From Bespoke Models to Shared Infrastructure: NICE ECD18 and the Future of HTA Modelling

HTA386

Gareth Wynne-McHardy[1], MBChB, Amanda Hansson Hedblom[1], MSc, Anna McCormick[1], DPhil, Gerdi Strydom[1], MBA, PharmD

[1]Valid Insight (Part of the Bioscript Group), Macclesfield, Cheshire, United Kingdom

Objectives

This study aims to:

- Describe the change from bespoke to standardised models in certain indications, from early pilots to the first disease-specific extension;
- Identify the elements of the analysis that stay open to company influence;
- Assess where GenAI tools can support these elements;
- And identify key uncertainties unearthed by reference case and model use.

Methods

In this targeted narrative review, NICE appraisals and methods documents were identified from January 2016 to October 2026, which discuss or show a move from bespoke to shared models. These studies were identified through ECD18 and its cited literature, NICE pilot pages and conference reports.[1]

Next, GenAI evidence was mapped from the ISPOR working group report and NICE HTA Lab outputs to each part of the analysis that stays open to company influence.[2, 3] Where a primary paper or abstract was available, GenAI results were checked against it.

Key terms in ECD18

Reference case extension

A NICE document that adds disease-level detail to the NICE reference case. It specifies what a model should have, but it is not a working model.[1] It can fix outcomes, comparators, subgroups, a conceptual model, input assumptions and datasets. Each element is 'required' or 'recommended'.[1] Example: PMG50.[4]

Reference model

A working (executable) model for a whole disease area. It is built independently of companies, filled with the best available evidence and validated. Companies can use it in their submissions.[1] Example: the TA964 pathway model. [1, 5]

Results

NICE's approach to health economic modelling is moving from bespoke to standardised models, but only in selected disease areas (Figure 1).[1]

Where an extension applies, NICE fixes the model design, so the company's effort moves to evidence and input justification (Figure 2). GenAI results vary most for model conceptualisation, the step that an extension fixes, and are nearer human level for extraction, NMA code and model coding (Figure 3). ECD18 names AI as one way to update or adapt reference models.[1]

From bespoke models to shared structure

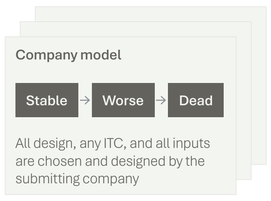
Default throughout: a separate model for each appraisal, usually built or commissioned by the submitting company[1]

First moves towards a shared structure: One model for several drugs		Pathways pilots: NICE-commissioned whole-pathway models		Being tested now: NICE specifies the design for selected diseases		
Jan 2016 TA375	2022 to Mar 2023 TA878	Apr 2024 TA964	2023 to 2026 NSCLC	Sep 2025 ECD18	May 2026 PMG50	2026 to 2027 MASH
Rheumatoid arthritis MTA. Six companies each sent their own model for seven drugs, and the analyses did not converge on the most cost-effective intervention.[6] The committee used the independent assessment group model. NICE reused it in TA715 (2021), and NICE's 2025 position statement on reference case extensions (ECD18) cites it as an early reference model. [1, 7]	**COVID-19 MTA.** Timelines were short, so the assessment group adapted an earlier model and compared nine treatments within it.[8] It did not save time compared to the standard process, and appeals on several treatments were upheld.[9]	**First NICE pathways pilot,** in advanced renal cell carcinoma. PenTAG built one open-source model of the full pathway: 12 treatments, 3 risk groups and up to 4 lines.[5, 10] After feedback, NICE said it would no longer "focus on NICE-built, owned and maintained pathways models", but on reference models that fit the standard appraisal process. [11]	**Second pathways pilot** in metastatic non-small cell lung cancer, with more than 50 NICE recommendations.[12] Bristol TAG built one updatable model for stage 4 disease, not eligible for first-line targeted therapy.[12, 13] With the change of focus following TA964, NICE cancelled the March 2024 committee meeting for this model.[13] Work on the model continued, and the report was published in 2026.[12, 13]	**NICE position statement.** NICE will write 'reference case extensions' for selected diseases. Each one specifies what a model should have (see Key terms). Companies must justify each deviation. NICE will consider building its own model only in some cases. It names model ownership, resources, upkeep and confidential company data as challenges.[1]	**The first reference case extension under ECD18** is for overweight and obesity in adults. It sets rules for population strata, comparators, model structure, clinical inputs, utilities and costs.[4] It applies to all NICE guidance on this topic.[1, 4]	**The second reference case extension** will cover metabolic dysfunction-associated steatohepatitis (MASH). In 2026 to 2027, NICE will adapt the conceptual model that its HTA Lab published in June 2025.[1, 14]

Figure 1. The default since 2000 is a separate model for each appraisal, however, multiple technology appraisals and pathways pilots were early moves to shared models and ECD18 makes this a policy, but only for selected disease areas.[1]

The company's role in bespoke versus shared models

Bespoke
The default



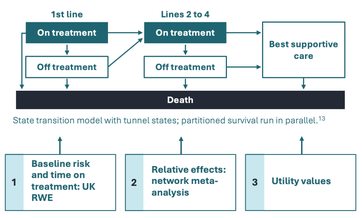
Each company designs its own model for each appraisal.

The company chooses the structure, assumptions and inputs, and defends them against EAG criticism.

Submissions for a single disease can use different designs, so inputs and assumptions can differ across appraisals.[1]

Shared
Example: TA964 - 2024

EAG pathway model (built by PenTAG)
12 treatments, 3 risk groups, prior adjuvant therapy or not[5]



One EAG model covered the whole pathway, but Ipsen disputed key assumptions:[5, 11, 15]

	Ipsen Preferred	Committee outcome
1	Time on treatment equal to PFS, not UK RWE (Ipsen could not see all UK data).	Kept UK RWE.
2	PH NMA throughout, not FP NMA at first line.	Kept FP NMA at first line; accepted PH NMA for later lines.
3	Utilities from more past appraisals, or from CheckMate 9ER.	Kept values accepted in past NICE appraisals; 9ER values judged implausibly high.

Ipsen also preferred 2 treatment lines (not 4) and naive safety comparisons; the committee kept its choices. Ipsen also raised surrogacy, PSA run time and dose intensity.[11, 15]

Figure 2. In a bespoke model, the company makes every design choice. In TA964, the committee kept its own choice on inputs 1 to 3, but a scenario that Ipsen pressed for supported a positive recommendation.[11, 15] An extension can also fix input assumptions and datasets, so less stays open to challenge.[1]

Current evidence: GenAI performance in supporting HEOR tasks

Element	Capability	Limitation
Screening and extraction	GPT-5 mini identified 96-100% of relevant records at the title-abstract stage.[16] GPT-5 Pro matched human reviewers on abstract screening.[17] GPT-5.2 and Gemini 3 Pro extracted 95.1% of data points correctly.[18]	GPT-5 Pro missed over half of relevant studies at full text.[17] Models omitted details and misassigned values in papers with many subgroup results.[18, 19]
ITC/NMA	GPT-4 extracted data with >99% success over 20 runs. Its R code ran with no or minor edits in most runs.[20]	Complex networks needed fixes.[20]
RWE	GPT-5.5 and Gemini 3.1 Pro turned published study designs into analysis code with up to 95-97% accuracy.[21] GPT-5-mini and similar models extracted data from hospital reports with F1 up to 0.99.[22]	Extraction from health records is tested in one hospital and one specialty.[22]
Model conceptualisation	GPT-5.2 proposed model structures within NICE-accepted choices in about 1h of computer time.[23-25]	GPT-5.2 missed a biomarker split that depended on patient-level data.[23, 25] The single study checked structure only, with unblinded experts.[23]
Model build in a set structure	Claude in Excel extracted over 90% of inputs correctly and built traceable formulas.[26] ChatGPT 5 replicated a simple published model with few corrections.[27]	Output accuracy for incremental QALYs reached at most 66%. Claude applied cycle length, half-cycle correction, and discounting inconsistently.[26] An experienced modeller still needed 4-8 h.[26]

Key: **Green:** several studies near human level **Amber:** case studies **Red:** variable or poor.

Conference abstracts and recently published, up-to-date evaluations of the capabilities of GenAI tools in model-related workflows demonstrate considerable potential. If indications with the greatest submission burden move towards a more standardised modelling process, the opportunity for GenAI-assisted model development increases. The HTA Lab is now evaluating an AI-assisted evidence submission.[3,28]

Figure 3. Targeted assessment of published and conference evidence on GenAI in HTA evidence and modelling tasks. Results were closest to human-level for title-abstract screening and data extraction, but not for full-text screening.[16-18] Evidence for NMA, RWE, and model conceptualisation came from a small number of case studies,[20, 21, 23] and results for model building varied between runs.[26]

Upcoming milestones

Obesity extension Published 6 May 2026	PMG50 is NICE's first reference case extension, for overweight and obesity in adults.[4] The role of company influence within the constraints of the extension is of particular interest, and in time the evaluation of any effect on recommendation outcome.[1]
MASH extension In development, 2026 to 2027	NICE will adapt its HTA Lab conceptual model into the second extension.[1, 14] ECD18 gives no date for consultation or publication.
Review of ECD18 After the obesity and MASH extensions	NICE will review the position statement once both extensions are in use, and judge whether the benefits outweigh the challenges.[1] The first review could change what NICE expects in company submissions.
Choice of future indications Process in development	NICE will develop a process, with stakeholder input, to choose indications in a transparent and proportionate way.[1] Priority goes to high-priority clinical areas with weak or inconsistent models, new technologies that will change care, or many products in parallel.[1] NICE will announce each planned extension, so that companies can align their models.[1] A framework presented at ISPOR 2026 in Vienna suggests advanced cancer as a plausible candidate.[29]

Conclusions

Where an extension applies, NICE has set a model design for companies to build towards, thereby shifting aspects of modelling initiative from companies to NICE.[1, 4] If the trend towards greater standardisation of model design continues, company influence will then increasingly rest on evidence quality, ITC methods and parameter justification, as TA964 shows.[11, 15] NICE will review ECD18 after its first two extensions, so the scale of the impact of this exploration will become clearer following their completion.[1]

GenAI now performs near-human-level performance in title-abstract screening and data extraction, but evidence for NMA, RWE, and model conceptualisation is limited to case studies.[16-18, 20, 21, 23] Model building yields the least consistent results because modelling choices vary between runs.[26] GenAI can speed up each task, but expert oversight is essential, and a NICE requirement.[3, 30] NICE's HTA Lab now calls for tests of AI model building from a model specification, which an extension provides.[3]

References

1. NICE. ECD18; 2025. https://www.nice.org.uk/corporate/ecd18.
2. Fleurence RL, et al. Value Health. 2025;28(2):175-183. doi:10.1016/j.jval.2024.10.3846.
3. NICE HTA Lab. GenAI in health economic modelling report. https://www.nice.org.uk/what-nice-does/our-research-work/hta-lab.
4. NICE. PMG50; 2026. https://www.nice.org.uk/process/pmg50.
5. Lee D, et al. Pharmacoecon Open. 2024;8(5):645-649. doi:10.1007/s41669-024-00490-x.
6. NICE. TA375; 2016. https://www.nice.org.uk/guidance/ta375.
7. NICE. TA715; 2021. https://www.nice.org.uk/guidance/ta715.
8. NICE. TA878; 2023 (updated 2026). https://www.nice.org.uk/guidance/ta878.
9. Tosh J. ISPOR Europe 2023; Copenhagen, Denmark.
10. Lee D, et al. Health Technol Assess. 2026;30(1). doi:10.3310/GJDL0327.
11. NICE. TA964 committee B slides; 2024. https://nice.org.uk/guidance/ta964/documents/1-2.
12. Pedder H, et al. Health Technol Assess. 2026;30(46). doi:10.3310/GJNW3422.
13. NICE. ID6234 (GID-TA11289); 2024. https://www.nice.org.uk/guidance/gid-ta11289/documents/html-content-5.
14. NICE HTA Lab. MASH model concept. https://www.nice.org.uk/what-nice-does/our-research-work/hta-lab.
15. NICE. TA964; 2024. https://www.nice.org.uk/guidance/ta964.
16. EasySLR. ISPOR Europe 2026; Vienna, Austria.
17. SAGE study. JMIR Preprints. Preprint 108163; 2026.
18. Vivekanantha P, et al. Knee Surg Sports Traumatol Arthrosc. 2026;34(8):3011-3029. doi:10.1002/ksa.70412.
19. Yang X, et al. medRxiv. Preprint 2026. doi:10.64898/2026.09.06.26362399.
20. Reason T, et al. Pharmacoecon Open. 2024;8(2):205-220. doi:10.1007/s41669-024-00476-9.
21. Kim H, et al. JAMIA Open. 2026;9(4):ooag131. doi:10.1093/jamiaopen/ooag131.
22. Albakkar A, et al. Clin Res Cardiol. 2026;115:683. doi:10.1007/s00392-026-02870-1.
23. Srivastava T, et al. Pharmacoecon Open. 2026. doi:10.1007/s41669-026-00687-2.
24. NICE. TA687; 2021. https://www.nice.org.uk/guidance/ta687.
25. NICE. HST10; 2019. https://www.nice.org.uk/guidance/hst10.
26. Gupta S, et al. ISPOR Europe 2026; Vienna, Austria. Abstract EE656.
27. Hawkins N, et al. Eur Psychiatry. 2026;69(suppl 1):S931. doi:10.1192/j.eurpsy.2026.11693
28. NICE HTA Lab. AI-assisted evidence submission project. Accessed October 2026. https://www.nice.org.uk/what-nice-does/our-research-work/hta-lab.
29. Claxton L, et al. ISPOR Europe 2026; Vienna, Austria.
30. NICE. ECD11; 2024. https://www.nice.org.uk/corporate/ecd11.

Abbreviations

AI: artificial intelligence; **EAG:** external assessment group; **FP:** fractional polynomial; **GenAI:** generative artificial intelligence; **HEOR:** health economics and outcomes research; **HTA:** health technology assessment; **ISPOR:** The Professional Society for Health Economics and Outcomes Research; **ITC:** indirect treatment comparison; **MASH:** metabolic dysfunction-associated steatohepatitis; **MTA:** multiple technology appraisal; **NICE:** National Institute for Health and Care Excellence; **NMA:** network meta-analysis; **NSCLC:** non-small cell lung cancer; **PenTAG:** Peninsula Technology Assessment Group; **PFS:** progression-free survival; **PH:** proportional hazards; **PSA:** probabilistic sensitivity analysis; **QALY:** quality-adjusted life year; **RWE:** real-world evidence; **TA:** technology appraisal; **TAG:** Technology Assessment Group.

Disclosures

Funding: none. No financial disclosures.
AI statement: editing.

VALID INSIGHT®

PART OF THE BIOSCRIPT GROUP

validinsight.com

Valid Insight, Greenway House, Larkwood Way, Tytherington Business Park, Tytherington, Macclesfield, Cheshire, SK10 2XR

gareth.wynne-mchardy@bioscriptgroup.com

ISPOR, 11 11 2026©
2026 Valid Insight Ltd