPE (re OR ar)

2.3.2. PubMed search strategy (MeSH and Title/Abstract):

("Models, Economic"[MeSH] OR "Cost-Benefit Analysis"[MeSH] OR "health economic evaluation"[tiab] OR "HEOR"[tiab]) AND ("Markov Chains"[MeSH] OR "Markov model"[tiab] OR "microsimulation"[tiab] OR "system dynamics"[tiab]) AND ("Quality-Adjusted Life Years"[MeSH] OR QALY[tiab] OR DCEA[tiab] OR "distributional cost-effectiveness"[tiab]) AND ("limitations"[tiab] OR "assumptions"[tiab] OR "machine learning"[tiab]) AND (2015:2025[pdat]) AND (systematicreview[Filter] OR review[Filter])

2.4. Selection of Evidence Sources

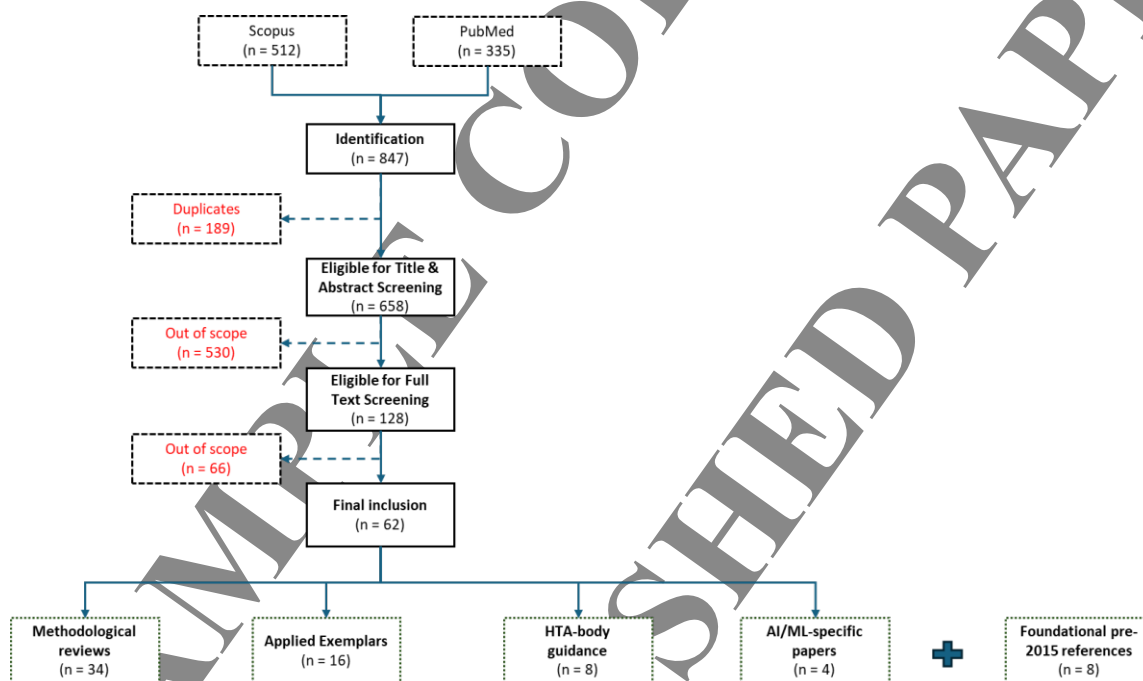


Figure 1: PRISMA-ScR Flow Diagram

A total of 847 records were identified across Scopus (n=512) and PubMed (n=335). Following removal of 189 duplicates, the title and abstract of 658 records were screened by the author and co-reviewer against the inclusion criteria, resulting in exclusion of 530 records. 128 of the total records were eligible for full-text search and were reviewed by the author and the co-reviewer.

Inclusion: Following full-text review, 62 records were retained for data charting and synthesis, consisting of 34 methodological reviews/good-practice statements, 16 applied exemplars, 8 HTA-body guidance documents, and 4 AI/ML-specific methodological papers. Additionally, 8 foundational pre-2015 references were included that are cited as anchoring methodological standards in the post-2015 literature. Any conflict between the author and the co-reviewer was resolved through mutual consultation.

2.5. Data Charting and Synthesis

To extract data, a structured data chart was developed to capture the following:

- Model architecture type;
- Disease area;
- Structural-uncertainty handling method;
- Validation domain [9].
- Equity-sensitivity method (DCEA, ECEA, severity modifier);
- RWE integration approach;
- AI/ML application domain (per PALISADE taxonomy); and reported limitations.

For synthesis, three strategies were used:

1. Narrative-thematic analysis (grouped by limitation domain);
2. Systems-mapping analysis (leveraging causal loop diagrams and the Iceberg Model to explore links between economic models and wider health-system dynamics [20].
3. Conceptual-architectural synthesis to build a future-ready operational framework.

3. Results

3.1. Characteristics of Included Sources

Author (Year)	Model type	Key limitation identified	Equity approach	RWE integration	AI/ML use
Siebert et al. (2012) [3]	Cohort Markov (good practice)	Memorylessness; time-homogeneity	None (QALY default)	Trial-based	None
Hawkins et al. (2005) [10]	Cohort Markov	Forced homogeneity in chronic disease	None	Trial extrapolation	None
Alarid-Escudero et al. (2015) [21]	Markov calibration	Non-identifiability of transition probabilities	None	Calibration to aggregate data	None
Krijkamp et al. (2018) [22]	Microsimulation (R tutorial)	Computational cost; data demands	None (QALY)	Trial + registry	None
Incerti & Jansen (2021) [23]	hesim (cohort + individual STM)	Software fragmentation; reproducibility	QALY default	Modular RWE input	None
Hamilton et al. (2022) [24]	Open-source CHEM framework	Closed proprietary models; lack of transparency	None	Framework for linkage	None
Cookson et al. (2017) [15]	DCEA framework	QALY distributional neutrality	DCEA (equity subgroups)	Not addressed	None
Verguet et al. (2016) [16]	ECEA framework	Omission of financial-risk protection	ECEA	Survey + admin data	None
Padula et al. (2022) [25]	PALISADE (ML in HEOR)	Parameter uncertainty; cohort selection bias	Not addressed	ML-augmented RWE	5 HEOR domains
Elvidge et al. (2024) [26]	CHEERS-AI reporting	Under-reporting of AI model specifics	Not addressed	Not addressed	AI reporting standards
Jackson et al. (2009) [27]	Bayesian model averaging	Structural uncertainty under-addressed	None	Trial-based	None
Bojke et al. (2009) [28]	Structural uncertainty review	Informal scenario analysis dominant	None	Mixed	None
Strong & Oakley (2014) [29]	Metamodelling / VOI	Under-utilisation of EVPPI/EVSI	None	Not addressed	None
Tracy et al. (2018) [13]	Agent-based models	Validation challenges; parameter under-identification	None	Mixed	None
Jit et al. (2019) [30]	Dynamic transmission	Static models overestimate vaccine CE	None	Registry + surveillance	None
Reason et al. (2024) [31]	GenAI model construction	Quality assurance for LLM-generated models	Not addressed	Not addressed	LLM code generation
Squires et al. (2016) [32]	Public health economic models	Disease siloing; omission of capacity constraints	Partial	Mixed	None
De Clercq et al. (2023) [33]	ML moral choice analysis	Algorithmic bias in value judgements	Implicit	Not addressed	ML classification

Table 1: Data Charting: Representative Extraction from Included Sources

Cohort Markov models are dominant across the included literature and are benchmarks for methodological critiques. Equity-sensitive methods, such as DCEA and ECEA, have been discussed in specialized papers; however, not been integrated into applied economic evaluations. RWE integration remains largely trial-based with limited use of target trial emulation or linked administrative records. Post-2022, the reference to AI/ML applications is significant; however, it primarily addresses parameter estimation and not structural or equity dimensions. Handling structural uncertainty is mostly casual (scenario analysis), with Bayesian model averaging done in only two dedicated methodological papers.

3.2. Deconstructing the Memoryless Constraint

In cohort state-transition modelling, the Markov property assumes that the probability of progressing to a future state is dependent only on the current state and not on the sequence of states preceding it [2,3]. For a well-staged chronic condition with a treatment-naïve patient population and time-homogeneous transition probabilities, these assumptions have produced consistent results. However, when such conditions are violated, three typical systematic distortions emerge.

1. **Ahistorical transition estimation:** In diseases where progression is dependent on time spent in a state, total treatment exposure, or past events (for instance, multi-line oncology, chronic kidney disease, heart failure with preserved ejection fraction, or relapsing-remitting multiple sclerosis), the memoryless property forces modellers to give every patient in a state the same transition odds, no matter their path there. Or else they build tunnel states and time-varying matrices that only partly fix the core problem. In the current model, modelers are forced to assign similar transition probabilities across all patient population in a state irrespective of their arrival pathway, even for conditions where cost and prognosis are depended on duration-in-state, accumulated treatment exposure, or prior event history (such as multi-line oncology, chronic kidney disease, heart failure with preserved ejection fraction, relapsing-remitting multiple sclerosis). Alternatively, they construct tunnel states and time-dependent matrices that only partially resolve the underlying violation. While Hawkins, Sculpher and Epstein (2005), demonstrated that this forced homogeneity leads to systematic bias in cost-effectiveness estimates; Alarid-Escudero et al. (2015) concluded that such aggregations lead to non-identifiable parameter solutions [10,21].
2. **State-space explosion:** When treatment sequencing matters clinically, modelers must exponentially expand the state space to capture all possible treatment history combinations within a Markov framework—for n treatment lines with m health states per line, this scales as $O(m^n)$, quickly becoming unmanageable. Microsimulation solves this by tracking individual treatment histories as patient attributes, but at the expense of much higher computational demands and data needs [22,34]. The hesim R package and the DARTH workgroup tutorials have reduced these barriers, however, the institutional default remains the cohort Markov [23,35,36].

3. **Suppression of patient-level variance:** Cohort models push average expected values – such as mean costs and QALYs – through transition matrices for the whole population. This hides the full range of individual patient outcomes essential for budget impact risk assessment, subgroup effects and/or equity analysis. While PSA manages parameter uncertainty, however, it overlooks outcome variations within the cohort. Solutions, such as microsimulation which recovers this distributional information, and partitioned survival models which captures variance in time-to-event distributions, have not displaced the cohort Markov as the default tool for institutions.

3.3. The QALY's Distributional Architecture: Equity Limitations

The QALY utilizes preference-based utility weights to integrate length and quality of life. Grounded in a strictly utilitarian paradigm, the metric treats a QALY gained by any individual as equivalent—irrespective of baseline health status, socioeconomic position, or lifetime health trajectory [5]. This inherent distributional neutrality is a structural feature of the metric, directly generating the equity limitations that are now extensively documented.

In a systematic review related to QALY effectiveness and controversies, Neumann, Cohen and Ollendorf (2021) documented related core objections in four distinct areas [20]:

1. **Severity insensitivity:** the social value assigned in a QALY is identical to a health gain irrespective of the patient's health status and does not differentiate between mild impairment or severe disability. However, across multiple public-preference studies, higher social weight has been assigned to health gains for the severely ill – a preference the underweighted QALY cannot capture [37,38].
2. **Distributional blindness:** The framework further values one QALY distributed across 100 patients identically to one QALY concentrated in a single patient.
3. **Age-weighting:** By design, maximizing lifetime QALYs structurally favors younger patients who naturally have more potential life-years to gain.
4. **Disability paradox:** Higher utilities are reported by individuals with chronic disabilities than the general population valuations assign to their health state [39].

There are three analytics responses which have matured over a period of time:

- DCEA (distribution cost-effectiveness analysis) proposed by Asaria, Griffin and Cookson, 2016; Cookson et al., 2017 decomposes population health gains by equity-relevant subgroups (socioeconomic quintiles or geographic strata), making the equity-efficiency trade-off explicit [14,15].
- ECEA (equity cost-efficiency analysis) by Verguet, Kim and Jamison, 2016; Verguet et al., 2017 expands evaluation beyond health outcomes to incorporate financial-risk protection as an advantage to incorporate the insurance value of publicly financed interventions [16,40].
- Another model, severity-weighted analysis, provides a within-framework adjustment to account for social value of QALY

gain according to the baseline disease severity, operationalised through either proportional shortfall (the fraction of remaining QALYs lost to disease, as adopted in the Dutch system) or absolute shortfall (the absolute number of QALYs lost, as formalised in the 2022 NICE Methods Review; NICE, 2022) [41,42].

While none of the models above resolves the tension between utilitarian aggregation and prioritarian distributive approaches, they do demonstrate that the underweighted QALY is insufficient for decision support related to equity considerations [43].

3.4. The Epistemic Gap: from RCTs to Real-World Evidence

Many of the health economic models leverage treatment-effect estimates from RCTs – done in a controlled setting over a selected population with high adherence and limited follow-ups – which are extrapolated to real-world population, characterized by comorbidity, polypharmacy and variable adherence. The current paradigm for population-level decision making is based on the friction between the trial internal validity and the external validity [44,45].

To overcome this friction methodologically, Hernan and Robins, 2016 proposed target trial emulation[46]. To recover treatment-effect estimates valid both internally and externally, they defined a protocol of a hypothetical trial and emulated it using observational data with causal interference methods (inverse probability weighting, doubly-robust estimation, or TMLE). Further, the PALASIDE checklist by Padula et al., 2022 established causal interference as one of the five domains where ML can enhance HEOR [25]. However, extrapolation of target trials to standard HTA workflow remains underdeveloped and most of the analysis is based on trial-derived base-case estimates.

Additionally, cost transferability further compounds the epistemic gap since the cost is context-specific and depends on system structure, unit-cost methodology, etc. [45,47,48]. In LMICs, the analysis is further complicated by the absence of routing cost-accounting data, forcing reliance on extrapolated HIC cost structures [49,50]. Further, empirically grounded cost-effectiveness thresholds yield substantially lower values than the critiqued WHO-CHOICE rule-of-thumb [51-53].

3.5. Structural Uncertainty: The Under-Theorised Dimension

While parameter uncertainty is routinely quantified and propagated through probabilistic sensitivity analysis (PSA) as mandated by most global HTA agencies, structural uncertainty remains the most consequential and least addressed epistemic challenge [54]. In their foundational taxonomy, Bojke et al. (2009) illustrated the origin to be the architecture of the model itself and not from the data inputs [28]. It covers the subjective choices modelers must make regarding state definitions, cycle lengths, survival extrapolations, comparator selection, evidence-synthesis methodologies, and the overarching analytical perspective.

To propagate structural uncertainty, Bayesian Model Averaging (BMA) offers a formal mechanism. A foundational example was

provided by Jackson, Thompson and Sharples (2009), who applied BMA to health economics models evaluating different surgical techniques [27]. Instead of committing to a single assumed framework, they assigned posterior probabilities to several competing model structures, conditioned strictly on the observed data. The outcome led to a final cost-effectiveness estimate reflecting uncertainty distributed across multiple structural choices, rather than just the parameter variance within one. Mechanically, BMA evaluates the outputs of each candidate model according to how well the evidence supports each model, ultimately blending them into a single composite posterior distribution for the incremental net benefit. Building on this paradigm, Strong and Oakley (2014) explored efficient exploration of structural sensitivity by extending this approach through metamodeling [29].

However, BMA adoption remains limited largely due to computational demands, institutional inertia and reviewers' familiarity. These operational barriers are operational and methodological and can be addressed at the institutional level.

While BMA quantifies the structural uncertainty across competing models, a strategic immediate question for policymakers remains unanswered – 'does this uncertainty threatens optimal resource allocation? Will investing in new research to resolve it yield worthy outcomes?'

The Value-of-information (VOI) analysis bridges this gap between statistical variance and health policy. It provides a decision-theoretic framework for determining the financial implications of current uncertainty. Matrices—such as EVPI (expected value of perfect information), EVPPI (expected value of partial perfect information), and EVSI (expected value of sample information)—are methodologically mature, computationally tractable with recent advances, and directly policy-relevant. Yet VOI analysis is rarely commissioned in practice [55-57].

4. Discussion and Synthesis

4.1. The Model as System Fragment

The evidence discussed above suggests a structural disconnect of the standard health economic model from the health system in which its outputs should operate. The Markov model calculates the incremental cost-effectiveness of the interventions based on fixed assumptions about care pathways, resource availability, workforce capacity, and behavioural adherence. Each of these variables is dynamic and subject to different constraints across settings and systems. Because of the paradigm's implicit ontology, which treats an intervention as a discrete input to a static system, the model treats these factors as exogenous parameters.

The practical consequences are obvious. Models demonstrating cost-effectiveness under trial conditions many times fail to deliver projected health gains at scale because of system and stakeholder-led factors, such as workforce constraints, diagnostic bottlenecks, behavioural adaptation, etc. The model is—as described by Meadows' (2008)—operating at the level of 'events' without having mapped the 'patterns', 'structures', or 'mental models' that

govern system response [20].

4.2. Systems Thinking as Pre-Analytic Infrastructure

If deployed as a pre-analytics layer before economic qualification, systems-level thinking tools can augment economic models and enhance their decision-making cost-effectiveness analysis.

- Causal loop diagramming (CLD) identifies reinforcing and balancing feedback loops bridging the gap between the intervention and the broader system. For instance, in a national vaccination program, CLDs can identify relationships/dependence of reinforcing loops (herd immunity reducing disease burden) and balancing loops (vaccine hesitancy responding to adverse-event reports, workforce fatigue constraining delivery capacity). These interdependencies can help decision-makers determine if linear extrapolation of trial-derived efficacy will hold at scale. Pitman et al. (2012) and Jit et al. (2019) established that static models systematically overestimate cost-effectiveness for vaccines by omitting the reinforcing loop of herd immunity [30,58].
- Stock-and-flow mapping identifies capacity constraints—workforce, equipment, treatment slots, funding envelopes—that can limit implementation and national-level roll-outs. Marshall et al. (2015) and Squires et al. (2016) argued for incorporating capacity dynamics into public health economic evaluation [12,32].
- The Iceberg Model provides a diagnostic hierarchy: events (the ICER) are driven by patterns (uptake and adherence trends), which are driven by structures (care-pathway design, reimbursement rules), which are driven by mental models (assumptions about behaviour, incentives, political priorities) [20,59]. An economic model operating only at the event level is, by construction, an incomplete decision tool.

4.3. Why the Current Paradigm Persists

Multiple factors are driving a widespread application of the cohort Markov and unweighted QALY despite the limitations documented above.

1. Simplicity and transparency: Cohort Markov models are interpretable by non-specialist committee members who must make reimbursement decisions. The state-transition logic is visually intuitive, parameter inputs can be traced to identifiable sources, and the cost-effectiveness plane and cost-effectiveness acceptability curves provide decision-ready outputs. More complex architectures—microsimulation, ABMs, system dynamics—sacrifice this transparency, creating a tension between methodological sophistication and decision-maker comprehension.
2. Regulatory acceptance and institutional infrastructure: Majority of regulatory bodies, such as NICE, CADTH, IQWiG, and PBAC, have built their processes including submission templates, technical assessment group training, and committee appraisal, around state-transition models. Transitioning to alternative structures may require additional

investment in submission guidance, reviewer training, and quality-assurance infrastructure.

3. Data limitations: Many of the alternatives to the cohort Markov—such as microsimulation, dynamic transmission models and system dynamics models—require substantially higher data demands. In disease areas or geographies where required data is not available, the Markov may represent the best option.
4. Computational feasibility: Despite advances in computational power globally, there is limited investment to optimize and meet computational requirements for advanced modelling tools, such as probabilistic microsimulation or agent-based models, at a national program level scale. The hesim package and C++-backend simulation address this partially, but the gap remains significant for time-constrained HTA timelines [23].

Therefore, the evolution of calculating the impact of any intervention must be supported by institutional investment and not just by methodological advocacy.

5. The Future-State HEOR Framework: A Conceptual Architecture

5.1. From Static Publication to Living Model

Most of the economic models constructed for a specific HTA submission are limited by the current paradigm, which treats them as a fixed artefact and therefore, are rarely updated. However, current requirements, including adaptive licensing, coverage-with-evidence-development agreements, and post-market effectiveness monitoring, increasingly challenge them.

An alternative is a ‘living model’ - a modular, version-controlled, continuously updated modelling platform that ingests new evidence streams and dynamically revises transition probabilities, cost estimates, and cost-effectiveness conclusions.

The technical infrastructure is developing and can support proposed living models. Hamilton et al. (2022) developed open-source software frameworks for transparent, reusable computational health economic models [24]. Incerti and Jansen (2021) developed ‘Hesim’ with a modular submodel architecture [23]. The DARTH workgroup has published reproducible microsimulation tutorials [22,35]. However, the institutional architecture is missing. There is a need for API-driven data pipelines linking health-data infrastructure to modelling platforms, governance protocols for model-version management, and regulatory frameworks to recognize living models as valid decision inputs.

5.2. The SHAN Strategic House: a Conceptual, Forward-Looking Framework

The proposed ‘SHAN Strategic House’ is a conceptual, forward-looking framework for national-level outcome and performance calculations. It is an aspirational architecture against which current capabilities can be benchmarked and investment priorities can be identified.

Layer	Function	Methods and tools	Current maturity
1. Systems map-ping (foundation)	Map endogenous vari-ables, feedback loops, capacity constraints, and mental models before economic quan-tification.	Causal loop diagrams; stock-and-flow models; Iceberg Model; stakeholder co-production.	Low. Rarely integrat-ed into HTA work-flow.
2. Modular model engine	Execute economic simulations on in-teroperable dis-ease/costing compo-nents sharing a demo-graphic backbone.	Whole-disease models; mi-crosimulation/DES (hesim, DARTH); version-controlled open-source code [60].	Medium. Frameworks emerging; institutional adoption limited.
3. Evidence inte-gration	Continuously ingest trial, registry, EHR, and surveillance data; update parameters via Bayesian calibration.	API-driven pipelines; target trial emulation; Bayesian updating; TMLE/double-debiased ML (PALISADE).	Low–Medium. Meth-ods mature; data link-age lagging.
Layer	Function	Methods and tools	Current maturity
4. Equity and dis-tributional analysis	Decompose costs/outcomes by equity subgroups; quan-tify equity–efficiency trade-offs.	DCEA; ECEA; severity-weighted analysis; health-inequality impact assessment.	Low. Published but rarely operationalised in HTA.
5. Decision-support interface (roof)	Present real-time indi-cators, updated ICERs/NMB, VOI priorities, and budget-impact forecasts.	Dashboard interfaces; scenario tools; EVPI/EVSI displays; API integration with national health platforms.	Very low. No national HTA body operates a real-time interface of this kind.

Table 2. The SHAN Strategic House: A Five-Layer Conceptual Architecture for National-Level HEOR

The foundation—systems mapping—is the base layer and determines the model boundaries. It helps the ‘modular model engine’ layer to avoid the similar isolation problem that effects current static models. The ‘decision-support interface’ (layer 5) repositions the economic model from a static HTA artefact to a continuously operating national health-system governance transitioning to a preventive-care command and decision-support center for national regulators with capabilities of monitoring leading indicators, dynamically updating transition probabilities as post-market data accumulates, and presenting VOI-informed research prioritization.

Moving forward, future research will focus on validating each layer through empirical evidence available.

5.3. Governed AI Augmentation

AI is proposed to support the architecture across multiple layers. At the evidence-integration layer, machine learning (ML) can support cohort selection, feature identification, predictive analytics, and causal inference [25]. At the model-engine layer, GenAI can accelerate model construction and sensitivity-analysis automation [31]. While the AI and ML can optimize future work, argued in favor of human-in-the-loop governance to mitigate the risk associated with ML-based approaches and algorithmic biases [33]. CHEERS-AI provides reporting infrastructure; PALISADE provides quality assurance, but binding HTA regulatory guidance remains minimal [26].

5.4. Closing the HIC–LMIC Maturity Gap

The SHAN Strategic House is context-adaptable. In HIC settings, emphasis will be on Layers 3–5. In LMICs, the priority will be establishing Layers 1–2. The maturity gap will not be closed by importing HIC model structures and will require investment

in LMIC modelling capacity, federated data infrastructure, and empirically grounded thresholds [49,52,53,61].

6. Conclusion

For three decades, the cohort Markov model and the unweighted QALY have been the benchmarks for economic modelling in healthcare. They successfully established a common analytic language to enable systematic comparison of health technologies and interventions, and provided a decision-theoretic grounding for resource allocation. However, the validity of these tools is now challenged as the complexity of the decision landscape has shifted and expanded. These memoryless properties are mathematically constrained by precision-medicine treatment sequencing. The QALY’s distributional neutrality is subject to growing methodological critique for equity-sensitive decision-making.

These closed population models are static and disease-siloed, and therefore may be insufficient for system-level decision-making, which will require feedback, needs to acknowledge capacity constraints and ensure behavioral adaptation. To transition from memoryless models to history-aware architectures, further research is warranted. The need is to move away from utilitarian aggregation towards distribution analysis; from static extrapolation to continuous Bayesian updating; from disease silos to system-embedded platforms. The SHAN Strategic House framework aims to provide a conceptual target whose empirical validation and institutional operationalization will lead to the next generation’s research program. Technical methods and capacities are available. The institutes need to commit and follow. While the first generation of health economic modelling demonstrated that systematic analysis of costs and outcomes could improve resource allocation, the second generation will demonstrate that it can do so

equitably, dynamically, and at the system level [62-66].

Acknowledgements and Disclosures

1. **Kushagra Gupta, student – GEMS, Dubai:** Co-reviewer for identification and shortlisting of relevant published and unpublished research.
2. **Grammarly:** Used for language edits and grammar iterations
3. **Google Gemini:** Used to uncover unpublished research, including Google Scholar

References

1. Beck, J. R., & Pauker, S. G. (1983). The Markov process in medical prognosis. *Medical decision making*, 3(4), 419-458.
2. Sonnenberg, F. A., & Beck, J. R. (1993). Markov models in medical decision making: a practical guide. *Medical decision making*, 13(4), 322-338.
3. Siebert, U. et al. (2012) 'State-transition modeling: ISPOR-SMDM Task Force-3', *Value in Health*, 15(6), pp. 812-820.
4. Caro, J.J. et al. (2012) 'Modeling good research practices—overview: ISPOR-SMDM Task Force-1', *Value in Health*, 15(6), pp. 796-803.
5. Weinstein, M.C., Torrance, G.W. and McAlister, A. (2009) 'QALYs: the basics', *Value in Health*, 12(Suppl 1), S5-S9.
6. Drummond, M.F. et al. (2005) *Methods for the economic evaluation of health care programmes*. 3rd edn. Oxford: OUP.
7. Briggs, A., Sculpher, M., & Claxton, K. (2006). *Decision modelling for health economic evaluation*. Oup Oxford.
8. Briggs, A.H. et al. (2012) 'Model parameter estimation and uncertainty: ISPOR-SMDM Task Force-6', *Value in Health*, 15(6), pp. 835-842.
9. Eddy, D. M., W. Hollingworth, J. J. Caro, J. Tsevat, K. M. McDonald, and J. B. Wong. "ISPOR- SMDM Modeling Good Research Practices Task Force. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7." *Value Health* 15, no. 6 (2012): 843-50.
10. Hawkins, N., Sculpher, M. and Epstein, D. (2005) 'Cost-effectiveness analysis of treatments for chronic disease', *Health Economics*, 14(12), pp. 1229-1236.
11. Frederix, G.W.J. et al. (2017) 'The use of health economic models for oncology drugs', *Value in Health*, 20(A498).
12. Marshall, D. A., Burgos-Liz, L., IJzerman, M. J., Osgood, N. D., Padula, W. V., Higashi, M. K., ... & Crown, W. (2015). Applying dynamic simulation modeling methods in health care delivery research—the SIMULATE checklist: report of the ISPOR simulation modeling emerging good practices task force. *Value in health*, 18(1), 5-16.
13. Tracy, M., Cerdá, M., & Keyes, K. M. (2018). Agent-based modeling in public health: current applications and future directions. *Annual review of public health*, 39, 77-94.
14. Asaria, M., Griffin, S., & Cookson, R. (2016). Distributional cost-effectiveness analysis: a tutorial. *Medical decision making*, 36(1), 8-19.
15. Cookson, R., Mirelman, A. J., Griffin, S., Asaria, M., Dawkins, B., Norheim, O. F., ... & Culyer, A. J. (2017). Using cost-effectiveness analysis to address health equity concerns. *Value in Health*, 20(2), 206-212.
16. Verguet, S., Kim, J. J., & Jamison, D. T. (2016). Extended cost-effectiveness analysis for health policy assessment: a tutorial. *Pharmacoeconomics*, 34(9), 913-923.
17. Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., ... & Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Annals of internal medicine*, 169(7), 467-473.
18. Arksey, H., & O'malley, L. (2005). Scoping studies: towards a methodological frame-work. *International journal of social research methodology*, 8(1), 19-32.
19. Levac, D., Colquhoun, H., & O'brien, K. K. (2010). Scoping studies: advancing the methodology. *Implementation science*, 5(1), 69.
20. Meadows, D.H. (2008) *Thinking in systems: a primer*. White River Junction: Chelsea Green.
21. Neumann, P.J., Cohen, J.T. and Ollendorf, D.A. (2021) 'Controversy over QALYs: a systematic review', *Health Affairs*, 40(9), pp. 1402-1410.
22. Alarid-Escudero, F., MacLehose, R. F., Peralta, Y., Kuntz, K. M., & Enns, E. A. (2018). Nonidentifiability in model calibration and implications for medical decision making. *Medical Decision Making*, 38(7), 810-821.
23. Krijkamp, E. M., Alarid-Escudero, F., Enns, E. A., Jalal, H. J., Hunink, M. M., & Pechlivanoglou, P. (2018). Microsimulation modeling for health decision sciences using R: a tutorial. *Medical Decision Making*, 38(3), 400-422.
24. Incerti, D., & Jansen, J. P. (2021). hesim: Health economic simulation modeling and decision analysis. *arXiv preprint arXiv:2102.09437*.
25. Hamilton, M.P. et al. (2022) 'A prototype software framework for transferable CHEMs', *PLOS ONE*, 17(2), e0263299.
26. Padula, W. V., Kreif, N., Vanness, D. J., Adamson, B., Rueda, J. D., Felizzi, F., ... & Crown, W. (2022). Machine learning methods in health economics and outcomes research—the PALISADE checklist: a good-practices report of an ISPOR task force. *Value in health*, 25(7), 1063-1080.
27. Elvidge, J., Hawksworth, C., Aşar, T. S., Zempenyi, A., Chalkidou, A., Petrou, S., ... & Wilson, E. (2024). Consolidated health economic evaluation reporting standards for interventions that use artificial intelligence (CHEERS-AI). *Value in Health*, 27(9), 1196-1205.
28. Jackson, C. H., Thompson, S. G., & Sharples, L. D. (2009). Accounting for uncertainty in health economic decision models by using model averaging. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 172(2), 383-404.
29. Bojke, L., Claxton, K., Sculpher, M., & Palmer, S. (2009). Characterizing structural uncertainty in decision analytic models: a review and application of methods. *Value in Health*, 12(5), 739-749.
30. Strong, M., & Oakley, J. E. (2014). When is a model good enough? Deriving the expected value of model improvement via specifying internal model discrepancies. *SIAM/ASA Journal on Uncertainty Quantification*, 2(1), 106-125.
31. Jit, M. et al. (2019) 'Multi-model comparison of vaccination strategies', *Vaccine*, 37(19), pp. 2515-2520.

31. Reason, T., Rawlinson, W., Langham, J., Gimblett, A., Malcolm, B., & Klijn, S. (2024). Artificial Intelligence to Automate Health Economic Modelling: A Case Study to Evaluate the Potential Application of Large Language Models: T. Reason et al. *PharmacoEconomics-Open*, 8(2), 191-203.
32. Squires, H. et al. (2016) 'A framework for developing economic models of public health interventions', *Public Health Research*, 4(4).
33. Smeele, N. V., Chorus, C. G., Schermer, M. H., & de Bekker-Grob, E. W. (2023). Towards machine learning for moral choice analysis in health economics: A literature review and research agenda. *Social science & medicine*, 326, 115910.
34. Karnon, J., Stahl, J., Brennan, A., Caro, J. J., Mar, J., & Möller, J. (2012). Modeling using discrete event simulation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-4. *Medical decision making*, 32(5), 701-711.
35. Jalal, H., Pechlivanoglou, P., Krijkamp, E., Alarid-Escudero, F., Enns, E., & Hunink, M. M. (2017). An overview of R in health decision sciences. *Medical decision making*, 37(7), 735-746.
36. Alarid-Escudero, F., Krijkamp, E., Enns, E. A., Yang, A., Hunink, M. M., Pechlivanoglou, P., & Jalal, H. (2023). An introductory tutorial on cohort state-transition models in R using a cost-effectiveness analysis example. *Medical Decision Making*, 43(1), 3-20.
37. Nord, E. (2005). Concerns for the worse off: fair innings versus severity. *Social science & medicine*, 60(2), 257-263.
38. Dolan, P. et al. (2005) 'QALY maximisation and people's preferences', *BMJ*, 331, pp. 685-688.
39. Grosse, S. D., Lollar, D. J., Campbell, V. A., & Chamie, M. (2009). Disability and disability-adjusted life years: not the same. *Public Health Reports*, 124(2), 197-202.
40. Verguet, S. et al. (2017) 'Health gains and financial risk protection afforded by public financing', *The Lancet*, 390(10104), pp. 1480-1489.
41. Stolk, E. A., van Donselaar, G., Brouwer, W. B., & Busschbach, J. J. (2004). Reconciliation of economic concerns and health policy: illustration of an equity adjustment procedure using proportional shortfall. *PharmacoEconomics*, 22(17), 1097-1107.
42. NICE (2022) NICE health technology evaluations: the manual. Process and methods [PMG36]. London: *National Institute for Health and Care Excellence*.
43. Norheim, Ole F., Rob Baltussen, Mira Johri, Dan Chisholm, Erik Nord, DanW Brock, Per Carlsson et al. "Guidance on priority setting in health care (GPS-Health): the inclusion of equity criteria not captured by cost-effectiveness analysis." *Cost Effectiveness and Resource Allocation* 12, no. 1 (2014): 18.
44. Sculpher, M.J. et al. (2004) 'Subgroup analysis alongside clinical trials', *Health Economics*, 13(5), pp. 461-475.
45. Drummond, M., Barbieri, M., Cook, J., Glick, H. A., Lis, J., Malik, F., ... & Severens, J. (2009). Transferability of economic evaluations across jurisdictions: ISPOR Good Research Practices Task Force report. *Value in health*, 12(4), 409-418.
46. Hernán, M. A., & Robins, J. M. (2016). Using big data to emulate a target trial when a randomized trial is not available. *American journal of epidemiology*, 183(8), 758-764.
47. Welte, R., Feenstra, T., Jager, H., & Leidl, R. (2004). A decision chart for assessing and improving the transferability of economic evaluation results between countries. *Pharmacoeconomics*, 22(13), 857-876.
48. Goeree, R., He, J., O'Reilly, D., Tarride, J. E., Xie, F., Lim, M., & Burke, N. (2011). Transferability of health technology assessments and economic evaluations: a systematic review of approaches for assessment and application. *ClinicoEconomics and Outcomes Research*, 89-104.
49. Vassall, A. et al. (2017) 'Reference case for estimating costs of global health services', *Global Health Action*, 10(sup1), 1364022.
50. Hutubessy, R., Chisholm, D., Edejer, T. T. T., & Who-choice. (2003). Generalized cost-effectiveness analysis for national-level priority-setting in the health sector. *Cost effectiveness and resource allocation*, 1(1), 8.
51. Marseille, E. et al. (2015) 'Thresholds for cost-effectiveness: alternative approaches', *Bulletin of the WHO*, 93(2), pp.118-124.
52. Woods, B., Revill, P., Sculpher, M., & Claxton, K. (2016). Country-level cost-effectiveness thresholds: initial estimates and the need for further research. *Value in health*, 19(8), 929-935.
53. Ochalek, J., Lomas, J., & Claxton, K. (2018). Estimating health opportunity costs in low-income and middle-income countries: a novel approach and evidence from cross-country data. *BMJ global health*, 3(6), e000964.
54. Claxton, K., Sculpher, M., McCabe, C., Briggs, A., Akehurst, R., Buxton, M., ... & O'Hagan, T. (2005). Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. *Health economics*, 14(4), 339-347.
55. Strong, M., Oakley, J. E., & Brennan, A. (2014). Estimating multiparameter partial expected value of perfect information from a probabilistic sensitivity analysis sample: a nonparametric regression approach. *Medical Decision Making*, 34(3), 311-326.
56. Heath, A., Kunst, N., Jackson, C., Strong, M., Alarid-Escudero, F., Goldhaber-Fiebert, J. D., ... & Jalal, H. (2020). Calculating the expected value of sample information in practice: considerations from 3 case studies. *Medical Decision Making*, 40(3), 314-326.
57. Ades, A. E., Lu, G., & Claxton, K. (2004). Expected value of sample information calculations in medical decision modeling. *Medical decision making*, 24(2), 207-227.
58. Pitman, R., Fisman, D., Zaric, G. S., Postma, M., Kretzschmar, M., Edmunds, J., ... & ISPOR-SMDM Modeling Good Research Practices Task Force. (2012). Dynamic transmission modeling: a report of the ISPOR-SMDM modeling good research practices task force-5. *Value in health*, 15(6), 828-834.
59. Senge, P.M. (1990) *The fifth discipline*. New York: Doubleday.
60. Tappenden, P., Chilcott, J., Brennan, A., Squires, H., & Stevenson, M. (2012). Whole disease modeling to inform

resource allocation decisions in cancer: a methodological framework. *Value in Health*, 15(8), 1127-1136.

61. Chisholm, D., & Evans, D. B. (2007). Economic evaluation in health: saving money or improving care? *Journal of Medical Economics*, 10(3), 325-337.
62. Bertram, M. Y., Lauer, J. A., De Joncheere, K., Edejer, T., Hutubessy, R., Kieny, M. P., & Hill, S. R. (2016). Cost-effectiveness thresholds: pros and cons. *Bulletin of the World Health Organization*, 94(12), 925.
63. Jin, H., Tappenden, P., Ling, X., Robinson, S., & Byford, S. (2023). A systematic review of whole disease models for informing healthcare resource allocation decisions. *PLoS One*, 18(9), e0291366.
64. Peters, M.D.J. et al. (2020) 'Updated methodological guidance for scoping reviews', *JBI Evidence Synthesis*, 18(10), pp. 2119–2126.
65. Shillcutt, S. D., Walker, D. G., Goodman, C. A., & Mills, A. J. (2009). *Cost effectiveness in low-and middle-income countries: a review of the debates surrounding decision rules. Pharmacoeconomics*, 27(11), 903-917.
66. Reckers-Droog, V.T., van Exel, N.J.A. and Brouwer, W.B.F. (2018) 'Looking back and moving forward: on the application of proportional shortfall in healthcare priority setting in the Netherlands', *Health Policy*, 122(6), pp. 621–629.

SAMPLE COPY
UNPUBLISHED PAPER

Copyright: ©2026 Anuj Gupta. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.