



NICE's Pathways Pilot: Pursuing Good Decision Making in Difficult Circumstances

Dawn Lee¹ · Darren Burns^{1,2} · Ed Wilson¹

Accepted: 27 March 2024 / Published online: 13 April 2024
© The Author(s) 2024, corrected publication 2024

1 Introduction

The National Institute for Health and Care Excellence (NICE) announced their pathways pilot at the end of 2022 as part of a wider work programme aiming for a more proportionate approach to technology appraisals [1]. The pilot was proposed as a means for NICE to cope with the high concentration of topics within a limited number of disease areas (almost half of NICE health technology assessments [HTAs] lie within just 10). NICE's core concept was the production of reusable reference models for each disease area to reduce repetition and improve consistency in decision making. NICE also used this pilot as a mechanism to test (1) the use of real-world evidence (RWE) in decision making; and (2) innovative ways of making decisions in disease areas with multiple comparators that impact on multiple lines of treatment [2]. This was an ambitious project with multiple aims and objectives, some of which were potentially conflicting.

The Peninsula Technology Assessment Group (PenTAG) at the University of Exeter collaborated with NICE working on the production of the first pathways pilot model and appraisal in advanced renal cell carcinoma (RCC), known as the Exeter Oncology Model: RCC edition (EOM-RCC). The pilot involved (1) the production of a sequencing model capable of handling a multi-comparator decision space with 12 possible treatments, three different risk subgroups, the possibility of prior adjuvant therapy or not, and four possible lines of active therapy; and (2) using that model to appraise a first-line treatment for RCC: cabozantinib plus nivolumab. Work began in December 2022; the preliminary assessment report was scheduled to be delivered to NICE in March 2023, with the final assessment report in July 2023.

These timelines were extremely ambitious: construction of a previous model with a similar scope took a team at the Innovation Value Initiative 2 years [3]. This left very little scope to deal with difficulties arising from the nature of this project as a pilot incorporating not only multiple new processes and technical areas but also requiring application to a live appraisal.

This editorial presents PenTAG's learnings from the pilot process and some thoughts for consideration for HTA bodies looking to build on these for future work.

2 Our Experience of Building a Reference Model

Conceptually, a reference model promotes good-quality decision making. Treatments are compared consistently rather than in an inconsistent piecewise manner, while duplication of effort through the generation of so-called *de novo* yet structurally identical models is eliminated. Such an approach should help ensure the right treatments are made available to National Health Service (NHS) patients at the right price, contributing to technical and allocative efficiency of the NHS budget.

The PenTAG team developed the reference model in R [4], publicly hosted on Github [5]. R was ideal for this model due to its ability to handle the extensive computations required. With 744 potential treatment sequences across various populations, the model efficiently executed block-diagonal sparse matrix multiplications for sequencing calculations. This would be infeasible to implement efficiently in Microsoft Excel. Unfortunately, we were not able to add a graphical user interface in the timeframes available for the model build [6]; instead, we built the front-end in Excel, allowing the stakeholders a familiar and flexible means to interact with it. The model extracted all inputs directly from Excel, separating all sensitive data and inputs from the code.

✉ Dawn Lee
d.lee7@exeter.ac.uk

¹ PenTAG, University of Exeter Medical School, University of Exeter, St Luke's Campus, Exeter EX1 2LU, UK

² Delta Hat, Long Eaton, UK

Consequently, no confidential information was contained in the code.

When run, the model would:

- extract cost, resource use, utility, relative effectiveness, and treatment sequence setting inputs from Excel front-end;
- compute possible treatment sequences for each population;
- load patient-level data and conduct survival analyses;
- load network meta-analyses;
- populate and propagate relative efficacy network for all treatments at all lines;
- compute patient flow for all possible sequences at all possible lines and apply cost and utility weights;
- for the live appraisal, compute weighted average patient flow by first-line treatment and calculate the impact of patient access schemes (confidential discounts on the published list price) in increments of 1 from 1 to 100% for all treatments;
- output results as files to store and as a fully automated Word document following formatting requirements for NICE.

Previous appraisals in RCC highlighted issues with subsequent treatments in trials not being available in UK practice and difficulties in matching cost and effectiveness data when trying to compensate [7–11]. Consequently, we built a state transition model with tunnel states to incorporate time-dependency at later lines [12]. This approach simplified incorporation of the sequencing features that arose within the scope of the pilot. For prudence, we incorporated a partitioned-survival (PartSA) modelling approach in parallel. This allowed comparison with models following implementation precedent in advanced RCC HTA.

There were several key issues that we encountered in the process of model building. The fundamental tools for implementing the model already existed across several R packages [13–17]; however, we encountered a paucity of previous health economic cost-effectiveness models suitable for addressing our decision problem. Our modelling approach included time-dependency at all lines of treatment, resulting in a need for large (but sparse) matrix multiplications of up to nearly 15,000 rows and columns. Another issue was the sheer scale of the decision problem in terms of the systematic review, network meta-analyses, and clinical consultation work required. This in turn led to a complex, computationally expensive model: runtime was around 90 processor minutes to simulate hundreds of treatment pathways for tens of thousands of health states for thousands of time cycles for each pathway. By contrast, the PartSA version of the model took less than 5 min, although without addressing any of the issues of that approach. Compromises

were required when it came to probabilistic analysis, specifically removal of time dependency for second-line onwards during probabilistic analysis. Greater focus was given to testing of structural uncertainty in the more than 90 scenario analyses eventually required during the appraisal process. Use of the University of Exeter high performance computer, ISCA, facilitated parallelisation and dramatically reduced overall computational time.

Although the gold standard in principle, version control (in our case using Git and GitHub) was a double-edged sword [5]. Git allows developers to build pieces of the model simultaneously and then integrate their work alongside continuous quality control, while always having full accountability and the ability to trace back throughout the entire history of the project. The PenTAG team had varied experience with Git, ranging from none to several years of use in multi-developer settings. In parallel to extremely challenging timelines, Git version control use issues created some difficulties when two people were working on the same part of the code, although likely prevented many much more serious problems such as accidental deletion or code changes. Furthermore, the other stakeholders had little to no experience with Git and varying experience with R, and consequently found it difficult to navigate.

3 The Difficulties of Making Decisions in a Multi-line, Multi-treatment Decision Space

Making decisions using NICE's single technology appraisal (STA) process in a multi-line, multi-treatment decision space is extremely difficult and can lead to perverse outcomes. The RCC space is a perfect example of this: tyrosine kinase inhibitor monotherapies were recommended first (2009–2018), and nivolumab plus ipilimumab was then introduced via the Cancer Drugs Fund in 2019 without comparison to one of these (cabozantinib monotherapy, recommended in 2018; TA542) due to the timing of the appraisals. Upon exiting the Cancer Drugs Fund in 2022, this comparison was still not made. The consequences became apparent in a subsequent multiple technology appraisal (MTA; TA858, 2023), whereby neither nivolumab plus ipilimumab nor the new treatment pembrolizumab plus lenvatinib were found to be cost effective versus cabozantinib. NICE compromised, providing a positive recommendation for pembrolizumab plus lenvatinib if, and only if, nivolumab plus ipilimumab would otherwise be offered. Similarly, in the atopic dermatitis MTA (TA814), sequencing of treatments was a key consideration. New therapies were likely to add additional lines to the pathway but NICE struggled to address the problem fully due to a lack of data. The committee considered that cost-effectiveness analyses for sequences

should ideally be taken into account in decision making. In addition, NICE often receives enquiries around how its guidance fits within the broader pathway, how it relates to other available treatments, and how it applies in different clinical situations [18].

There is no agreed basis on which a decision-making committee can recommend more than a single option and be confident that its guidance represents an effective use of NHS resources. Statements such as 'options A and B are **both** cost effective' or 'options A and B are **similarly** cost effective' simply have no meaning, at NICE or in broader health economic literature. NICE's current piecewise modus operandi is essentially "once cost effective, always cost effective." There are problems with this; price changes (e.g. when coming off patent, which occurred during our work), displacements due to license changes, or new entrants potentially affect the cost effectiveness of *all* drugs in a pathway. The previously most cost-effective strategy at any line may change as a consequence. The decision problem needs revisiting every time the state of the world changes. Examining the incremental cost effectiveness (or equivalently net health benefit) of possible *sequences* of treatments may be one approach to take with decision rules made around that [19, 20]. Thus far, defining 'similar' has proven elusive, even in terms of clinical effectiveness. NICE's cost comparison route lacks a clear definition of similarity. Moreover, obtaining the necessary clinical data to accurately assess the impact of drug ordering can be challenging [21–23]. This is because sequencing models often rely on heroic assumptions, such as independence of effects, or require access to patient-level data.

4 What Incorporating Real-World Evidence Really Means

The identification, assessment and incorporation of RWE into our economic model was a key challenge. At the outset, the intention was to work with a vendor willing to provide such evidence to NICE. This arrangement fell through and we were consequently required to use evidence identified during our own literature review performed in accordance with the NICE RWE framework [24]. Fortunately, we identified a retrospective review of cases produced by the UK Renal Oncology Collaborative (ROC) [25]. This covered 17 UK centres, providing information on overall survival (OS), progression-free survival (PFS) and time on treatment for up to five lines of therapy, alongside the key disease and demographic variables to a very high level of completeness. The data available from the UK ROC were a much richer data source than the original planned source, which did not include information on PFS or risk status; however, with RWE, there often comes a catch. In this case, some data

were to be kept confidential from the companies involved while the data owners completed their publications. This led to protest from several of the company stakeholders, who argued that if they refused to provide data to NICE they would face negative consequences [26]. While of course this is not an ideal situation, we would note that as the external assessment group (EAG), neither we nor NICE received complete patient-level data in Analysis Data Model (ADaM) format from any of the involved companies and that, as with many oncology submissions, a large volume of critical data were redacted (utility values from the trial, data on time on treatment, relative dose intensities, etc.).

We found that when compared with clinical trials, patients in the real-world had less favourable outcomes due to treatments being given to people who did not meet restrictive trial inclusion criteria, reflecting the well-known differential between efficacy and effectiveness. We also found that subsequent therapies used in the trials differed considerably from those used in the real-world. This led to lower estimated OS when using RWE, less absolute OS gain for a given relative efficacy, and therefore less favourable (but more realistic) cost-effectiveness estimates. If NICE move to regular use of RWE to assess baseline risk, one could expect the need for larger price discounts to ensure cost effectiveness.

5 What Now?

Having completed the RCC pilot, it is clear that having a reference model of this nature achieves standardisation goals and highlights nicely the issues and shortcomings of the piecewise STA process (where structural uncertainties are often glossed over). However, it represents a considerable investment in time and resource. Technical efficiency gains will naturally follow only after a series of appraisals applying the reference model.

If NICE move to increased use of reference models, it will be important for academics to take the lead in developing these in partnership with industry, following the example of reference models in diabetes, as well as frequently re-used/re-built models such as the CORE Diabetes Model and Project HERCULES model in Duchenne's [27, 28]. This way, models can be developed that make the best use of all available data and consider all companies value propositions without bias towards particular companies. Our model is available open source now that the final guidance has been published (<https://github.com/nice-digital/NICE-model-repo>). Given the similarity of model structures used across the majority of oncology applications, it would form the perfect basis for an adaptable generic oncology model template.

NICE's pilot programme is to be commended for allowing the 'norms' of HTA to be challenged and allowing a

framework in which issues with current process can be addressed and new solutions tested. We would perhaps offer our view that these types of pilots take time and that de-risking them by increasing timelines and initially de-linking them from the heat of a live appraisal may be advisable in future.

Acknowledgements The authors would like to acknowledge the use of the University of Exeter High-Performance Computing (HPC) facility in carrying out this work. This project utilised high-performance computing funded by the UK Medical Research Council (MRC), Clinical Research Infrastructure Initiative (award number MR/M008924/1).

Funding Our work for NICE, as part of the pathways pilot, was funded by the National Institute for Health and Care Research (NIHR) Evidence Synthesis Programme as project number NIHR136008. No funding was received to produce the editorial.

Data availability The model used within NICE's pathways pilot can be found at: <https://github.com/nice-digital/NICE-model-repo>. This contains all data that was not redacted for confidentiality during the appraisal.

Declarations

Conflicts of Interest/Competing Interests Dawn Lee has provided consultancy services to a number of companies unrelated to possible intervention, comparators, intervention or comparator companies in the RCC appraisal. She was previously employed as the Chief Scientific Officer at Lumanity until September 2022 and is a very small shareholder in Value Demonstration Group (who are the holding company for Lumanity). She has also been a member of Fiecon's strategic council from November 2023. Darren Burns is employed as an analyst at Delta Hat Limited. He has worked on projects, and for companies, unrelated to possible intervention, comparators, intervention or comparator companies in this RCC appraisal. He has provided consultancy services for Pfizer on a technology unrelated to RCC, and Servier on multilocular cystic RCC (MCRCC), which is not considered a possible comparator in this appraisal. He has also worked for the Scottish Medicines Consortium on multiple projects. Ed Wilson has no conflicts of interest to declare.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

References

1. National Institute for Health and Care Excellence (NICE). Taking a proportionate approach to technology appraisals. Available at: <https://www.nice.org.uk/about/what-we-do/proportionate-approach-to-technology-appraisals>. Accessed 19 Feb 2024.
2. PenTAG. Treatments for renal cell carcinoma [ID6186]: Final Analysis Plan. 2023. Available at: <https://www.nice.org.uk/guidance/gid-ta11186/documents/assessment-report-3>. Accessed 19 Feb 2024.
3. Incerti D, Curtis JR, Shafrin J, Lakdawalla DN, Jansen JP. A Flexible Open-Source Decision Model for Value Assessment of Biologic Treatment for Rheumatoid Arthritis. *Pharmacoeconomics*. 2019;37(6):829–43. <https://doi.org/10.1007/s40273-018-00765-2>.
4. R Core Team. A Language and Environment for Statistical Computing. Available at: <https://www.R-project.org>. Accessed 19 Feb 2024.
5. GitHub. Available at: <https://github.com/>. Accessed 19 Feb 2024.
6. Hart R, Burns D, Ramaekers B, et al. R and Shiny for cost-effectiveness analyses: why and when? A hypothetical case study. *Pharmacoeconomics*. 2020;38(7):765–76. <https://doi.org/10.1007/s40273-020-00903-9>.
7. National Institute for Health and Care Excellence (NICE). Nivolumab with ipilimumab for untreated advanced renal cell carcinoma. Technology appraisal guidance [TA780]. Available at: <https://www.nice.org.uk/guidance/ta780>. Accessed 19 Feb 2024.
8. National Institute for Health and Care Excellence (NICE). Pembrolizumab with axitinib for untreated advanced renal cell carcinoma. Technology appraisal guidance [TA650]. Available at: <https://www.nice.org.uk/guidance/TA650>. Accessed 19 Feb 2024.
9. National Institute for Health and Care Excellence (NICE). Avelumab with axitinib for untreated advanced renal cell carcinoma. Technology appraisal guidance [TA645]. Available at: <https://www.nice.org.uk/guidance/ta645>. Accessed 19 Feb 2024.
10. National Institute for Health and Care Excellence (NICE). Tivozanib for treating advanced renal cell carcinoma. Technology appraisal guidance [TA512]. Available at: <https://www.nice.org.uk/guidance/ta512>. Accessed 19 Feb 2024.
11. National Institute for Health and Care Excellence (NICE). Nivolumab for previously treated advanced renal cell carcinoma. Technology appraisal guidance [TA417]. Available at: <https://www.nice.org.uk/guidance/ta417>. Accessed 19 Feb 2024.
12. Woods B, Sideris E, Palmer S, Latimer N, Soares M. TSD 19: Partitioned survival analysis as a decision modelling tool. 2017. Available at: <https://www.sheffield.ac.uk/media/34205/download?attachment>. Accessed 19 Feb 2024.
13. Bates D, Maechler M, Jagan M. Matrix: Sparse and Dense Matrix Classes and Methods, R package version 1.6-3. Available at: <https://CRAN.R-project.org/package=Matrix>. Accessed 19 Feb 2024.
14. Bengtsson H. A Unifying Framework for Parallel and Distributed Processing in R using Futures. *The R Journal*. 2021;13(2):208–27.
15. Bengtsson H. matrixStats: Functions that Apply to Rows and Columns of Matrices (and to Vectors), R package version 1.2.0. Available at: <https://CRAN.R-project.org/package=matrixStats>. Accessed 19 Feb 2024.
16. Dowle M, Srinivasan A. data.table: Extension of 'data.frame', R package version 1.14.8. Available at: <https://CRAN.R-project.org/package=data.table>. Accessed 19 Feb 2024.
17. Baio G, Berardi A, Heath A. Bayesian Cost-Effectiveness Analysis. V2.4.6. Available at: <https://CRAN.R-project.org/package=BCEA>. Accessed 19 Feb 2024.
18. National Institute for Health and Care Excellence (NICE). Public Board Meeting. Wednesday 18 September 2019 at 1.30pm in Sheffield Town Hall, Pinstone St, Sheffield S1 2HH. 2019. Available at: <https://www.nice.org.uk/Media/Default/Get-involved/Meetings-In-Public/Public-board-meetings/agenda-and-papers-sep-19.pdf>. Accessed 19 Feb 2024.
19. Alarid-Escudero F, Enns EA, Kuntz KM, Michaud TL, Jalal H. "Time Traveling Is Just Too Dangerous" but some methods are worth revisiting: the advantages of expected loss curves over

- cost-effectiveness acceptability curves and frontier. *Value in Health*. 2019;22(5):611–8. <https://doi.org/10.1016/j.jval.2019.02.008>.
20. Eckermann S, Willan AR. Presenting evidence and summary measures to best inform societal decisions when comparing multiple strategies. *Pharmacoeconomics*. 2011;29(7):563–77. <https://doi.org/10.2165/11587100-000000000-00000>.
 21. Tappenden P, Chilcott J, Brennan A, Squires H, Stevenson M. Whole disease modeling to inform resource allocation decisions in cancer: a methodological framework. *Value in Health*. 2012;15(8):P1127–1136.
 22. Lord J, Willis S, Eatock J, et al. Economic modelling of diagnostic and treatment pathways in National Institute for Health and Care Excellence clinical guidelines: the Modelling Algorithm Pathways in Guidelines (MAPGuide) project. *Health Technol Assess*. 2013;17(58):v–vi. <https://doi.org/10.3310/hta17580>.
 23. Hawkins N, Sculpher M, Claxton K. A method for identifying optimum treatment sequences in chronic diseases. In: 27th Annual Meeting of the Society for Medical Decision Making. 2005.
 24. National Institute for Health and Care Excellence (NICE). NICE real-world evidence framework. Corporate document [ECD9]. Available at: <https://www.nice.org.uk/corporate/ecd9/chapter/overview>. Accessed 19 Feb 2024.
 25. McGrane J, Frazer R, Challapalli A, et al. A multicenter real-world study reviewing systemic anticancer treatment choices and drop off rates between treatment lines for metastatic renal cell carcinoma in the United Kingdom: in the immunotherapy era. *J Clin Oncol*. 2022;40(6 Suppl):358–358. https://doi.org/10.1200/JCO.2022.40.6_suppl.358.
 26. National Institute for Health and Care Excellence (NICE). Cabozantinib with nivolumab for untreated advanced renal cell carcinoma [TA964]. Available at: <https://www.nice.org.uk/guidance/ta964>. Accessed 19 Feb 2024.
 27. Brown J, Palmer A, Bisgaard P, Chan W, Pedula K, Russell A. The Mt. Hood challenge: cross-testing two diabetes simulation models. *Diabetes Res Clin Pract*. 2000;50(3):S57–64.
 28. Duchenne UK. Project HERCULES. Available at: <https://www.duchenneuk.org/project-hercules/>. Accessed 19 Feb 2024.