



Bayesian Hierarchical Modelling for Histology-Independent Time-to-Event Outcomes in the NICE Single Technology Appraisal of Pembrolizumab for Solid Tumours with MSI-H/dMMR: External Assessment Group Perspective

Bradley M. Sugden¹ · Sabine E. Grimm¹ · Robert Wolff² · Nigel Armstrong² · Thomas Otten¹ · Teebah Abu-Zahra¹ · Mark Perry² · Mubarak Patel² · Jiongyu Chen² · Caro Noake² · Manuela Joore¹ · Willem J. A. Witlox¹

Accepted: 7 April 2024 / Published online: 7 May 2024
© The Author(s) 2024

1 Introduction

Oncological treatments, traditionally based on pathologic classification and organ of origin, are increasingly shifting towards histology-independent targeted therapies, classified based on specific genomic or molecular alterations. This is based on the idea that tumour types with a shared genomic/molecular alteration potentially respond in a similar way to such treatments that aim to interact or bind with the target molecule [1]. The first histology-independent marketing authorisation was granted by the European Medicines Agency in 2019 [2]. The latest European Medicines Agency guidance revision addresses the emergence of indications defined by a common biomarker and histology-independent basket trial designs (i.e. a trial investigating a targeted therapy for multiple histological subtypes with a shared biomarker or mutation) [3]. The European Medicines Agency identified two possible roles of basket trials, including one in early-phase trials. If considered as evidence for licensing decisions, the need to demonstrate sufficient homogeneity is specified: “... sponsors must justify and make it convincingly plausible by clinical and/or pre-clinical data that the interaction with tumour site or histology is negligible and this should also be supported by the final data” [3]. For reimbursement decisions, the National Institute for Health Research HTA Programme commissioned a report in 2020 assessing modelling approaches for histology-independent

cancer drugs to inform the National Institute of Health and Care Excellence (NICE) technology appraisals. The report highlighted the greater levels of heterogeneity within the licensed population, use of surrogate endpoints and the usual lack of comparators as possible challenges for using histology-independent basket trials to inform evidence within the appraisal process. For heterogeneity, Bayesian hierarchical modelling (BHM) is considered particularly suited to the assumption that inter-tumour site efficacy is similar within basket trials, representing a middle ground between assuming complete homogeneity (i.e. pooling all tumour sites) and complete heterogeneity (i.e. independent modelling of tumour sites). Bayesian hierarchical modelling therefore accounts for heterogeneity whilst simultaneously leveraging information available from different tumour sites [2].

Bayesian hierarchical modelling allows for the borrowing of information regarding treatment effects across histological subtypes, which is particularly useful in the context of small sample sizes in individual histological subtypes. As such, BHM provide a foundation to allow for the treatment effect in a given histology to be informed by all histologies, increasing the utilisation of the available data [4]. However, sufficient homogeneity is still required: “... the BHM is advantageous only if it is considered reasonable to allow such borrowing” [2]. Nevertheless, the approach does maintain the possibility of assessing histological subtypes individually, in addition to pooling the assessment. Such dual reporting is particularly useful in facilitating transparency in histology-independent technology appraisals [2, 5, 6].

Experience with BHM in the context of histology-independent cancer treatments and time-to-event outcomes has so far been lacking. Pembrolizumab for previously treated solid tumours with high microsatellite instability (MSI-H) or DNA mismatch repair deficiency (dMMR) represents the

✉ Bradley M. Sugden
Bradley.sugden@mumc.nl

¹ Department of Clinical Epidemiology and Medical Technology Assessment, Maastricht University Medical Centre+ (MUMC+), Maastricht, The Netherlands

² Kleijnen Systematic Reviews Limited, York, UK

first NICE technology appraisal submission (TA 914) based on evidence from an immunotherapy basket trial, and the first submission to utilise BHM to evaluate time-to-event outcomes. This commentary underscores the key learnings regarding basket trials and use of BHMs from the perspective of the External Assessment Group (EAG).

2 Case Study: Pembrolizumab for Solid Tumours with MSI-H/dMMR Synopsis

As part of NICE's Single Technology Appraisal process, NICE invited the manufacturer (Merck Sharp and Dohme UK) of pembrolizumab (Keytruda®) to submit evidence for the clinical and cost effectiveness of this drug for the following populations:

1. Adults with unresectable or metastatic MSI-H or dMMR colorectal cancer previously treated with fluoropyrimidine-based combination therapy.
2. Adults with advanced or recurrent MSI-H or dMMR endometrial cancer, whose disease has progressed on or following treatment with a platinum-containing therapy and who are not candidates for curative surgery or radiation.
3. Adults with unresectable or metastatic MSI-H or dMMR gastric, small-intestine or biliary cancer, whose disease has progressed on or following at least one prior therapy.

The company submission (CS) considered National Health Service standard of care as the comparator, which differed per tumour site and mostly consisted of baskets of pharmacological treatments.

The CS utilised the KEYNOTE-158 single-arm basket trial to inform the treatment effectiveness for endometrial, biliary, gastric and small-intestine cancer with MSI-H or dMMR [7]. For colorectal cancer, another trial, KEYNOTE-164, was used [8]. The KEYNOTE-158 trial was an ongoing, phase II, open-label, non-randomised, multicentre, single-arm basket trial evaluating the effects of 200 mg of pembrolizumab once every 3 weeks, in adults with advanced (unresectable and/or metastatic) tumours having MSI-H and/or dMMR status. Patients had tumours from four sites: endometrial ($n = 83$), small intestine ($n = 27$), gastric ($n = 51$) or biliary ($n = 22$). Outcomes of overall survival (OS), progression-free survival (PFS), objective response rate, duration of objective response and health-related quality of life were reported [7]. The KEYNOTE-164 trial was a completed, phase II, open-label, non-randomised, multicentre, single-arm trial evaluating the effects of 200 mg of pembrolizumab once every three weeks on adults with advanced (unresectable and/or metastatic) colorectal solid tumours ($n = 124$). Outcomes of OS, PFS, objective response rate and

duration of objective response were reported, but health-related quality of life was not included as the trial was not originally designed as a registration study [8].

Problems arose with the basket trial approach of KEYNOTE-158, patients having mixed histology with only a solitary biomarker in common. Heterogeneity between tumour sites was substantial, including median PFS, which varied from 4.1 months for gastric cancer to 23.4 months for small-intestine cancer [5]. The lack of a comparator in the trial was also problematic, requiring consideration of some form of unanchored indirect treatment comparison using other trials [9]. This was compounded by the MSI-H and dMMR status for most comparator trials being unknown. Consequently, population adjustment methodology, such as matching adjusted indirect comparison, was identified to be insufficient in reducing the risk of bias observed. Furthermore, reporting of adverse events was also a cause for concern with adverse events being combined across the four tumour sites of interest. It is therefore conceivable that such adverse event aggregation could obscure a high prevalence at specific tumour sites. Indeed, at the request of the EAG, rates per tumour site were provided before the committee meeting, which did reveal a substantial variation, such as a much higher rate of vomiting in the biliary cancer group than the gastric cancer group (27.3% vs 15.7%) [5].

The main treatment effectiveness outcomes within the CS were OS, PFS and time to death. To explore and capture heterogeneity between tumour sites while leveraging all data to inform each tumour site, the company used BHM alongside standard parametric modelling independent of tumour sites. For the BHM approach, data from the KEYNOTE-158 basket trial and KEYNOTE-164 colorectal cancer trial were pooled to inform pembrolizumab OS and PFS time-to-event analyses [7, 8]. The committee concluded that, whilst neither the BHM nor the standard parametric modelling (independent of tumour site) was ideal, both were plausible to inform decision making and acknowledged the usefulness of having both approaches [10].

2.1 BHM for Histology-Independent Technology Appraisals: EAG Considerations

The CS justified utilisation of the BHM approach with reference to a balance between assuming complete independence between tumour sites and complete homogeneity through pooling all tumour sites. The EAG, whilst acknowledging the advantage of BHM in the context of small individual sample sizes, questioned the suitability of the approach in this case, considering that borrowing across all tumour sites would only be appropriate under the assumption that each site can be justifiably considered to be subgroups of an overarching MSI-H/dMMR population. Indeed, the EAG noted the substantial heterogeneity between tumour sites, as observed in

Table 1 Median overall survival and progression-free survival for pembrolizumab, by tumour site [5] Source: External Assessment Group Report sections 3.2.5.1 and 3.2.5.2 [5]

Tumour site	Median (months): AsaT population (95% CI)	
	Overall survival	Progression-free survival
Colorectal	36.1 (24.0–NR)	4.0 (2.1–7.4)
Endometrial	NR (48.0–NR)	13.1 (4.9–25.7)
Gastric	26.9 (6.6–NR)	4.1 (2.1–24.6)
Small intestine	NR (16.2–NR)	23.4 (4.3–NR)
Biliary	19.4 (6.5–44.8)	4.2 (2.1–24.6)

ASaT all subjects as treated, CI confidence interval, NR not reached

the differences in survival outcomes (OS and PFS) (Table 1) [5]. Through applying BHM, tumour site-specific survival estimates are pulled towards an overall average (dependent on the permitted level of borrowing), potentially biasing survival estimates in individual tumour sites. However, in the presence of small sample sizes, complete independent modelling of individual tumour sites is likely to lead to imprecise estimates [4, 11]. The EAG thus recommended only sharing information between comparable tumour sites, justified and supported by clinical arguments and evidence. Specifically, the EAG recommended modelling the KEYNOTE-164 data for colorectal cancer independently, only allowing information to be shared using a BHM approach across tumour sites included in the KEYNOTE-158 trial. The company's scenario analyses using this approach and modelling all tumour sites independently based on individual parametric survival models had a relatively minor impact on cost-effectiveness results, suggesting the modelling approach was unlikely to be a key model driver [5].

The committee was particularly concerned with the lack of previous applications of BHM to time-to-event outcomes, and that there had been no peer review of the applied methodology. However, provided that the modelling approach had a minimal impact on cost-effectiveness estimates, with incremental cost-effectiveness ratios remaining below a £30,000/quality-adjusted life-year willingness-to-pay threshold, the committee considered both BHM and independent modelling approaches for informing decision making under the premise that neither approach was ideal [10]. Nevertheless, significant uncertainty remains for the application of BHM methods to address future histology-independent technology appraisals. Bayesian hierarchical modelling remains a suitable approach only if it is considered appropriate to allow for information borrowing between histological subtypes. Further, regarding concerns for using BHM to model time-to-event outcomes, Murphy et al. concluded that, despite response endpoints being unreliable surrogates for PFS or OS, utilising a surrogate-based modelling approach,

informed by meta-analysis predictions, may still be preferred for NICE appraisals, rather than extrapolating heavily censored PFS and OS data [2].

2.2 Key Learnings

Bayesian hierarchical modelling is a useful approach in the context of histology-independent basket trials, although only under the assumption that each histological subtype can be justifiably considered to be subgroups of an overarching population. Further research is required to provide guidance regarding reasonable justification of this, and as to determining the extent to which borrowing should be allowed in such models. For now, in the face of uncertainty regarding whether information borrowing is appropriate, the use of BHM can be supplemented with standard parametric modelling to aid in informing decision making.

Using BHMs to directly model time-to-event outcomes remains highly uncertain. Utilising a surrogate-based modelling approach, informed by meta-analysis predictions, may still be preferred for NICE appraisals.

Basket trials with a mixed histology population might be sufficient for licensing, but not for reimbursement decisions given the lack of homogeneity in final outcomes such as PFS and OS. Further evidence collection should therefore be considered, if possible in the relevant context. Most basket trials being single arm further complicates indirect comparisons, not only because of the heterogeneity in tumour histology, but also discrepancies in the status of the specific genomic or molecular alterations between the intervention and comparator trials.

3 Conclusions

Pembrolizumab was the first NICE technology appraisal based on evidence from an immunotherapy basket trial, and to utilise BHM to evaluate time-to-event outcomes. It can be concluded that EAGs and committees must be conscious of the potential for complete pooling or, in the case of small samples, no pooling to bias cost-effectiveness results. Furthermore, when considering BHMs, the appropriateness of information sharing across individual histologies should be sufficiently supported with evidence. Exploring different levels of borrowing and the use of standard parametric modelling can be used to further inform decision making.

Acknowledgements We gratefully acknowledge the expert advice input to the EAG report from Veerle Coupe, Department of Epidemiology and Data Science, Amsterdam Public Health, Amsterdam University Medical Center, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands.

Declarations

Funding This project was funded by the National Institute for Health Research Health Technology Assessment programme. Please visit the HTA programme website for further project information (<https://www.nihr.ac.uk/explore-nihr/funding-programmes/health-technology-assessment.htm>). This summary of the EAG report was compiled after NICE issued the final appraisal document. The views and opinions expressed herein are those of the authors and do not necessarily reflect those of NICE or the Department of Health.

Conflict of interest Bradley M. Sugden, Sabine E. Grimm, Robert Wolff, Nigel Armstrong, Thomas Otten, Teebah Abu-Zahra, Mark Perry, Mubarak Patel, Jiongyu Chen, Caro Noake, Manuela Joore and Willem J.A. Witlox have no conflicts of interest that are directly relevant to the content of this article.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Availability of data and material Not applicable.

Code availability Not applicable.

Authors' contributions The draft of this manuscript was written by BS, WW, SG, RW, NA, MPa, JC and CN. All authors had the chance to comment on previous versions of the manuscript and provided their approval for the final version to be published. RW, NA, MPe, MPa and CN critiqued the clinical effectiveness data reported, and the literature search conducted by the company and contributed to the writing of the EAG report. BS, SG, TO, TA-Z, MJ and WW critiqued the mathematical model provided and the cost-effectiveness analyses submitted by the company and contributed to the writing of the EAG report. BS acts as the overall guarantor for the article.

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

- Mansinho A, Fernandes RM, Carneiro AV. Histology-agnostic drugs: a paradigm shift: a narrative review. *Adv Ther.* 2023;40(4):1379–92.
- Murphy P, Glynn D, Dias S, Hodgson R, Claxton L, Beresford L, et al. Modelling approaches for histology-independent cancer drugs to inform NICE appraisals: a systematic review and decision-framework. *Health Technol Assess.* 2021;25(76):1–228.
- European Medicines Agency (EMA). Guideline on the evaluation of anticancer medicinal products in man: EMA/CHMP/205/95 Revision 6. Amsterdam: EMA; 2023. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-clinical-evaluation-anticancer-medicinal-products-revision-6_en.pdf. Accessed 15 Feb 2024.
- Murphy P, Claxton L, Hodgson R, Glynn D, Beresford L, Walton M, et al. Exploring heterogeneity in histology-independent technologies and the implications for cost-effectiveness. *Med Decis Mak.* 2021;41(2):165–78.
- Wolff R, Witlox W, Grimm S, Sugden B, Abu-Zarah T, Otten T, et al. Pembrolizumab for previously treated solid tumours with high microsatellite instability or mismatch repair deficiency [ID4036]: a single technology assessment. York: Kleijnen Systematic Reviews Ltd; 2023.
- Weymann D, Pollard S, Lam H, Krebs E, Regier DA. Toward best practices for economic evaluations of tumor-agnostic therapies: a review of current barriers and solutions. *Value Health.* 2023;26(11):1608–17.
- Maio M, Ascierto PA, Manzyuk L, Motola-Kuba D, Penel N, Cassier PA, et al. Pembrolizumab in microsatellite instability high or mismatch repair deficient cancers: updated analysis from the phase II keynote-158 study. *Ann Oncol.* 2022;33(9):929–38.
- Le DT, Diaz LA Jr, Kim TW, Van Cutsem E, Geva R, Jäger D, et al. Pembrolizumab for previously treated, microsatellite instability-high/mismatch repair-deficient advanced colorectal cancer: final analysis of KEYNOTE-164. *Eur J Cancer.* 2023;186:185–95.
- Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR, Welton NJ. NICE DSU Technical Support Document 18: methods for population-adjusted indirect comparisons in submission to NICE. Sheffield: Decision Support Unit, ScHARR; 2016. [http://www.nicedsu.org.uk/Populationadjusted-ICs-TSD\(3026862\).htm](http://www.nicedsu.org.uk/Populationadjusted-ICs-TSD(3026862).htm). Accessed 21 Feb 2024.
- National Institute for Health and Care Excellence. Pembrolizumab for previously treated endometrial, biliary, colorectal, gastric or small intestine cancer with high microsatellite instability or mismatch repair deficiency: NICE technology appraisal guidance 914. London: NICE; 2023. www.nice.org.uk/guidance/ta914. Accessed 15 Feb 2024.
- Mackay EK, Springford A. Evaluating treatments in rare indications warrants a Bayesian approach. *Front Pharmacol.* 2023;14:1249611.