Original Article

MDM Medical Decision Making

Medical Decision Making
1–14

DOI: 10.1177/0272989X20980327
journals.sagepub.com/home/mdm
SAGE

Exploring Heterogeneity in Histology-Independent Technologies and the Implications for Cost-Effectiveness

Peter Murphy, Lindsay Claxton*, Robert Hodgson*, David Glynn, Lucy Beresford, Matthew Walton, Alexis Llewellyn, Stephen Palmer, and Sofia Dias

Background. The National Institute for Health and Care Excellence and a number of international health technology assessment agencies have recently undertaken appraisals of histology-independent technologies (HITs). A strong and untested assumption inherent in the submissions included identical clinical response across all tumour histologies, including new histologies unrepresented in the trial. Challenging this assumption and exploring the potential for heterogeneity has the potential to impact upon cost-effectiveness. **Method**. Using published response data for a HIT, a Bayesian hierarchical model (BHM) was used to identify heterogeneity in response and to estimate the probability of response for each histology included in single-arm studies, which informed the submission for the HIT, larotrectinib. The probability of response for a new histology was estimated. Results were inputted into a simplified response-based economic model using hypothetical parameters. Histology-independent and histology-specific incremental cost-effectiveness ratios accounting for heterogeneity were generated. **Results.** The results of the BHM show considerable heterogeneity in response rates across histologies. The predicted probability of response estimated by the BHM is 60.9% (95% credible interval 16.0; 91.8%), lower than the naively pooled probability of 74.5%. A mean response probability of 56.9% (0.2; 99.9%) is predicted for an unrepresented histology. Based on the economic analysis, the probability of the hypothetical HIT being cost-effective under the assumption of identical response is 78%. Allowing for heterogeneity, the probability of various approval decisions being cost-effective ranges from 93% to 11%. **Conclusions.** Central to the challenge of reimbursement of HITs is the potential for heterogeneity. This study illustrates how heterogeneity in clinical effectiveness can result in highly variable and uncertain estimates of cost-effectiveness. This analysis can help improve understanding of the consequences of histology-independent versus histology-specific decisions.

Keywords
bayesian hierarchical model, economic evaluation, heterogeneity, histology-independent

Date received: March 19, 2020; accepted: November 17, 2020

The National Institute for Health and Care Excellence (NICE) and a number of international health technology assessment (HTA) agencies have recently undertaken appraisals of histology-independent technologies (HITs) for the treatment of cancer.[1–5] HITs are approved on the basis of a target genetic mutation, rather than on tumor histology, type, or location (hereafter referred to as "histology" for simplicity). Larotrectinib and entrectinib have

*These authors contributed equally to this work.

Corresponding Author
Peter Murphy, Centre for Reviews and Dissemination, University of York, A/B Block, York, Yorkshire, YO10 5DD, UK (peter.murphy@york.ac.uk).

recently gained regulatory approval as HITs for treating patients with advanced cancer with fusions involving the neurotrophic receptor tyrosine kinase (*NTRK*) genes.[6–8] Therefore, any patient with advanced cancer harboring an *NTRK* gene fusion is eligible for treatment, subject to other criteria in the marketing authorization.[6–8]

It is important to ensure that HIT reimbursement decisions are supported by systematic and robust assessments of clinical and economic evidence (i.e., how well the medicine or treatment works and its value for money) when considering their use in practice. Assessment of the clinical and cost-effectiveness of such treatments, however, creates a number of challenges given the breadth of the population covered by a histology-independent approval (which is likely to cover many individual histologies) and the types of evidence generated in support of these technologies.

Evidence of the effectiveness of HITs is commonly generated using basket trials. These largely early-stage, phase II exploratory trials recruit small cohorts to baskets defined by a common genetic mutation/marker or by tumor histology, based on a master protocol common to all baskets.[9–11] Despite recommendations in the literature for considering potential differences in response across baskets,[9,12–15] the clinical efficacy evidence used for decision making for larotrectinib[16,17] consisted of an average pooled overall response rate (ORR). Response is defined by tumor shrinkage, measured across histologies included in the basket trial, but with insufficient data to assess response by individual histology. However, the use of an average ORR to represent all histologies covered by the marketing authorization implicitly assumes identical clinical effectiveness across all histologies. This assumption fails to allow for heterogeneity in clinical effectiveness across histologies, that is, the fact that different groups of patients may obtain different treatment benefits based on observed characteristics (i.e., histology). In the context of HITs, there may be clinical and scientific arguments for heterogeneous treatment efficacy across tumor histologies[18] as well as across other clinical characteristics such as age, fusion type, and position in the treatment pathway.

This can present novel challenges to the decision making of reimbursement bodies, whose determinations typically consider only a technology's clinical and cost-effectiveness in a single indication. Such decisions operate on the assumption that a single, expected incremental cost effectiveness ratio (ICER) adequately represents the cost-effectiveness of a technology across the whole eligible population, including patients not represented in the available evidence. Failure to account for heterogeneity in cost-effectiveness across histologies may result in the reimbursement of a HIT for histologies in which it is not cost-effective.

Furthermore, if the assumption of homogeneous clinical effectiveness fails to hold across the histologies present in the clinical evidence, this casts further doubt on the assumption that homogeneity extends to histologies for which there is no evidence. The potential cost and health consequences of this uncertainty (i.e., of making an "incorrect" decision) could be significant. Thus, the consequences of heterogeneity for decision uncertainty should be quantified to allow for informed and accountable decision making.[19–21]

Bayesian hierarchical modeling (BHM) frameworks, which have been more typically used in adaptive basket trial designs,[22–25] can be used to overcome some of the limitations and assumptions highlighted above. Estimates of the level of heterogeneity across histologies, as well as pooled treatment effects for each histology, can be produced. They work on the assumption that treatment effects across histologies are exchangeable (i.e., drawn from the same distribution of effects) rather than identical—a more reasonable assumption in the absence of evidence to the contrary. In addition, these frameworks allow the prediction of the clinical effect in unrepresented histologies as long as they can also be assumed to be exchangeable with the included histologies.

Although, in theory, the BHM can be applied to dichotomous (e.g., histology response) and to time-to-event (TTE) outcomes (e.g., progression-free survival [PFS] and overall survival [OS]),[26] the assumption of exchangeability of the effects of treatment on survival outcomes across histologies is harder to justify than the equivalent assumption made for the effects of treatment

Centre for Reviews and Dissemination, University of York, York, Yorkshire, UK (PM, LC, RH, LB, MW, AL, SD); Centre for Health Economics, University of York, York, Yorkshire, UK (DG, SP). The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Financial support for this study was provided entirely by a contract with the National Institute for Health Research (NIHR). The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report. This article presents independent research funded by the NIHR. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care. This article is based on work completed for the Single Technology Assessments of Entrectinib and Larotrectinib commission by the NIHR Systematic Reviews Programme, project numbers 12/87/63 and 12/78/95. This work is also based on ongoing work carried out for the NIHR Systematic Reviews Programme commissioned project: modeling approaches for site-agnostic cancer drugs to inform NICE appraisals, project number 127582.

on response. Prognostic heterogeneity usually means pooled survival is harder to disentangle than pooled response,[10,27] and survival data tend to be immature at the time of reimbursement applications, making estimation of a hierarchical model for time-to-event outcomes even more challenging.

The aim of this study is to consider heterogeneity in response rates across histologies and to characterize its implications to the cost-effectiveness of HITs. A BHM framework is used to analyze published clinical evidence of an existing HIT to demonstrate the potential for heterogeneity in response rates across histologies. An economic model, using a simplified model structure and hypothetical data, considers how such heterogeneity may affect estimates of cost-effectiveness and decision uncertainty faced by HTA and reimbursement agencies. Finally, some of the issues raised by these approaches and possible alternatives are discussed.

Methods

Clinical Data

The analysis is based on the response rates used in the regulatory approval of the histology-independent Trk-inhibitor, larotrectinib.[16,17] The data set consists of a post hoc pooling of 55 patients covering 12 histologies and can be considered typical of the type of data that will be available when appraising the value of histology-independent drugs.[8,16,28–30] The ORR observed across the 55 patients is 74.5% and ranges from 0% to 100% across histologies. For a breakdown of the individual histology response rates from the trial, see Supplementary Material Table S1.

Predictive Response Using the BHM Framework

The BHM assumes that for each of the histologies j, the log-odds of response, θ_j (i.e., the measures of treatment effects), are exchangeable and follow a normal distribution[31]:

$$\theta_j \sim Normal(\mu, \sigma^2)$$

where σ is the standard deviation quantifying the between-histology heterogeneity and μ is the pooled mean effect across all histologies. Prior distributions must be selected for μ and σ and are likely to have some influence on the posterior estimates,[31,32] particularly when a small number of groups, each containing few patients, are included. A normal prior distribution for μ is used, centered on a probability of response of 0.3, with a variance of 10 across all histologies. A relatively conservative uniform prior distribution for σ (i.e., a priori assuming limited sharing of information across histologies) is used, which was found to be robust in a simulation study.[32] The sensitivity of the results to alternative priors presented in the literature is assessed.[31,32] The prior distributions used for the base-case analysis are

$$\mu \sim Normal(-0.8473, 10)$$

$$\sigma \sim Uniform(0, 5)$$

When the outcome is binary, the probability of response in each site, p_j, is recovered as

$$p_j = \frac{\exp(\theta_j)}{1 + \exp(\theta_j)}$$

Because the evidence does not reflect every histology that could be eligible for larotrectinib under the marketing authorization, the predictive distribution for the response rate in a new histology is calculated to reflect the full degree of uncertainty both due to the sample size and the observed heterogeneity in effects across the observed histologies. The resulting distribution is the probability of response in a "new," that is, unrepresented, histology.

To illustrate the impact of assuming identical response across histologies, a scenario in which identical response rates are assumed is implemented. This is achieved by analyzing the response data through a fixed-effects version of the BHM to ensure consistency with the methods of estimation.

The model is adapted from Thall et al.[31] and estimated using Markov chain Monte Carlo in OpenBUGS,[33] implemented in R[34] (version 3.6.0) using R2OpenBUGS[35] (version 3.2.3.2). Code and implementation details are presented in the supplementary material.

Model fit is assessed by plotting individual histology contributions to the residual deviance (in a well-fitting model, these are expected to be close to 1) and by comparing the total residual deviance to the number of histologies, G.

For all analyses, 55,000 iterations are run on 2 parallel chains, and the first 5000 iterations are discarded as "burn-in." Convergence is assessed by visual inspection of the Brooks-Gelman-Rubin plots and assessment of the $\hat{R}$ statistic.[36,37]

Economic Evaluation

To assess the economic implications of characterizing heterogeneity in clinical effectiveness, the cost-effectiveness

Table 1 Input parameters included in the economic model

Parameter	Value	95% CI	Source
Effectiveness			
Response rate	See Results section		BHM[16]
Median progression-free survival			
Responders	24 mo	[21.6; 26.4]	Assumed
Nonresponders	6 mo	[5.4; 6.6]	Assumed
Median overall survival			
Responders	36 mo	[32.4; 39.6]	Assumed
Nonresponders	12 mo	[10.8; 13.2]	Assumed
Utilities			
Progression-free survival			
Hypothetical HIT	0.79	[0.71;0.87]	Assumed
SoC	0.72	[0.65;0.79]	Assumed
Postprogression survival			
Hypothetical HIT	0.64	[0.57;0.71]	Assumed
SoC	0.64	[0.57;0.71]	Assumed
Costs (£)			
Drug acquisition costs			
Hypothetical HIT	Value-based price	—	Assumed
SoC	£20	—	Assumed
Health state costs			
Progression-free survival	£350	[£315; £385]	Assumed
Postprogression survival	£500	[£450; 550]	Assumed
Terminal care cost	£6,878	—	[39]
Distribution of eligible patients			[40]
Soft tissue sarcoma	1.80%	—	
Appendix	5.80%	—	
Breast	1.80%	—	
Cholangiocarcinoma	0.00%	—	
Colorectal	8.30%	—	
GIST	1.40%	—	
IFS	9.80%	—	
Salivary gland	0.70%	—	
Melanoma	1.10%	—	
Lung	6.20%	—	
Pancreatic	6.20%	—	
Thyroid	2.20%	—	
Unrepresented	55.10%	—	

CI, confidence interval; BHM, Bayesian hierarchical model; HIT, histology-independent technology; SoC, standard of care; GIST, gastrointestinal stromal tumour; IFS, Infantile fibrosarcoma.

of a hypothetical HIT for the treatment of solid tumors harboring an *NTRK* gene fusion is assessed. The simplified economic model draws on evidence from an existing Trk-inhibitor (larotrectinib) in the form of response outcomes from the BHM but otherwise uses hypothetical inputs and assumptions. The results of the model are therefore for purely illustrative purposes. In line with the NICE reference case,[38] the model considers a National Health Service and Personal Social Services perspective. Costs and quality-adjusted life-years (QALYs) are discounted using a 3.5% discount rate, and results are presented over a lifetime (30-y) time horizon. All parameters used in the economic model are shown in Table 1.

The economic model uses a landmark response-based structure that incorporates PFS and OS distributions, conditional on response, as presented in Ouwens et al.[41] The model structure consists of 3 mutually exclusive health states: 1) progression-free disease, 2) progressed disease, and 3) death. State occupancy is derived using the partitioned survival technique, which uses PFS curves to partition OS into those patients with progression-free and progressed disease.

In the context of HITs, the general challenges of generating an appropriate control are complicated by the need to cover multiple histologies and potentially use multiple data sets, each requiring adjustment for important

prognostic factors. A number of approaches to generating controls can be considered, including using nonresponders as a proxy for patients not receiving an active treatment,[42] differences in time-to-progression in the previous line of treatment,[43] literature estimates of postprogression survival,[43] the use of "big data,"[44] and elicitation methods.[45] For the purpose of this study, nonresponders are used as a proxy for controls, that is, the standard-of-care (SoC) arm. This approach is considered appropriate because of the ease with which potentially important prognostic factors are matched between arms (i.e., the number of prior lines of therapy within histologies and the presence of *NTRK* gene fusions).

Survival in the hypothetical HIT arm is calculated as a weighted average of the responder and nonresponder survival curves based on the percentage response. In model scenarios in which unrepresented tumors are included, the response rate of the unrepresented tumors is based on the predictive distributions of the BHM model. Survival in the SoC arm is modeled assuming a 0% response. This assumption is considered appropriate, as the hypothetical HIT is assumed to be used as an end-of-line treatment and therefore not displacing any active treatments. The plausibility of this assumption and the appropriateness of basing the control on nonresponders would need to be considered in a real economic analysis.

It is assumed that the survival functions of responders and nonresponders follow an exponential distribution; parametric survival curves are fitted to hypothetical estimates of median PFS and OS for responders and nonresponders. Hypothetical estimates of median PFS and OS were used. In practice, the trial data are likely the most appropriate source of PFS and OS data. However, the reliance on surrogate outcomes and lack of a concurrent randomized control arm represents a key limitation of the trial designs for HITs. As a result, additional external data (e.g., surrogate evidence based on landmark response outcomes) may be required to more appropriately inform extrapolations, particularly where there is very short follow-up or there is significant confounding.

It is assumed that patients are treated with the hypothetical HIT until progression and a monthly drug acquisition cost is applied while patients receive treatment. The hypothetical HIT price is set at a level that results in an ICER close to NICE's cost-effectiveness threshold. This is done using a version of the economic model in which the unrepresented tumors are included in the analysis (see the Decision Options section).

For the purpose of this analysis, the hypothetical HIT was assumed to meet NICE's end-of-life criteria, allowing a cost-effectiveness threshold of £50,000 per QALY gained. To represent forthcoming one-off treatment modalities such as gene therapies, a scenario analysis in which a technology with a one-off drug cost applied at the start of treatment is modeled. This scenario also sets the price such that the ICER is close to NICE's end-of-life threshold.

Hypothetical health state utility values, monthly health state costs, and monthly costs of SoC are used in the economic model. It is assumed that the hypothetical HIT is associated with a benefit to health-related quality of life while in the progression-free health state. It is also assumed there is no cost of identifying patients eligible for the HIT. A one-off terminal care cost, obtained from Georghiou and Bardsley,[39] is applied upon transition from the progressed disease state to the death state. The distribution of patients eligible for hypothetical HIT by histology is estimated using the approach outlined in the literature[40] and is used to reweight the histology-specific results.

To assess the uncertainty surrounding the variables included in the cost-effectiveness model, a probabilistic sensitivity analysis is undertaken using 10,000 samples. Uncertainty in the response rates is captured through inputting 10,000 iterations from the BHM into the economic model. To reflect uncertainty in the utility values, health state costs, and survival, standard errors are assumed to be 10% of the mean. All results are calculated as the mean average of the 10,000 iterations.

Decision Options

To compare the economic implications of allowing for heterogeneity in the response rates, the economic analysis considers 4 alternative approaches to generating ICERs of the hypothetical HIT compared with SoC. Three HTA decision options (decisions 1–3) can be considered true histology-independent decisions in which a single ICER is used to represent the cost-effectiveness of the technology across all histologies covered by a histology-independent marketing authorization, although they differ in how this common ICER is obtained. A fourth decision (decision 4) shows the range of ICERs for histology-specific decisions. The decision options are as follows:

- Decision 1: Uses the response rate produced by the fixed-effects version of the BHM. This assumes homogeneity in response across all histologies.
- Decision 2: Uses the individual histology response rates generated by the BHM to generate incremental costs and QALYs for each individual histology included in the clinical evidence. A single ICER is

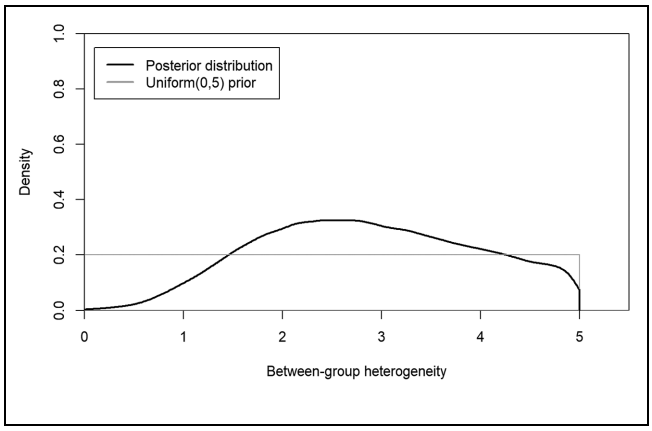

Figure 1 Prior and posterior distributions for the between-group heterogeneity standard deviations.

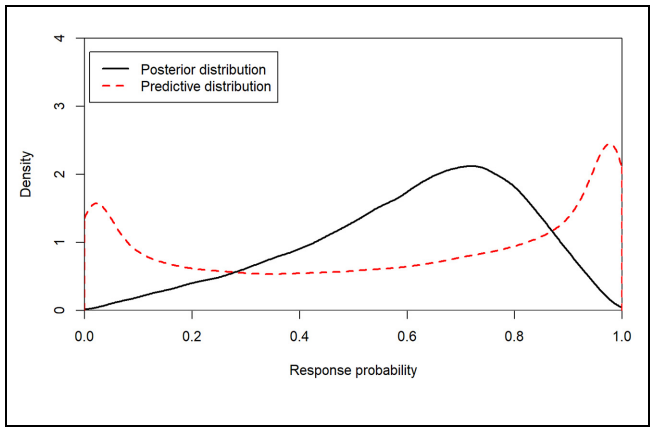

Figure 2 Posterior and predictive distributions of response probability.

then generated as a weighted average according to the distribution of patients in each histology making up the eligible population.

- Decision 3: Uses the approach in decision 2 but also including the unrepresented histologies in the calculation of the single ICER.
- Decision 4: Uses individual histology response rates produced by the BHM to show the range of histology-specific ICERs. These demonstrate the potential range of individual histology-specific ICERs when moving away from a histology-independent recommendation.

Results

Calculation of Response

The BHM estimates substantial heterogeneity in response between the histologies presented in the clinical evidence. The posterior median of σ was 2.86 on the log-odds scale, although there is considerable uncertainty (95% credible interval [CrI], 0.92 to 4.83; Figure 1).

The posterior distribution of the response probability, accounting for heterogeneity, is presented in Figure 2 and has an estimated mean response rate across all histologies of 60.9% with 95% CrI (16.0% to 91.8%). This value is lower than the naïve pooled mean response rate of 74.5% (i.e., the value obtained under a homogeneity model).

The mean response probability predicted for an unrepresented histology is 56.9%; however, the 95% CrI is wide, meaning this probability could be as low as 0.2% or as high as 99.9%. The uncertainty in the response of a histology unrepresented in the trial population represents both the underlying uncertainty in the mean response and the estimated heterogeneity across observed histologies (Figure 2).

The predictive and posterior probabilities of response were insensitive to the use of a half-normal prior, an inverse gamma prior, and a uniform prior centered on a probability of response of 0.5, the result of which can be seen in Table S2 of the supplementary material. Model fit statistics for the base-case and the sensitivity analyses are presented in the Supplementary Table S3.

The estimated probabilities of response for each histology are shown in Table 2. The estimated mean response of the fixed-effects version of the BHM shows a small difference in the probability of response compared with the observed pooled response: 74.2% compared with the observed response of 74.5%. This is likely a result of simulation error.

The effect of allowing borrowing of information across the histologies is to shrink the observed response probabilities toward the pooled mean response probability. Histologies with few patients borrow more information than histologies with more patients. Although the observed response suggested that cholangiocarcinoma, cancer of the appendix, breast cancer, and pancreatic cancer did not respond to larotrectinib, results of the BHM suggest the estimated mean response is greater than zero, by borrowing information from other histologies with more allocated patients and more promising response rates. The posterior distributions for these histologies are very wide, illustrating that there is very little information in the data to obtain estimates of the response rate with a sample size of 1 for these histologies.

Table 2 Probabilities of response for all histologies

Histology	Observed Response	Estimated Mean Response Based on BHM (%)	95% CrI
Fixed effects			
Pooled	41/55 = 74.5%	74.20%	62.0%−84.7%
Random effects			
Soft-tissue sarcoma	10/11 = 90.9%	88.10%	66.0%–99.1%
Salivary gland	10/12 = 83.3%	81.80%	58.0%−96.8%
IFS	7/7 = 100%	93.30%	70.5%−100%
Thyroid	5/5 = 100%	91.60%	63.0%−100%
Lung	3/4 = 75.0%	72.60%	30.4%−97.8%
Melanoma	2/4 = 50.0%	52.50%	12.4%−89.4%
Colon	1/4 = 25.0%	32.00%	2.6%−75.5%
GIST	3/3 = 100%	88.30%	49.3%−100%
Cholangiocarcinoma	0/2 = 0%	21.00%	0.0%−75.7%
Appendix	0/1 = 0%	30.00%	0.1%−89.7%
Breast	0/1 = 0%	30.00%	0.1%−90.1%
Pancreas	0/1 = 0%	29.80%	0.1%−89.7%
Unrepresented	—	56.90%	0.2%−99.9%

Economic Evaluation

Drug acquisition cost. Assuming continuous treatment, the cost of the hypothetical HIT is estimated to be £2200 per month; for a one-off cost, it was estimated to be £50,000.

Economic evaluation results. Relaxing the assumption of homogeneity in histology response has an impact on the ICER of the hypothetical HIT compared with SoC (Table 3). In the base case, the mean histology-independent ICER produced from the fixed-effects BHM model (decision 1) is almost £4000 under the cost-effectiveness threshold: £46,137 (£41,206 to £64,733). This is in contrast to the histology-independent decisions, which allow for heterogeneity in response (decisions 2 and 3): the ICERs are approximately at the cost-effectiveness threshold; however, the CrIs around the ICERs have increased: £50,339 (£42,899 to £71,959) and £50,009 (£41,487 to £83,857), respectively. When using the BHM response rates to generate histology-specific decisions (decision 4), the mean ICERs show considerable variability in the cost-effectiveness across histologies. The ICERs range from £43,639 (£39,356 to £60,540) per QALY for infantile fibrosarcoma (IFS) to £73,446 (£43,985 to £389,728) for cholangiocarcinoma. This variation can also be seen in the incremental costs and incremental QALYs. The 95% CrIs around the ICERs reveal the increase in uncertainty in the economic results of histologies in which the response was based on small patient numbers. For example, the 95% CrI around the ICER of the hypothetical HIT compared with SoC for appendiceal cancer ranges from £41,760 to £373,368.

The results of the scenario analysis in which a one-off cost of the hypothetical HIT was modeled at the start of treatment can be seen in Table 4. The mean histology-independent ICERs for the scenario analysis show a similar trend to the monthly cost scenario; however, the 95% CrIs around the ICERs are larger (£32,167 to £133,343 compared with £41,487 to £83,857 for decision 3). The range of histology-specific ICERs has increased as a result of the scenario analysis. The mean ICERs range from £33,530 to £115,526, considerably larger than was generated under the monthly cost of the hypothetical HIT model assumption. The results of this scenario analysis also show much larger 95% CrIs around the ICERs compared with the base case.

The influence of the uncertainty on each of the histology-independent decisions can be seen in the probability of decisions being cost-effective in Table 3. For decision 1, the probability of the hypothetical HIT being cost-effective at a cost-effectiveness threshold of £50,000 is 78%. This drops to 48% for decisions 2 and 3, despite only a small change in the mean ICER. Table 3 also shows the range of probabilities of individual histologies being cost-effective at a cost-effectiveness threshold of £50,000: approximately 93% for IFS to 11% for cholangiocarcinoma. A similar trend can be seen for the scenario of a one-off cost of treatment (Table 4). Cost-effectiveness acceptability curves illustrating the probability of the hypothetical HIT being cost-effective at different cost-effectiveness thresholds for the base case and

Table 3 Summary of the average incremental costs, incremental QALYS and ICERs for various histology-independent and histology-specific decision options under the base case assumption

Decision	Description	Incremental Costs (95% CrI)	Incremental QALYs (95% CrI)	ICER (95% CrI)	Probability of Being Cost-Effective at £50,000/QALY
Histology independent					
Decision 1	Histology independent–homogenous response	£66,290 (£52,777; £81,311)	1.44 (0.93; 1.96)	£46,137 (£41,206; £64,733)	78%
Decision 2	Histology independent–heterogeneous response	£54,161 (£41,487; £71,761)	1.08 (0.64; 1.66)	£50,339 (£42,899; £71,959)	48%
Decision 3	Histology independent–heterogeneous response including unrepresented histologies	£54,939 (£31,570; £79,507)	1.10 (0.39; 1.90)	£50,009 (£41,487; £83,857)	48%
Histology specific					
Decision 4	Soft-tissue sarcoma	£75,151 (£56,892; £93,111)	1.70 (1.07; 2.33)	£44,192 (£39,681; £61,503)	90%
	Salivary gland	£71,091 (£52,716; £89,755)	1.58 (0.96; 2.22)	£45,004 (£40,085; £62,689)	85%
	IFS	£78,283 (£59,487; £96,373)	1.79 (1.15; 2.43)	£43,639 (£39,356; £60,540)	93%
	Thyroid	£77,170 (£56,343; £95,910)	1.76 (1.07; 2.41)	£43,829 (£39,450; £60,612)	91%
	Lung	£65,404 (£38,349; £89,569)	1.41 (0.57; 2.21)	£46,366 (£40,156; £70,082)	72%
	Melanoma	£52,883 (£27,417; £80,303)	1.04 (0.28; 1.91)	£50,969 (£41,473; £101,176)	43%
	Colon	£40,011 (£21,162; £69,344)	0.65 (0.10; 1.56)	£61,098 (£43,711; £213,810)	17%
	GIST	£75,029 (£48,724; £95,567)	1.70 (0.89; 2.40)	£44,221 (£39,477; £62,014)	88%
	Cholangiocarcinoma	£33,174 (£17,971; £68,732)	0.45 (0.05; 1.53)	£73,446 (£43,985; £389,728)	11%
	Appendix	£38,622 (£18,330; £79,025)	0.61 (0.05; 1.86)	£62,915 (£41,760; £373,368)	21%
	Breast	£38,609 (£18,346; £78,912)	0.61 (0.05; 1.86)	£62,917 (£41,881; £371,002)	21%
	Pancreas	£38,588 (£18,304; £79,127)	0.61 (0.05; 1.85)	£63,037 (£41,842; £373,030)	21%
	Unrepresented	£55,562 (£19,223; £92,190)	1.12 (0.06; 2.29)	£49,755 (£39,887; £341,618)	52%

The base case assumes patients are treated with histology-independent technology until progression. A monthly drug acquisition cost of £2200 is modeled while the patient is receiving treatment. CrI, credible interval; QALY, quality-adjusted life-year; ICER, incremental cost-effectiveness ratio; IFS, Infantile fibrosarcoma; GIST, gastrointestinal stromal tumour.

Table 4 Summary of the average incremental costs, incremental QALYS and ICERs for various histology-independent and histology-specific scenarios – Scenario analysis of a one-off drug payment

Decision	Description	Incremental Costs (95% CrI)	Incremental QALYs (95% CrI)	ICER (95% CrI)	Probability of Being Cost-Effective at £50,000/QALY
Histology independent					
Decision 1	Histology independent–homogenous response	£58,030 (£54,941; £62,011)	1.44 (0.93; 1.96)	£40,388 (£31,474; £59,518)	88%
Decision 2	Histology independent–heterogeneous response	£55,889 (£53,194; £59,915)	1.08 (0.64; 1.66)	£51,945 (£36,056; £84,060)	39%
Decision 3	Histology independent–heterogeneous response including unrepresented histologies	£56,024 (£51,715; £61,481)	1.10 (0.39; 1.90)	£50,997 (£32,167; £133,343)	49%
Histology specific					
Decision 4	Soft-tissue sarcoma	£59,598 (£55,631; £64,445)	1.70 (1.07; 2.33)	£35,046 (£27,407; £52,677)	96%
	Salivary gland	£58,879 (£55,020; £63,719)	1.58 (0.96; 2.22)	£37,273 (£28,494; £57,740)	91%
	IFS	£60,148 (£56,129; £65,088)	1.79 (1.15; 2.43)	£33,530 (£26,566; £49,441)	98%
	Thyroid	£59,952 (£55,686; £64,919)	1.76 (1.07; 2.41)	£34,050 (£26,719; £52,846)	96%
	Lung	£57,874 (£52,841; £63,512)	1.41 (0.57; 2.21)	£41,028 (£28,617; £93,325)	75%
	Melanoma	£55,662 (£51,086; £61,444)	1.04 (0.28; 1.91)	£53,647 (£31,995; £184,431)	42%
	Colon	£53,389 (£50,106; £59,034)	0.65 (0.10; 1.56)	£81,527 (£37,695; £497,405)	14%
	GIST	£59,572 (£54,558; £64,896)	1.70 (0.89; 2.40)	£35,111 (£26,890; £61,973)	92%
	Cholangiocarcinoma	£52,181 (£49,819; £58,911)	0.45 (0.05; 1.53)	£115,526 (£38,349; £1,027,916)	10%
	Appendix	£53,145 (£49,822; £60,781)	0.61 (0.05; 1.86)	£86,572 (£32,555; £974,846)	20%
	Breast	£53,149 (£49,822; £60,923)	0.61 (0.05; 1.86)	£86,611 (£32,841; £966,066)	20%
	Pancreas	£53,139 (£49,823; £61,006)	0.61 (0.05; 1.85)	£86,808 (£32,765; £971,990)	20%
	Unrepresented	£56,132 (£49,840; £63,952)	1.12 (0.06; 2.29)	£50,265 (£27,736; £853,091)	53%

This scenario assumes patients are treated with a one-off treatment of histology-independent technology. A one-off drug acquisition cost of £50,000 is modeled at the start of treatment. CrI, credible interval; QALY, quality-adjusted life-year; ICER, incremental cost-effectiveness ratio; IFS, Infantile fibrosarcoma; GIST, gastrointestinal stromal tumour.

the alternative pricing scenario can be seen in the Supplementary Figure S1 and Figure S2.

Discussion

In this article, we used a BHM framework to explore the potential for heterogeneity in the clinical effectiveness evidence supporting a HIT. The results of the BHM analysis identified substantial evidence of heterogeneity in clinical effects across histologies, demonstrating the importance of appropriately accounting for heterogeneity when estimating clinical effects.

Using a simplified model based on hypothetical data, we illustrate how alternative approaches to characterizing the clinical effectiveness of a hypothetical HIT affect the cost-effectiveness and the uncertainty around the ICER. The histology-independent ICERs ranged from £46,137 (£41,206 to £64,733) per QALY under the assumption of homogenous response rates to £50,009 (£41,487 to £83,857) per QALY once we accounted for heterogeneity and included unrepresented histologies.

In addition, histology-specific estimates of cost-effectiveness may vary substantially, ranging from £43,639 (£39,356 to £60,540) per QALY in IFS to £73,446 (£43,985 to £389,728) in cholangiocarcinoma because of the different response rates between histologies; in our analysis, 7 of the histologies would be considered cost-effective at a threshold of £50,000 per QALY in the base case, while a further 6 histologies would not. If a histology-independent reimbursement decision was made, a large proportion of patients under the marketing authorization would not be treated cost-effectively. This has consequences for reimbursement agencies considering HITs such as entrectinib and larotrectinib, as they have the opportunity to optimize recommendations by limiting reimbursement to where benefits are greatest, increasing allocative efficiency.[19,20]

The results of our analysis demonstrate the importance of appropriately accounting for uncertainty when considering histology-specific estimates of cost-effectiveness and how point estimates of the ICERs independent of the uncertainty may result in misleading conclusions. For example, the mean estimated ICERs for melanoma and the unrepresented histologies are both approximately £50,000 per QALY (Table 3). However, comparison of the uncertainty reveals that the unrepresented histology has a 95% CrI 5 times the width, with an upper limit of £301,731 per QALY.

The results also demonstrate the potential masking of heterogeneity by considering only the ICERs for different decisions. In the base case, the results of the economic model show large variation in the incremental QALYs and less variation in the ICERs. Because time on treatment is linked to PFS, a large variation in PFS results in an almost proportionate variation in costs. This effect was demonstrated in the scenario in which a fixed, one-off cost of treatment was modeled. This could have important implications for HIT reimbursement decision making. Currently, approved HITs (such as entrectinib and larotrectinib) link time on treatment (and therefore cost) to PFS. However, given the ongoing basket trial evaluating a gene therapy,[46] it is plausible that a technology requiring a one-off treatment cost could seek a histology-independent recommendation in the future.

Strengths and Limitations of the Proposed Analytical Framework

The principal strength of the BHM framework is that it allows the restrictive assumption of identical effectiveness across histologies to be relaxed and for individual histology response rates to be estimated even when patient numbers are small or where no evidence exists. These histology-specific response rates are required for a number of reasons. First, the NICE reference case recommends the exploration of heterogeneity in clinical and cost-effectiveness.[38] Exploration of clinical heterogeneity would not be possible for each histology separately given limited patient numbers. Second, generating a single ICER requires the distribution of patients eligible for treatment in the trial to reflect the distribution expected in the population. Any reweighting of the results to match the trial distribution to the expected real-world distribution (to improve the external validity) requires individual histology results. Third, histologies not included in the evidence are still included in a histology-independent decision. Without the predictive distribution provided by the BHM, decisions are being made about a potentially substantial number of histologies with no empirical evidence.

The BHM approach relies on the assumption that response rates are exchangeable across histologies (i.e., that they are similar to one another rather than being either equal or completely different). Hierarchical designs have been criticized when there is insufficient information in the outcome data to determine whether borrowing across subgroups is appropriate.[13,47] Alternative forms of BHM that restrict the borrowing of information to similar baskets while avoiding optimistic borrowing from extreme baskets can be used.[47] This, however, requires judgments to be made (based on clinical and/or empirical criteria) about the set of baskets within which information can be borrowed.

The use of a landmark response framework presents a further limitation, as it makes strong assumptions about the relationship between survival outcomes (PFS and OS) and response. Specifically, it assumes that response is not only a good predictor of future survival outcomes but also that this relationship is constant across histologies. In reality, it is unlikely that this assumption will ever hold and will frequently be far from being a reasonable approximation, as there will be a degree of variability in the strength of response as a predictor and the nature of this relationship.[48–50] When applying a response-based modeling approach for the evaluation of HITs, it is therefore important to consider the strength of evidence supporting the surrogacy assumptions and whether the ability to generate histology-specific ICERs outweighs any associated uncertainties. Where there are substantive concerns regarding the potential for heterogeneity in survival endpoints, alternative analysis methods may be more appropriate.

The application of flexible parametric models and mixture models may allow for heterogeneity in survival outcomes to be reflected. However, in the context of HITs, this has a number of limitations. First, the use of a single full-population ICER across multiple tumor sites with potentially different treatment effectiveness, comparators, costs, and health-related quality of life will be difficult to interpret. Given the amount of heterogeneity associated with HITs, estimating the average cost-effectiveness for the full patient population covered in the scope may not provide enough information to decision makers. Second, the approaches rely on extrapolations of the observed survival data, which will potentially be immature, resulting in highly uncertain predictions. This is likely to be the case for trials that are powered on response endpoints, such as larotrectinib and entrectinib.[6–8]

Heterogeneity in PFS/OS could also be explored using the BHM in a similar way to response.[26] However, given the immaturity of the survival data and restrictions around the requirement of a common parametric distribution across histologies, it is unclear whether this type of model would provide useful results. To address concerns regarding the maturity of the TTE endpoints, the BHM could alternatively be applied to specific landmark survival time points (e.g., 6 or 12 mo) for which more robust data exist, with surrogate relationships employed to predict longer-term survival conditional on survival up to these specific time points.

A further alternative would be to apply the BHM response assessments to conditional PFS and OS distributions from the overall population or to link them to external surrogate data. Although such an approach is less desirable than having robust TTE data for the overall population and each specific subgroup of interest, it may provide a basis for initial explorations of the potential impact and importance of heterogeneity as well as guide further data collection and help prioritize specific subgroups.

Limitations of the Exemplar Analysis

Because of a number of simplifying assumptions, it is possible that our results underestimate potential heterogeneity, as we focused only on heterogeneity in clinical effects. In reality, there may be a number of other input parameters that vary across histologies. For example, in our hypothetical economic analysis, health state utility values, health state costs, and comparator costs are all assumed to be identical across histologies. It is, however, likely that these have the potential to vary substantially, along with important patient characteristics such as line of therapy and *NTRK*-fusion partner, further contributing to heterogeneity in the estimates of cost-effectiveness. These simplifying assumptions are also likely to mean that the predicted estimates of uncertainty are underestimates because several parameters were excluded from the probabilistic analysis. For example, uncertainty around the predictive value of response is not included in the model nor is uncertainty in the response rate for SoC which is assumed to be zero.

Importantly, our model ignores testing costs associated with identifying patients who are *NTRK* fusion positive. Testing costs may vary substantially across histologies because of variation in current testing availability and in the prevalence of *NTRK* fusions. Such variations are likely to further exacerbate the heterogeneity in cost-effectiveness estimates across histologies.

This article focused on histology as the main source of heterogeneity. However, heterogeneity could be explored using a range of alternative characteristics and subgroups. To move from histology as the main source of heterogeneity to considering a wider range of characteristics requires an understanding of how different characteristics can be used and combined in different ways in decision making. This is a complex question that requires further research.

Further, although this study emphasized the importance of characterizing the uncertainty associated with heterogeneity, we do not consider the consequences of uncertainty. An exploration of the value of distinguishing between different types of patients, known as the value of heterogeneity,[19–21] would inform reimbursement

decision makers of the consequences of alternative policy options and be an appropriate subsequent exploration of heterogeneity in the context of HITs. Furthermore, given the decision uncertainty identified in this analysis, value of information could also be conducted to help provide reimbursement decision makers with information on the drivers of decision uncertainty.

Conclusion

Histology-independent treatments represent a potentially important shift in the treatment of cancer. However, it is important to properly address the clinical and cost-effectiveness of these technologies. This study found considerable heterogeneity in response rates across histologies, which can result in highly heterogeneous histology-specific estimates of cost-effectiveness. This study calls into question the assumption of homogeneity in HIT response rates across different histologies, which undermines the appropriateness of histology-independent reimbursement decisions. Where there is evidence of heterogeneity, decision makers may consider making more optimized recommendations in which a HIT is approved for only specific subgroups or histologies for which efficacy evidence is sufficiently robust.

Contributors and Guarantor of Information

PM and SD carried out the statistical analysis. PM, LC, and LB created and implemented the economic model. PM and RH drafted the manuscript. All authors discussed the direction of the research and commented on the early drafts and the final version of the manuscript.

ORCID iDs

Peter Murphy https://orcid.org/0000-0001-8864-1416
Matthew Walton https://orcid.org/0000-0003-1932-3689
Sofia Dias https://orcid.org/0000-0002-2172-0221

Availability of Data and Materials

All data used in this analysis are provided in the main body of text and the supplementary material.

Code Availability

See supplementary material.

Supplemental Material

Supplementary material for this article is available on the *Medical Decision Making* Web site at http://journals.sagepub.com/home/mdm.

References

1. National Institute for Health and Care Excellence. Entrectinib for treating NTRK fusion-positive solid tumours [ID1512]. London: NICE; 2020 Available from: https://www.nice.org.uk/guidance/indevelopment/gid-ta10414. Accessed February 20, 2020.
2. National Institute for Health and Care Excellence. Larotrectinib for treating advanced solid tumours with TRK fusions [ID1299]. London: NICE; 2020. Available from: https://www.nice.org.uk/guidance/indevelopment/gid-ta10229. Accessed February 20, 2020.
3. Canadian Agency for Drugs and Technologies in Health (CADTH). Entrectinib (TBD) for neurotrophic tyrosine receptor kinase (NTRK) fusion-positive solid tumours. 2019 Available from: https://www.cadth.ca/entrectinib-tbd-neurotrophic-tyrosine-receptor-kinase-ntrk-fusion-positive-solid-tumours. Accessed February 20, 2020.
4. Canadian Agency for Drugs and Technologies in Health (CADTH). Larotrectinib for neurotrophic tyrosine receptor kinase (NTRK) locally advanced or metastatic solid tumours. 2019. Available from: https://cadth.ca/larotrectinib-neurotrophic-tyrosine-receptor-kinase-ntrk-locally-advanced-or-metastatic-solid. Accessed February 20, 2020.
5. Institute for Quality and Efficiency in Health Care (IQWiG). Larotrectinib (solid tumours with neurotrophic tyrosine receptor kinase [NTRK] gene fusion)—benefit assessment according to §35a Social Code Book V 2020. Available from: https://www.iqwig.de/en/projects-results/projects/drug-assessment/a19-90-larotrectinib-solid-tumours-benefit-assessment-according-to-35a-social-code-book-v.12640.html. Accessed February 20, 2020.
6. US Department of Health and Human Services. Food and Drug Administration. Drug approval package: Vitrakvi (larotrectinib). Rockville (MD): US Food and Drug Administration; 2018. Available from: https://www.accessdata.fda.gov/drugsatfda_docs/nda/2018/210861Orig1s000_21171Orig1s000TOC.cfm. Accessed January 17, 2020.
7. US Department of Health and Human Services. Food and Drug Administration. Center for Drug Evaluation and Research. *Approval package for application number 212725Orig1s000, 212726Orig1s000*. Trade name: ROZLYTREK® capsules, 100mg and 200 mg (entrectinib). Rockville (MD): US Food and Drug Administration; 2019.
8. European Medicines Agency. Assessment report: VITRAKVI. International non-proprietary name: larotrectinib: London: EMA; 2019.
9. Blagden SP, Billingham L, Brown LC, et al. Effective delivery of complex innovative design (CID) cancer trials—a consensus statement. *Br J Cancer*. 2020;122:473–82.
10. Renfro LA, Mandrekar SJ. Definitions and statistical properties of master protocols for personalized medicine in oncology. *J Biopharm Stat*. 2018;28(2):217–28.
11. Woodcock J, LaVange LM. Master protocols to study multiple therapies, multiple diseases, or both. *N Engl J Med*. 2017;377(1):62–70.

12. Antonijevic Z, Beckman R, eds. *Platform Trial Designs in Drug Development: Umbrella Trials and Basket Trials*. Boca Raton (FL): CRD Press; 2019.
13. Freidlin B, Korn EL. Borrowing information across subgroups in phase II trials: Is it useful? *Clin Cancer Res*. 2013;19(6):1326–34.
14. Simon R. Critical review of umbrella, basket, and platform designs for oncology clinical trials. *Clin Pharmacol Ther*. 2017;102(6):934–41.
15. London WB, Chang MN. One- and two-stage designs for stratified phase II clinical trials. *Stat Med*. 2005;24(17): 2597–611.
16. Center for Drug Evaluation and Research. NDA multidisciplinary review and evaluation NDA 210861 and NDA 211710 VITRAKVI (larotrectinib). Rockville (MD): US Food and Drug Administration; 2016.
17. Drilon A, Laetsch TW, Kummar S, et al. Efficacy of larotrectinib in TRK fusion-positive cancers in adults and children. *N Engl J Med*. 2018;378(8):731–9.
18. Hyman DM, Puzanov I, Subbiah V, et al. Vemurafenib in multiple nonmelanoma cancers with BRAF V600 mutations. *N Engl J Med*. 2015;373(8):726–36.
19. Coyle D, Buxton MJ, O'Brien BJ. Stratified cost-effectiveness analysis: a framework for establishing efficient limited use criteria. *Health Econ*. 2003;12(5):421–7.
20. Espinoza MA, Manca A, Claxton K, Sculpher MJ. The value of heterogeneity for cost-effectiveness subgroup analysis: conceptual framework and application. *Med Decis Making*. 2014;34(8):951–64.
21. Basu A, Meltzer D. Value of information on preference heterogeneity and individualized care. *Med Decis Making*. 2007;27(2):112–27.
22. Herson J. Predictive probability early termination plans for phase II clinical trials. *Biometrics*. 1979;35(4):775–83.
23. Thall PF, Simon R. Practical Bayesian guidelines for phase IIB clinical trials. *Biometrics*. 1994;50(2):337–49.
24. Heitjan DF. Bayesian interim analysis of phase II cancer clinical trials. *Stat Med*. 1997;16(16):1791–802.
25. Tan S-B, Machin D. Bayesian two-stage designs for phase II clinical trials. *Stat Med*. 2002;21(14):1991–2012.
26. Thall PF, Wathen JK, Bekele BN, Champlin RE, Baker LH, Benjamin RS. Hierarchical Bayesian approaches to phase II trials in diseases with multiple subtypes. *Stat Med*. 2003;22(5):763–80.
27. Simon R. New designs for basket clinical trials in oncology. *J Biopharm Stat*. 2018;28(2):245–55.
28. Center for Drug Evaluation and Research. NDA multidisciplinary review NDA 212725 ROZLYTREK (entrectinib). Rockville (MD): US Food and Drug Administration; 2019.
29. Health Canada. Summary basis of decision (SBD) for VITRAKVI. Government of Canada, 2019. Updated December 18, 2019; May 13, 2020. Available from: https://hpr-rps.hres.ca/reg-content/summary-basis-decision-detailTwo.php?linkID=SBD00455
30. Health Canada. Rozlytrek—notice of compliance with conditions—qualifying notice. Government of Canada, 2020. Updated March 18, 2020; May 13, 2020. Available from: https://www.canada.ca/en/health-canada/services/drugs-health-products/drug-products/notice-compliance/conditions/rozlytrek-qualifying-notice.html
31. Thall PF, Wathen JK, Bekele BN, Champlin RE, Baker LH, Benjamin RS. Hierarchical Bayesian approaches to phase II trials in diseases with multiple subtypes. *Stat Med*. 2003;22(5):763–80.
32. Cunanan KM, Iasonos A, Shen R, Gönen M. Variance prior specification for a basket trial design using Bayesian hierarchical modeling. *Clin Trials*. 2019;16(2):142–53.
33. Lunn D, Jackson C, Best N, Thomas A, Spiegelhalter D. *The BUGS Book*. Boca Raton (FL): CRC Press; 2013.
34. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna (Austria): R Foundation for Statistical Computing; 2019. Available from: https://www.R-project.org/
35. Sturtz S, Ligges U, Gelman A. R2WinBUGS: A package for running WinBUGS from R. *J Stat Softw*. 2005;12(3): 1–16. Available from: http://www.jstatsoft.org/v2/i03
36. Brooks SP, Gelman A. General methods for monitoring convergence of iterative simulations. *J Comput Graph Stat*. 1998;7(4):434–55.
37. Gelman A, Rubin DB. Inferences from iterative simulation using multiple sequences. *Stat Sci*. 1992;7:457–72.
38. National Institute for Health and Care Excellence (NICE). *Guide to the Methods of Technology Appraisal*. London: NICE; 2013.
39. Georghiou T, Bardsley M. *Exploring the Cost of End of Life Care*. London: Nuffield Trust; 2014.
40. Hodgson R, Llewellyn A, Murphy P, et al. Entrectinib for treating NTRK fusion-positive solid tumours: a single technology appraisal. York (UK): CRD and CHE, University of York, Technology Assessment Group; 2019.
41. Ouwens M, Mukhopadhyay P, Zhang Y, Huang M, Latimer N, Briggs A. Estimating lifetime benefits associated with immuno-oncology therapies: challenges and approaches for overall survival extrapolations. *Pharmacoeconomics*. 2019;37(9):1129–38.
42. Hatswell AJ, Thompson GJ, Maroudas PA, Sofrygin O, Delea TE. Estimating outcomes and cost effectiveness using a single-arm clinical trial: ofatumumab for double-refractory chronic lymphocytic leukemia. *Cost Eff Resour Alloc*. 2017;15:8.
43. Hatswell AJ, Sullivan WG. Creating historical controls using data from a previous line of treatment: two non-standard approaches. *Stat Methods Med Res*. 2019;29(6): 1563–72.
44. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol*. 2016;183(8):758–64.
45. Bojke L, Claxton K, Bravo-Vergel Y, Sculpher M, Palmer S, Abrams K. Eliciting distributions to populate decision analytic models. *Value Health*. 2010;13(5):557–64.
46. ClinicalTrials.gov. National Library of Medicine (U.S.). EGFR806 CAR T cell immunotherapy for recurrent/

refractory solid tumors in children and young adults. Identifier NCT03618381. 2018. Available from: https://ClinicalTrials.gov/show/NCT03618381. Accessed February 22, 2020.
47. Neuenschwander B, Wandel S, Roychoudhury S, Bailey S. Robust exchangeability designs for early phase clinical trials with multiple strata. *Pharm Stat*. 2016;15(2):123–34.
48. Kaufman HL, Schwartz LH, William WN Jr., et al. Evaluation of classical clinical endpoints as surrogates for overall survival in patients treated with immune checkpoint blockers: a systematic review and meta-analysis. *J Cancer Res Clin Oncol*. 2018 Nov;144(11):2245–61.
49. Brown LC, Sonpavde G, Armstrong AJ. Can RECIST response predict success in phase 3 trials in men with metastatic castration-resistant prostate cancer? *Prostate Cancer Prostatic Dis*. 2018;21(3):419–30.
50. Moriwaki T, Yamamoto Y, Gosho M, et al. Correlations of survival with progression-free survival, response rate, and disease control rate in advanced biliary tract cancer: a meta-analysis of randomised trials of first-line chemotherapy. *Br J Cancer*. 2016;114(8):881–8.